# Interactions with Pseudo-Sapiens

## User perception of anthropomorphism, mind, and trust in humanlike social agents

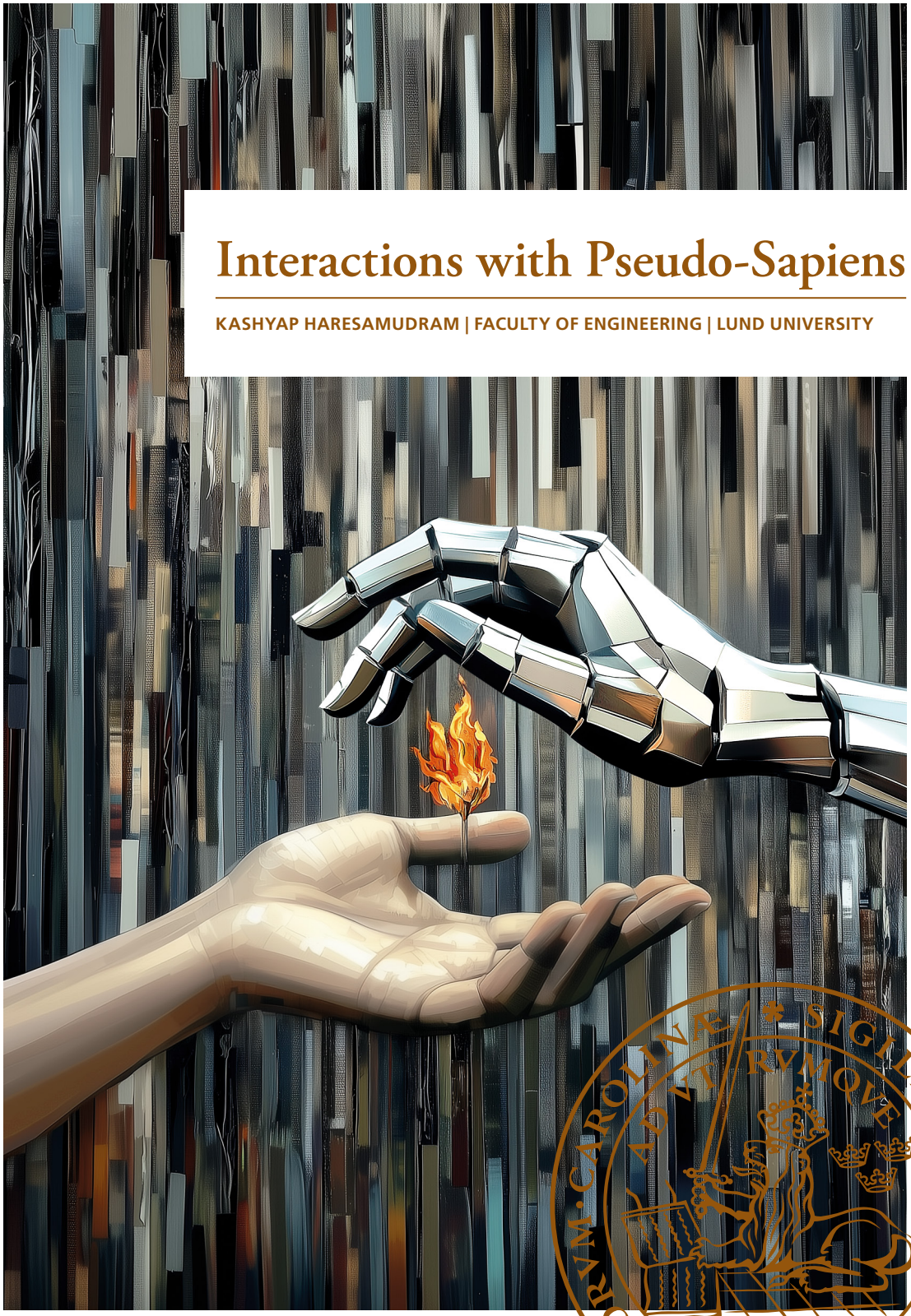
Haresamudram, Kashyap

2025

*Document Version:*
Publisher's PDF, also known as Version of record

# Interactions with Pseudo-Sapiens

KASHYAP HARESAMUDRAM | FACULTY OF ENGINEERING | LUND UNIVERSITY



Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

Faculty of Engineering
Department of Technology and Society

ISBN 978-91-8104-270-2
ISSN 1652-4810


LUND
UNIVERSITY

# Interactions with Pseudo-Sapiens

User perception of anthropomorphism, mind, and trust in humanlike social agents

Kashyap Haresamudram

LUND UNIVERSITY

## DOCTORAL DISSERTATION

Presented with the permission of the Faculty of Engineering, Lund University,
for public defence on Friday, the 14th of February 2025, at 09:00
in E:A, floor 1, E-huset, LTH, Lund University.

**Thesis advisors:**

Assoc. Prof. *Stefan Larsson*, Prof. *Fredrik Heintz*, Asst. Prof. *Ilaria Torre*

**Faculty opponent:**

Asst. Prof. *Minha Lee*

**Organisation:** Lund University
**Date of Issue:** 2025-02-14
**Document Name:** Doctoral Dissertation
**Sponsoring Organisation:** Wallenberg AI, Autonomous Systems and Software Program - Humanity and Society

**Author:** Kashyap Haresamudram

**Title and subtitle:** Interactions with Pseudo-Sapiens – User perception of anthropomorphism, mind, and trust in humanlike social agents

**Abstract:** Advancements in AI and Robotics have made it possible, at least to some extent, for technology to interact with humans in humanlike ways, such as being able to use natural language. Some of these technologies have rapidly overtaken the consumer market in the form of services such as ChatGPT, which in 2022 became the fastest growing user-base in history by acquiring 100 million users within two months of launching. Beyond chatbots, several consumer products such as personal assistant "smart" speakers like Amazon Alexa or Apple Siri and personal robots such as Amazon Astro have been available to consumers for some time now. The common thread between these products is their use of "humanlikeness" of their appearance or behaviour (or both) to facilitate interaction. Humanlikeness in technology design, in one sense, is not new, however, interaction with technologies that explicitly resemble or mimic humans is rapidly developing in ways that have previously been unachievable. Research in interaction with such technologies is essential to understand how these technologies impact humans in interaction and society at large. This thesis takes a user-centred focus on such technologies and examines interaction with humanlike social agents, with a focus on user perception of anthropomorphism, mind and trust in them. The thesis is comprised of a compilation of research articles that each examine interactions with different types of agents, such as robots, chatbots and voice assistants, particularly contrasting embodied agents with disembodied agents, and text-based agents with voice-based agents, in order to study the effect of humanlikeness on the perception of the agent in interaction. Employing video-based methods, the thesis finds that users may be less likely to form trust perceptions regarding an agent based on its humanlike physical or behavioural characteristics compared to its performance. Additionally, users broadly perceive agents to possess similar "mind" to one another irrespective of their physical or behavioural traits, with this "mind" being distinct from that of humans or other biological agents. The thesis advocates for further research on humanlikeness, collectively referring these agents as "Pseudo-Sapiens".

**Keywords:** human-agent interaction, social robotics, artificial intelligence, anthropomorphism, trust, mind perception, explainability, transparency, pseudo-sapiens

**Language:** English
**Number of pages:** 112
**ISBN (print):** 978-91-8104-270-2
**ISSN:** 1652-4810
**ISBN (electronic):** 978-91-8104-271-9
**ISRN:**



**Signature:**
**Date:** 2025-01-21

# Interactions with Pseudo-Sapiens

User perception of anthropomorphism, mind, and trust in humanlike social agents

## Attributions and Disclosures:

**Thesis Description:** A doctoral thesis in Sweden takes either the form of a single, cohesive research document (monograph thesis), or a summary of research articles (compilation thesis). In the latter case the thesis consists of two parts, (1) an introductory text (kappa) that puts the research into context and summarises the articles, and (2) the research articles themselves, along with a description of the individual contributions. The articles may already be published, or in various stages of development (accepted for publication, in review, submitted, or in draft). A completed thesis is publicly defended against a faculty opponent, and the defence is evaluated by a committee. This thesis is one such compilation thesis, defended against the following opponent and evaluation committee:

**Faculty Opponent:**
Asst. Prof. *Minha Lee*, Dept. of Industrial Design, Eindhoven University of Technology, NL

**Evaluation Committee:**
Prof. *Kerstin Fischer*, Dept. of Design, Media and Educational Science, University of Southern Denmark, DK
Prof. *Jonas Ivarsson*, Dept. of Applied Information Technology, University of Gothenburg, SE
Assoc. Prof. *Valentina Fantasia*, Dept. of Philosophy (Cognitive Science), Lund University, SE
Prof. *Christian Balkenius*, Dept. of Philosophy (Cognitive Science), Lund University, SE (suppleant)

**Cover Illustration:** Image created by Kashyap Haresamudram using *Midjourney* V.6.1.



**AI Disclosure:** Authorship of this thesis is attributed to Kashyap Haresamudram. It represents authentic and original research output. ChatGPT-4 has been utilized as a copy-editing tool to correct spelling, grammar, and punctuation on the completed manuscript.

**Funding Disclosure:** The work was funded by the Wallenberg AI, Autonomous Systems and Software Program - Humanity and Society (WASP-HS). WASP-HS is funded by the *Marianne and Marcus Wallenberg Foundation* and the *Marcus and Amalia Wallenberg Foundation*.


Faculty of Engineering, Department of Technology and Society, Lund University
**ISBN:** 978-91-8104-270-2 (print) | 978-91-8104-271-9 (electronic) | **ISSN:** 1652-4810
Printed by Media-Tryck, Lund University, Lund, Sweden 2025

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

*to Mummy and Pappa*

# Contents

# List of included publications

i **Tasks Over Traits: User perception of humanlike features in goal-oriented chatbots (in-review)**
Haresamudram, K., van As, N. and Larsson, S.
*N/A; Manuscript*

ii **Talking Body: the effect of body and voice anthropomorphism on perception of social agents (published)**
Haresamudram, K., Torre, I., Behling, M., Wagner, C. and Larsson, S.
*Frontiers in Robotics and AI, 11, 1456613*

iii **Mind and Body: dimensions of mind perception across agent types in human-agent interaction (accepted, forthcoming)**
Haresamudram, K.
*Social Robotics. ICSR+AI 2024. Lecture Notes in Computer Science, vol 15237*

iv **Three levels of AI transparency (published)**
Haresamudram, K., Larsson, S. and Heintz, F.
*Computer, 56(2), 93-100*

## Author contributions for included publications

i **Tasks Over Traits: User perception of humanlike features in goal-oriented chatbots (in-review)**

Haresamudram, K.: *Conceptualization, Project administration, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing*; van As, N.: *Conceptualization, Writing – review & editing*; Larsson, S.: *Writing – review & editing, Supervision, Funding acquisition*

ii **Talking Body: the effect of body and voice anthropomorphism on perception of social agents (published)**

Haresamudram, K.: *Conceptualization, Project administration, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing*; Torre, I.: *Writing – review & editing, Supervision*; Behling, M.: *Formal analysis, Visualization, Writing – review & editing*; Wagner, C.: *Formal analysis, Visualization, Writing – review & editing*; Larsson, S.: *Writing – review & editing, Supervision, Funding acquisition*

iii **Mind and Body: dimensions of mind perception across agent types in human-agent interaction (accepted, forthcoming)**

Haresamudram, K.: *Conceptualization, Project administration, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing*

iv **Three levels of AI transparency (published)**

Haresamudram, K.: *Conceptualization, Project administration, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing*; Larsson, S.: *Writing – review & editing, Supervision, Funding acquisition*; Heintz, F.: *Writing – review & editing, Supervision, Funding acquisition*

## Popular-science summary

The advent of AI and robotics has given rise to increasingly humanlike technology. Engaging with technology that can interact with humans in a humanlike manner, whether through natural language, voice, body language and gestures, or other forms that mimic human interactions, is a new phenomenon. These technologies have social capabilities and some limited agency to perform tasks independently, and can therefore be called social agents. How these AI-enabled social agents can affect human social interaction is not fully understood. It is therefore important to study their impact, starting with how they affect human perception. Whether people perceive social agents as (on some level) human, or as something else, has implications for the design and implementation of these technologies. Trust, transparency and explainability appear to be key factors in the ethical and beneficial development of social agents. In previous research, a feature of human perception called anthropomorphism has been linked to the human ability to understand even non-human attributes in human terms. Anthropomorphism has also been shown to influence trust. It is also significant for the perception of the existence of a mind in seemingly intelligent non-humans. However, whether people perceive different types of social agents differently, and how different humanlike features contribute to this perception, is unclear. This thesis studies how different humanlike features of three types of agents (chatbot, voice assistant speaker and robot) influence people's perceptions of anthropomorphism, mind and trust through a series of experiments. The results indicate that people have a distinct perception of agents in terms of mind, where they are seen as humanlike in terms of so-called task-oriented cognition, but significantly lower in terms of more complex human attributes such as emotional cognition and reflection. It is more positive for trust when agents complete tasks successfully compared to when they have human attributes. The perception of anthropomorphism is significantly influenced by human characteristics, and voice appears to be a very influential characteristic. The results led to the coining of the term *Pseudo-Sapiens*, which is found in the title of the thesis, to collectively refer to these anthropomorphic technologies.

## Populärvetenskaplig sammanfattning

Tillkomsten av AI och robotik har gett upphov till allt mer människoliknande teknologi. Teknik som kan interagera med människor på ett mänskligt sätt, oavsett om det är genom naturligt språk, röst, kroppsspråk och gester, eller andra former som efterliknar mänskliga interaktioner, är ett nytt fenomen. Dessa teknologier har sociala förmågor och en viss begränsad handlingskraft för att utföra uppgifter självständigt, och kan därför kallas sociala agenter. Hur AI-understödda sociala agenter kan påverka mänsklig social interaktion är inte helt förstått.. Det är därmed viktigt att studera vilken inverkan de har, med utgångspunkt i hur de påverkar människans perception. Huruvida människor uppfattar sociala agenter som (på någon nivå) mänskliga, eller som något annat har konsekvenser för utformningen och implementeringen av dessa teknologier. Förtroende, transparens och förklarbarhet framstår som nyckelfaktorer i etisk och fördelaktig utveckling av sociala agenter. I tidigare forskning har ett särdrag i mänsklig perception som kallas antropomorfism kopplats till människans förmåga att förstå även icke-mänskliga attribut i mänskliga termer. Antropomorfism har också visat sig påverka tillit. Det är också betydande för uppfattningen av existensen av ett sinne hos till synes intelligenta icke-människor. Huruvida människor uppfattar olika typer av sociala agenter olika, och hur olika människoliknande egenskaper bidrar till denna uppfattning, är dock oklart. I denna avhandling studeras hur olika människoliknande egenskaper hos tre typer av agenter (chatbot, röstassistent i högtalare och robot) påverkar människans uppfattning om antropomorfism, sinne och tillit genom en serie experiment. Resultaten indikerar att människor har en distinkt uppfattning om agenter i termer av sinne, där de är rankade som människolika när det gäller s.k. uppgiftsorienterad kognition, dvs. under ett styrt syfte, men betydligt lägre när det gäller mer komplexa mänskliga attribut som emotionell kognition och reflektion. Det är mer positivt för tilliten när agenterna slutför uppgifter framgångsrikt jämfört med att de har mänskliga attribut. Uppfattningen om antropomorfism påverkas avsevärt av mänskliga egenskaper, och rösten framstår som en mycket inflytelserik egenskap. Resultaten ledde till myntandet av begreppet *Pseudo-Sapiens*, som finns i avhandlingens titel, för att kollektivt referera till dessa antropomorfa teknologier.

# Acknowledgements

In life, I have been incredibly fortunate. Among the greatest evidence of this fortune is the guidance, mentorship, and friendship I have received from remarkable individuals over the past four years. I owe immense gratitude to Stefan Larsson, my main supervisor, for his unwavering support. From mobilizing resources to bring me to Sweden from India during the height of the pandemic lockdowns and closed borders, to fostering an environment that allowed me to explore my research interests freely, and for his patience with my proclivities and idiosyncrasies, Stefan has profoundly shaped me and my doctoral journey. I am equally indebted to my co-supervisors — Fredrik Heintz for his insightful advice, and Ilaria Torre for stepping in to guide me when I needed it most. My appreciation also extends to my former co-supervisor, Maliheh Ghajargar, whose early contributions helped shape my work before her departure abroad.

I have been privileged to be part of an exceptional team headed by Stefan, with Charlotte Högberg, Katarzyna Söderlund, James White, Laetitia Tanqueray, and Ellinor Blom Lussi — the AI and Society Group at the Department of Technology and Society, Lund University. Their kindness and encouragement have enriched my experience immeasurably. I am also deeply appreciative of the Fastighetsvetenskap division at LTH, which has been our team's home, and particularly to its head, Ingemar Bengtsson, for believing in me and offering his support. A special mention goes to Monika Baranowska, whose assistance with countless administrative requests and friendly conversations over the years have made a world of difference. I am also thankful to Ericka Johnson, head of the WASP-HS graduate school, and Eva Sjöstrand, WASP-HS admin, for their invaluable efforts in ensuring the graduate school and its network continue to flourish. My heartfelt gratitude also goes to Nena van As at Boost.ai, who has been both a colleague and a friend, contributing greatly with her knowledge and experience. To Christian Balkenius and André Pereira, thank you for your contribution through the midway and final seminars. I would be remiss not to acknowledge the many teachers, professors, and mentors throughout my schooling and college years who have shaped who I am today.

The opportunity provided by WASP-HS to learn from incredible researchers, build a diverse network, and form lasting friendships with brilliant individuals has been a true gift. These four years would have been infinitely more difficult without the solidarity of my fellow WASP-HS doctoral colleagues and dear friends: Amandus, Rachael, Pasko, Nika, Emelie, Mafalda, Silvia...and many more.

Equally, I am beyond grateful for those who created a sense of home for me in Sweden. To my flatmate Nathalie and her mother Monica, you are like family to me; your kindness and generosity know no bounds. To my previous flatmates Renos and Jesse, you made my early days in Sweden far easier and warmer than they might otherwise have been. And to my newfound circle of friends in Malmö — Daniel, Eugéne, Geoff, Jonathan, and Rebecka — you've made my life fuller.

I am deeply fortunate to have lifelong friendships with Aishvvarya, Dilip, Bhargav, and Arpita, who continue to be a source of strength and joy. Your presence in my life is a part of who I am and always will be.

Above all, I owe everything to my parents, Vijaya and Ramesh, whose unwavering support and unconditional love have been the bedrock upon which my life is built.

To everyone else who has touched my life in ways big or small, time long or short, I am truly grateful. Thank you.

Lastly, to Denmark and Sweden, where I found myself, I am profoundly indebted.

love and luck to all,
Kash

# Interactions with Pseudo-Sapiens

*"There is nothing to be known about anything except an initially large, and forever expandable, web of relations to other things. Everything that can serve as a term of relation can be dissolved into another set of relations, and so on for ever. There are, so to speak, relations all the way down, all the way up, and all the way out in every direction: you never reach something which is not just one more nexus of relations."*

- Richard Rorty, 1989

# Chapter 1

# Introduction

*The myth of Prometheus* from ancient Greece recounts the tale of the eponymous Titan, who created humans from clay and defied the Olympian gods by stealing fire (symbolising knowledge) from Mount Olympus to give to humanity. Zeus (King of the Olympian gods) had denied humans fire, fearing the empowerment and potential independence it would grant. Prometheus' gift of fire enlightened humans, enabling the development of technology, arts, and culture. As a punishment for the defiance, Zeus ordered Hephaestus (the god of fire and craftsmanship) to create the first human woman, Pandora, beautiful and charming, with curiosity and deceitfulness bestowed upon her by the Olympian gods. She was given a box containing all the evils of the world, such as sickness, death, and suffering, with instruction never to open it. However, bound by her preordained curiosity, she opened the box, unleashing all the suffering that would afflict humanity. The immortal Prometheus was bound to a mountainside, condemned to have his liver devoured by an eagle each day, only for it to regenerate every night, enduring torment for all eternity. Myths such as this have played an important role in shaping early ideas about the nature of existence, knowledge, creation, and the relationship between the creator and their creations. Prometheus' story has been recontextualised often, to reflect shifting zeitgeists of societies over time. Now, it has been reinterpreted again to reflect both the opportunities and the risks that artificial intelligence (AI) and robotics present, for humans as their creators, and humanity as a whole (Starmans, 2020).

Whether out of curiosity about our creation, or an innate desire to understand and reflect the human condition, we have seemingly always envisioned humanlike artificial entities. Stories of mechanical beings with human characteristics have existed since antiquity, long before modern technology. The Greek *myth of Talos* tells of a giant, bronze, often humanoid, automaton 1.1, created and animated by Hephaestus at the request of Zeus, to protect Europa (Zeus' consort) on Crete by patrolling the shores of the island and defending it from invaders (Iavazzo et al., 2014). As one of the earliest examples of a humanoid automaton[1], Talos exemplifies humanity's long-standing inclination to imagine artificial beings created in our own likeness.


**Figure 1.1:** Talos depicted on a silver didrachm from Phaistos, Crete (ca. 300/280-270 BC)

While the myth of Talos involves gods as the creators, stories where humans are portrayed as the creators of artificial beings have also existed long before such creations were considered feasible, even in theory. An ancient Chinese legend recounts the story of Yen Shih, an engineer who is said to have crafted a mechanical man that was capable of lifelike movements and even singing, that was designed

[1]A mechanical, self-operating machine, especially one designed to mimic human actions or perform tasks autonomously.

to entertain the emperor (LaGrandeur, 2012). Similarly, an ancient Indian story speaks of a serving girl made of wood that could move like a human, make eye contact, and serve wine – described as being real in every way except her inability to speak (LaGrandeur, 2012). Other ancient Indian texts speak of the concept of *yantra-purusha* or mechanical man, demonstrating early ideas about artificial humanoid beings (Iavazzo et al., 2014). These myths, legends and stories describe entities that we would recognise as humanoid social robots today, millennia before the term 'robot' was coined in 1921 by the Czech writer Karel Čapek in his play *R.U.R. (Rossum's Universal Robots)* (Christoforou and Müller, 2016).

Significant technological leaps have been made in the fields of AI and Robotics in the last 50-or-so years, getting us closer today to those early ideas of humanlike artificial entities than we have ever been before. While ancient stories may have had early ideas about what would eventually become humanoid social robots, today we are also capable of creating entirely digital entities, some with virtual bodies, and others with none at all, concepts that could never have been imagined even a century ago. These robots and digital entities, such as chatbots and voice assistants, are often autonomous, relatively intelligent, and designed for social interaction with humans – they are collectively referred to as *humanlike social agents*.

Several agents created over the past few decades exemplify the technological breakthroughs made in this area. In 1966, ELIZA, developed by Joseph Weizenbaum, became the first ever chatbot to simulate a conversation in *natural language*, and was highly influential in subsequent chatbot developments (Berry, 2023). Not long after, in 1973, Waseda University developed Wabot-1 (Hashimoto and Takanishi, 2015), the first full-scale humanoid robot (with limited autonomy) capable of *bipedal walking* on flat surfaces, *communicating* with humans, and *moving objects*. In 1995, ALICE (Artificial Linguistic Internet Computer Entity), developed by Richard Wallace, improved upon earlier chatbot systems using AIML (Artificial Intelligence Markup Language) to create more *dynamic, context-sensitive conversations* (AbuShawar and Atwell, 2015). And in 2000, ASIMO, a robot developed by Honda, capable of *walking, running, and interacting* with humans, represented a major leap in humanoid robotics by demonstrating its ability to *move fluidly* in human environments (Shigemi et al., 2018).

The late 2000's brought several consumer-oriented chatbots and voice assistants such as Apple Siri 1.2, Google Assistant, and Amazon Alexa, making social, *humanlike conversations* with technology more widespread. Increasing collaboration

between AI and Robotics led to the creation of robots that could *navigate environments, recognise objects and people, and even interact with humans* in meaningful ways. In 2017, Jibo launched as one of the first consumer-oriented robots to feature natural human-robot interaction, using its *expressive face* and *body movements* to engage with people. Jibo could *recognise faces, track people's movements, and hold conversations* (Rane et al., 2014). In 2019, Replika, an AI chatbot designed to serve as a personal companion, was released. It uses machine learning to *learn* from user interactions and develop a more *individualised, empathetic conversation style* (Possati, 2023). And in 2021, Ameca was developed by Engineered Arts. With highly *realistic facial expressions* and body movements, it can mirror human expressions and engage in *emotional, humanlike communication* (ElDiwiny, 2023). It is touted as one of the most advanced humanoid robots today.

**Figure 1.2:** Apple HomePod mini smart home speaker with Siri voice assistant.

This exceptional pace of advancement has resulted in humanlike social agents becoming increasingly viable in consumer-facing contexts today. The motivation for their development is rooted in the promised potential of these technologies, which are expected to increase productivity and efficiency (Johannsen et al., 2021; Al Naqbi et al., 2024), increase profits or reduce business costs (Cernetic, 2003; Heo et al., 2018), improve quality of services (Meskó et al., 2018; Akdim and

Casaló, 2023), alleviate human resource shortages in essential domains (Edwards and Cheok, 2018; Hurmuz et al., 2023; Carioli et al., 2024), and enhance overall human well-being (Ta et al., 2020; Jeong et al., 2023). In order to achieve this, several different implementations of these agents are being studied in various contexts. Broadly, chatbots, voice assistants, and social robots represent three of the most dominant types of humanlike social agents, stemming from significant investments in research and development, growing adoption, and a proliferation of consumer-oriented products and services.

