Masterarbeit gemäß § 10 der Studien- und Prüfungsordnung vom 17.08.2009

im Masterstudiengang Strategic Information Management

an der Hochschule für angewandte Wissenschaften Neu-Ulm

The Impact of Anthropomorphic Design on Social Presence and Chatbot Gratitude Practices

Erstkorrektor: Prof. Dr. Andy Weeger

Zweitkorrektorin: Prof. Dr. Michaela Eßbach

Verfasser: Jonas Regending (Matrikel-Nr.: 334806)

Thema erhalten: 20.12.2024

Arbeit abgeliefert: 17.04.2025

Abstract

In an era marked by rising demand for accessible mental health interventions that take a preventive approach, digital solutions such as chatbot-facilitated gratitude practices offer scalable and user-friendly approaches to emotional well-being. This study critically examines whether anthropomorphic design features, such as conversational memory and human-like avatars, increase the effectiveness of gratitude interventions delivered via chatbots. Grounded in the Computers Are Social Actors (CASA) framework, the research explores the role of social presence as a mediator between chatbot design and psychological outcomes.

Through a randomized longitudinal study, participants engage in a three-day gratitude intervention delivered by either a socially enriched ("Rosa") or a neutral ("RoBot") chatbot. Key variables including gratitude expression, affect balance, positive and negative affect, and perceived social presence are measured at multiple time points. Contrary to theoretical expectations, the results do not reveal significant improvements in gratitude expression or well-being across the general sample. While participants interacting with Rosa report marginally higher social presence, these differences do not reach statistical significance. Likewise, the anthropomorphic design does not yield superior well-being outcomes compared to the neutral condition.

However, there are noteworthy nuances. Short-term improvements in affect balance are observed immediately after the intervention across both groups, suggesting a temporary emotional uplift. Furthermore, exploratory analyses indicate a potential trend toward increased gratitude among participants with initially low baseline gratitude, though this does not achieve statistical significance. These findings reflect broader challenges in digital health interventions, where short-duration and automated formats may fall short of producing sustained psychological change.

This study contributes to the literature by offering a nuanced evaluation of how anthropomorphic chatbot design interacts with user experience and psychological outcomes. It challenges the assumption that human-like features inherently enhance intervention efficacy and underscores the importance of considering user characteristics, intervention length, and contextual factors. While anthropomorphism may enrich interaction quality, its impact on emotional outcomes appears limited without deeper, sustained engagement.

By highlighting both the potential and limitations of chatbot-based gratitude practices, this research provides valuable insights for designers, practitioners, and researchers seeking to develop emotionally resonant and evidence-based digital mental health tools. Future studies should explore longer interventions, hybrid models with human facilitation, and adaptive personalization to better harness the relational potential of digital agents.

Keywords: anthropomorphism, chatbot, gratitude, digital intervention, social presence

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CA......Conversational Agent
CASA......Computers Are Social Actors
CBT......Cognitive Behavioral Therapy
CMC......Computer-Mediated Communication
DiGA......Digital Health Application
DMHIs......Digital Mental Health Intervention
GQ-6......Gratitude Questionnaire 6
HAI......Human-Agent Interaction
HCI......Human-Computer Interaction
HRI......Human-Robot Interaction
LLMs......Large Language Models
LMM......Linear Mixed Models
LTM......Long-Term Memory
mHealth......mobile Health
NLP......Natural Language Processing
NLU......Natural Language Understanding
PAI......Positive Activity Intervention
SPANE......Scale of Positive and Negative Experience
SPRES......Social Presence Scale
STM......Short-Term Memory
VR......Virtual Reality

1 Introduction

Society is experiencing a pervasive yet often understated mental health crisis of epidemic proportions. Nearly 1 in 8 people worldwide - approximately 970 million individuals - are living with a mental disorder as of 2019 (World Health Organization 2022). The most recent Eurobarometer report from 2023 revealed that nine in ten Europeans (89%) regard mental health promotion as equally important as physical health promotion. However, almost half (46%) of the respondents reported experiencing emotional or psychosocial problems, such as anxiety or depression, in the past year, with more than half (54%) of those affected not seeking professional help (European Commission 2023). This mental health crisis is not confined to Europe at large; its impacts are particularly visible in countries such as Norway and Germany. In Norway, longitudinal studies indicate that the prevalence of anxiety and depression among adolescents has nearly doubled over the past two decades, with 30% of adolescents reporting such symptoms during the period of 2017–2019 (Krokstad et al. 2022). Similarly, in Germany, the COVID-19 pandemic exacerbated existing mental health challenges, with depressive symptoms among adults peaking at record levels by 2022 (Mauz et al. 2023). These trends highlight the need for innovative, evidence-based approaches to the prevention of mental health issues. One promising avenue lies in the realm of positive psychology, particularly through gratitude-based interventions. Gratitude, often described as the recognition and appreciation of positive aspects in life, has been shown to foster emotional resilience, alleviate symptoms of depression and anxiety, and strengthen social bonds (Bohlmeijer et al. 2021). Regular gratitude practices, such as maintaining gratitude journals or reflecting on positive experiences, have been demonstrated to improve life satisfaction and reduce stress (Emmons and Stern 2013). These characteristics make gratitude an appealing, low-cost, and easily implementable strategy for addressing the mental health crisis. Despite the proven benefits of gratitude interventions, their scalability and accessibility remain significant challenges. Technological advancements offer a compelling solution. Digital tools, particularly chatbots powered by artificial intelligence (AI), provide a scalable and accessible platform for delivering gratitude interventions. Chatbots equipped with natural language processing (NLP) can mimic human-like interactions, fostering trust and engagement (Bae Brandtzæg et al. 2021). By delivering interventions in a user-friendly and nonjudgmental manner, these tools have the potential to reach individuals who might otherwise avoid traditional mental health services. The efficacy of chatbot-based gratitude interventions has already been demonstrated. Lee et al. conducted a study in which a chatbot delivered gratitude-focused interactions to participants over three days, incorporating activities such as recalling grateful moments, expressing appreciation, and reflecting on personal relationships. Results revealed significant increases in self-reported gratitude and positive affect, as well as reductions in negative emotions (Lee et al. 2024). The integration of anthropomorphic cues, such as names, visual avatars, and conversational memory, into chatbot design may further enhance their effectiveness. Research suggests that these human-like features can create a sense of social presence, fostering stronger emotional connections between users and the chatbot. This increased engagement can amplify the emotional depth and perceived authenticity of gratitude interventions (Pereira et al. 2014; Van Der Goot 2022).