With applications spanning customer service, e-commerce, and even mental health support, chatbots have emerged as a primary interface for businesses to interact with consumers. Technavio, a technology market research firm, states that the global chatbot market is expected to grow significantly, with a projected increase of USD 5.37 billion between 2023 and 2028, at a compound annual growth rate (CAGR) of 35.27% (Technavio, 2024). This growth is driven by an increasing demand for automated customer support, web self-service options, and improvements in customer relationship management (CRM). Apart from chatbots (which may also be voice enabled), voice assistants like Apple Siri, Google Assistant, and Amazon Alexa have become household staples with 75.7 million, 88.9 million, and 84.2 million users respectively according to the market research firm EMARKETER, who claim that 132.9 million users will engage with various voice assistants by the end of 2024 in the US alone (EMARKETER, 2024).

A lot of this growth in chatbots and voice assistants can be attributed to advancements in Conversational AI technologies, such as Natural Language Processing/Understanding (NLP/NLU), particularly in the form of Large Language Models (LLMs) and Generative AI. In November 2022, OpenAI launched *ChatGPT*, a Generative AI chatbot based on the GPT-3 Large Language Model (LLM), which became the fastest growing consumer application in history, acquiring over 100 million monthly active users within two months of launch (Reuters, 2023). At the time, ChatGPT was considered by many to be a revolutionary shift in the field owing to its ability to have highly natural, humanlike, coherent, and contextually relevant conversations (Jo, 2023).

In many ways, ChatGPT was the first easily accessible, AI-powered, consumer-facing technology that was *seemingly* both humanlike and intelligent. It sparked the latest wave of AI hype, an increased awareness of AI among the general public, and the subsequent development of a slew of competitors to ChatGPT, such as

Google's Bard – based on their LaMDA and PaLM language models, Anthropic's Claude, and Meta's open-source LLaMA model. ChatGPT's widespread adoption, along with its competitors, also marked a turning point in making humanlike interactions with AI a commonplace experience for everyday consumers. Unlike earlier chatbots, LLM-based chatbots excel at maintaining coherent, multi-turn dialogues, remembering context, and tailoring their tone and phrasing to suit the emotional tenor of the conversation. AI is also making voice assistants more viable with increasingly naturalistic voice integrations, evidenced by ChatGPT's recently added voice functionality. These technologies are perceived as being so humanlike that it has been suggested ChatGPT may be able to pass the Turing Test with naïve users (Gams and Kramar, 2024).

While AI has enabled software services in the form of chatbots and voice assistants, physical robots as consumer technologies are also growing fast. Non-humanlike service robots such as robotic vacuums and delivery bots are already becoming commonplace in homes. Anthropomorphic[2] companion robots such as Amazon's Astro, Anki's Vector, and Energize Lab's Eilik have been available for some time now. Robots are also poised to become more humanlike in areas such as companionship and healthcare, where research suggests humanlikeness may be important (Coghlan, 2022; Ahmed et al., 2024). Research Nester, another market research firm, claims that the social robot[3] market was valued at USD 5.58 billion in 2024, and projected to reach USD 128.65 billion by 2037, at a compound annual growth rate (CAGR) of 27.3% (Nester, 2024).

The push for humanlike consumer robots designed for social functions has already begun. In October 2024, Elon Musk announced that the *Tesla Optimus* humanoid robot 1.3 would go into limited production in 2025, and may be commercially available as early as 2026 (Techopedia, 2024). While there is general scepticism regarding the timeline, Tesla Optimus has captured the public imagination in a way that previous humanoid robots have not, largely due to the brand equity of Tesla and the cult personality of Elon Musk himself. Unlike earlier humanoid robots such as ASIMO or Ameca, which were often research projects or niche products, Tesla Optimus is being marketed as a potential mass-market, consumer-grade, affordable humanoid robot. Musk's vision for the Tesla robot includes not

[2]Chapter 2 elaborates on definitions and distinctions.

[3]Including humanlike and non-humanlike social robots.

only assisting in factories or helping with labour but also performing household tasks and potentially serving as personal assistants, suggesting a near-future where humanoid robots could integrate into people's homes, workplaces, and daily lives.

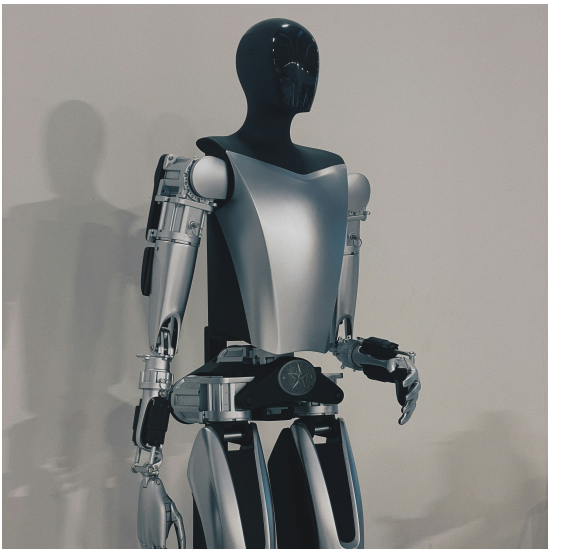


**Figure 1.3:** Tesla Optimus robot (Gen 1) at Tesla Center Verona, Verona, Italy.

Evidently, more and more individuals are going to come in contact with chatbots, voice assistants, and robots, in the not-so-distant future, if not already. Simultaneously, many of these technologies are likely to get more humanlike, for instance, the implementation of Conversational AI in voice assistants and social robots will enable them to have increasingly naturalistic conversations (Dong et al., 2023; Lykov and Tsetserukou, 2023). In this context, how humans perceive humanlikeness in these agents and whether and the role humanlikeness plays in interaction are highly consequential, both to inform design decisions, and to ensure ethical development of these technologies.

## 1.1 Research Frontier

While humanlike social agents are becoming increasingly commonplace in everyday life, research in human interaction with these agents still has several open questions that need to be studied in order to understand the dynamics of human-agent interaction. Firstly, while anthropomorphism – the phenomenon by which humans attribute humanlike traits to non-human entities such as technology (Epley et al., 2007) – is generally regarded as a facilitator of human-agent interaction, its nature and effects are not clearly understood. Regarding nature, Kühne and Peter (2023) elaborate on two issues:

> First, anthropomorphism has not been consistently distinguished from related psychological phenomena. Notably, it is not evident whether certain cognitions and behaviors present precursors, consequences, or constitutive elements of anthropomorphism. For instance, it is not clear whether perceptions of a robot's shape, movement, and behavior lead to the anthropomorphization of the robot or whether they are an aspect of anthropomorphization...Second, no consensus exists about the dimensional structure of anthropomorphism in HRI. Scholars disagree on whether anthropomorphism is best conceptualized as uni- or multidimensional, that is, whether anthropomorphism is a monolithic concept (i.e., unidimensional) or whether it has multiple facets, or dimensions, that constitute the overall concept (i.e., multidimensional) (Kühne and Peter, 2023).

Regarding effects, in a meta-analysis of the phenomenon in human-robot interaction research, Roesler et al. (2021) find that while anthropomorphism always has a positive effect on perception in social contexts, the effect does not transfer over to reciprocal interactional outcomes, and even that the positive correlation in perception in social interaction with robots may not be applicable in other contexts such as task-related settings. This leads them to question the usefulness of anthropomorphism in these domains. They write that:

> Whereas social HRI consistently benefits from anthropomorphic robot design, a mixed picture emerges for other application domains... Most of all, our results suggest that interaction quality between humans

> and robots can particularly be promoted by implementing anthropomorphic communication features, by multiple implementations of anthropomorphism, and by implementing task-relevant anthropomorphism. (Roesler et al., 2021).

They further mention a limitation of the meta analysis that did not consider degrees of anthropomorphism, as it is difficult to measure objectively – with most included empirical findings contrasting only two levels – noting it as an opportunity for future research. Similar themes are echoed by Li and Suh (2022) in a literature review of anthropomorphism in AI-enabled technologies (AIET), where they make the following (subset of) recommendations:

> Future studies can consider to separately or jointly measure the psychological, visual, and verbal aspects of anthropomorphism... In conceptualizing anthropomorphism as a technological stimulus, researchers should operationalize it considering not only the psychological features of AIET but also the visual and verbal features to specifically understand anthropomorphism and its consequences in the AIET context... visual, verbal, and psychological aspects may be considered three dimensions of anthropomorphism (Li and Suh, 2022).

Apart from anthropomorphism, trust is generally seen as another central concept that enables human-agent interaction. The relationship between trust and anthropomorphism, though broadly considered positive, remains unclear. Several studies suggest a positive effect of anthropomorphism on trust (Waytz et al., 2014; De Visser et al., 2016), some find a positive but indirect relationship (Chen and Park, 2021), some have linked trust specifically to the mind perception aspect of anthropomorphism (Mou et al., 2020), others argue anthropomorphism may lead to misjudged trustworthiness (Placani, 2024), or overtrust (Aroyo et al., 2021). Transparency mechanisms have been proposed as a means for trust calibration in other autonomous agent domains (Wang et al., 2021). Generally, due to being subject to contextual and individual differences, more studies on trust and anthropomorphism in varying contexts, agents, and individuals are needed to develop a deeper understanding of the relationship (the central concepts are each expanded on in Chapter 2 and 3). There is specifically a need for comparative studies across agent types. In terms of mind perception, Koban and Banks (2024) write that:

> Although existing research suggests that disembodied machine agents are often understood as remarkably similar to embodied agents, others have stressed that embodiment and corporeality have meaningful impact on people's perception, evaluation, and ascription. Until agent-comparative research addresses this issue, we call for caution when attempting to draw implications for disembodied machine agents (Koban and Banks, 2024).

The following research gaps can be distilled from the above: (1) studying the dimensional nature of anthropomorphism, specifically visual, verbal, and psychological aspects, (2) studying the effects of anthropomorphism in task-oriented settings, (3) studying the effect of levels of anthropomorphism, (4) studying the relationship between trust and anthropomorphism, and (5) comparing the perception of mind in different types of agents.

## 1.2 Aim and Scope of the Thesis

The thesis draws from four articles (Article I, II, III, and IV)[4] with *the overarching aim is to understand how the perception of humanlikeness—through the lenses of anthropomorphism, mind, and trust—varies across chatbots, voice assistants, and social robots*. The studies that the thesis draws from contribute, in different ways, to the research gaps outlined in the previous section, split between the following research questions.

### Research Questions

**RQ 1:** How do different humanlike attributes such as body, voice, personality, input method, and communication medium affect the extent to which users *anthropomorphise* chatbots, robots, and voice assistant speakers during interaction?

[4]The apparent discrepancy in the order of the papers and the research questions arises from the questions being informed by different papers and the conceptual framing of the thesis. In the thesis the progression follows a sequence from anthropomorphism (the phenomenon) to mind perception (a subset of anthropomorphism), and finally to trust (the outcome).

**RQ 2:** How does agent type – chatbot, voice assistant, social robot – affect *mind perception* in human-agent interaction?

**RQ 3:** How do different humanlike attributes such as body, voice, personality, input method, and communication medium influence user *trust* in chatbots, robots, and voice assistant speakers during interaction?

In informing the results, Article IV serves as a pre-study, outlining a theoretical framework for transparency in AI, and identifying 'interaction transparency' as an integral level of transparency. Article II compares the perception of trust and anthropomorphism between a voice assistant and humanoid robot, with four levels of humanlikeness of voice. It also measures two types of trust (human-trust and technology-trust), and is designed to highlight whether body and voice constitute two dimensions of anthropomorphism. Article I studies the effect of humanlikeness of personality, input method, and communication medium (voice) on anthropomorphism and trust. While personality and input method do not directly address the research gaps, communication medium is a comparative experiment, and together they serve the aim. Both the studies are conducted in the same task-oriented context, with nearly identical interaction possibilities, and similar methodology. Article III takes advantage of the study design by comparing mind perception between all three agent types. While other variables – animacy, likeability, perceived intelligence, and perceived safety – were measured in the former two studies, the scope of this thesis is narrowed specifically to anthropomorphism, mind perception, and trust. This is done for two reasons, (1) they are directly relevant to the research gaps and aim, and (2) they were the primary focus of the original studies even though other variables were measured. Findings related to other variables will be addressed briefly in the discussion.

## 1.3 Thesis Structure

The kappa (introductory text) contextualises and summarises the articles that are part of this thesis. Articles I, II and III address the research questions and make empirical contributions to HAI, while Article IV addresses the broader ethical aspects, specifically transparency, of AI and makes theoretical contributions to the field.

**Section 2:** Outlines the core concepts in HAI relevant to the research questions and defines specific uses of the concepts within this thesis.

**Section 3:** Outlines various theoretical positions on the three human propensities central to the thesis, highlights current empirical research on them within HAI, and identifies relevant theoretical perspectives that inform this thesis.

**Section 4:** Outlines methodologies used to conduct the studies presented in Articles I, II and III.

**Section 5:** Briefly overviews the results of the individual studies presented in Articles I, II and III.

**Section 6:** Synthesises the results from Article I, II, and III, within the context of the research questions outlined in the thesis, discusses broader implications including Article IV, and outlines future research avenues.

# Chapter 2

# Definitions

## 2.1 Humanlike, Social, and Agent: A New Paradigm

The label 'humanlike social agents' may be considered an oxymoron in lay terms: these technologies are, as of today, only marginally humanlike, can merely simulate social behaviour, and do not truly possess agency[1] – at least not in the same sense as human agency. This semantic scrutiny however, seemingly does not appear to align with the experiential reality of interacting with some[2] of these agents. Their humanlikeness, ability to simulate social behaviour, and apparent agency are convincing enough that we are beginning to witness the emergence of parasociality[3] (Maeda and Quan-Haase, 2024), a phenomenon that is likely to grow. We do not as yet fully understand the implications of such agents for human relationships and human social behaviour in general (Wu, 2024). While humans have long interacted with other agents, such as animals, and have for some time interacted

[1] "Sense of agency is the phenomenology associated with the responsibility we feel over voluntary actions and their effects" (Silver et al., 2021).

[2] AI-driven chatbots are leading the way, as elaborated previously.

[3] "It refers to an asymmetrical, one-sided relationship between individuals and media personalities, real/fictional characters, or celebrities... wherein the individual experiences a personal connection with the media figure despite having little-to-no interpersonal interactions with them... " (Maeda and Quan-Haase, 2024).

with human-centric technologies like computers, they have never thus far, outside of imagination, been able to engage with entities that not only possess human-like characteristics but are also capable of replicating human language, learning, reasoning, and complex problem solving.

Etymologically, the term agent[4] comes from the Latin word *agens*, derived from the verb *agere*, meaning "to do" or "to act"; the root of the word focuses on the concept of action. As the concept of agency evolved, the term agent began to imply not just someone who acts, but someone who chooses, decides, and exercises control in their actions (Schlosser, 2019). Swanepoel (2021) writes that, "In its simplest form, an agent is a thing which performs *intentional actions*. What constitutes an intentional action is, roughly, that the action is something an agent wishes or desires to do" (Swanepoel, 2021). Philosophically, much has been written about the nature of agency, the role of autonomy, and their relationship to the human experience (Emirbayer and Mische, 1998). Regarding technology however, the term has been defined based on functional criteria. With the advent of computers that could perform tasks autonomously, displaying some degree of agency, the term agent came to be applied in this context. Franklin and Graesser (1996) describe an autonomous agent as,

> A system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future (Franklin and Graesser, 1996).

Discussions continue about what it means to ascribe agency to technological artefacts (Swanepoel, 2021). Bradshaw (1997) writes that,

> 'Agent is that agent does' is a slogan that captures, albeit simplistically, the essence of the insight that agency cannot ultimately be characterized by listing a collection of attributes but rather consists fundamentally as an attribution on the part of some person (Van de Velde 1995)...This insight helps us understand why coming up with a once-and-for-all definition of agenthood is so difficult: one person's 'intelligent agent' is another person' s 'smart object'; and today's

[4]As per Merriam-Webster in December 2024

> 'smart object' is tomorrow's 'dumb program'. The key distinction is in our expectations and our point of view (Bradshaw, 1997).

Social agency in humans has been defined as a, "sense of agency when the voluntary action's effect is the direct or indirect reaction of a conspecific that we perceive as an independent agent" (Silver et al., 2021). Extrapolating from Franklin and Graesser's (1996) definition of an autonomous agent, and considering social agency in the context of an autonomous social agent (referred to henceforth simply as a social agent), we may define it as *'a system situated within and a part of a social environment that senses that environment and acts on it socially, over time, in pursuit of a shared agenda, and so as to influence both what it and others sense in the future'*. This definition emphasises that a social agent is not isolated but is embedded in a larger social context, is capable of perceiving or sensing (i.e., gathering data about) the environment (whether it be physical or digital), and interacting with other agents or individuals in the environment, over an extended period, through ongoing engagement and adaptation, towards shared common objectives with other agents or individuals, influencing not only its own behaviour and perception but also that of others in the environment. A *humanlike* social agent takes that further by incorporating human characteristics within the social agent.

It is helpful here to distinguish between the concepts of anthropomorphism and humanlikeness in relation to agents. *Anthropomorphism* refers to the human trait of attributing human characteristics, emotions, intentions, or behaviours to non-human entities (Epley et al., 2007). *Humanlikeness*[5][6] refers to an agents' attributes, and the degree to which these attributes – such as behaviours, appearance, interactions, and capabilities – resemble those of a human being (Law et al., 2022). In this sense, the two concepts are interconnected: anthropomorphism enables humanlikeness, and humanlikeness elicits anthropomorphism. In the literature, the terms 'anthropomorphic' and 'humanlike' are sometimes used interchangeably, but there is a critical distinction: 'anthropomorphic' implies qualities that an agent may or may not possess, depending on human perception, whereas 'humanlike' refers to qualities intentionally embedded in the agent by design. It is the latter concept – humanlikeness – that this thesis concerns itself with.

[5]Including its variants human-likeness, humanlike, and human-like.

[6]The term and its intended usage exist in literature but are often inconsistently applied.

## 2.2 Human-Agent Interaction: Minds Meet Machines

The concept of *interaction* is central to Human-Agent Interaction (HAI). In the field of Human-Computer Interaction (HCI), where it is similarly a "field-defining" concept, Hornbæk and Oulasvirta (2017) argued that the concept itself had remained largely undefined. In an effort to "move from everyday concepts to sharper, scientific concepts", they analysed different formulations across literature, and proposed the following definition:

> Interaction concerns two entities that determine each others' behaviour over time. In HCI, the entities are computers (ranging from input devices to systems) and humans (ranging from end-effectors to tool users). Their mutual determination can be of many types, including statistical, mechanical, and structural. But their causal relationship is teleologically determined: Users, with their goals and pursuits, are the ultimate metric of interaction. From here on, how researchers construe that determination influences the phenomena they can attend to, what they think good interaction is, and what tools they have to offer evaluation and design (Hornbæk and Oulasvirta, 2017).

This definition is largely applicable to HAI, where the entities are agents (instead of computers[7]) and humans. However, in the context of social agents, the concept of mutual determination may also take on a *relational* dimension. Relational determinants – such as communicative, emotional, and behavioural factors – become central to defining the interaction, reflecting the inherently *social* and *dynamic* nature of these agents.

Humans have shown a remarkable propensity for social interactions with non-human entities, particularly with technology. Reeves and Nass (1996) argued that humans do not instinctively distinguish between real-life interactions and mediated representations, and that they engage in *mindless* (instinctive or unreflective) and *social* responses when interacting with mediated representations that mimic human social characteristics (Reeves and Nass, 1996). From this foundation, the

[7]Agents can be considered a type of computer, but interacting with them differs significantly enough to warrant a separate categorisation and an expanded definition appropriate for this context.

CASA paradigm (Computers Are Social Actors) emerged, focusing more specifically on human interactions with technologies perceived to exhibit social cues and some level of agency (Nass et al., 1994; Nass and Moon, 2000). Apart from mindless behaviour, several studies have linked this phenomenon to anthropomorphism (Gong and Nass, 2007; Wang, 2017), however, the evidence is inconsistent regarding this explanation (Lee, 2010).

Numerous studies provide evidence supporting CASA. For instance, Nass, one of the theory's co-proposers, demonstrated through several studies that humans respond socially to computers exhibiting social cues. People rated computers labelled as teammates more positively than those without such labels (Nass et al., 1994), preferred computers that flattered them (Fogg and Nass, 1997), and applied gender stereotypes to computers with gendered voices (Nass et al., 1997). Beyond computers, CASA has found support across a wide range of technologies, such as chatbots (Adam et al., 2021), voice assistants (Schneider and Hagmann, 2022), robots (Kim et al., 2013), and autonomous vehicles (Waytz et al., 2014). Even subtle social cues in these technologies can elicit social responses from humans. However, there has been some critique of the consistency and reproducibility of the phenomenon (Schaumburg, 2001; Gambino et al., 2020).

While CASA paradigm serves as an explanation for humans' social responses to computers, the interactions studied under CASA are not necessarily complex (although they can be) – as the phenomenon is rooted in understanding human responses to subtle or minimal social cues. In terms of meaningful, complex interactions with humans, Krämer et al. (2012) identified perspective-talking, common ground, imputing one's knowledge to others, and theory of mind, as core aspects in human-human interaction and prerequisites for agents. It is theorised that such capabilities would by default elicit natural human social behaviour. Current technology, while being able to simulate some of these qualities, does not fully meet these requirements. Despite this, studies – such as with CASA – have found that humans seem to readily apply human interaction norms in their interactions with agents that display (or simulate) even some of these characteristics. As a result, it is hypothesised that human-agent interactions are likely to follow a human-human interaction (HHI) paradigm, especially as technology advances and more of these prerequisites are built into social agents (Krämer et al., 2012).

In human interaction, trust is often considered one of the most fundamental concepts. It plays a crucial role in various kinds of relationships from interpersonal

and group dynamics, to institutional and societal. It is often referred to as the "glue that holds society together" (Schilke et al., 2021). Baier (2014) writes,

> Trust, the phenomenon we are so familiar with that we scarcely notice its presence and its variety, is shown by us and responded to by us not only with intimates but with strangers, and even with declared enemies (Baier, 2014).

Human trust also extends to technology (Mcknight et al., 2011). Humans apply similar mechanisms of establishing trustworthiness in technology as they do in other humans, where reliability plays an important role (Mcknight et al., 2011). With regard to trust in technology, the concept of transparency is often seen as an antecedent of trust (Brunk et al., 2019), which likely informs assessments of reliability (although this connected has not explicitly been explored in literature and is an open question). As a consequence, trust and transparency are together (along with perceived risk) viewed as significant factors influencing technology acceptance (Märtins et al., 2022). Research on trust has historically had a broader and more multidisciplinary scope than transparency, which was primarily rooted in governance literature. However, with the advent of technology, particularly the internet, and now AI, the scope and implications of the concept have widened (Larsson and Heintz, 2020), and multidisciplinary frameworks are needed to address this (Weller, 2019).

Evidently, if the HHI paradigm should successfully apply in HAI, trust will play an equally central role in this context. Research on anthropomorphism and trust in HAI has been steadily growing; however, several open questions remain regarding how these complex and often debated phenomena influence human interactions with agents.

## 2.3 Key Human Propensities in HAI

Human perception of agents can have a significant impact on the subsequent interaction. Three related propensities – anthropomorphism, mind perception, and trust – have received considerable attention in research on human-agent interaction (Roesler et al., 2021; Li and Sung, 2021), and are considered central concepts.

**Anthropomorphism** in HAI enables humans to attribute humanlike qualities to agents such as chatbots and robots (Roesler et al., 2021). As a result, it can help facilitate interaction by enabling humans to use human-human interaction knowledge in human-agent interaction (Fink, 2012). The use of anthropomorphism as an intentional design tool to create humanlike agents results in pro-social responses from humans, such as cooperation and trust (Fink, 2012). At the same time it has been found that anthropomorphism may lead to negative consequences due to expectation mismatch caused by the association users make between the agent and a human (Złotowski et al., 2015). Making an agent too humanlike can trigger the uncanny valley phenomenon and cause negative emotions (Song and Shin, 2024). The role of anthropomorphism, within the context of HAI, is still actively debated, with varying conceptions, fragmented uses, and a lack of nuances about its dimensionality (Kühne and Peter, 2023), requiring further research.

**Mind perception** in humans is a phenomenon whereby humans perceive others to possess mental states (Waytz et al., 2014). In non-humans, such as social agents, anthropomorphism can lead to mind perception (Waytz et al., 2014). That is to say, perception of mind in non-humans is an aspect of anthropomorphism. Ascribing mind to agents does not however mean that they are seen as mindful beings (Koban and Banks, 2024). The perception of an explicit mind in agents has been shown to mitigate instances of robot abuse (Keijsers et al., 2022). Robots performing humanlike actions or depicting biological need have been perceived as possessing agency, which is a an aspect of mind perception. Anthropomorphic features such as face have been shown to also increase the perception of mind (Broadbent et al., 2013). Hortensius et al. (2021) note that, "In order to better understand the way humans attribute socialness and even form social relationships with non-human agents and objects, a better understanding of the role anthropomorphism and theory-of-mind play in these new interactions is warranted" (Hortensius et al., 2021).

**Trust** is also inherently linked to both anthropomorphism and mind perception (Epley et al., 2007). The physical appearance of an agent, especially robot, has an impact on trust (Hancock et al., 2011), with anthropomorphic robots being viewed as more trustworthy (Natarajan and Gombolay, 2020). When humans overestimate an agent's capacity and it under-delivers, they lose trust (Kwon et al., 2016). Different individuals have different propensities to trust in agents, just as with humans (Bernotat et al., 2021). Agents that are perceived as being transparent, providing explanations, or simply more information, are seen as more trustworthy

(Kok and Soh, 2020). Several researchers have called for additional research on anthropomorphism and trust in human-agent interaction over the years as trust dynamics with humanlike agents are not fully understood, particularly with regard to interaction context and individual differences (Blut et al., 2021; Sheehan et al., 2020; Kühne and Peter, 2023).

## 2.4 Key Social Agent Attributes in HAI

Agent attributes can be characterised in several ways. Broadly, Kim and Im (2023) identify two types of factors affecting social agents perception, morphological features and intelligence features. They categorise agent appearance, behaviour, and movement as morphological features, and cognitive and emotional intelligence as intelligence features (Kim and Im, 2023). These features could be slightly reorganised into appearance, behaviour (including movement), and intelligence (including cognitive and emotional intelligence) for simplification.

**Appearance** of a social agent can be very varied from embodied robot to disembodied chatbot. Appearance has a significant impact on shaping perception. It is the first feature of an agent that is noticed, and first impressions seem to matter (Bergmann et al., 2012). Generally, there is a large overlap between appearance and anthropomorphism. Anthropomorphic appearance can enhance perceptions of social presence, trust, and satisfaction (Chen et al., 2024). It can positively affect social presence (Letheren and Glavas, 2017). Appearance can shape expectations. Anthropomorphic robots are perceived as more humanlike, resulting in overestimated capabilities, whereas non-human-but-biological appearance may elicit realistic behavioural expectations (Haring et al., 2013). Generally, embodied agents are perceived as more humanlike, trustworthy, conscientious, agreeable, and intelligent when compared to disembodied agents (Carolus and Wienrich, 2022).

**Behaviour** is also closely associated with anthropomorphism. Humanlike agent behaviours generally seem to be preferred by users. Humanlike robot behaviour can lead to increased perceived animacy, improved emotional state, and self-disclosure (Rosenthal-von der Pütten et al., 2018). Extroversion and submissiveness are also strongly preferred by users (Mileounis et al., 2015). Humanlike language style improves perceived experience (Jenneboer, 2022). Functional intelligence, sincerity, and creativity empower consumers in voice assistants (Poushneh, 2021). Chatbot

personality has a positive effect on authenticity and intended engagement (Kuhail et al., 2022). Voice of a disembodied agent can still lead to anthropomorphism and perception of social partnership (Lee and Jeon, 2024). Seaborn et al. (2021) note that "There is great need and opportunity to consider how the voice of the machine...is actually perceived in relation to the intended effects and goals of the larger system, service, or experience" (Seaborn et al., 2021).

**Intelligence** in an agent as perceived by a user can significantly impact adoption and usage. Perceived intelligence in a robot influences perceived usefulness and trust (Tusseyeva et al., 2024). Perception, action, and learning dimensions of perceived intelligence significantly affect consumer adoption of voice assistants (Bawack, 2021). In chatbots, users may have difficulties perceiving elements of social intelligence (Mariacher et al., 2021). Making a chatbot more anthropomorphic improves perception of reliability, because it also improves perceived intelligence (Lee and Yoon, 2022). The perception of autonomy, adaptability, reactivity, multifunctionality, cooperativeness and humanlike interaction all have an impact on perceived intelligence of an agent (Tusseyeva et al., 2024). Zhao et al. (2024) note that, "...future work should further compare the impact of perceived intelligence on anthropomorphism across different types of software and hardware" (Zhao et al., 2024).

## 2.5 Concepts in Context

In this thesis, the key human propensities are measured across a range of key social agent attributes in the three studies.

**Artile I** primarily focuses on chatbot behaviour, particularly chatbot personality – where a chatbot using personal pronouns, friendly languge, and emojis (humanlike personality) is compared against a chatbot that does not use any of those social cues (less humanlike). Input method – where free text input is compared against button-based input, framed as different levels of humanlikeness, with free-text being more humanlike. And communication medium, where a text-only chatbot is compared to a text-based chatbot with voice output, and a voice-only voice assistant. Perception of anthropomorphism and trust are measured.

**Article II** focuses on both appearance and behaviour, where robot with a humanoid body/head is compared with a voice assistant speaker, each with four levels of humanlikeness of voice. Perception of anthropomorphism and trust are measured.

**Article III** where the chatbots from Article I are compared against the robot and voice assistant from Article II. Perception of mind are measured.

Intelligence is the only social agent characteristic that is not manipulated or measured in any of the studies, remaining constant across all agents.

# Chapter 3

# Theory

*Sensation* and *Perception* are key, synergistic mechanisms in cognition[1] that modulate our experiences. Sensation is a physiological process that involves the detection of stimuli through our sensory organs, providing the raw data necessary for our experiences, while perception is the psychological process that interprets and assigns meaning to this sensory information (Proctor and Proctor, 2012). In the case of light detection, the eye's photoreceptors are responsive only to wavelengths within the visible spectrum. As a result, we can physically detect only a narrow band of wavelengths, encompassing the colours we recognise as red, green, blue, and so forth. Generally, humans are recognised to possess five primary senses[2]: vision (sight), audition (hearing), olfaction (smell), gustation (taste), and somatosensation (touch). Each with their own limitations. These senses serve as the main channels for perception, providing the raw data that the brain interprets to form an understanding of the environment (Proctor and Proctor, 2012).

Perception, as a fundamental concept in metaphysics, plays a central role in human experience and meaning-making (Hoffman, 2019). It functions as a lens that colours our internal experience of an external reality[3]. Sensory limitation has pro-

[1]Different theories place varying levels of significance for each of these processes as seen below.

[2]While this is a widely accepted categorisation, other variations exist, including additional senses such as proprioception (body awareness) and nociception (pain sensation).

[3]Whether there is such a thing as an external reality, if it is objective, and whether such an

found implications for perception, as our understanding and interpretation of the world are intrinsically shaped by what our sensory systems can detect. Since our vision is limited to the visible spectrum, we experience the world as though it contains only the colours within that range, while ultraviolet or infrared wavelengths, though present, remain entirely outside our perceptual experience. Consequently, we perceive an environment defined by this sensory scope, as perception uses sensory data to create a coherent, actionable experience of reality. An important function of perception is to compensate for incomplete or varying sensory data in order to create the experience of a stable and consistent environment. This is the case with the colour purple. Since purple isn't found on the visible spectrum as a wavelength, it is *created* by the brain to fill in for combinations of light that don't fit traditional spectrum colours. This unique response allows us to perceive a wide range of stimuli, expanding our sensory and interpretive ability to navigate our environment.

Perception filters the vast array of various kinds of raw sensory input, selecting certain signals while disregarding others, to form the coherent impressions that guide our thoughts and actions. For instance, our photoreceptors constantly receive an endless stream of various wavelengths and intensities of light, despite this overwhelming amount of sensory input, we don't perceive every single ray of light individually or equally. Instead, our perception (coupled with attention) filters out irrelevant details, allowing us to focus on essential elements that produce meaningful experience. Attention acts as a filter within perception, selecting specific stimuli to focus on while downplaying or ignoring others. This results in, for example, us being able to hear a song, isolated from all other auditory data that is simultaneously received by our ears.

While perception constructs meaning by selectively focusing on particular details, the selection is also shaped by factors such as past experiences, beliefs, and even current moods (Leopold and Logothetis, 1999), serving as the mechanism that gives rise to subjectivity. Thus, perception does not passively mirror the world but actively interprets it. Consequently, each individual's perceptual experience is inherently unique and (as-yet[4]) inaccessible to others. This process is central to our ability to interpret complex environments, reinforcing that perception is not just

objective reality is accessible, are matters of ongoing philosophical debate - see discussion.

[4]Brain–computer interfaces (BCIs) could, in theory, make it possible to access in the future.

about seeing the world but about shaping an experience that is coherent, adaptive, and primed for interaction. Whether navigating a dark room, recognising emotions in others, or communicating with a digital assistant, perception enables us to filter and organise sensory information into actionable insights that guide our interactions.

The theoretical concept of perception has a rich history, and is widely accepted in scientific literature; however, its nature, origin, and functioning remain highly debated. Broadly, two opposing views have emerged, *representationalism* and *relationalism* (Nanay, 2015). Representationalism has been the normative view in contemporary mainstream psychology and cognitive science, and is a central concept in *cognitivism*, which emerged as part of the 'cognitive revolution' in psychology, marking a shift away from behaviourism's focus on observable behaviour towards understanding the internal processes of the mind (Miller, 2003). The influence of computational and information-processing models shaped cognitive psychology in the 20th century, and as a result cognitivist perspectives view the brain as an information-processing system, much like a computer (Glenberg et al., 2013). *The computational theory of mind* (CTM) holds that the brain processes sensory input like a computer processes data; just as a computer generates, decodes, stores, and manipulates digital representations in the form of binary code, the brain is thought to generate, decode, store, and manipulate *mental representations* based on sensory information. These representations are believed to serve as *mental models* that correspond to external reality (cf. Destéfano, 2021). This theoretical perspective generally views cognition as a hierarchy, where sensation and perception are distinct yet sequential processes: sensation comes first, followed by perception. Together, they are considered 'lower-order' processes, handling the initial stages of information processing that provide structured input for 'higher-order' cognitive functions.

Charles (2017) notes that most traditional approaches to perception overlook the fundamental question of how perception occurs, focusing instead on studying the organism's response to perceived objects and events without addressing the underlying mechanisms that enable perception itself (Charles, 2017). Evolutionary Psychology often falls short in this regard, primarily concentrating on behavioural adaptations without fully exploring the perceptual processes that underlie them, while Cognitive Psychology tends to treat perception and behaviour as separate, conceptually independent areas, often viewing perception merely as the initial step leading to higher cognitive functions (Charles, 2017). Researchers continue to

grapple with classic philosophical debates on how organisms 'know' the world, framing the problem as organisms forming 'internal representations' of the world from incomplete sensory input, reviving nature-versus-nurture debates and reinforcing mind-body dualism (Charles, 2017). As Charles (2017), echoing some of the general critiques of the mainstream theories of perception, states,

> There is little to no concern for understanding adaptation; little to no consideration of the structure of the environment; and the more 'perception' is removed from 'sensation', the closer it gets to imagination. The overall lack of evolutionary logic in traditional theories of perception is not totally surprising, as theories of perception developed for thousands of years before Darwin's time[5] (Charles, 2017).

Relationalism, in contrast, posits that knowledge, meaning, and perception are fundamentally shaped by the relationships between organisms and their environments (cf. Nanay, 2015). It may be argued that this perspective aligns with Darwinian evolutionary theory by proposing that perceptual abilities evolve within the context of specific ecological interactions (cf. Charles, 2017). The idea that cognition and perception emerge through interactions supports the view that these traits are adaptive responses to particular environmental conditions. Relationalism also rejects mind-body dualism, much like Darwin, who noted in his personal notebooks (1836–1844) that "experience shows the problem of the mind cannot be solved by attacking the citadel itself – the mind is a function of the body – we must bring some stable foundation to argue from" (Sheets-Johnstone, 2011, p.435). This view also tends to take an evolutionary perspective in addressing the experience of reality. Perception is not framed as a limited or partial window onto an objective reality; rather, it is seen as a way of engaging with and acting upon a world that is relevant and sufficient for human purposes, with the associated mechanisms and limitations having evolved to facilitate our needs.

*Ecological Psychology* and *Enactivism* are two of the leading relationalist paradigms, and are considered important perspectives within the broader framework of Embodied/Radical Embodied Cognitive Science. In the 1960s and 1970s, James J. Gibson developed Ecological Psychology in contrast to cognitivist theories, chal-

[5]Although cognitivism is a relatively new paradigm, it does not address evolutionary perspectives directly

lenging the view that perception is a constructive process (that reality is constructed through representations). Instead, he argued that perception is *direct* and grounded in the *affordances* offered by the environment (Dotov et al., 2012). An 'affordance' refers to the action possibilities provided by the environment relative to an organism (Dotov et al., 2012). According to Gibson, we perceive affordances directly, without the need to construct internal representations or process sensory data in the way that cognitivist theories suggest. In this view, sensation is not central to how perception is understood (Read and Szokolszky, 2020). Perception is more about directly detecting affordances – opportunities for action that the environment offers, like a surface that affords sitting or a handle that affords grasping. Rather than constructing meaning from sensations, perception involves immediately recognising these affordances in the environment, guiding action and interaction seamlessly. Enactivism, on the other hand, emerged in the 1990s, with foundational ideas articulated in the book *The Embodied Mind* (1991) by Francisco Varela, Evan Thompson, and Eleanor Rosch (Varela et al., 2017). This perspective proposes that cognition and perception are active, embodied processes arising from the continuous interaction between an organism and its environment (Ward et al., 2017). Enactivism places greater emphasis on the active role of the organism in 'bringing forth' its world, viewing perception as a participatory process where the organism co-creates its experience through action within an environment (Ward et al., 2017). In this view, sensorimotor capacities that enable action are central to how perception is understood (Read and Szokolszky, 2020). Enaction implies that (1) perception entails action guided by perception, and (2) cognitive structures arise from repeated sensorimotor patterns that support perceptually-guided action (Read and Szokolszky, 2020).

Proponents of representationalism generally put forth two main criticisms of relationalist perspectives (Chemero, 2009). Chemero (2009) summarises these as, "First, they can say that it will be impossible to explain truly cognitive phenomena without mental gymnastics (See, e.g., Clark and Toribio, 1994; Adams and Aizawa, 2008). Second, they can say that the models and theories used in radical embodied cognitive science actually do attribute representations to cognitive systems (Clark, 1997; Markman and Dietrich, 2000a,b; Wheeler, 2005)" (Chemero, 2009). These criticisms and their responses represent the leading edge of theoretical research in theory of mind and cognitive science in general.

## 3.1 Anthropomorphism Perception

Anthropomorphism is a fundamental phenomenon in human perception where humans attribute human characteristics, such as appearance, behaviours, emotions, and cognitive abilities, to non-human entities (Złotowski et al., 2015). Arguably, anthropomorphism can be seen as a dominant (in humans) subset of animism, which is the attribution of intentional action and life to non-living things such as objects, abstract concepts or phenomena (Airenti, 2018). Anthropomorphism manifests as a widespread behaviour, observed across cultures and contexts, through individuals projecting human traits onto animals, objects, and even abstract concepts (Złotowski et al., 2015). As detailed in the introduction, from ancient myths to modern narratives, anthropomorphism has been central to how humans create meaning and relate to the unknown. However, neither the mechanism by which the phenomenon occurs, not the extent to which it is prevalent, are fully understood.

### 3.1.1 Perspective on Anthropomorphism

From a cognitive science perspective, anthropomorphism is generally seen as a cognitive bias (or in some perspectives, a belief (Airenti, 2018)), and as a result it is framed as an error of perception or reasoning, because it involves projecting human traits onto entities that do not possess them in reality (Dacey, 2017). There are several simultaneous and overlapping phenomena at play in human perception that bias our perception to anthropomorphise. Firstly, as elaborated above regarding perception, humans are fundamentally predisposed to pattern recognition (Pi et al., 2008), a cognitive mechanism that evolved to help us quickly identify meaningful patterns from the stimuli we receive in order to comprehend, interact with, and navigate in an environment. Secondly, this tendency for pattern recognition is so strong in human cognition that sometimes it leads to the perception of meaningful patterns between unrelated or random stimuli, a phenomenon known as *apophenia* (Zhou and Meng, 2020). When this occurs in visual stimuli, it leads to the perception of familiar objects or entities within abstract or ambiguous shapes, like seeing animals, faces, or other recognisable forms in the shapes of clouds, rock formations, or shadows, known as *pareidolia* (Liu et al., 2014). Thirdly, a particularly strong subset of pareidolia involves the perception of faces, known as *facial pareidolia*. Humans are highly attuned to detecting faces, likely due to the evo-

lutionary importance of recognising and interpreting facial expressions for social communication and survival. For example, emoticons are interpreted as faces and some power outlets are perceived as faces too. Lastly, *own-species bias*, which refers to our heightened sensitivity to perceiving humanlike characteristics over those of other species (Scott and Fava, 2013), seems to play an important role. We are more likely to perceive something to be a human face as opposed to an animal, and we humanise even animal faces and expressions. Similarly, we ascribe even objects such as cars or robots humanlike emotional expressions based on their design, interpreting them as "happy," "sad," or "angry". Collectively, these phenomena may result in (or at least play a significant role in) anthropomorphic or humanlike perception of the environment.

Not all researchers agree with the framing of anthropomorphism as an error of perception. From a psychological perspective, Epley et al. (2008) note that,

> ...considering an inference anthropomorphic only when it is clearly a mistake is itself a mistake...People conceive of gods, gadgets, and an entire gaggle of nonhuman animals in humanlike terms. Although interesting, whether such inferences are accurate is orthogonal to a psychological understanding of the conditions under which people are likely to make an anthropomorphic inference (Epley et al., 2008).

Epley et al (2007) proposed one of the most substantial perspectives on anthropomorphism called 'A Three-Factor Theory of Anthropomorphism' elaborating on three psychological determinants for anthropomorphism, (1) Elicited agent knowledge, (2) Effectance, and (3) Sociality. Elicited agent knowledge pertains to their claim that knowledge about humans and oneself serves as the basis for induction about the attributes of an unknown agent, giving way to anthropomorphism. Primarily because such knowledge is readily accessible. As more knowledge is gained about the agent, it replaces the human or oneself as the induction basis. They write that, "anthropomorphism itself involves a generalisation from humans to nonhuman agents through a process of induction, and the same mental processes involved in thinking about other humans should also govern how people think about nonhuman agents" (Epley et al., 2007). Effectance and Sociality, they claim, are motivation driven. Effectance refers to the human need for effective interaction with non-human agents, understand their functioning, and make accurate predictions about them. To do this, they argue that humans are likely to

anthropomorphise an unknown agent as it is their only source of testable hypothesis about expected behaviour. Sociality refers to the human desire to make social connections, which may extend to anthropomorphised non-human agents.

An Embodied Cognition perspective on anthropomorphism has some superficial commonality with Epley et al's (2007) psychological account, in that both place the self as an important source of comparison, although they fundamentally differ on why. If cognition is highly embodied and situated, the process of attributing humanlike characteristics to non-human entities arises from the human tendency to interpret the world through their own embodied experiences, grounded in sensorimotor systems rather than abstract reasoning (Strongman, 2008).

Individual differences in anthropomorphism can have significant implication for both the human and the non-human entity in interaction. Waytz et al. (2010) demonstrate that individual difference in the tendency to anthropomorphise can predict the level of moral care and concern they show towards a non-human entity, the level of responsibility and trust they place on the entity, and how much the entity serves as a source of social influence on them (Waytz et al., 2010a).

Rejecting the notion of anthropomorphism as a belief, Airenti (2018) argues that it is instead grounded in interaction, where a non-human entity takes the place of a human interlocutor in interaction (Airenti, 2018), making it necessarily a social/relational phenomenon. She proposes that this perspective explains inconsistencies arising from doxastic perspectives of anthropomorphism where, adults may treat entities as if they have thoughts even when they know they do not actually have a mind, the same entity may be anthropomorphised or seen just as an object based on the situation, different entities may be anthropomorphised in different ways with no consistency, anthropomorphism is variable based on affective states rather than knowledge about an entity or naive of the human (Airenti, 2018).

Evidently, the nature and function of anthropomorphism are still debated. Li and Su (2022) in a literature review on anthropomorphism in AI-enabled technology find that across different studies anthropomorphism has been conceptualised as (1) a tendency, (2) a technological stimulus, (3) a perception, (4) a process, and (5) an inference, with the first three being the most widely used (Li and Suh, 2022). Furthermore, some research shows that the current theoretical explanations of anthropomorphism are rooted in Western ontological paradigms, overlooking cultural and individual differences (Spatola et al., 2022). This highlights a need for

more comprehensive and inclusive research to better understand the complexities of anthropomorphism across diverse contexts.

### 3.1.2 Dimensions of Anthropomorphism

Not only does the current understanding of anthropomorphism need conceptual development in terms of its breadth, as evidenced by the lack of engagement with cultural and individual differences, but it also needs development in terms of depth, as anthropomorphism is often conceptualised as a unidimensional phenomenon. The limited existing literature on the dimensions of anthropomorphism come from Human-Robot interaction studies, particularly from a need to introduce nuance to how anthropomorphism and perception of humanlikeness are measured in empirical studies.

Złotowski et al. (2014) employ the orthogonal concepts of anthropomorphisation and dehumanisation to develop two-dimensional measures to distinguish between different approaches to enhancing a robot's humanlike perception. Combining a model of dehumanisation differentiating between two distinct senses of humanness, uniquely human and human nature (where denying uniquely human characteristics leads to animal-like perception and denying human nature characteristics leads to machine-like perception), and dimensions of mind perception of human and non-human agents, they show empirical evidence for anthropomorphism as a multidimensional phenomenon (Złotowski et al., 2014).

While not directly referring to anthropomorphism or its dimensions, von Zitzewitz et al. (2013) introduce 'Parameters of Human Likeness', comprised of two parameters, appearance and behaviour, each comprised of five parameter fields; visual appearance, sound, smell, haptic appearance, and taste, for the appearance parameter, and movement, interactive behaviour, social behaviour, verbal communication, and nonverbal communication, for the behaviour parameter (von Zitzewitz et al., 2013). These parameters have been referred to in the context of anthropomorphism in subsequent research (Wagner and Schramm-Klein, 2019). Seeing as humanlikeness is a property of the entity, while anthropomorphism is a cognitive process in the observer that enables perception of humanlikeness, it may be argued that the parameters might reflect a similar mechanism in anthropomorphism that is enabling this perception distinctly across the parameters.

## 3.2 Mind Perception

Mind perception, Theory of Mind (ToM) and Anthropomorphism are related concepts. Mind perception involves recognising that another entity has a mind with the capacity for thought, emotions, and intentions (Waytz et al., 2010b). ToM builds on mind perception – once an entity is recognised as having a mind, ToM enables us to infer specific mental states like beliefs, desires, and intentions (Waytz et al., 2010b). Mind perception is an integral aspect of anthropomorphism which involves actively attributing humanlike characteristics, including mental states, to non-human entities (Złotowski et al., 2015). As such, it has been argued that anthropomorphism can be seen an extending of ToM to non-human entities (Atherton and Cross, 2018). The three concepts are integral to social interaction. Mind perception is the starting point of social interaction, allowing us to recognise that others have mental states, and enabling us to treat them as intentional agents. ToM is crucial for interpreting their thoughts, beliefs, and intentions, in order to predict their behaviour, which allows for meaningful and coordinated social interactions (Epley et al., 2010). And anthropomorphism extends these human traits to non-human entities, enabling humans to interact socially with pets, robots, or even abstract phenomena (Hortensius et al., 2021).

### 3.2.1 Mind Perception in Theory

Several evolutionary accounts for mind perception have been proposed over the years, Epley and Waytz (2010) highlight the following: (1) That the bias towards humanlike mental states favours identifying intention agents even if one isn't present as the cost of not identifying one when it is present is greater for reproductive fitness (Guthrie, 1995). (2) that attributing humanlike mental states to non humans provides a useful analogy to reason about the natural and the artificial necessary for survival (Mithen, 1996). (3) that mind perception evolved to maximise the primary drivers of natural selection, survival and sexual reproduction, as the ability to infer others' minds would increase the likelihood of both (Nichols and Stich, 2003). And (4) that mind perception facilitates survival in large groups and societies (Herrmann et al., 2007).

Epley and Waytz (2010) conducted an extensive review of state of art in mind perception, and highlighted several key themes. They highlight that mind perception

is important as it facilitates three key functions in humans, (1) the ability to infer mental states such as beliefs, desires, and intentions enables explanations for others' actions and makes them more comprehensible and meaningful, (2) reasoning about the mental states of others enables one to align knowledge, overcome linguistic ambiguity, and communicate effectively, and (3) mind reading enables one to reason about relational and temporal aspects which aids in coordination. They overview the two main theoretical positions on mind perception aim to explain how it works, theory theory and simulation theories. Theory theory comprises inferential theories – that people reason about others' minds using a theory about how minds work to make inferences irrespective of their own perspective. Theory theory includes observations such as – understanding that other people have thoughts and feelings different from our own, that we are separate individuals from others, and that people's actions might sometimes be misleading and not reflect what they truly believe. Three types of findings are cited as evidence for theory theory, (1) weak or ambiguous introspective signals lead to people reasoning about own internal state using same theoretical inference they use to reason about others, (2) Adults' capacity to understand the outcomes of others holding false beliefs, and (3) relying on individuating or categorical information instead of egocentric simulations when reasoning about different others. Simulation theories on the other hand hold that people makes inferences about others' mind based off of their own by simulating and reasoning with themselves. Four types of findings are cited as evidence for simulation theories, (1) mind perception exhibits systematic biases consistent with a simulation mechanism, (2) simulations can also be revealed in the features that are absent in the outputs of mind perception, (3) people are more often than not egocentric about other' mental states, (4) people appear to simulate in their whole bodies, not just minds (Epley et al., 2010). Epley and Waytz (2010) note that hybrid models have since tried to reconcile the two positions, and propose that people use both types of strategies based on context, cognitive cost, and motivation. They also show that the processes that allow individuals to comprehend minds of other individuals seem to work in similar ways for understanding different kinds of minds, such as those of other people, past and future versions of oneself, and non-human entities like animals, technological agents, or supernatural beings. Lastly, they underscore that despite extensive research on mind perception, there is still no comprehensive account of the phenomenon (Epley et al., 2010).

The discovery of the so called mirror neurons has brought fresh interest in mind perception. Unlike other neurons, these neurons respond both when action is per-

formed as well as observed, forming a cortical system matching observation and execution of goal-related motor actions (Gallese and Goldman, 1998). Experimental evidence suggests that these neurons may be involved in the perception of mind in others by way of responding to and recreating the observations of others, giving weight to simulation theories; although theory theory explanations have also been proposed (Epley et al., 2010).

The predominant perspective on mind perception is representationalist. Embodied Cognition challenges 'mindreading' (inferring mental states of others) with two main oppositions, (1) challenging the developmental perspective on mindreading, and (2) challenging the notion of ubiquity of mindreading, however, this opposition is refuted (Spaulding, 2010). Embodied Cognition proposes the mirror neuron theory, specifically pointing to mirror neurons firing not only during activation but also during observation, being present in an area of the brain associated with motor function, and they carry details specific to the motor modality, however, empirical evidence does not seem to support this argument (Caramazza et al., 2014).

### 3.2.2 Dimensions of Mind Perception

Mind, for a long time, was conceived as a unidimensional phenomenon. In their seminal paper, Gray et al. (2007) presented a survey study conducted with 2040 respondents, for 78 pairwise ratings of 13 characters on 18 mental capacities, showing that mind perception is a two-dimensional phenomenon (Gray et al., 2007). The 13 characters they employed were baby, chimp, dead woman, dog, fetus, frog, girl, god, man, permanent vegetative state man, robot, woman, and you (the respondent). Their results, derived through factor analysis explaining 97% of the variance, divided the 18 mental capacities into two factors Experience and Agency. Experience accounted for 88% of the variance, and was comprised 11 capacities – hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy. Agency accounted for 8% of the variance, and was comprised of 7 capacities – self-control, morality, memory, emotion recognition, planning, communication, and thought. The two dimensions captured different aspects of morality, relating to Aristotle's moral agents (Agency) and moral patients (Experience). The results showed that respondents perceived babies to have little to no experience and low agency, God to have extremely high agency but little to no ex-

perience, all living adult characters to have both high experience and agency (Gray et al., 2007).

Since then, several studies have replicated these results and some have expanded on the study. Malle (2019) similarly four studies with a wider range of mental capacities, different question types, and many evaluated agents. The Results consistently found three dimensions of perceived mind: Affect (A), Moral and Mental Regulation (M), and Reality Interaction (R). Rather than grouping similar features, the dimensions reflected the psychological functions of the mind – interacting with own processes, with other minds, and with the social and physical world. (Malle, 2019).

## 3.3 Anthropomorphism and Mind in Social Agents

Several studies have found that people anthropomorphise several different characteristics of social agents, with some leading to positive and others leading to negative effects on perception of the agent. People also attribute mind to agents to varying degrees.

*Anthropomorphism as a Driver of Positive User Perceptions and Usage Intentions:* Multiple studies demonstrated that anthropomorphic cues enhanced positive attitudes, trust, usage intentions, and overall acceptance of AI agents. For voice assistants, Li and Sung (2021) found that anthropomorphism improved evaluations, positive attitudes, and satisfaction (Li and Sung, 2021). Mishra et al. (2022) showed that anthropomorphism shaped utilitarian attitudes, influencing usage and word-of-mouth recommendations (Mishra et al., 2022). Similarly, Blut et al. (2021) reported that in service robots, anthropomorphism strongly increased customer intention to use a robot (Blut et al., 2021). In chatbots, Konya-Baumback et al. (2023) demonstrated that anthropomorphism positively influenced trust, purchase intention, word of mouth, and shopping satisfaction (Konya-Baumbach et al., 2023). Sheehan et al. (2020) found that a consumer's need for human interaction correlated with stronger anthropomorphism and adoption intent (Sheehan et al., 2020).

*Enhancing Social Presence, Interaction, and Emotional Connection:* Anthropomorphism frequently emerged as a key factor in fostering social presence, intimacy,

and emotional bonds. In voice assistants, Fernandes and Oliveira (2021) showed that perceived social presence, influenced by humanlike attributes, was a driver of acceptance (Fernandes and Oliveira, 2021), while Aw et al. (2022) found that perceived anthropomorphism predicted parasocial interactions (Aw et al., 2022). In robots, Spatola et al. (2020) observed that social presence effects were mediated by anthropomorphism (Nichols and Stich, 2003). For chatbots, Christoforakos et al. (2021) showed that interaction duration and intensity predicted social connectedness, mediated by perceived anthropomorphism and social presence (Christoforakos et al., 2021), and Lee et al. (2023) found that chatbots expressing humanlike emotions increased willingness to donate, mediated by anthropomorphism and social presence (Lee et al., 2023).

*Influence on Perceptions of Warmth, Competence, and Service Quality:* Anthropomorphic features shaped perceptions of warmth, competence, and related attributes across technologies. For voice assistants, Wienrich et al. (2022) found that humanlike visualisations increased perceptions of anthropomorphism and humanlike characteristics (Wienrich et al., 2022). In robots, Yoganathan (2021) reported that anthropomorphising service robots increased warmth/competence inferences (Yoganathan et al., 2021), while Belanche et al. (2021) noted that humanlikeness positively affected service value expectations, including utilitarian and relational dimensions (Belanche et al., 2021). In chatbots, Pizzi et al. (2023) observed that competence perceptions, influenced by anthropomorphism, reduced consumer skepticism (Pizzi et al., 2023), and Roy and Naidoo (2021) found that preference for warm or competent chatbot conversations depended on consumers' temporal orientation (Roy and Naidoo, 2021).

*Mediators, Moderators, and Functional Outcomes:* Anthropomorphism operated through various mediators and was influenced by moderators. In voice assistants, Yu et al. (2024) identified social presence, performance expectancy, and customer value as mediators between cuteness and usage intention, with perceived risk moderating these effects (Yu et al., 2024). Calahorra (2024) highlighted humanity embedded in the voice and perceived safety as key factors influencing voice shopping acceptance (Calahorra-Candao and Martín-de Hoyos, 2024). For robots, Blut et al. (2021) found that animacy, intelligence, likeability, safety, and social presence served as mediators, while robot type and service type were moderators (Blut et al., 2021). For chatbots, Lee and Toon (2022) showed that anthropomorphism increased chatbot reliability, mediated by chatbot intelligence and moderated by individual need for human interaction (Lee and Yoon, 2022), and Pentina et al.

(2023) identified AI anthropomorphism and authenticity as antecedents leading to AI attachment, with social interaction as a mediator and usage motivations as a moderator (Pentina et al., 2023).

*Complex or Negative Responses to Anthropomorphism:* Some studies noted that anthropomorphism could lead to mixed or negative reactions. For voice assistants, Hsu and Lee (2023) found that more humanlike traits increased trust and enjoyment, without explicit mention of negative effects (Hsu and Lee, 2023). However, in robots, Akdim et al. (2021) observed that realistic robots could evoke negative attitudes (Akdim et al., 2023), and Mende et al. (2019) found that humanoid service robots elicited discomfort and compensatory consumption (Mende et al., 2019). For chatbots, Schanke et al. (2021) noted that while anthropomorphism improved transaction outcomes, it also heightened offer sensitivity and fostered a fairness evaluation mindset (Schanke et al., 2021). Araujo (2018) found that humanlike language reduced mindless anthropomorphism for machine-like agents, but did not enhance social presence (Araujo, 2018).

*Context-Dependent Effects and Personal User Characteristics:* Studies indicated that the impact of anthropomorphism varied depending on situational factors and individual differences. For voice assistants, Wienrich et al. (2023) showed that assigning social roles influenced empathy and enjoyment (Wienrich et al., 2023). In robots, Lu et al. (2021) demonstrated that a humanlike voice and language style influenced service outcomes through emotion and cognition (Lu et al., 2021), while Zhang et al. (2021) reported that different robot appearances affected performance expectancy, positive emotions, and effort expectancy, moderated by humour (Zhang et al., 2021). In chatbots, Lee et al. (2023) showed effects on donation willingness contingent on the chatbot's expression of humanlike emotions (Lee et al., 2023), and Araujo (2018) indicated that adopting an intelligent frame mattered for reducing mindless anthropomorphism (Araujo, 2018).

*Mind Perception in Shaping Social and Emotional Support:* Perceiving a mind in chatbots and robots increased feelings of closeness, helpfulness, and social support. Lee et al. (2020) found that the more participants perceived a mind behind the chatbot, the more co-presence and interpersonal closeness they experienced (Lee et al., 2020). Similarly, Lee and Hahn (2024) showed that explicitly perceiving a humanlike mind in the chatbot made its support more helpful for resolving stressful events (Lee and Hahn, 2024). Alimardani and Qurashi (2020) shows that elderly participants attributed higher mind perception scores to the robot and

treated it more as a human social partner, and a significant positive correlation emerged between mind perception and attitude in both the elderly and young adult groups (Alimardani and Qurashi, 2020).

*Mind Perception and the Attribution of Agency and Intentionality:* Mind perception influences human responses to the actions and decisions of artificial agents. Saltik et al. (2021) found that human-specific actions by a robot led to higher agency ratings compared to manipulative actions, though the robot's appearance had no effect on agency or experience scores (Saltik et al., 2021). Lee et al. (2021) showed that manipulating machines' mind dimensions influenced human responses differently in simple economic games compared to more complex negotiations, indicating that agents were perceived not only as social actors but also as intentional actors (Lee et al., 2021).

## 3.4 Trust Perception

Trust is the foundation of human social behaviour (Kumar et al., 2020). It has often been conceptualised as an emergent property of social life (Robbins, 2016). The benefits of social cooperation enabled by trust outweigh the risks, which in evolutionary terms is cited as reason for its development in human social behaviour (Clément, 2020). Trust is a concept that, on the surface, is seemingly experienced and intuitively understood by most individuals simply through lived experience. Despite this, trust remains an elusive concept to systematically define, with fragmented literature across various disciplines, and no consensus on its origin or function (Robbins, 2016). Simpson (2012) writes that, "there is a strong prima facie case for supposing that there is no single phenomenon that 'trust' refers to, nor that our folk concept has determinate rules of use" (Simpson, 2012). Part of what makes trust challenging to define is its pervasive role across every level of human social interaction—individual, group, institutional, and societal (Christov-Moore et al., 2022). Furthermore, trust can be conceptualised in diverse ways depending on psychological, sociological, political, economic, and philosophical perspectives within any given context (Hupcey et al., 2001; Hudson, 2004). Nevertheless, attempts at a unified concept of trust continue to be made, as there is a lack of consensus even on the notion that trust cannot be conceptualised as a singular phenomenon.

### 3.4.1 Nature of Trust

Broadly, existing perspectives on trust can be categorised into two opposing accounts, doxastic and non-doxastic. *Doxastic* accounts hold that trust involves a form of *belief* held by the truster, either that the trustee is trustworthy or that the trustee will do as they are trusted to do; *non-doxastic* accounts reject the notion that trust necessarily involves belief, and hold instead that it involves a *mental attitude* which may, according to different perspectives, involve a moral, dispositional, or emotional/affective component (Keren, 2014). The two accounts of trust have significantly different implications for understanding the rationality, functioning, and value of trust. Karen (2020) writes that,

> if trust just is a belief, then we should be able to derive the conditions for the rationality of trust from the epistemological study of rational belief. Evidential considerations would have primary place in the evaluation of the rationality of trust even if trust is not a belief, but merely entails one. In contrast, if we accept a non-doxastic account of trust, then evidence should be no more central for the justification of trust than ethical or instrumental reasons (Keren, 2020).

### 3.4.2 Trust, Reliance and Trustworthiness

Trust is often regarded as a species of reliance. Characterised as a supposition (which may or may not be a belief) to act on. Reliance can be understood as – (for a 'rely-er') to rely on a 'rely-ee' to X is (for the 'rely-er') to act on the supposition that the 'rely-ee' will indeed X (Goldberg, 2020). The encapsulated interest account of trust proposed by Hardin (1993) incorporates this construct and holds that: A truster trusts a trustee to X only when the truster relies on the trustee to X, and does so on the basis of the belief that the trustee will X because the trustee's incentives regarding whether to X encapsulate the truster's relevant interests (Hardin, 1993). While an important definition in highlighting the role of reliance in trust, the definition has been critiqued for its lack of a moral dimension, not accounting for mismatches in reliance-trust attributions, and not incorporating nuances in reactive attitudes of the truster in the event of misplaced trust. Jones (2004) building on her own previous work, in turn built on Annette Baier's earlier work, proposes that: Trust is accepted vulnerability to the trustee's power over something

the truster cares about, where (1) the truster forgoes (in the moment) searching for ways to reduce such vulnerability, and (2) the truster maintains normative expectations of the trustee that they not use that power to harm what is entrusted (Jones, 2004).

Most definitions of trust are predicated on a relationship between trust, trustworthiness, and reliance (often including vulnerability), broadly along the lines that the truster is vulnerable or reliant on the trustee and trusts them to act in their interest. Trustworthiness, in turn, involves the trustee acting in ways that safeguard the truster's vulnerability or justify their reliance. Trust thus involves an implicit ethical responsibility for the trustee to protect the truster's interests and not exploit their reliance, and trustworthiness is established through upholding this responsibility. As a result, the relationship between trust and trustworthiness is often characterised as a normative one, governed by ethical and rational expectations, where each influences what is right or reasonable for the other party to do – the demonstrated trustworthiness of a trustee creates a moral or rational basis for the truster to extend trust, and conversely, if the truster fails to trust a trustee that has clearly demonstrated trustworthiness, this refusal may be regarded as an injustice (Scheman, 2020). It has been suggested that the normative nature of the relationship leads the truster to judge themselves in the event of unwisely misplaced trust or unjustly refused trust (Scheman, 2020). Rather than characterise trustworthiness from the point of view of the trustor as is typical, Jones (2012) characterises it from the point of view of the trustee as the willingness and ability of the trustee to signal to the truster the domains in which the truster can rely on them (Jones, 2012). This conception specifically highlight the role of the trustee's competence in trust, rationalising the truster's reliance.

### 3.4.3 Factors of Trustworthiness

A truster's evaluation of a trustee's trustworthiness is a crucial step in establishing trust. Through an extensive literature review, Mayer et al. (1995) identify three characteristics of the trustee, *ability* (sometimes referred to as competence), *benevolence*, and *integrity*, as being crucial in the evaluations of trustworthiness by the truster (Mayer, 1995). McKnight et al. (2002) arrive at the same three characteristics as being important, and as generally encompassing/overlapping other alternative conceptualisations, based on a categorisation of trusting beliefs in 32 trust

articles/books, and provide empirical evidence for this conceptualisation (McKnight et al., 2002). The ability/competence, benevolence, and integrity grouping has ben used extensively to design measurement instruments for trust and trustworthiness (Lankton et al., 2015).

**Ability**, also referred to as competence, refers to the truster's perception/belief that the skills, knowledge, and characteristics possessed by a trustee within a given domain make them capable of fulfilling an expected role or task effectively (Mayer, 1995). The domain-specificity means that a trustee proficient in one domain may not inspire trust in a different one Bhattacherjee (2002). Perceptions of ability can be shaped by prior experience, endorsements, or institutional validation, such as credentials or reputations for quality and innovation. This factor emphasises that trust is directed toward areas where the trustee demonstrates relevant aptitude and reliability.

**Benevolence** refers to the truster's perception/belief that a trustee genuinely desires to act in the best interest of the truster, independent of self-serving motives or profit (Cazier, 2007). It involves the trustee's willingness to invest time, effort, and resources to help the truster, driven by altruistic intentions Bhattacherjee (2002). Benevolence is often crucial in emotionally charged or relational decisions, where the truster values the trustee's concern for their well-being over technical competence alone (Cazier, 2007). This factor is emphasises that selflessness, emotional investment, and genuine care are important considerations for trust.

**Integrity** refers to the truster's perception/belief that the trustee adheres to a set of principles that are not only consistent but also acceptable to the truster (Cazier, 2007). It is about alignment between what the trustee says, thinks, and does, ensuring that actions follow through on promises. Integrity involves values such as fairness, honesty, reliability, and dependability Bhattacherjee (2002). This factor is emphasises that trusters rely on the belief that a trustee will act in a morally consistent manner, and their perception of integrity is shaped by the alignment between the trustee's stated intentions and their actual behaviour(Mayer, 1995).

These factors are interdependent. A trustee might demonstrate competence but lack integrity, undermining trust; conversely, benevolence without competence can inspire goodwill but may fail to establish effective reliance. Together, competence, benevolence, and integrity provide a framework for understanding and fostering trust across personal, professional, and institutional relationships.

### 3.4.4 Trust: Psychology, Sociology, and Cognitive Science

The psychological account of trust highlights unconscious processes underlying social exchanges; this account holds that trust is the result of an unconscious resulting in a certain attitude (Clément, 2020). The human brain uses various fast and frugal heuristic mechanisms, that are implicit and unconscious, to evaluate the trustworthiness of a potential trusee. These evaluations trigger affective responses that guide behaviour. For example, neutral faces of people that are evaluated as trustworthy are perceived as happier, and those evaluated as less trustworthy are perceived as angrier (Clément et al., 2013). In-group/out-group phenomenon also seems to play a significant role in quick evaluations of trustworthiness, with a positive bias for same group members and a negative bias for other groups (Tanis and Postmes, 2005). Language plays a significant role in acquiring information through testimonials and observations that may contribute to epistemic trust (Quine and Ullian, 1978). Studies show that many of these heuristics are already present at humans at a very young age. Taken together, Clément (2020) writes that, "trust can be understood as the intimate resonance of an evaluation process that is most often left opaque to us. With time and the development of a theory of mind i.e. the ability to reflexively represent one's mental states and those of others, more reflexive evaluations can take place" (Clément, 2020).

A distinguishing feature of sociological accounts of trust is that they emphasise the relational aspects of trust, characterising not just the individual actors involved but also the state of a relationship – enabling the application of such an account to understand trust relationships even of a non-interpersonal nature, such as in groups, organisations, institutions, and society (Cook and Santana, 2020). Developing on Hardin's (1993) encapsulated interest account of trust described earlier, Cook et al. (2005) develop a relational account, claiming that, "trust exists when one party to the relation believes the other party has incentive to ask in his or her interest or to take his or her interest to heart" (Cook et al., 2005). Trustworthiness is determined through judgements made by the truster regarding the competence and integrity of a trustee in a given context, and in more complex networks, social capital, reputation, and status play an important role in determining trustworthiness. (Cook and Santana, 2020). Internet-based platforms, being anonymous and decentralised, abstract the traditional methods of assessing trustworthiness, instead reputational mechanisms such as ratings and reviews may need to serve as signals of reliability (Diekmann et al., 2014). At a macro-level in institutions and soci-

eties, Zucker (1986) identifies three modes of trust production, (1) process-based trust tied to the history of previous or expected exchange, (2) characteristic-based trust tied to the characteristics of an individual such as personal values or background, and (3) institutional-based trust tied to formal structures and practices that support coordination (Zucker, 1986).

Cognitive Science interacts with various disciplines such as psychology, neuroscience, linguistics, computer science, and philosophy. As a result, it offers a multidisciplinary perspective on trust. Drawing from various sources, Castelfranchi and Falcone (2020) propose a doxastic account of trust under four attributes:

1. **Dispositional:** where trust is an attitude by the truster towards the world or agents/trustees, and the attitude has two characteristics:
    - The attitude is hybrid with both affective and cognitive components
    - The attitude is composite, based on the truster's beliefs, expectations, and evaluations, plus different dimensions of the trustee's qualities, content of the trustee's action, and dimensions of external conditions
2. **A mental and pragmatic process:** where the process of trust has several steps and is a multilayered process resulting in the establishment of a social relation, conceptualised on (at least) three levels:
    - The decision to trust the trustee
    - The intention formation to trust
    - The act of trusting the trustee
3. **Multilayered and recursive:** if the truster trusts the trustee based on information, belief, signs etc., then the truster must also trust the source of such information, which entails trusting in the information, beliefs, signs, etc., about those source itself
4. **Dynamic:** trust can change and evolve over time and have complex relationships with other trusted sources, where:
    - Trust evaluations, decisions, and relations are subject to change based on new insight
    - Trust can derive from trust

In their attempt at a unified conception of trust, they stress that trust is complex, and requires a non-reductive definition which accounts for interpersonal as well as broader (such as group and institutional) notions of trust (Castelfranchi and Falcone, 2020).

### 3.4.5 Trust in Technology

Trust in technology (or technology trust) shares many of the characteristics of trust in humans. The fundamental premise of the truster being vulnerable and relying on the trustee, and the assessment of trustworthiness in the trustee operate similarly, only the trustee in this case is a technology (Mcknight et al., 2011). Researchers have argued that the more humanlike a technology is, the more human characteristics of trust apply (Lankton et al., 2015). In the context of dyads that constitute a primary decision maker, comparing a human 'advisor' or an intelligent automated decision support system, Madhavan and Wiegmann (2004) claim that when both the human and the automated aid behave similarly and signal similar reliability, the process of trust development is comparable, however cognitive biases of the user may produce different verbal assessments of trust and distrust. They write that, "the process of trust development in a decision aid is primarily a function of the cognitive and psychological biases and response tendencies of the user" (Madhavan and Wiegmann, 2004). Similar to trust in humans, an assessment of trustworthiness is made about the technology before placing trust. McKnight (2011) proposes three factors analogous to the factors of trustworthiness, functionality, helpfulness, and predictability, corresponding to ability, benevolence, and integrity respectively (Mcknight et al., 2011).

**Functionality** refers to the truster's perception/belief that the technology has the features, mechanisms, or functions, within a given domain, that enable it to fulfil an expected role or utility effectively.

**Helpfulness** refers to the truster's perception/belief that the technology is able to help them when needed, through prompts, guides, and other means, in an adequate, effective, and responsive manner, in order to complete a task.

**Predictability**, also referred to as reliability, refers to the truster's perception/belief that the technology is able to function smoothly and consistently in a manner that is expected, with no unforeseen breakdowns.

This conceptualisation has been validated (Lankton and McKnight, 2011), and extended in other research on technology trust (Lankton et al., 2015).

### 3.4.6 Transparency and Explainability

Transparency in itself is not a human factor, however, there is an overlap between transparency and trust. It is fundamental to informational notions of trust, where trust is an informed assessment. Conceptually, transparency is closely linked to accountability and openness, which naturally lend themselves to discourse on governance aspects. Bentham's 'panopticon' refers to transparency through the idea that watching the actions of individuals leads to 'correct' or 'norm compliant' behaviour, and as a consequence he argues transparency in government would prevent 'conspiracy'. Rousseau advocated for transparency, viewing opaqueness as evil, and argued that civil servants should act transparently to avoid destabilising intrigues (Meijer, 2014). Our modern day understanding of transparency is rooted in these earlier notions, with Sweden becoming the first nation to introduce access to information legislation in 1766. Several countries, particularly western-style democracies, have since followed Sweden's lead later in the 20th century (Ekkilä 2012). Business and management studies carried these notions into practice outside of governance context through organisational management and business practice, giving rise to various thoughts relating to corporate accountability, stakeholder management and, public relations. Today, transparency is a normative concept, important not only in law and governance, but also in business studies, communication studies, economics, political science, and increasingly in AI and Computer Science. Transparency is often considered a central concept in 'trustworthy AI'. The European Commission High-Level Expert Group on AI highlight transparency as one of the 7 key facets in their Ethics Guidelines for Trustworthy AI (HLEG AI, 2019). This view is also echoed in other AI ethics frameworks; one study found that the concept of transparency was represented in 84 different AI ethics guidelines around the world (Jobin et al, 2019).

Transparency is inherently linked to explainability in the context of AI. Explanation is a natural part of human conversation. It is how we convey causal relationships between phenomena (Lipton 2001). Explanations facilitate alignment and cooperation. They are seen as an important communication mechanism that help build trust. With the advent of machine learning and neural networks (tech-

niques that underpin most AI today) it became possible to create large, complex algorithms that were able to produce highly complex results, including (but not limited to) the ability to communicate (NLP) and make decisions (decision-assist systems). These algorithms became increasingly good, and more efficient, at performing complex and sensitive tasks such as cancer identification and law enforcement. However, the complexity of the algorithms meant that these were 'black-box models' whose decision-making was inaccessible even to its creators (Miller, 2019). Several early cases of biased AI algorithms, such as the Amazon HR algorithm (Lewis, 2018), and the US recidivism algorithm (Brennan et al, 2009), highlighted the need for AI to explain its decisions. This gave rise to the concept of explainability in AI, which formalised into Explainable Artificial Intelligence (XAI). However, XAI leaned more heavily into its roots in computer science and approached explainability from an algorithmic and computational standpoint, focusing on mathematical or statistical causal relationships. These explanations were neither intuitive nor accessible to non-experts. Critics of XAI have pointed out that existing research on human explanations within the social sciences has largely been ignored by the early efforts in XAI, and that this knowledge should be used to build more human-compatible explanations (Miller, 2019). At the same time, from a user point of view, there is no clear consensus on what constitutes a good explanation from an AI/social agent. Interaction research on explanations is a newly emerging area, and the complexity of interaction modalities, the situated nature of interactions, and variability of explanation possibilities makes it particularly challenging to study. Precisely how trust relates to transparency and explanations, how they operate in humans, what elements apply to trust in technology, and particularly in social agents that are capable of interacting in human language and producing human behaviours, is far from clear.

## 3.5 Perception of Trust in Humanlike Social Agents

*Antecedents and Influences on Trust:* Goodman and Mayhorn (2023) found that participants trusted female-voiced assistants more, with pitch and gender influencing trust, but individual differences accounted for most variance (Goodman and Mayhorn, 2023). Hsu and Lee (2023) showed that voice assistants exhibiting humanlike linguistic traits and positive behaviour traits increased trust (Hsu and Lee, 2023). Malodia et al. (2023) found that convenience and status-seeking enhanced trust, while risk perceptions reduced it (Malodia et al., 2023), and Malo-

dia et al. (2024) showed that social identity and personification related strongly to usefulness and playfulness, with trust moderating the relationship between usefulness and usage (Malodia et al., 2024). Al Shamsi et al. (2022) indicated that trust and perceived ease of use positively influenced perceived usefulness, with trust in technology affecting perceived ease of use (Al Shamsi et al., 2022). Wienrich et al. (2021) demonstrated that perceiving a voice assistant as a specialist increased trustworthiness ratings (Wienrich et al., 2021). Cheng et al. (2022) found that perceived warmth and competence increased trust in chatbots, while communication delay reduced it (Cheng et al., 2022), and Baek and Kim (2023) showed that while personalisation improved trust, increased creepiness lowered continuance intention (Baek and Kim, 2023). Mostafa and Kasamani (2020) found that compatibility, perceived ease of use, and social influence increased initial trust in chatbots (Mostafa and Kasamani, 2022). Tussyadiah et al. (2020) reported that cognitive trust formation in intelligent robots depended on negative attitudes and a propensity to trust technology (Tussyadiah et al., 2020), and Ullman and Malle (2019) found that manipulating robot roles and contexts influenced perceptions on trust subscales, highlighting reliable-capable and ethical-sincere dimensions (Ullman and Malle, 2019).

*Trust Development and Variability Over Time and Context:* Skjuve et al. (2021) found that trust developed gradually as human-chatbot relationships evolved from superficial curiosity into deeper affective engagements, positively influencing users' perceived wellbeing (Skjuve et al., 2021). Alarcon et al. (2021) observed that distrust behaviours over time reduced trustworthiness perceptions, trust intentions, and trust behaviours, and that trust violations by anthropomorphised robots did not differ meaningfully from those by humans (Alarcon et al., 2021).

*Outcomes and Consequences of Trust:* Song and Shin (2024) showed that making a chatbot more humanlike increased eeriness, thereby lowering trust, purchase intentions, and willingness to reuse (Song and Shin, 2024). Park et al. (2024) found that humanlike chatbot representation enhanced compliance with mental health recommendations through trust (Park et al., 2024). Cheng et al. (2022) noted that greater trust in chatbots reduced consumers' intention to switch to a human agent (Cheng et al., 2022), while Malodia et al. (2023) showed that trust increased the intention to use virtual assistants for transactional services (Malodia et al., 2023). Baek and Kim (2023) revealed that improved trust through personalisation and task efficiency contributed to continuance intention (Baek and Kim, 2023).

## 3.6 Theory Selection and Use

The studies employ specific theories by way of the implicit theoretical perspectives inherent in the measurement instruments used. These remain largely implicit but can be elaborated as follows:

**Anthropomorphism** is measured using the Godspeed anthropomorphism questionnaire (Bartneck et al., 2009a). The questionnaire was developed based off of Powers and Kiesler's (2006) work in the paper 'The advisor robot: tracing people's mental model from a robot's physical attributes' (Powers and Kiesler, 2006). Inherent in the use of mental models here is a representationalist view of cognition. By consequence, Epley et al's. (2007) psychological conception of anthropomorphism, and Epley and Waytz's (2010) conception of **mind perception** (Epley et al., 2010) are adopted as they are compatible with this view of cognition, and with each other. However, embodied cognition perspectives are also used in the discussion.

**Trust and reliability** are measured using the trust scales developed by Lankton et al. (2015), which are based on the Ability, Benevolence, and Integrity framework in humans, and the analogous Functionality, Helpfulness and Predictability framework in technology (Lankton et al., 2015). This makes them doxastic accounts of trust. The cognitive science account of trust by Castelfranchi and Falcone (2020) is adopted as it is both doxastic and compatible with the perspectives on cognition, anthropomorphism, and mind as discussed above (Castelfranchi and Falcone, 2020).

# Chapter 4

# Methods

Methodological choices pose significant trade-off's in human-agent interaction research. Human experiences are highly complex, dynamic, and context-dependent, which makes them difficult to study. Phenomena like perception, particularly of anthropomorphism, mind and trust pertinent to this thesis are inherently subjective, and as seen in the previous section, are themselves active topics of research with no standardised definitions or universally accepted frameworks. This makes them difficult to measure consistently, as the available tools and methods, which are often only indirect measures of perceptive experience, are limited by current knowledge. Standardising the variables risks reducing the richness of human experience, however, allowing variability reduces the generalisability of the findings. This trade-off leads researchers to choose between breadth and depth in their approach, which often presents as a choice between quantitative or qualitative methods. Quantitative methods offer objectivity and generalisability, but they may oversimplify complex human experiences. Qualitative methods provide rich, detailed insights into individual experiences, but they are harder to generalise and replicate. Experimental methods offer greater control by enabling researchers to isolate distinct factors, such as appearance or specific behavioural characteristics of an agent, resulting in more reliable data on cause-and-effect relationships, but these experiments sacrifice ecological validity since these interactions likely differ in real-world settings. Naturalistic studies in real-life settings can yield more realistic data but at the expense of control over variables, making it harder to specify the factors that drive observed behaviours or attitudes in humans.

While AI and Robotics have made significant advances in recent years, current technology available to use as research objects are often unreliable and inconsistent in their performance. For example, robots frequently fail during operation. Additionally, the research questions in this thesis necessitate using certain characteristics, such as – natural language, personality and humanlike voice – that were not possible to reproduce through technology at the time of the study design[1]. As a result, a controlled experimental setting was deemed most suitable in this instance. The Wizard of Oz method, that has been used in similar circumstances, allows researchers to simulate a part, or whole, of the agent as well as the interaction, through real or mock computer interfaces (Maulsby et al., 1993). In this thesis, simulating interactions through videos allowed for greater control of the various agents being studied, with the added benefit of maintaining consistency across interactions. As a result, a video-based, simulated interaction was chosen as the method. This method naturally allowed the possibility of online data collection through surveys, resulting in a series of quantitative online experiments. Three studies were conducted in order to examine user perception of anthropomorphism, mind and trust across chatbots, voice assistant speakers, and robots with different levels of humanlike characteristics.

Study I, henceforth referred to as the 'Chatbot study' (corresponding to Article I), examines whether chatbot personality, input method (button and text), and communication medium (text and voice), affect user perception of trust, trusting intention, reliability, and the Godspeed series[2] which includes anthropomorphism, animacy, likeability, perceived intelligence and perceived safety (Bartneck et al., 2009b). Study II, henceforth referred to as the 'Robot-VA study' (corresponding to Article II), examines whether agent embodiment[3] and levels of humanlikeness of voice have an effect on user perception of human-trust, technology-trust, and Godspeed series in robots and voice assistant speakers. Study III, henceforth referred to as the 'Mind Perception study', (corresponding to Article III) examines mind perception in chatbots, voice assistant speakers, and robots, in contrast to biological beings. Table 4.1 provides an overview of all three study designs.

[1]Naturalistic interactions with chatbots rapidly evolved during this time with the release of ChatGPT, however integrations were relatively new, voice was not yet a feature, and standardising these elements across three agents was (and still remains) a challenge.

[2]Only the anthropomorphism scale is in focus here as explained in the thesis scope.

[3]referring to agent body rather than interaction.

**Table 4.1:** Experimental design and methods employed for Study I, II, and III.

| | Chatbot study | Robot-VA study | Mind Percp. study |
|---|---|---|---|
| **Agents** | Chatbot | Robot & Voice Asst. Speaker | Chatbot, robot & Voice Asst. Speaker |
| **Setting** | Online | Online | Online |
| **Format** | Vignette | Vignette | Vignette |
| **Interaction** | Direct interaction & Video simulation | Video simulation | Video simulation |
| **Design** | Randomised controlled trial | 2x4 factorial analysis | Exploratory factor analysis |
| **Structure** | Between-subjects | Within & Between-subjects | Between-subjects |
| **Measurement** | Likert-scale survey | Likert-scale survey | Likert-scale survey |
| **Analysis** | Non-parametric | Parametric | Non-parametric |

## 4.1 Agents

Several simulated interaction videos of chatbots, voice assistant speakers, and robots were created for the three studies. The agents used in the videos remained consistent across the studies (for example the same voice assistant speaker and voice was used in both the Chatbot study as well as the Robot-VA study).

The chatbot study had 3 direct-interaction chatbots, 2 chatbot interaction videos, and a voice assistant interaction video (as described in Article I). The 3 direct-interaction chatbots were created using the boost.ai platform. A screen recording of one of these chatbots was created for the video conditions. The AI-driven text-to-speech (TTS) generator NaturalRedaer (www.naturalreaders.com) was used to generate audio that was edited onto the video to create the simulated interaction. Lastly, the same smart speaker condition from the robot-VA study (below), using the same generated audio, was used for the smart speaker condition.

The Robot-VA study had 8 videos, 4 of each agent (robot and voice assistant speaker), using 4 different levels of humanness of voice. The Epi robot from LUCS, Lund University (see Figure 4.1) was used to create a robot interaction

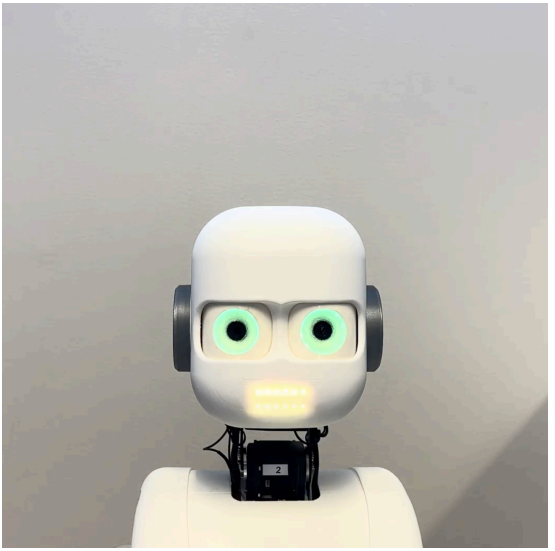


**Figure 4.1:** The Epi robot platform at LUCS, Lund University, Sweden.

video. A generic speaker (see Figure 4.2) was used for the voice assistant speaker interaction video. The lights in the robot and speaker were manipulated to convey speaking and listening. Audio for the four voices was generated based on a script using Apple MacOS' built-in TTS generator, NaturalRedaer AI TTS generator (www.naturalreaders.com), a real human voice, and a modified version of the real human voice using the audio editing software Audacity (www.audacityteam.org). These formed the four levels of humanness of voice[4] (as described in Article II) that were then edited onto the two videos of agents to create the final eight conditions.

## 4.2 Setting and Format

All three studies were conducted online. Sunet Surevy (www.sunet.se) was used to create the surveys and collect the data. Participants were recruited on Prolific (www.prolific.com), and were paid the average hourly rate.

[4]Perception of four distinct levels was confirmed in a pre-test as explained in Article II.

It is important to handle interaction context with care when studying human-agent interaction because it defines the purpose of the interaction, and influences how humans expect to communicate, significantly shaping the dynamics, expectations, and outcomes of the interaction. In order to derive meaningful insights across the three studies, the interaction context was maintained as a constant, with the interaction taking place in a service context, where the agent was a travel assistant/planner. The 'travel assistant' context was chosen in order to make the interaction less high-stakes, such as banking or insurance, where privacy concerns may hinder free interaction.

To elicit this context in the online setting, all three studies utilized a 'vignette' (described in Articles I and II). A vignette is a brief, descriptive scenario that outlines a specific situation, designed to immerse participants in a particular real-world context relevant to the research objectives. Vignettes have been used in similar studies (Nørskov et al., 2020; Law et al., 2021). The use of vignettes is particularly effective in survey-based research because it bridges the gap between hypothetical questions and real-world application, ensuring that participants' responses are aligned with the practical, situational nuances of the behaviours or perceptions being studied.

For the video-based experiments (robots, smart speakers), participants were asked to watch a video of a user interaction where a user plans a trip using a travel assistant agent that they were told was capable of helping plan and make all reservations pertaining to a trip to Aarhus, Denmark. For the direct-interaction experiments (chatbot), participants were asked to imagine that they were planning a vacation to Aarhus, Denmark, and that the chatbot was a prototype travel assistant for the city that could help plan all aspects of their trip. They were provided with specific instructions for tasks to complete, such as booking a hotel, recreational activities, and transportation (the same tasks performed in the videos). Interactions in both types of experiments were simulated as per a pre-written script in order to maintain consistency, thus combining the Wizard of Oz method with Vignettes.

## 4.3 Interaction

Two methods of testing interaction are employed in different studies. Primarily, a video-based interaction method is used, where a video recording of an interaction between a user and an agent is shown to the participants. The video is recorded in a way such that only the agent is in the frame, and the user is seemingly off the screen. For this method the survey is modified so the participants answer the question in a hypothetical direct interaction based on the video they watched. In two of the chatbot experiments (personality and input method), direct interaction method is used where the participants directly interact with a chatbot to perform certain tasks (described below).

**Figure 4.2:** The voice assistant speaker.

## 4.4 Design and Structure

The chatbot study is comprised of 3 experiments. The first experiment is a direct-interaction, between-subjects, randomised controlled trial between two chatbots

with chatbot behaviour (personality) as the independent variable, where a chatbot with and without personality are compared. The second experiment is a direct-interaction, between-subjects, randomised controlled trial between two chatbots with input method as the independent variable, where button-based input and free text input are compared. The third experiment is a video-based, between-subjects, randomised control trial between a chatbot with voice output and a voice assistant speaker, with communication medium being the independent variable. In this study Trust, trusting intention, anthropomorphism, animacy, likeability, perceived intelligence and perceived safety are the dependent variables. This experimental design gives rise to the following conditions:

1. **Experiment 1: Personality**
    - **NP-B:** No-personality with button input
    - **P-B:** Personality with button input
2. **Experiment 2: Input Method**
    - **P-B:** Personality with button input
    - **P-T:** Personality with free-text input
3. **Experiment 3: Voice Output**
    - **C-NV:** Chatbot with no voice
    - **C-V:** Chatbot with voice
    - **S-V:** Speaker with voice

The robot-VA study is a single experiment comparing interaction videos of a robot and a voice assistant speaker with four different voice outputs. As a result, it is a 2x4 factorial design with the agent body and voice being two independent variables. In this study, the type of trust exhibited towards an agent, effect of embodiment on trust, and effect of embodiment on anthropomorphism are examined. Human-trust, technology-trust, anthropomorphism, animacy, likeability, perceived intelligence and perceived safety are the dependent variables. Type of trust (human-trust and technology-trust) are examined as a within-subjects variable, while all other variables are between-subjects. The study design gives rise to the following conditions:

**Table 4.2:** Robot-VA study design and conditions.

| | Text-to-speech Voice (TSV) | AI Generated Voice (AIV) | Authentic Human Voice (AHV) | Modified-Human Voice (MHV) |
|---|---|---|---|---|
| **Robot Body (RB)** | RB-TSV | RB-AIV | RB-AHV | RB-MHV |
| **Speaker Body (SB)** | SB-TSV | SB-AIV | SB-AHV | SB-MHV |

The mind perception study is comprised of a factor analysis on participant ratings of 18 mental capacities across 12 characters – which included 3 biological and 9 social agents. The biological agents were adult human, infant human and dog. The social agents were 3 chatbots, 3 voice assistant speakers, and 3 robots. Of the 12 agents, participants were asked to image 6 of them, while they experienced the other 6 across different conditions of the previous two experiments (elaborated below). This gave rise to the following set of agents. (1) imagine adult human - AH, (2) imagined infant human - IH, (3) imagined dog - DG, (4) imagined robot - RB, (5) imagined smart speaker - SS, (6) imag- ined chatbot - CB, (7) experienced robot with real human voice - RHV, (8) experienced robot with artificially generated voice - RAV, (9) experienced voice assistant speaker with real human voice - SHV, (10) experienced voice assistant speaker with artificially generated voice - SAV, (11) experienced chatbot with artificially generated voice - CAV, and (12) experienced chatbot without voice - CNV. Based on the factor scores from the factor analysis, participant perception of mind between the different agents was compared.

## 4.5 Data Collection

The data for the Chatbot study and Robot-VA study were collected separately on different sets of participants. The data for the mind perception study was collected as part of both the Chatbot and Robot-VA studies by measuring perception of mind (as described above) in each of the conditions in the two experiments.

## 4.6 Measurement and Analysis

The chatbot study and robot-VA study both employed the same measurement scales, the trust scale by Lankton et al (Lankton et al., 2015), and the Godspeed series questionnaire by Bartneck et al (Bartneck et al., 2009b). However, different subscales of the trust scale were used in the two studies. The chatbot study used the human-trust, trusting intention, and reliability subscales, whereas robot-VA study used the human-trust and technology-trust subscales. All variables were measured on a 7-point Likert scale. The mind perception study employed the 18 mental capacities identified by Gray et al (Gray et al., 2007), as a 7-point Likert scale instead of the pairwise rating from the original study.

All three studies had issues with non-normality and heteroscedasticity in the data, which is common for Likert-scale ordinal data. As a result, for the chatbot study, non parametric tests were used, particularly, the Mann-Whitney U test and the Kruskal-Wallis test with Holm adjusted pairwise Dunn post-hoc tests. For the robot-VA study, both parametric and non-parametric tests were used, for parametric tests, t-tests and robust two-way ANOVAs were employed, and in case of significant result, non-parametric, particularly the Mann-Whitney U test and the Kruskal-Wallis test, were used to confirm the result. The mind perception study employed a Confirmatory Factor Analysis (CFA) to test the factor structure identified by Gray et al. (Gray et al., 2007) which was not replicated, as a result an Exploratory Factor Analysis (EFA) was performed to identify a new factor structure. The Kruskal-Wallis test with Holm-Bonferroni adjusted Dunn post-hoc tests were employed to compare mind perception across agents.

# Chapter 5

# Results

In the context of AI and Robotics, transparency and explainability have been identified as significant factors contributing to trust (Larsson and Heintz, 2020). Building on this notion, the 'Three Levels of AI Transparency' paper (Article IV) elaborated on the levels within which transparency operates, namely (1) algorithmic transparency, (2) interaction transparency, and (3) social transparency. The levels are loosely based on Meijer's (2014) transparency framework. In this paper, the concepts of transparency and explainability are defined distinctly from one another, framing explainability as an aspect of transparency, particularly algorithmic transparency. This conception of transparency is disentangled from explanations more generally in humans, and it is argued that explainability is a technical concept in the field of AI that is narrowly focused on making the black-box nature of algorithms and their resultant decisions explainable to humans. The final level, social transparency pertains to broader societal and institutional notions of transparency. It pertains to AI governance, AI ethics, and corporate responsibility in developing AI. The aspects of transparency emerging through interaction are framed under the new concept of interaction transparency. This level of transparency is more concerned with interaction and agent design that facilitate and prioritise transparency in interactions. The conception of this level made way to the subsequent studies on humanlikeness. From the empirical studies, the results of all variables measured are reported here, but only anthropomorphism, mind, and trust remain central in the discussion.

## 5.1 Chatbot Study (Article I)

The chatbot study assessed whether humanlikeness of chatbot through chatbot behaviour, input method, and communication medium had an impact on perception in human-chatbot interaction respectively. The results indicate that all three independent variables had a statistically significant impact on interaction on some of the dependent variables (trust, trusting intention, reliability, anthropomorphism, animacy, likeability, perceived intelligence and perceived safety), but none on all.

Humanlikeness of chatbot behaviour (here mainly referred to as 'personality') had a significant effect on anthropomorphism, animacy and likeability, with the chatbot with personality being rated higher on all three counts, but had no significant effect on trust, reliability, perceived intelligence or perceived safety.

Humanlikeness of interaction modality (buttoned-based vs free-text) only had a significant effect on trusting intention, with participants having greater trusting intention for the chatbot with free-text input. There was no significant difference in trust.

Humanlikeness of communication medium (text chatbot vs chatbot with voice output vs voice assistant speaker) had a significant effect of trust, anthropomorphism, and animacy, with both the chatbot being rated higher on all three counts than the voice assistant, but had no significant effect on trusting intention, reliability, likeability, perceived intelligence, or perceived safety. There was no difference between both chatbots.

## 5.2 Robot-Voice Assistant Study (Article II)

The Robot-VA study assessed whether humanlikeness of body (embodiment) and voice had an effect on perception of trust (whether they elicited human or technology conceptions of trust), anthropomorphism, animacy, likeability, perceived intelligence and perceived safety. Humanlikeness of body was represented through a humanoid robot, and the lack thereof by a voice assistant speaker. Four levels of humanness of voice were used with both agents.

There are three main findings from this study:

1. Body and voice both contribute to anthropomorphism perception independently from one another, serving as empirical evidence for the multidimensionality of anthropomorphism.
2. Voice has a stronger impact on the perception of Godspeed series characteristics of an agent compared to body.
3. Provided successful interaction, users perceive both human and technology conceptions of trust in agents with no significant difference and irrespective of the physical or voice attributes.

## 5.3 Mind Perception Study (Article III)

The mind perception study assessed whether dimensions of mind perception apply similarly across various types of social agents, different from perception of mind in biological beings, and study whether agent types result in significantly different perception of mind.

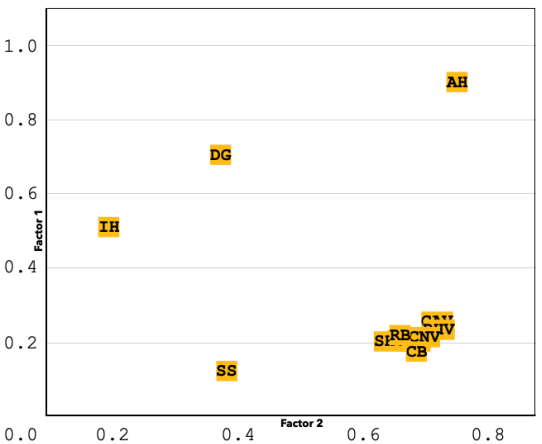



**Figure 5.1:** Perception of mind across characters.

The results indicate that the perception of mind differs significantly between biological beings and social agents. Perception of mind is broadly split along the lines of task-oriented cognition and affective-reflective cognition, with no significant difference of perception between adult humans and social agents in terms of task-oriented cognition. Agents are perceived significantly differently to adult humans in terms of affective-reflective cognition, with adult humans being rated higher. Infant humans and dogs are perceived significantly lower than both adult humans and social agents in terms of task-oriented cognition. And perception of agents does not differ from when individuals imagine an agent and experience an agent, except in the case of smart speaker where imagined agent is rated significantly poorly compared to when the agent is experienced. The results also produce a new factor structure (see Table 5.1) compared to Gray et al. (Gray et al., 2007).

**Table 5.1:** Dimensions of mind perception for social agents

| | Factor 1 | Factor 2 |
|---|---|---|
| **Personhood** | **0.935** | -0.042 |
| **Emotion** | **0.888** | -0.297 |
| **Emotion Recognition** | **0.865** | 0.119 |
| **Thought** | **0.753** | 0.329 |
| **Morality** | **0.738** | 0.385 |
| **Self Control** | **0.663** | 0.438 |
| **Planning** | -0.062 | **0.923** |
| **Communication** | 0.106 | **0.879** |
| **Memory** | 0.224 | **0.788** |

# Chapter 6

# Discussion

The conception of the articles and studies in this thesis began with the work on transparency and trust, which was a consequence of the concept being a central theme in the research project *AI Transparency and Consumer Trust* that the doctoral position is tied to. Identifying that transparency and explainability are *realised* in interaction, and conceptualising this relationship as one of the levels of transparency in AI – interaction transparency – in Article IV, resulted in the gradual shift toward interaction research. The concept of transparency itself moved to the periphery as trust became the central concept of interest in the subsequent work, given its significant role in interaction. The move to interaction research on trust was realised through a collaboration on an extended abstract exploring trust and failures in social robot interaction, submitted to a workshop at the Human-Robot Interaction conference in 2022. Simultaneous discussions with boost.ai regarding chatbot interactions resulted in the eventual merging of various research directions into the three related studies that comprise this thesis. In this manner, the articles included represent the doctoral journey, from transparency, to robot studies and to chatbot studies.

In the subsequent discussion, the first section will focus specifically on the research questions asked in this thesis, and results of the studies. The remainder of the discussion will focus on reflections and broader aspects of the field of study.

## 6.1 Synthesis of Results

### RQ1: How do different humanlike attributes such as body, voice, personality, input method, and communication medium affect the extent to which users anthropomorphise chatbots, robots, and voice assistant speakers during interaction?

Across both studies, a complex and inconsistent picture of anthropomorphism emerges, where some aspects that are less humanlike are perceived as more anthropomorphic and vice versa.

#### Chatbot Study

The results on the effect of chatbot personality on anthropomorphism (experiment 1) were in line with expectations. Chatbots with personality (friendly, use of personal pronouns, use of emoji) were rated as more anthropomorphic than chatbots without personality (robotic language, no personal pronouns, no emoji use). Previous studies have found a similar effect (Araujo, 2018; Jenneboer, 2022), and have further found that this resulted in greater user experience, which was not measured here in this experiment. However, the results on the effect of input method on anthropomorphism (experiment 2) were contrary to expectations, since it was hypothesised that free-text input method simulates natural interaction and thus would be perceived as more humanlike. But no difference in anthropomorphism was found between between button-based and free-text input methods. In a previous study, Haugeland et al. (2022) similarly found counter-intuitive results that button-based chatbot interaction was rated higher than free-text for *both* hedonic and pragmatic qualities, but had no difference in perceived anthropomorphism. They extrapolated from interviews that free-text input may not directly be associated with increased hedonic quality, rather the flexibility and adaptability potential of free-text may be the important underlying factors (Haugeland et al., 2022). Since the simulated chatbot interactions used in the input method experiment offered limited flexibility for interaction outside the intended path, this may explain the result. With the exception of LLM-based chatbots, it is generally not possible to achieve this level of flexibility and adaptability in chatbot interactions. At the same time, LLM based chatbots are are better suited for general-purpose applications, rather than goal-oriented domains (Deng et al., 2023). This may have implications for specific use-cases in task-oriented domains where button input

may be preferable. It is unclear whether an LLM-based chatbots that could hypothetically be more goal-oriented would be perceived as more anthropomorphic than button-based systems. Even if they are, it is unclear whether they would be preferable in light of other considerations, such as transparency.

Results on the effect of chatbot communication medium (experiment 3) were also contrary to expectations, where both the text-only chatbot, as well as the chatbot with voice, were perceived as significantly more anthropomorphic than the voice assistant, with no difference between the two chatbots. Ischen et al. (2022), similarly, found that text-based assistants were perceived as more humanlike than voice-based assistants (Ischen et al., 2022). On the other hand, the insignificant result between the two chatbots is contrary to Cohn et al's. (2024) study on pseudo-LLM chatbots that found that a multimodal chatbot presenting both text and voice outputs was perceived as more anthropomorphic compared to a text-only chatbot (Cohn et al., 2024). The context of interaction between an LLM chatbot and a goal-oriented chatbot are different, and voice may be a more important factor in hedonic contexts than pragmatic ones. In terms of methodology, the experience of viewing a screen recording of a text-only chatbot interaction and a chatbot with voice interaction may not be sufficiently different in terms of anthropomorphism, and additionally, it may be that it is easier to envision a text-only chatbot, due to familiarity, than a multimodal one, when answering a survey after watching a video.

**Robot-VA Study**

Both agent body and voice had a significant effect on anthropomorphism. In terms of voice, agents with least humanlike voice (robotic text-to-speech) were consistently rated as less anthropomorphic than agents with more humanlike voice (AI generated and real human), irrespective of agent type. However, there was no difference between the AI generated and the real human voice. Eyssel et al. (2012) similarly found that more humanlike voice in robots led to more perceived anthropomorphism compared to robot-like voice (Eyssel et al., 2012). The lack of difference in perception of anthropomorphism between the AI generated and real human voices indicates that the AI generated voice was perceived as being more or less similar to a human. Craig and Schroeder (2017) demonstrated that modern synthetic voices performed similarly to real human voices in an online learning setting (Craig and Schroeder, 2017). The result contributes to the notion expressed by Seaborn et al. (2021) that modern synthetic voices are reaching the

level of human voices (Seaborn et al., 2021).

However, in terms of agent body, where it was expected that a robot would be perceived as more anthropomorphic than a voice assistant speaker, the reverse was true. It has to be noted though that the robot in the study was not much more than a speaker in robot form (it did not move), and this may have caused an expectation mismatch. Anthropomorphism can lead to higher expectations regarding capabilities, against which amassments of performance are made, when the expectations are dis-confirmed, it can lead to poor outcomes (Grazzini et al., 2023). While this explanation requires the agent to be initially assessed as more anthropomorphic, reassessment upon mismatched expectations may lead to lower perception of anthropomorphism in the end. This may have negatively affected the perception of the robot; whereas the speaker may have better matched the expectations of the participants. Additionally, the lack of embodiment given the online study design inherently disallows an arguably important dimension of interaction, which may have also contributed to the result. Regardless, the result may indicate that the human form alone may not directly result in greater anthropomorphism in comparison to other anthropomorphic characteristics such as voice, especially in virtual interactions.

### RQ2: How does agent type – chatbot, voice assistant, social robot – affect mind perception in human-agent interaction?

People seemingly possess a distinct notion of mind in humanlike agents in general, irrespective of agent type or characteristics, compared to biological characters such as adult human, infant human, or dog. Generally participants seem to perceive all agents as possessing more or less a similar form of mind. This mind is seemingly capable of *task-oriented cognition*, similar to human minds, but not *affective-reflective cognition* which humans are capable of. The two dimensions Agency and Experience, consisting of 18 mental capacities from Gray et al. (2007) could not be replicated using an exploratory factor analysis on the data. Instead two new dimensions, as mentioned above, emerged with slightly different factor structures through an exploratory factor analysis. Previous studies have employed the dimensions outlined by Gray et al. (2007) and found further nuances in mind perception, for example that the factor Experience may be split by an affective dimension (Kamide et al., 2013), and that perception of mental capacities may differ based on interaction style (Cucciniello et al., 2023). These studies were conducted on robots alone and were able to more or less reproduce the original factor

structure. The different factor structure emerging here may be attributed to the inclusion of biological agents along with various types of humanlike social agents highlighting the different conceptions of mind between the two.

The original study included 13 characters, with 11 of them being biological, and only two non-human characters, god and robot. Of the 11 biological agents, 7 were human (Gray et al., 2007). As a result, the dimensions represented a human conception of mind, which was the intention of the study. In this study, there were 12 characters, with 3 being biological, of which 2 were human, and 9 were agents. This factor structure captures the concept of agent mind in contrast to biological mind (in terms of human mental capacities) since factor analysis reflects the underlying variance in the data, and a dataset dominated by agents might emphasize different dimensions of mind compared to one dominated by human characters. This also explains the high multicollinearity between several of the mental capacities that were more strongly associated with biological characters than the agents, necessitating compound variables to account for this skew. Interestingly, the coordinate points for factor scores of the three biological agents in this study, when plotted on the two factors (see Figure 4.1) closely align with the points for the same characters in the original study by Gray et al. (2007), indicating that despite the rearrangement of the factor structure, the result on a broader level still reflects the same aspects of mind perception as the original study. The new factor structure is indicative of the aspects of mind that are perceived as distinctly human/biological, and those that are shared with agents.

Perceiving a mind in non-living entities is inherently an act of anthropomorphism, as it involves ascribing humanlike mental states and capacities to something that does not inherently possess them, for example, objects like computers, abstract concepts like god, or animals. Unlike this type of anthropomorphism, the dynamics of anthropomorphism seen in humanlike social agents is slightly different, since these agents are already designed to mimic human characteristics. This means users are not projecting human traits onto a completely non-human other but rather evaluating the degree to which a humanlike entity appears to possess truly human qualities. In this context, the results of this study suggest the possibility of a distinct conception of mind in agents – differentiated by the subjective sense of (an agent not having) *deep-human* attributes such as agency, consciousness, and intentionality (better characterised as humanness) – as a starting point to study its evolution in comparison to a human conception of mind. While the dimensions identified in this study lay the groundwork for such a conception, further research

is needed to understand whether direct interaction, varying context, and individual differences influence this distinct conception of agent mind. It is important to note that the lack of affective-reflective mind may be explained by the fact that the agents in this study were not 'emotional agents', in that the experiment was not designed to capture the emotional dimensions. It is possible, perhaps likely, that an emotive agent may rank higher in that capacity than the agents in this study. Another dimension that is not considered is embodied interaction. Direct interaction may produce significantly different perception of mind, however this is difficult to capture. Regardless, the finding, which may be regarded as initial finding, of a common notion of agent mind, calls for further research with regard to the aforementioned factors.

### RQ3: How do different humanlike attributes such as body, voice, personality, input method, and communication medium influence user trust in chatbots, robots, and voice assistant speakers during interaction?

Across both studies, trust remained relatively stable, and no significant differences in trust were found between different conditions (humanlike attributes) for the same agent in either study.

#### Chatbot Study

The results on the effect of chatbot personality on trust (experiment 1), show that there was no significant difference between trust. Existing research shows that chatbot personality has a significant effect on trust (Müller et al., 2019; Kuhail et al., 2022), however, context plays a role (Smestad and Volden, 2019). Følstad and Brandtzaeg (2020) highlight pragmatic and hedonic aspects of chatbot interaction as differently applicable for goal-oriented and general-purpose chatbots, noting that for goal-oriented chatbots pragmatic considerations such as performing expected tasks successfully and efficiently are the most important factors, and that hedonic qualities such as emoji use and friendliness may improve user experience but may not be essential (Følstad and Brandtzaeg, 2020). In this study, the goal-oriented chatbots were simulated to perform tasks in the exactly same manner, and successfully, meeting the most important pragmatic aspects of chatbot interaction. Seemingly, this is more important for trust in this context than humanlike attributes such as personality. Law et al. (2022) found that humanlikeness and task performance have no interaction effect when predicting general trust, task-

specific trust and reliability (Law et al., 2022). This may mean that trust derived from pragmatic aspects are independent from hedonic aspects, and in the goal-oriented context, with pragmatic aspects being more important, they may have a stronger influence on trust. This also explains the results on the effect of chatbot input method on trust (experiment 2), which also had no significant difference in trust. Given successful task completion, trust remained stable. However, free-text input had greater perceived trusting intention than button-based input. This is a confounding result, as trusting intention pertains to the willingness to trust in technology. This may indicate that despite the difference in actual trust, free-text input has desirable characteristics that may enable people to depend on it in the future.

The results on the effect of chatbot communication medium on trust (experiment 3), similar to anthropomorphism, shows that both chatbots were perceived as more trustworthy than the voice assistant. One explanation here would be that anthropomorphism influences trust (Waytz et al., 2014), and since the chatbots were perceived as more anthropomorphic, they were also perceived as more trustworthy. However, given that both anthropomorphism and trust were counter-intuitively skewed in the direction of the chatbots, it may warrant some reflection on the experiment. The videos presented to the participants depicted screen recordings of chatbot interactions, and a recording of a chatbot (in frame) interacting with a human (not in frame). This results in very different types of videos, even if the script of the actual interaction remains the same, and the audio in the chatbot with voice condition is the exact same as the voice assistant. The chatbot screen recordings are simply more dynamic. It is possible that this difference in dynamism contributed to greater perception of anthropomorphism and trust in the chatbots compared to the voice assistant.

**Robot-VA Study**

Comparing whether participants displayed a human conception of trust or a technology conception of trust, and if these levels differed, in the robot and voice assistant with different levels of humanlikeness of voice, revealed that they displayed both types of trust, to a high extent, and without a significant difference in between the various conditions. Furthermore, both human and technology conceptions of trust were not influenced by either the body – robot and voice assistant – or the levels of humanlike voice. It may be that people exhibit both human and technology conceptions of trust in humanlike social agents, as they recognise them

to be both humanlike and technology-like. With high levels of trust as a consequence of the context and successful task completion as explained above. Lankton et al. (2011) find that this is also the case with Facebook (Lankton and McKnight, 2011), which is far less humanlike than the agents in question. Other researchers note that trust in automation, with some critical differences, develops largely similarly to the way it does in humans (Madhavan et al., 2006). However, similar levels of technology trust in the results may indicate that individuals continue to assess humanlike social agents as technologies despite anthropomorphism and humanlikeness.

## 6.2 Anthropomorphism Dimensions

Whether anthropomorphism is multidimensional, and if so, what these dimensions look like and how they should be conceptualised has been an open question in HAI (Kühne and Peter, 2023). The Robot-VA study results showed that humans anthropomorphised the agents based on both agent body and agent voice, but there was no interaction effect between the two, meaning that body and voice both influenced anthropomorphism independently. This provides evidence that individuals independently anthropomorphise different aspects of agent humanlikeness. There may be two possible interpretations of this, (1) the result may indicate a notion of sensory dimensionality to anthropomorphism, or (2) the results reflect multidimensionality of humanlikeness in agents. The former interpretation does not align with Kühne and Peter's (2022) multidimensional conceptualisation along five dimensions based on Theory-of-Mind – thinking, feeling, perceiving, desiring, and choosing (Kühne and Peter, 2023). In their conception of multidimensional anthropomorphism, body and voice are likely to be seen as precursors to anthropomorphism. They puts more emphasis on mind perception as the central element of anthropomorphism, with sensory-perceptive elements categorised as precursors, and social dimensions such as personality attribution or trust as consequences. In motivating their conception, they argue,

> The fact that anthropomorphism precedes the application of ToM elucidates why anthropomorphism and ToM entail corresponding subdimensions: The application of ToM to a robot logically presupposes that the robot possesses a specific set of human mental facul-

> ties: Only if the robot possesses these mental faculties can explanations of the robot's behavior be based on them. Anthropomorphism thus constitutes a cognition in which an individual evaluates whether and to which degree the robot possesses the respective human mental faculties. Each assessment of a distinct mental faculty, in turn, constitutes a dimension of anthropomorphism (Kühne and Peter, 2023).

This argument does not align with how the relationship between the various phenomena have been conceptualised in this thesis. As detailed in the theory section, anthropomorphism is understood as the attribution of humanlike qualities to nonhumans, of which mind perception is one aspect, and upon perceiving a mind, theory of mind may apply (Waytz et al., 2010b). Kühne and Peter (2022) seemingly argue that anthropomorphism *is* mind perception, as they do not distinguish between mind perception and attribution of theory of mind. The relationship between the three concepts is far from clear in literature, and as Waytz et al. (2010) note, there has been more work on theory of mind in cognitive science than what comes before:

> The inferences people make about minds comprise three basic research questions. First, do people think a particular entity has a mind? Second, what state is that other mind in? Third, what are the behavioral consequences of perceiving a mind in another entity? Most research has focused on the second question– the perception of mental states– that has come to be known as 'theory of mind' or 'mentalizing' [7,8]. Researchers are now expanding their attention to the first and third questions– about the causes and consequences of mind perception, respectively– a shift that represents an important emerging trend worthy of attention (Waytz et al., 2010b).

It is unclear whether anthropomorphism and mind perception are distinct, related, and if so, how closely related. On the other hand, Li and Suh (2022) seemingly argue for visual and vocal aspects as different dimensions of anthropomorphism, a view, it may be argued, is substantiated by the result of this study (Li and Suh, 2022).

Alternatively, an argument may be made regarding what may constitute different dimensions of anthropomorphism within the non-representationalist paradigm

of embodied cognitive science such as enactivism, where cognition is not centred around the mind, and the theory of mind or mind in general do not play the same central role in explanations as they do in representationalist paradigms. In this paradigm, voice and body humanlikeness may be considered as representing distinct, interaction-specific dimensions through which humans attribute human-like qualities to non-human agents. This interpretation shifts the focus from static traits (facets of humanlikeness) to dynamic processes of engagement and perception.

Lastly, the result also aligns with von Zitzewitz et al's. (2013) 'Parameters of Human Likeness', with two main parameters: appearance and behaviour, where appearance is comprised of five fields: visual appearance, sound, smell, haptic appearance, and taste (von Zitzewitz et al., 2013). Here, visual appearance and sound align with body and voice. In this interpretation, body and voice constitute parameters of humanlikeness. Although that does not inherently preclude them from also being dimensions of anthropomorphism.

## 6.3 Excluded Variables

Apart from defining a narrower scope for the thesis, one of the motivations for excluding the additional variables (animacy, likeability, perceived intelligence, and perceived safety) from the Godspeed series is that there is some overlap between some of the constructs. Bartneck (2023) the creator of the measurement instrument writes that, "We refer to the GQS as a series since it consists of five scales that can each be used as a stand-alone questionnaires. The GQS is not intended to be always used with all of its questions, particularly due to the overlap between anthropomorphism and animacy" (Bartneck, 2023). Additionally, the perceived safety scale which is comprised of only three items has a Cronbach's alpha below (0.6) (Bartneck, 2023), which indicates low internal consistency reliability, raising concerns about the scale's ability to measure the construct consistently across its items. Regardless, the results are briefly presented below.

**Chatbot Study**

In experiment 1, the chatbot with personality was perceived as having higher animacy and likeability, but there was no significant difference in perceived intelli-

gence and perceived safety. This suggests that while humanlikeness can improve the qualitative experiences of a chatbot, a lack of such characteristics might not negatively affect the perception of a chatbot's task-related abilities.

In experiment 2, no difference in the perception of animacy, likeability, perceived intelligence or perceived safety was found between button-based and free-text inputs, indicating that, given successful task-completion, there are no significant perceptive differences based on input method.

In experiment 3, both the chatbot with and without voice were perceived as more animate than the voice assistant, with no significant different in perception of likeability, perceived intelligence or perceived safety. There was also no significant difference in the perception of animacy, likeability, perceived intelligence or perceived safety between the two chatbots.

**Robot-VA Study**

No significant difference was found in the perception of animacy between the robot and the speaker, however there was a statistically significant difference in the perception of animacy of voice, with the real human voice being perceived as the most animate.

Similarly, no significant difference was found in the perception of likeability between the robot and the speaker, however there was a statistically significant difference in the perception of likeability of voice, where both the AI generated and real human voices were perceived as more likeable than the robotic text-to-speech voice. However, there was no significant difference in likeability between the AI generated and real human voices.

No significant difference was found in the perceived intelligence of neither the robot or speaker, nor the voices.

No significant difference was found in perceived safety between the robot and the speaker, however there was a statistically significant difference in the perceived safety between voices, where both the AI generated and real human voices were perceived as safer than the robotic text-to-speech voice. However, there was no significant difference in perceived safety between the AI generated and real human voices.

## 6.4 Transparency in HAI

The three levels of AI transparency highlighted in Article IV conceptualise transparency across algorithmic, interaction, and social dimensions. The three levels, loosely based on Meijer's (2014) conceptualisation of transparency in governance (Meijer, 2014), point to three distinct areas in contemporary AI development that require deeper evaluation of transparency mechanisms, particularly since transparency is key to establish trust in technology (Brunk et al., 2019). These levels also indicate for whom these levels of transparency are relevant. For example, algorithmic transparency is unlikely to be useful or relevant for a consumer, but essential for an auditor. In this conception, explanations are split between two levels, the more technical notion of explainable AI (XAI) that is part of algorithmic transparency, and the interactional notion of explanations that is part of interaction transparency, that are required for consumers to effectively understand and use these technologies. The social dimension of transparency has received considerable attention due to ongoing discussions about AI regulations in the EU (Larsson and Heintz, 2020). And algorithmic transparency, particularly in XAI, is also a growing field (Minh et al., 2022). Interaction transparency however, requires more attention in AI/HAI research.

In the context of HAI, particularly from the point of view of a human interacting with an agent, interaction transparency is the most immediately apparent. Interaction transparency can be broadly split into two notions/approaches, (1) informational, (2) behavioural. The informational notion pertains to the information that an agent directly communicates to the user. Research on this type of transparency in interaction is now being recognised (Bhaskara et al., 2020). The behavioural notion of transparency is less explicitly recognised. Building humanlikeness into agents can be understood as a means to enhance transparency. Humans intuitively understand objects that reference human paradigms, as they allow us to utilise our existing knowledge about the world to understand and use these objects (Fink, 2012), this is interaction transparency as elaborated in Article I. The same is true for agents. For example, it is much more intuitive for humans to relate to and predict how a humanoid robot arm might move when compared to an industrial robot arm. This intuitiveness in design helps make interaction more seamless, predictable, and safe. Such a transparent design may even mitigate loss of trust when agents fail to perform as expected, particularly when effective explanations are used (Kox et al., 2021).

Humanlikeness, however, may not always convey transparency. The result of the effect of input method on anthropomorphism (experiment 2) in the Chatbot study could be interpreted through an interaction transparency lens to illustrate this point. Button-based inputs prime users with interaction possibilities, in a sense the agent gives the user various options to progress an interaction towards a particular goal at any given stage, giving the user a clear sense of how they can engage with the agent, in doing so, make the interactional aspect of the agent transparent. This is not the same as an agent providing the user information about how they may interact, rather, the agent is by-design able to communicate its interaction capabilities at every give moment within an interaction. This is precisely what button-based input does. And as is true for transparency in other domains, the greater the "stakes", the greater the impact of interaction transparency. This would be the interaction transparency explanation for why button-based interactions are preferred in goal-oriented contexts. Further development of the concept is needed to identify transparency opportunities in various interaction mediums and leverage them.

However, there may also be a flip side to this type of transparency. Because, for example, humanlikeness gives a sense of interaction transparency, it also allows individuals to wrongly attribute greater humanlike competence to these objects. At best, this makes individuals perceive the agent as incompetent, at worst this could have grave consequences to safety. At the same time, there is potential here for corporations to manipulate users by utilising interaction transparency to create a false sense of being transparent while exploiting the users. Further development of the concept of interaction transparency can help researchers identify, contextualise, and research these behavioural and ethical considerations emerging from such transparency.

## 6.5 Limitations

### Embodiment and Presence

Embodiment is another one of the terms that is not clearly defined. The concept of embodiment is understood differently in different fields, leading to some conceptual obfuscation. In the context of this thesis, there are two notions of embodiment that are relevant, (1) Embodiment of the agent, that is, the physical

body or presentation of the agent, and (2) Embodiment in interaction, that is, direct in-person interaction with an agent possessing a physical form. In this thesis, the former notion of embodiment is applicable pertaining to the robot, and the latter, referring to physical presence, is out of scope considering the video-based methodology. The scarcity of direct interaction studies is often noted in literature (Złotowski et al., 2015; Blut et al., 2021), owing to the complications arising from unreliable technology, difficulty in controlling and isolating variables, and conducting large-scale studies. However, the results from controlled, simulated, and laboratory-confined studies may not directly be applicable in real world settings, limiting the ecological validity of such experiments. One of the significant limitations of the studies in this thesis is that they are all video-based and conducted online. And while the robot is described as embodied, it is in reference to the humanlike physical appearance of the agent, and not embodiment in interaction. A meta analysis on embodiment in social human-robot interaction found that embodied robots may influence subjective and objective measure, while depicted robots may only influence subjective measures (Roesler et al., 2023). Another review found that telepresent robots were perceived similar to virtual agents (Li, 2015). However, there seems to be nuances to the effect physically present agents have on perception, which are not clearly understood. Mollahosseini et al. (2018) find that:

> Eye gaze and certain facial expressions are perceived more accurately when the embodied agent is physically present than when it is displayed on a 2D screen either as a telepresent or a virtual agent. Conversely, we find no evidence that either the embodiment or the presence of the robot improves the perception of visual speech, regardless of syntactic or semantic cues. Comparison of our findings with previous studies also indicates that the role of embodiment and presence should not be generalized without considering the limitations of the embodied agents (Mollahosseini et al., 2018).

Regardless, the design of the three studies in this thesis may be regarded as a limitation in terms of ecological validity.

**Agents and Context**

As is the case with experimental studies, the generalisability of the studies is limited by the specific design decisions made. Studies on subjective measures are highly subject to variability. The specific agents used, the travel assistant context, the measurement instruments used, all have a significant implications on how participants perceive the agents and answer the surveys. Different agents may elicit different responses in different contexts. As a result, the external validity of the results may be regarded as a limitation.

**Mind Perception Study**

The mind perception data was collected across both the Chatbot and Robot-VA studies. Only the video-based, most humanlike conditions were chosen from both studies to compare the perception of mind across all types of agents. Although the two studies were structurally and methodologically very similar, including highly similar, scripted, interaction possibilities, there are still differences in how the agents in the two studies can interact and be perceived. This may be regarded as a limitation in terms of internal validity.

## 6.6 Pseudo-Sapiens: The Humanlike Social Agents

Generally, the categorisation of social agents typically focuses on their underlying technologies or the specific domains of their application. There are a number of different terms across literature that overlap and describe similar albeit differently framed concepts of different types of agents, such as Socially Interactive Agents (SIA), Conversational Agents (CA), Embodied Conversational Agents (ECA), Intelligent Virtual Agents (IVA), AI-Enabled Technologies (AIET) and Humanoid Social Robots, to name a few. These established scientific categorisations, originating in a technical context, often also drive research in interaction studies. On the other hand, functional categorisations of end user products such as chatbot, voice assistant, or social robot, are used to describe a broad and evolving category of intelligent social technologies with varying characteristics. These terms are generally vague, leaving ambiguity about what the defining qualities of the various agents that may fall under each category are. Furthermore, the terms themselves emerged from the technical landscape they were primarily designed for, and their defini-

tions tend to retain the technical demarcations of the specialised fields in which they were developed. None of these terms originate from interaction research or human experience of humanlike agents. There is a need for a concept in Human-Agent Interaction that takes a human-centric approach in its conception. Such a concept would frame research objectives from a human experience perspective rather than a technology design perspective, which is generally the case today.

From a human experience perspective, it may be argued that broadly, the interaction and user experience of humanlike AI and robots, regardless of their form and function, have far more in common with one another than with their non-humanoid counterparts. For instance, a humanoid robot like Pepper and a disembodied chatbot like ChatGPT may share no morphological similarities but they share experiential similarities, such as conversational interaction, emotional expression, and perceived personality. In contrast, an industrial robot like a robotic arm used in factories operates with completely different interaction paradigms compared to Pepper, despite sharing the same technological field of robotics. The argument herein is that this capacity for natural, humanlike interaction creates a shared category in human experience – grouping humanoid robots, chatbots, and voice assistants, and other humanlike AI together. The result from the mind perception study indicating a distinct and relatively similar type of perceived mind across agent types may be regarded as a preliminary finding in this direction. The current classification of these technologies does not reflect this broader perspective.

Consequently, this thesis introduces the term 'Pseudo-Sapien technologies' or simply 'Pseudo-Sapiens'[1] as a way to collectively refer to all humanlike social agents. A Pseudo-Sapien is defined as any technological entity that simulates or imitates human characteristics, behaviour, or intelligence, regardless of its technical or physical form. Whether embodied in a physical robot, present as a virtual assistant, or functioning as an AI-driven chatbot, the defining feature of a Pseudo-Sapien is its capacity to replicate humanlike interactions and behaviours. The use of "pseudo" here is not intended to carry the negative connotations of the term, such as "false" or "deceptive." Rather, it serves as a neutral descriptor to clarify that, while these technologies may convincingly mimic human characteristics, they are fundamentally artificial constructs, and the base word sapien highlights the inherent humanlikeness. Two characteristics seem to be particularly relevant in this regard,

[1]Plural: Pseudo-Sapiens; Singular: Pseudo-Sapien

the capacity for natural language, and humanlike appearance. One hypothesis that needs testing is that natural language may be the strongest signal of humanlikeness, followed by appearance. The various characteristics that comprise humanlikeness in perception, their relationship to one another, potential hierarchies, are all aspects that need further research, and benefit from conceptualising these individual technologies collectively as 'pseudo-sapien technologies'.

Looking into the future, it is likely that interaction capabilities, particularly with natural language, become increasingly similar across various types of agents, with embodiment remaining the primary difference. More speculatively, systems that are interconnected and operate across platforms may become a reality. Cross-platform AI assistants already exist in limited capacity. It is plausible that in the future, particularly with AI assistants, the same AI may present as a chatbot, voice assistant, or robot, depending on the context and need. Human interaction with, and perception of such broad-ranging, cross-platform, body-agnostic forms of technology cannot be fully understood within the current scheme of categorisation, and are better conceptualised by a concept such as Pseudo-Sapien. Research on this broader outlook on humanlikeness will also help better understand the implications for humans to interact with humanlike non-human others, which is particularly of relevance as technologies become more and more humanlike.

## 6.7 Phenomenal Humanlikeness

Discourse regarding the existence of an objective reality, its nature, and its availability to subjective human experience has been a prominent theme in philosophy since antiquity. The thought experiment, *'If a tree falls in a forest and no one is around to hear it, does it make a sound?'* serves to highlight two dominant, opposing ontological positions. *Realism*[2] would argue, '*Yes*, the tree makes a sound because the physical vibrations that produce sound, caused by the tree falling, occur regardless of whether anyone is there to hear it', while *Idealism* would argue, '*No*, the tree does not make a sound in the absence of an observer, as sound is not merely the physical vibrations, but rather the result of the mind's experience of those vibrations'. This distinction illustrates the central issue in this debate:

[2]This account is defined as 'direct realism', though other schools of realism, such as 'indirect realism', seek to address the role of perception (Platchias, 2004)

whether phenomena exist independently of the mind or whether they are, in some way, shaped or constructed by it. If we consider realism and idealism are two ends of a spectrum, various ontologies fall somewhere between them (there are many exceptions to this oversimplification) on whether they uphold the existence of a reality, and if so, whether they see it as accessible, even partly, to humans or not. These ontological positions have significant implications for how anthropomorphism is understood.

The contemporary mainstream view of anthropomorphism is that it is a cognitive bias, one that is pervasive, but not insurmountable (Epley et al., 2007), implying that there is a reality outside the anthropomorphic that we could, at least in part, access. This ontological assumption is the dominant one in most subsequent research on anthropomorphism, which includes a majority of the current studies in HAI. In this view, anthropomorphism has sometimes been regarded as an error. Duffy (2003) notes that,

> Few psychological experiments have rigorously studied the mechanisms underlying anthropomorphism where only a few have seen it as worthy of study in its own right...the role of anthropomorphism in science has more commonly been considering it as a hindrance when confounded with scientific observation rather than an object to be studied more objectively (Duffy, 2003).

The alternate view on anthropomorphism is that it is an innate, existential aspect of human embodied cognition (Strongman, 2008), that fundamentally informs and limits our interaction with the world. Grech (2019) writes that,

> Perhaps the anthropomorphic should not be thought of as a naive failure of human imagination, but as the condition and possibility of thought itself (Grech, 2019).

Motivating her argument by pointing to the inability of the human mind to conceive of a non-human existence, human nothingness, or perspectives outside an anthropomorphic paradigm, owing to what she calls *anthropomorphic spectrality* – the subtle, often implicit ways human traits and perspectives "haunt" or permeate conceptions of non-human entities or worlds (Grech, 2019).

In the traditional view that this thesis has also assumed, anthropomorphism is regarded as a phenomenon by which observers attribute humanlike qualities – characteristics of appearance, personality traits, emotions, intentions – to non-human entities (Epley et al., 2007). In AI and Robotics, this generally translates to identifying specific humanlike characteristics of an agent that give rise to anthropomorphism, and studying their effect on interaction (Złotowski et al., 2015). These humanlike characteristics are often viewed as *anthropomorphic design cues* that can be manipulated to change the degree to which an agent is anthropomorphised (cf Fink, 2012; Roesler et al., 2021). The focus in this framing is on what is being *attributed*: Which humanlike characteristics do individuals perceive the agent to have? How humanlike does it appear based on its features or behaviours?

This design-driven notion of humanlikeness is embedded within the historical development of humanlike abilities in AI and robotics, which, rightfully, focused on developing specific features for specific purposes. In this paradigm, these attributes are treated as objective features that can be manipulated to incorporate more or less humanlikeness (cf Gray et al., 2022; Prakash and Rogers, 2015; von Zitzewitz et al., 2013). While these objective measures of humanlike features of agents are useful in the context of agent design, assessing human perception of individual features is unlikely to account for the perceptual *experience* of humnalikeness. A subjective, phenomenal[3] notion of humanlikeness is necessary to understand the human experience of humanlike agents. Some existing studies hint at such a notion of humanlikeness, particularly when studying the uncanny valley phenomenon which pertains to the subjective experience of non-human characters in terms of humanlikeness, and has been shown to have a salient valence dimension (Cheetham et al., 2013; Burleigh et al., 2013), but do not engage in what it means for human experience.

Wang et al. (2015) note that, "Another fundamental yet understudied question is whether humanlikeness is a subjective or an objective construct", they further claim that, "In most studies, human likeness has been measured by means of subjective ratings. Human likeness was therefore defined subjectively" (Wang et al., 2015). This perspective is fundamentally rooted in different ontological and epistemological perspectives, as noted above. Many studies (including the ones in this thesis) do use subjective ratings of humanlikeness, however these subjective

[3]"In relation to phenomenal experience philosophers talk of the ineffable character and feel of its qualities" (Pharoah, 2018)

ratings are often treated as stand-ins for objective indicators, using participants' assessments to categorize features of an agent as more or less humanlike. In these cases, subjective reports were never intended to delve deeply into the nature of subjective experience itself; rather, they were used to approximate an underlying "objective" reality of how humanlike the agent was. If it is human experience that is being studied, then the core of anthropomorphism lies not simply in the attributes themselves, but in how these attributes are experienced and interpreted by human observers. Consequently, humanlikeness, in this context, becomes a subjective construct that varies widely across individuals, contexts, and moments of interaction. This notion of individual differences in the perception of anthropomorphism is established in literature (Epley et al., 2007). By shifting our focus from the agent's presented features to the observer's experience of meaningful humanlike presence, we stand to gain a more complete understanding of anthropomorphism – one that incorporates within human experience 'the other side of the coin' in humanlikeness.

In this phenomenal conception of humanlikeness, not only do individual characteristics of an agent shift to the periphery, the strong demarcation made in human-agent interaction research between various agent types does too – essential, in this context, the overarching concept of 'Pseudo-Sapien' proposed above becomes very relevant. Coeckelbergh (2011) writes that:

> Of course as a scientist, engineer or designer we 'know' that the robot is not itself an individuum, something that cannot be divided. We 'know' that it is an aggregate, composite, or system that consists of smaller parts. We also 'know' that there may be larger systems of which the robot is itself a part. And of course we 'know' there are 'other kinds of robots', which we could classify as 'non-social'. Some of us might even assert that 'in reality' all robots are 'mere machines'. The point of the approach I advocate here, however, is that there is no independent reason why one of these points of view takes priority; they are all different ways of knowing and seeing the robot, and which view we take will depend on our (social) role, training, culture, and interactions in particular contexts (Coeckelbergh, 2011).

While Coeckelbergh writes in the context of ethics of robotics, this is an important argument to consider when thinking about the experience of social agents in

general. The appropriate degree of abstraction is relative to the context. When studying the perception of humanlikeness in agents, the type of agent (or their individual characteristics) may be less relevant than the *phenomenal* experience of humanlikeness itself. Of course, the individual nature of an agent matters insofar as an individual account of experience with that particular agent, for example, the difference between an embodied robot and a disembodied chatbot matter when the subject in question is embodiment, but both the robot and the chatbot are still, on a higher level of abstraction, humanlike, which may be more salient in human experience than on the level of individual agent types of characteristics. This hypothesis will need to be tested.

## 6.8 Future Research

The perception of a distinct agent mind across several types of agents is an interesting and potentially important result were it found to be a broader phenomenon. Several types of studies could be pursued in that direction. First, studying whether emotional agents are perceived as possessing affective-reflective cognition, second, studying whether embodiment influences the conception of mine, and third, whether the result can be replicated with a wider range of agents.

There is a lot of scope to expand upon interaction transparency, and to apply theory from other transparency domains, particularly relating to reliability and trust in human-agent interaction. It may open avenues for novel design practices, but also, more importantly, more reflective perspective on the ethics of agent design.

The complementary concepts of Pseudo-Sapien and Phenomenal Humanlikeness put forth several research questions that can be pursued to form a deeper understanding of human interaction with humanlike agents. While the former concept may still be compatible with the dominant empirical approaches to study humanlikeness, the latter concept necessitates delving into phenomenology. The research questions asked in the thesis naturally evolved into what has been discussed in the Phenomenon of Humanlikeness section. In order to study cross-agent humanlikeness, and to understand how humans perceive interactions with humanlike non-human others, a qualitative approach is needed. Such a research study would also necessitate in-person, direct interaction studies with agents, which has been consistently highlighted as a research gap.

# Chapter 7

# Conclusion

Human-Agent Interaction (HAI) is a rapidly evolving field of research, driven by advancements in AI and robotics that have led to the development of increasingly humanlike technologies. Drawing from four articles, this thesis outlined three studies conducted on humanlike social agents – chatbots, voice assistants, and social robots – to study the perception of anthropomorphism, mind, and trust in these technologies. Highlighting contributions like identifying a distinct conception of perceived mind in agents, agent body and voice as two dimensions of anthropomorphism, and interaction transparency as a core aspect of transparency in autonomous systems, *'Pseudo-Sapiens'* was proposed as a collective concept to drive human-centred research on the phenomenon of humanlikeness in relation to humanlike social agents. This thesis represents an infinitesimally small contribution to the monumentally gargantuan task facing modern society: to understand and guide innovation in AI and Robotics in ways that are grounded in human perspectives and practices. As we navigate our relationship with these technologies that increasingly resemble us, bringing to life the age-old myths and stories about humanlike beings of our own making, we find ourselves in Prometheus' stead, wielding technology as the fire of knowledge. Through this fire, we endow our creations with the ability to 'think', the autonomy to 'do', and perhaps, one day, even true agency to 'be'. Yet, in creating these Pseudo-Sapiens, we must grapple with the consequences of their existence – ensuring that we not only understand their implications for us but also guide their evolution with responsibility, so we may avoid the fate that befell the tragic benefactor.

# Bibliography


AbuShawar, B. and Atwell, E. (2015). Alice chatbot: Trials and outputs. *Computación y sistemas*, 19(4):625–632.

Adam, M., Wessel, M., and Benlian, A. (2021). Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2):427–445.

Adams, F. and Aizawa, K. (2008). *The bounds of cognition*. Blackwell.

Ahmed, E., Buruk, O. ⊠., and Hamari, J. (2024). Human–robot companionship: Current trends and future agenda. *International Journal of Social Robotics*, 16(8):1809–1860.

Airenti, G. (2018). The development of anthropomorphism in interaction: Intersubjectivity, imagination, and theory of mind. *Frontiers in psychology*, 9:2136.

Akdim, K., Belanche, D., and Flavián, M. (2023). Attitudes toward service robots: analyses of explicit and implicit attitudes based on anthropomorphism and construal level theory. *International Journal of Contemporary Hospitality Management*, 35(8):2816–2837.

Akdim, K. and Casaló, L. V. (2023). Perceived value of ai-based recommendations service: the case of voice assistants. *Service Business*, 17(1):81–112.

Al Naqbi, H., Bahroun, Z., and Ahmed, V. (2024). Enhancing work productivity through generative artificial intelligence: A comprehensive literature review. *Sustainability*, 16(3):1166.

Al Shamsi, J. H., Al-Emran, M., and Shaalan, K. (2022). Understanding key drivers affecting students' use of artificial intelligence-based voice assistants. *Education and Information Technologies*, 27(6):8071–8091.

Alarcon, G. M., Gibson, A. M., Jessup, S. A., and Capiola, A. (2021). Exploring the differential effects of trust violations in human-human and human-robot interactions. *Applied Ergonomics*, 93:103350.

Alimardani, M. and Qurashi, S. (2020). Mind perception of a sociable humanoid robot: a comparison between elderly and young adults. In *Robot 2019: Fourth Iberian Robotics Conference: Advances in Robotics, Volume 2*, pages 96–108. Springer.

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in human behavior*, 85:183–189.

Aroyo, A. M., De Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., Jones, S., Lutz, C., Sætra, H., Solberg, M., et al. (2021). Overtrusting robots: Setting a research agenda to mitigate overtrust in automation. *Paladyn, Journal of Behavioral Robotics*, 12(1):423–436.

Atherton, G. and Cross, L. (2018). Seeing more than human: Autism and anthropomorphic theory of mind. *Frontiers in psychology*, 9:528.

Aw, E. C.-X., Tan, G. W.-H., Cham, T.-H., Raman, R., and Ooi, K.-B. (2022). Alexa, what's on my shopping list? transforming customer experience with digital voice assistants. *Technological Forecasting and Social Change*, 180:121711.

Baek, T. H. and Kim, M. (2023). Is chatgpt scary good? how user motivations affect creepiness and trust in generative artificial intelligence. *Telematics and Informatics*, 83:102030.

Baier, A. (2014). Trust and antitrust. In *Feminist Social Thought*, pages 604–629. Routledge.

Bartneck, C. (2023). Godspeed questionnaire series: Translations and usage. In *International handbook of behavioral health assessment*, pages 1–35. Springer.

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

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009b). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81.

Bawack, R. E. (2021). How perceived intelligence affects consumer adoption of ai-based voice assistants: An affordance perspective. In *PACIS*, page 178.

Belanche, D., Casaló, L. V., Schepers, J., and Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The humanness-value-loyalty model. *Psychology & Marketing*, 38(12):2357–2376.

Bergmann, K., Eyssel, F., and Kopp, S. (2012). A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12*, pages 126–138. Springer.

Bernotat, J., Eyssel, F., and 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.

Berry, D. M. (2023). The limits of computation: Joseph weizenbaum and the eliza chatbot. *Weizenbaum Journal of the Digital Society*, 3(3).

Bhaskara, A., Skinner, M., and Loft, S. (2020). Agent transparency: A review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, 50(3):215–224.

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of management information systems*, 19(1):211–241.

Blut, M., Wang, C., Wünderlich, N. V., and 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*, 49:632–658.

Bradshaw, J. M. (1997). An introduction to software agents. *Software agents*, 4:3–46.

Broadbent, E., Kumar, V., Li, X., Sollers 3rd, J., Stafford, R. Q., MacDonald, B. A., and Wegner, D. M. (2013). Robots with display screens: a robot with

a more humanlike face display is perceived to have more mind and a better personality. *PloS one*, 8(8):e72589.

Brunk, J., Mattern, J., and Riehle, D. M. (2019). Effect of transparency and trust on acceptance of automatic online comment moderation systems. In *2019 IEEE 21st Conference on Business Informatics (CBI)*, volume 1, pages 429–435. IEEE.

Burleigh, T. J., Schoenherr, J. R., and Lacroix, G. L. (2013). Does the uncanny valley exist? an empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in human behavior*, 29(3):759–771.

Calahorra-Candao, G. and Martín-de Hoyos, M. J. (2024). The effect of anthropomorphism of virtual voice assistants on perceived safety as an antecedent to voice shopping. *Computers in Human Behavior*, 153:108124.

Caramazza, A., Anzellotti, S., Strnad, L., and Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annual review of neuroscience*, 37(1):1–15.

Carioli, P., Czarnitzki, D., and Fernández, G. P. (2024). Evidence on the adoption of artificial intelligence: The role of skills shortage. *ZEW-Centre for European Economic Research Discussion Paper*, (24-013).

Carolus, A. and Wienrich, C. (2022). "imagine this smart speaker to have a body": An analysis of the external appearances and the characteristics that people associate with voice assistants. *Frontiers in Computer Science*, 4:981435.

Castelfranchi, C. and Falcone, R. (2020). Trust: Perspectives in cognitive science. In *The Routledge Handbook of Trust and Philosophy*, pages 214–228. Routledge.

Cazier, J. A. (2007). A framework and guide for understanding the creation of consumer trust. *Journal of International Technology and Information Management*, 16(2):4.

Cernetic, J. (2003). On cost-effectiveness of human-centered and socially acceptable robot and automation systems. *Robotica*, 21(3):223–232.

Charles, E. P. (2017). The essential elements of an evolutionary theory of perception. *Ecological Psychology*, 29(3):198–212.

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Frontiers in psychology*, 4:108.

Chemero, A. P. (2009). *Radical Embodied Cognitive Science*. The MIT Press.

Chen, J., Guo, F., Ren, Z., Li, M., and Ham, J. (2024). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International journal of human–computer interaction*, 40(14):3636–3654.

Chen, Q. Q. and Park, H. J. (2021). How anthropomorphism affects trust in intelligent personal assistants. *Industrial Management & Data Systems*, 121(12):2722–2737.

Cheng, X., Zhang, X., Cohen, J., and Mou, J. (2022). Human vs. ai: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management*, 59(3):102940.

Christoforakos, L., Feicht, N., Hinkofer, S., Löscher, A., Schlegl, S. F., and Diefenbach, S. (2021). Connect with me. exploring influencing factors in a human-technology relationship based on regular chatbot use. *Frontiers in digital health*, 3:689999.

Christoforou, E. G. and Müller, A. (2016). Rur revisited: perspectives and reflections on modern robotics. *International Journal of Social Robotics*, 8:237–246.

Christov-Moore, L., Bolis, D., Kaplan, J., Schilbach, L., and Iacoboni, M. (2022). Trust in social interaction: from dyads to civilizations. *Social and affective neuroscience of everyday human interaction: from theory to methodology*, pages 119–141.

Clark, A. (1997). *Being there*. The MIT Press.

Clark, A. and Toribio, J. (1994). Doing without representing? *Synthese*, 101:401–431.

Clément, F. (2020). Trust: Perspectives in psychology. In *The Routledge Handbook of Trust and Philosophy*, pages 205–213. Routledge.

Clément, F., Bernard, S., Grandjean, D., and Sander, D. (2013). Emotional expression and vocabulary learning in adults and children. *Cognition & Emotion*, 27(3):539–548.

Coeckelbergh, M. (2011). Is ethics of robotics about robots? philosophy of robotics beyond realism and individualism. *Law, Innovation and Technology*, 3(2):241–250.

Coghlan, S. (2022). Robots and the possibility of humanistic care. *International journal of social robotics*, 14(10):2095–2108.

Cohn, M., Pushkarna, M., Olanubi, G. O., Moran, J. M., Padgett, D., Mengesha, Z., and Heldreth, C. (2024). Believing anthropomorphism: Examining the role of anthropomorphic cues on trust in large language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Cook, K. S., Hardin, R., and Levi, M. (2005). *Cooperation without trust?* Russell Sage Foundation.

Cook, K. S. and Santana, J. J. (2020). Trust: perspectives in sociology. In *The Routledge handbook of trust and philosophy*, pages 189–204. Routledge.

Craig, S. D. and Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, 114:193–205.

Cucciniello, I., Sangiovanni, S., Maggi, G., and Rossi, S. (2023). Mind perception in hri: Exploring users' attribution of mental and emotional states to robots with different behavioural styles. *International Journal of Social Robotics*, 15(5):867–877.

Dacey, M. (2017). Anthropomorphism as cognitive bias. *Philosophy of Science*, 84(5):1152–1164.

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., and Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3):331.

Deng, Y., Lei, W., Huang, M., and Chua, T.-S. (2023). Rethinking conversational agents in the era of llms: Proactivity, non-collaborativity, and beyond. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, pages 298–301.

Destéfano, M. (2021). Cognitivism and the intellectualist vision of the mind. *Phenomenology and Mind*, (21):142–153.

Diekmann, A., Jann, B., Przepiorka, W., and Wehrli, S. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American sociological review*, 79(1):65–85.

Dong, X. L., Moon, S., Xu, Y. E., Malik, K., and Yu, Z. (2023). Towards next-generation intelligent assistants leveraging llm techniques. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5792–5793.

Dotov, D. G., Nie, L., and De Wit, M. M. (2012). Understanding affordances: history and contemporary development of gibson's central concept. *Avant: the journal of the philosophical-interdisciplinary vanguard*.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3-4):177–190.

Edwards, B. I. and Cheok, A. D. (2018). Why not robot teachers: artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence*, 32(4):345–360.

ElDiwiny, M. (2023). An interview with will jackson. *Robotics Reports*, 1(1):11–18.

EMARKETER (2024). Voice assistants: What they are, how the benefit marketers, and what the ai future holds for them. [Accessed 28-11-2024].

Emirbayer, M. and Mische, A. (1998). What is agency? *American journal of sociology*, 103(4):962–1023.

Epley, N., Waytz, A., Akalis, S., and Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social cognition*, 26(2):143–155.

Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4):864.

Epley, N., Waytz, A., et al. (2010). Mind perception. *Handbook of social psychology*, 1(5):498–541.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., and Hegel, F. (2012). 'if you sound like me, you must be more human' on the interplay of robot and

user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 125–126.

Fernandes, T. and Oliveira, E. (2021). Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research*, 122:180–191.

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, pages 199–208. Springer.

Fogg, B. J. and Nass, C. (1997). Silicon sycophants: the effects of computers that flatter. *International journal of human-computer studies*, 46(5):551–561.

Følstad, A. and Brandtzaeg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, 5(1):3.

Franklin, S. and Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. In *International workshop on agent theories, architectures, and languages*, pages 21–35. Springer.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12):493–501.

Gambino, A., Fox, J., and Ratan, R. A. (2020). Building a stronger casa: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1:71–85.

Gams, M. and Kramar, S. (2024). Evaluating chatgpt's consciousness and its capability to pass the turing test: A comprehensive analysis. *Journal of Computer and Communications*, 12(03):219–237.

Glenberg, A. M., Witt, J. K., and Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on psychological science*, 8(5):573–585.

Goldberg, S. C. (2020). Trust and reliance 1. *The routledge handbook of trust and philosophy*, pages 97–108.

Gong, L. and Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research*, 33(2):163–193.

Goodman, K. L. and Mayhorn, C. B. (2023). It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics*, 106:103864.

Gray, C. E., Chesser, A., Atchley, A., Smitherman, R. C., and Tenhundfeld, N. L. (2022). Humanlikeness and aesthetic customization's effect on trust, performance, and affect. In *2022 Systems and Information Engineering Design Symposium (SIEDS)*, pages 403–408. IEEE.

Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. *science*, 315(5812):619–619.

Grazzini, L., Viglia, G., and Nunan, D. (2023). Dashed expectations in service experiences. effects of robots human-likeness on customers' responses. *European Journal of Marketing*, 57(4):957–986.

Grech, M. (2019). Where'nothing ever was': anthropomorphic spectrality and the (im) possibility of the post-anthropocene. *New Formations*, 95(95):22–36.

Guthrie, S. E. (1995). *Faces in the clouds: A new theory of religion*. Oxford University Press.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527.

Hardin, R. (1993). The street-level epistemology of trust. *Politics & society*, 21(4):505–529.

Haring, K. S., Watanabe, K., and Mougenot, C. (2013). The influence of robot appearance on assessment. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 131–132. IEEE.

Hashimoto, K. and Takanishi, A. (2015). Biped robot research at waseda university. *J. Robotics Netw. Artif. Life*, 1(4):261–264.

Haugeland, I. K. F., Følstad, A., Taylor, C., and Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, 161:102788.

Heo, M., Lee, K. J., et al. (2018). Chatbot as a new business communication tool: The case of naver talktalk. *Business Communication Research and Practice*, 1(1):41–45.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., and Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, 317(5843):1360–1366.

Hoffman, D. (2019). Do we see reality? *New Scientist*, 243(3241):34–37.

Hornbæk, K. and Oulasvirta, A. (2017). What is interaction? In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 5040–5052.

Hortensius, R., Kent, M., Darda, K. M., Jastrzab, L., Koldewyn, K., Ramsey, R., and Cross, E. S. (2021). Exploring the relationship between anthropomorphism and theory-of-mind in brain and behaviour. *Human brain mapping*, 42(13):4224–4241.

Hsu, W.-C. and Lee, M.-H. (2023). Semantic technology and anthropomorphism: Exploring the impacts of voice assistant personality on user trust, perceived risk, and attitude. *Journal of Global Information Management (JGIM)*, 31(1):1–21.

Hudson, B. (2004). Trust: Towards conceptual clarification. *Australian Journal of Political Science,*, 39(1):75–87.

Hupcey, J. E., Penrod, J., Morse, J. M., and Mitcham, C. (2001). An exploration and advancement of the concept of trust. *Journal of advanced nursing*, 36(2):282–293.

Hurmuz, M. Z., Jansen-Kosterink, S. M., Flierman, I., del Signore, S., Zia, G., del Signore, S., and Fard, B. (2023). Are social robots the solution for shortages in rehabilitation care? assessing the acceptance of nurses and patients of a social robot. *Computers in Human Behavior: Artificial Humans*, 1(2):100017.

Iavazzo, C., Gkegke, X.-E. D., Iavazzo, P.-E., and Gkegkes, I. D. (2014). Evolution of robots throughout history from hephaestus to da vinci robot. *Acta medico-historica adriatica: AMHA*, 12(2):247–258.

Ischen, C., Araujo, T. B., Voorveld, H. A., Van Noort, G., and Smit, E. G. (2022). Is voice really persuasive? the influence of modality in virtual assistant interactions and two alternative explanations. *Internet Research*, 32(7):402–425.

Jenneboer, L. (2022). Interaction effects between anthropomorphic chatbot characteristics on the customer experience: appearance, language style, and emoji-use. Master's thesis, University of Twente.

Jeong, S., Aymerich-Franch, L., Alghowinem, S., Picard, R. W., Breazeal, C. L., and Park, H. W. (2023). A robotic companion for psychological well-being: A long-term investigation of companionship and therapeutic alliance. In *Proceedings of the 2023 ACM/IEEE international conference on human-robot interaction*, pages 485–494.

Jo, H. (2023). Understanding ai tool engagement: A study of chatgpt usage and word-of-mouth among university students and office workers. *Telematics and Informatics*, 85:102067.

Johannsen, F., Schaller, D., and Klus, M. F. (2021). Value propositions of chatbots to support innovation management processes. *Information Systems and e-Business Management*, 19(1):205–246.

Jones, K. (2004). Trust and terror. In DesAutels, P. and Walker, M. U., editors, *Moral Psychology: Feminist Ethics and Social Theory*, pages 3–18. Rowman & Littlefield.

Jones, K. (2012). Trustworthiness. *Ethics*, 123(1):61–85.

Kamide, H., Eyssel, F., and Arai, T. (2013). Psychological anthropomorphism of robots: measuring mind perception and humanity in japanese context. In *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings 5*, pages 199–208. Springer.

Keijsers, M., Bartneck, C., and Eyssel, F. (2022). Pay them no mind: the influence of implicit and explicit robot mind perception on the right to be protected. *International Journal of Social Robotics*, 14(2):499–514.

Keren, A. (2014). Trust and belief: a preemptive reasons account. *Synthese*, 191(12):2593–2615.

Keren, A. (2020). Trust and belief. In *The Routledge Handbook of Trust and Philosophy*, pages 109–120. Routledge.

Kim, J. and Im, I. (2023). Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior*, 139:107512.

Kim, K. J., Park, E., and Sundar, S. S. (2013). Caregiving role in human–robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior*, 29(4):1799–1806.

Koban, K. and Banks, J. (2024). It feels, therefore it is: Associations between mind perception and mind ascription for social robots. *Computers in Human Behavior*, 153:108098.

Kok, B. C. and Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1(4):297–309.

Konya-Baumbach, E., Biller, M., and von Janda, S. (2023). Someone out there? a study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, 139:107513.

Kox, E. S., Kerstholt, J. H., Hueting, T. F., and de Vries, P. W. (2021). Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous agents and multi-agent systems*, 35(2):30.

Krämer, N. C., von der Pütten, A., and Eimler, S. (2012). Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction. *Human-computer interaction: The agency perspective*, pages 215–240.

Kuhail, M. A., Thomas, J., Alramlawi, S., Shah, S. J. H., and Thornquist, E. (2022). Interacting with a chatbot-based advising system: Understanding the effect of chatbot personality and user gender on behavior. In *Informatics*, volume 9, page 81. MDPI.

Kühne, R. and Peter, J. (2023). Anthropomorphism in human–robot interactions: a multidimensional conceptualization. *Communication Theory*, 33(1):42–52.

Kumar, A., Capraro, V., and Perc, M. (2020). The evolution of trust and trustworthiness. *Journal of the Royal Society Interface*, 17(169):20200491.

Kwon, M., Jung, M. F., and Knepper, R. A. (2016). Human expectations of social robots. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 463–464. IEEE.

LaGrandeur, K. (2012). Robots, moving statues, and automata in ancient tales and history. In *Critical Insights: Technology and Humanity*, pages 99–111. Salem Press Carol Colatrella. Ipswich, MA.

Lankton, N. K. and McKnight, D. H. (2011). What does it mean to trust facebook? examining technology and interpersonal trust beliefs. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 42(2):32–54.

Lankton, N. K., McKnight, D. H., and Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, 16(10):1.

Larsson, S. and Heintz, F. (2020). Transparency in artificial intelligence. *Internet policy review*, 9(2).

Law, E. L.-C., Følstad, A., and Van As, N. (2022). Effects of humanlikeness and conversational breakdown on trust in chatbots for customer service. In *Nordic Human-Computer Interaction Conference*, pages 1–13.

Law, T., Chita-Tegmark, M., and Scheutz, M. (2021). The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study. *International Journal of Social Robotics*, 13(2):297–309.

Lee, E.-J. (2010). What triggers social responses to flattering computers? experimental tests of anthropomorphism and mindlessness explanations. *Communication Research*, 37(2):191–214.

Lee, H. K. and Yoon, N. (2022). Factors driving fashion chatbot reliability-focusing on the mediating effect of perceived intelligence and positive cognition. *Fashion & Textile Research Journal*, 24(2):229–240.

Lee, I. and Hahn, S. (2024). On the relationship between mind perception and social support of chatbots. *Frontiers in Psychology*, 15:1282036.

Lee, M., Lucas, G., and Gratch, J. (2021). Comparing mind perception in strategic exchanges: Human-agent negotiation, dictator and ultimatum games. *Journal on Multimodal User Interfaces*, 15:201–214.

Lee, S. and Jeon, J. (2024). Visualizing a disembodied agent: Young efl learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*, 37(5-6):1048–1073.

Lee, S., Lee, N., and Sah, Y. J. (2020). Perceiving a mind in a chatbot: effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human–Computer Interaction*, 36(10):930–940.

Lee, S., Park, G., and Chung, J. (2023). Artificial emotions for charity collection: A serial mediation through perceived anthropomorphism and social presence. *Telematics and Informatics*, 82:102009.

Leopold, D. A. and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in cognitive sciences*, 3(7):254–264.

Letheren, K. and Glavas, C. (2017). Embracing the bots: How direct to consumer advertising is about to change forever. *The conversation*.

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37.

Li, M. and Suh, A. (2022). Anthropomorphism in ai-enabled technology: A literature review. *Electronic Markets*, 32(4):2245–2275.

Li, X. and Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in user–ai assistant interactions. *Computers in Human Behavior*, 118:106680.

Liu, J., Li, J., Feng, L., Li, L., Tian, J., and Lee, K. (2014). Seeing jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, 53:60–77.

Lu, L., Zhang, P., and Zhang, T. C. (2021). Leveraging "human-likeness" of robotic service at restaurants. *International Journal of Hospitality Management*, 94:102823.

Lykov, A. and Tsetserukou, D. (2023). Llm-brain: Ai-driven fast generation of robot behaviour tree based on large language model. *arXiv preprint arXiv:2305.19352*.

Madhavan, P. and Wiegmann, D. A. (2004). A new look at the dynamics of human-automation trust: Is trust in humans comparable to trust in machines? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 581–585. SAGE Publications Sage CA: Los Angeles, CA.

Madhavan, P., Wiegmann, D. A., and Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human factors*, 48(2):241–256.

Maeda, T. and Quan-Haase, A. (2024). When human-ai interactions become parasocial: Agency and anthropomorphism in affective design. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1068–1077.

Malle, B. F. (2019). How many dimensions of mind perception really are there? In *CogSci*, pages 2268–2274.

Malodia, S., Ferraris, A., Sakashita, M., Dhir, A., and Gavurova, B. (2023). Can alexa serve customers better? ai-driven voice assistant service interactions. *Journal of Services Marketing*, 37(1):25–39.

Malodia, S., Islam, N., Kaur, P., and Dhir, A. (2024). Why do people use artificial intelligence (ai)-enabled voice assistants? *IEEE Transactions on Engineering Management*, 71:491–505.

Mariacher, N., Schlögl, S., and Monz, A. (2021). Investigating perceptions of social intelligence in simulated human-chatbot interactions. *Progresses in Artificial Intelligence and Neural Systems*, pages 513–529.

Markman, A. B. and Dietrich, E. (2000a). Extending the classical view of representation. *Trends in cognitive sciences*, 4(12):470–475.

Markman, A. B. and Dietrich, E. (2000b). In defense of representation. *Cognitive Psychology*, 40(2):138–171.

Märtins, J., Kehl, R., Westmattelmann, D., and Schewe, G. (2022). A multidimensional theory of transparency, trust, and risk in technology adoption. In *International Conference on Information Systems (ICIS) Proceedings*, volume 7.

Maulsby, D., Greenberg, S., and Mander, R. (1993). Prototyping an intelligent agent through wizard of oz. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 277–284.

Mayer, R. (1995). An integrative model of organizational trust. *Academy of Management Review*.

Mcknight, D. H., Carter, M., Thatcher, J. B., and Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, 2(2):1–25.

McKnight, D. H., Choudhury, V., and Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3):334–359.

Meijer, A. (2014). Transparency. In *The Oxford Handbook of Public Accountability*, pages 507–524. Oxford University Press, London, U.K.

Mende, M., Scott, M. L., Van Doorn, J., Grewal, D., and 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.

Meskó, B., Hetényi, G., and Győrffy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC health services research*, 18:1–4.

Mileounis, A., Cuijpers, R. H., and Barakova, E. I. (2015). Creating robots with personality: The effect of personality on social intelligence. In *Artificial Computation in Biology and Medicine: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2015, Elche, Spain, June 1-5, 2015, Proceedings, Part I 6*, pages 119–132. Springer.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in cognitive sciences*, 7(3):141–144.

Minh, D., Wang, H. X., Li, Y. F., and Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, pages 1–66.

Mishra, A., Shukla, A., and Sharma, S. K. (2022). Psychological determinants of users' adoption and word-of-mouth recommendations of smart voice assistants. *International Journal of Information Management*, 67:102413.

Mithen, S. (1996). The prehistory of the mind: a search for the origins of art, science and religion. *London and New York: Thames and Hudson*.

Mollahosseini, A., Abdollahi, H., Sweeny, T. D., Cole, R., and Mahoor, M. H. (2018). Role of embodiment and presence in human perception of robots' facial cues. *International Journal of Human-Computer Studies*, 116:25–39.

Mostafa, R. B. and Kasamani, T. (2022). Antecedents and consequences of chatbot initial trust. *European journal of marketing*, 56(6):1748–1771.

Mou, W., Ruocco, M., Zanatto, D., and Cangelosi, A. (2020). When would you trust a robot? a study on trust and theory of mind in human-robot interactions. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 956–962. IEEE.

Müller, L., Mattke, J., Maier, C., Weitzel, T., and Graser, H. (2019). Chatbot acceptance: A latent profile analysis on individuals' trust in conversational agents. In *Proceedings of the 2019 on Computers and People Research Conference*, pages 35–42.

Nanay, B. (2015). The representationalism versus relationalism debate: Explanatory contextualism about perception. *European Journal of Philosophy*, 23(2):321–336.

Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103.

Nass, C., Moon, Y., and Green, N. (1997). Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of applied social psychology*, 27(10):864–876.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78.

Natarajan, M. and Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 33–42.

Nester, R. (2024). Social robots market size share, by component (hardware, software, service); application (designing, consulting, support maintenance); end-user (healthcare, education, entertainment, retail); technology (machine learning, computer vision, context awareness, natural language processing) - global supply demand analysis, growth forecasts, statistics report 2025-2037.

Nichols, S. and Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.

Nørskov, S., Damholdt, M. F., Ulhøi, J. P., Jensen, M. B., Ess, C., and Seibt, J. (2020). Applicant fairness perceptions of a robot-mediated job interview: a video vignette-based experimental survey. *Frontiers in Robotics and AI*, 7:586263.

Park, G., Chung, J., and Lee, S. (2024). Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. *Current Psychology*, 43(5):4352–4363.

Pentina, I., Hancock, T., and Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140:107600.

Pharoah, M. (2018). Qualitative attribution, phenomenal experience and being. *Biosemiotics*, 11(3):427–446.

Pi, Y., Liao, W., Liu, M., and Lu, J. (2008). Theory of cognitive pattern recognition. *Pattern recognition techniques, technology and applications*, page 626.

Pizzi, G., Vannucci, V., Mazzoli, V., and Donvito, R. (2023). I, chatbot! the impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. *Psychology & Marketing*, 40(7):1372–1387.

Placani, A. (2024). Anthropomorphism in ai: hype and fallacy. *AI and Ethics*, pages 1–8.

Platchias, D. (2004). The veil of perception and contextual relativism. *Sorites*, page 76.

Possati, L. M. (2023). Psychoanalyzing artificial intelligence: The case of replika. *AI & SOCIETY*, 38(4):1725–1738.

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services*, 58:102283.

Powers, A. and Kiesler, S. (2006). The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 218–225.

Prakash, A. and Rogers, W. A. (2015). Why some humanoid faces are perceived more positively than others: effects of human-likeness and task. *International journal of social robotics*, 7(2):309–331.

Proctor, R. W. and Proctor, J. D. (2012). Sensation and perception. *Handbook of human factors and ergonomics*, 61:59–94.

Quine, W. V. O. and Ullian, J. S. (1978). *The web of belief*, volume 2. Random House New York.

Rane, P., Mhatre, V., and Kurup, L. (2014). Study of a home robot: Jibo. *International journal of engineering research and technology*, 3(10):490–493.

Read, C. and Szokolszky, A. (2020). Ecological psychology and enactivism: perceptually-guided action vs. sensation-based enaction. *Frontiers in psychology*, 11:1270.

Reeves, B. and Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10(10).

Reuters (2023). Chatgpt sets record for fastest-growing user base - analyst note. [Accessed 28-11-2024].

Robbins, B. G. (2016). What is trust? a multidisciplinary review, critique, and synthesis. *Sociology compass*, 10(10):972–986.

Roesler, E., Manzey, D., and Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, 6(58):eabj5425.

Roesler, E., Manzey, D., and Onnasch, L. (2023). Embodiment matters in social hri research: Effectiveness of anthropomorphism on subjective and objective outcomes. *ACM Transactions on Human-Robot Interaction*, 12(1):1–9.

Rosenthal-von der Pütten, A. M., Krämer, N. C., and Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics*, 10:569–582.

Roy, R. and Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126:23–34.

Saltik, I., Erdil, D., and Urgen, B. A. (2021). Mind perception and social robots: The role of agent appearance and action types. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 210–214.

Schanke, S., Burtch, G., and Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research*, 32(3):736–751.

Schaumburg, H. (2001). Computers as tools or as social actors?—the users'perspective on anthropomorphic agents. *International journal of cooperative information systems*, 10(01n02):217–234.

Scheman, N. (2020). Trust and trustworthiness. In *The Routledge handbook of trust and philosophy*, pages 28–40. Routledge.

Schilke, O., Reimann, M., and Cook, K. S. (2021). Trust in social relations. *Annual Review of Sociology*, 47(1):239–259.

Schlosser, M. (2019). Agency. [Accessed 02-12-2024].

Schneider, F. and Hagmann, J. (2022). Assisting the assistant: how and why people show reciprocal behavior towards voice assistants. In *International Conference on Human-Computer Interaction*, pages 566–579. Springer.

Scott, L. S. and Fava, E. (2013). The own-species face bias: A review of developmental and comparative data. *Visual Cognition*, 21(9-10):1364–1391.

Seaborn, K., Miyake, N. P., Pennefather, P., and Otake-Matsuura, M. (2021). Voice in human–agent interaction: A survey. *ACM Computing Surveys (CSUR)*, 54(4):1–43.

Sheehan, B., Jin, H. S., and Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115:14–24.

Sheets-Johnstone, M. (2011). *The primacy of movement*. John Benjamins Publishing Company.

Shigemi, S., Goswami, A., and Vadakkepat, P. (2018). Asimo and humanoid robot research at honda. *Humanoid robotics: A reference*, 55:90.

Silver, C. A., Tatler, B. W., Chakravarthi, R., and Timmermans, B. (2021). Social agency as a continuum. *Psychonomic Bulletin & Review*, 28(2):434–453.

Simpson, T. W. (2012). What is trust? *Pacific Philosophical Quarterly*, 93(4):550–569.

Skjuve, M., Følstad, A., Fostervold, K. I., and Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149:102601.

Smestad, T. L. and Volden, F. (2019). Chatbot personalities matters: improving the user experience of chatbot interfaces. In *Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5*, pages 170–181. Springer.

Song, S. W. and Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human–Computer Interaction*, 40(2):441–456.

Spatola, N., Marchesi, S., and Wykowska, A. (2022). Different models of anthropomorphism across cultures and ontological limits in current frameworks the integrative framework of anthropomorphism. *Frontiers in Robotics and AI*, 9:863319.

Spaulding, S. (2010). Embodied cognition and mindreading. *Mind & Language*, 25(1):119–140.

Starmans, R. (2020). Prometheus unbound or paradise regained: the concept of causality in the contemporary ai-data science debate. *Journal de la société française de statistique*, 161(1):4–41.

Strongman, L. (2008). The anthropomorphic bias: How human thinking is prone to be self-referential. *The Open Polytechnic of New Zealand Working Papers*.

Swanepoel, D. (2021). Does artificial intelligence have agency? *The mind-technology problem: Investigating minds, selves and 21st century artefacts*, pages 83–104.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., Loggarakis, A., et al. (2020). User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research*, 22(3):e16235.

Tanis, M. and Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European journal of social psychology*, 35(3):413–424.

Technavio (2024). Chatbot market analysis: Us, canada, china, germany, uk - size and forecast 2023-2027.

Techopedia (2024). Tesla's optimus robot: All you need to know. [Accessed 28-11-2024].

Tusseyeva, I., Sandygulova, A., and Rubagotti, M. (2024). Perceived intelligence in human-robot interaction–a review. *IEEE Access*.

Tussyadiah, I. P., Zach, F. J., and Wang, J. (2020). Do travelers trust intelligent service robots? *Annals of Tourism Research*, 81:102886.

Ullman, D. and Malle, B. F. (2019). Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 618–619. IEEE.

Varela, F. J., Thompson, E., and Rosch, E. (2017). *The embodied mind, revised edition: Cognitive science and human experience*. The MIT press.

von Zitzewitz, J., Boesch, P. M., Wolf, P., and Riener, R. (2013). Quantifying the human likeness of a humanoid robot. *International Journal of Social Robotics*, 5:263–276.

Wagner, K. and Schramm-Klein, H. (2019). Alexa, are you human? investigating anthropomorphism of digital voice assistants-a qualitative approach. In *ICIS*.

Wang, J., Liu, Y., Yue, T., Wang, C., Mao, J., Wang, Y., and You, F. (2021). Robot transparency and anthropomorphic attribute effects on human–robot interactions. *Sensors*, 21(17):5722.

Wang, S., Lilienfeld, S. O., and Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4):393–407.

Wang, W. (2017). Smartphones as social actors? social dispositional factors in assessing anthropomorphism. *Computers in Human Behavior*, 68:334–344.

Ward, D., Silverman, D., and Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, 36:365–375.

Waytz, A., Cacioppo, J., and Epley, N. (2010a). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3):219–232.

Waytz, A., Gray, K., Epley, N., and Wegner, D. M. (2010b). Causes and consequences of mind perception. *Trends in cognitive sciences*, 14(8):383–388.

Waytz, A., Heafner, J., and Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology*, 52:113–117.

Weller, A. (2019). Transparency: motivations and challenges. In *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 23–40. Springer.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. The MIT Press.

Wienrich, C., Carolus, A., Markus, A., Augustin, Y., Pfister, J., and Hotho, A. (2023). Long-term effects of perceived friendship with intelligent voice assistants on usage behavior, user experience, and social perceptions. *Computers*, 12(4):77.

Wienrich, C., Ebner, F., and Carolus, A. (2022). Giving alexa a face-implementing a new research prototype and examining the influences of different human-like visualizations on the perception of voice assistants. In *International Conference on Human-Computer Interaction*, pages 605–625. Springer.

Wienrich, C., Reitelbach, C., and Carolus, A. (2021). The trustworthiness of voice assistants in the context of healthcare investigating the effect of perceived expertise on the trustworthiness of voice assistants, providers, data receivers, and automatic speech recognition. *Frontiers in Computer Science*, 3:685250.

Wu, J. (2024). Social and ethical impact of emotional ai advancement: the rise of pseudo-intimacy relationships and challenges in human interactions. *Frontiers in Psychology*, 15:1410462.

Yoganathan, V., Osburg, V.-S., Kunz, W. H., and Toporowski, W. (2021). Check-in at the robo-desk: Effects of automated social presence on social cognition and service implications. *Tourism Management*, 85:104309.

Yu, X., Liu, X., and Xu, Z. (2024). Adorable interactions: investigating the influence of ai voice assistant cuteness on consumer usage intentions. *Frontiers in Communication*, 9:1380775.

Zhang, M., Gursoy, D., Zhu, Z., and Shi, S. (2021). Impact of anthropomorphic features of artificially intelligent service robots on consumer acceptance: moderating role of sense of humor. *International Journal of Contemporary Hospitality Management*, 33(11):3883–3905.

Zhao, Y., Chen, Y., Sun, Y., and Shen, X.-L. (2024). Understanding users' voice assistant exploration intention: unraveling the differential mechanisms of the multiple dimensions of perceived intelligence. *Internet Research*.

Zhou, L.-F. and Meng, M. (2020). Do you see the "face"? individual differences in face pareidolia. *Journal of Pacific Rim Psychology*, 14:e2.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., and Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics*, 7:347–360.

Złotowski, J., Strasser, E., and Bartneck, C. (2014). Dimensions of anthropomorphism: from humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 66–73.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior/JAI Press*.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., and Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, 7:347–360.