However, while the role of anthropomorphic design in digital interactions has been explored in other domains, empirical research specifically examining these effects in the context of gratitude interventions remains scarce. The application of chatbots in gratitude interventions is still a relatively new field, and to date, few studies have investigated whether human-like chatbot features significantly improve user engagement, emotional depth, and overall intervention effectiveness (Lee et al. 2024)

This gap in the literature underscores the need for a more nuanced understanding of how anthropomorphic design elements can humanize AI, optimizing its role in mental health support. By transforming gratitude-based digital interventions into more engaging and impactful experiences, this study aims to bridge this research gap and provide empirical evidence on the effectiveness of anthropomorphic chatbot design in this context. Accordingly, this study seeks to investigate the following research question:

RQ: Can human-like design elements in chatbots enhance the effectiveness and emotional depth of gratitude interventions?

After the Introduction, the thesis continues with a discussion of the Background, providing an overview of the relevant theoretical foundations. It first explores the significance of Mental Health, emphasizing the growing prevalence of psychological distress and the need for innovative, accessible interventions. Following this, the role of Gratitude in mental well-being is examined, highlighting its psychological benefits and potential as an prevention strategy. The discussion then shifts to Digital Health Solutions, outlining the ways in which technological advancements are transforming mental health care. The concept of Conversational Agents is also introduced, focusing on their applications, advantages, and challenges in the context of digital mental health interventions. Furthermore, key theoretical constructs such as Social Presence, Anthropomorphic Cues, and Conversational Memory are explored to establish the psychological mechanisms underlying chatbot-based interactions.

Subsequently, the Methodology is outlined, detailing the Experimental Setup, including participant recruitment, study design, and intervention structure. The selection of Measures used to assess gratitude, emotional well-being, and social presence is also described, ensuring methodological rigor. Additionally, the Chatbot Implementation is explained, differentiating between a standard chatbot and an anthropomorphic chatbot designed to enhance user engagement. The Intervention Procedure is then described, outlining how gratitude exercises were structured and delivered.

Following this, the Results of the study are presented, including a Statistical Analysis of gratitude expression, emotional well-being, and perceived social presence. These findings are then discussed in relation to the research question, offering insights into the effectiveness of chatbot-based gratitude interventions. Building on these findings, the Discussion section contextualizes the results within existing literature. Theoretical contributions, practical implications, and limitations are considered, along with recommendations for future research.

The study employs a randomized quantitative longitudinal design to explore the impact of anthropomorphic features in chatbot-based gratitude interventions. Participants are randomly assigned to engage with either a neutral chatbot, which provides gratitude exercises in a straightforward manner, or an anthropomorphic chatbot designed with human-like attributes such as conversational memory and a visual avatar. These features are intended to foster a sense of social presence and enhance emotional engagement. The interventions take place over three consecutive days, with varied prompts to ensure sustained user interest. Emotional and psychological outcomes are measured at several stages, offering a thorough analysis of the effects of human-like design elements on user experience and emotional well-being.

2 Background

2.1 Mental Health

Mental health is a fundamental component of overall well-being, influencing how individuals think, feel, and interact with the world around them (Galderisi et al. 2015). The World Health Organization (WHO) defines mental health as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (World Health Organization 2004). This definition underscores the holistic nature of mental health, emphasizing that it is not just the absence of psychiatric disorders but a dynamic state that enables individuals to cope with life's stresses, work productively, and contribute to their communities. Furthermore, research highlights that mental health is shaped by a complex interplay of biological, psychological, and social determinants. Factors such as genetics, socioeconomic status, access to healthcare, and social support systems significantly influence an individual's mental well-being (Kottke et al. 2016; McGovern 2014; Schroeder 2007). Despite growing awareness of mental health's importance, global health policies and investments remain disproportionately focused on disease treatment rather than prevention and well-being promotion, leading to persistent disparities in mental health outcomes (Abernethy et al. 2022). Alternative definitions have been proposed that challenge the WHO's perspective. Galderisi et al. argue that defining mental health primarily through well-being and functionality risks exclude individuals who experience distress or impairment due to external conditions. Their alternative definition describes mental health as "a dynamic state of internal equilibrium which enables individuals to use their abilities in harmony with universal values of society." This definition places greater emphasis on resilience, emotional regulation, and adaptability to life's challenges rather than an idealized state of well-being. It acknowledges that individuals in good mental health may experience sadness, anger, or stress while still maintaining functional emotional regulation and coping mechanisms. This perspective shifts the focus from an absolute state of well-being to a more nuanced, adaptable model of mental health that accounts for contextual and cultural variations (Galderisi et al. 2015). Understanding these diverse perspectives on mental health is crucial for comprehending the growing challenges associated with mental health disorders in modern societies. With rising rates of anxiety and depression observed across demographics, mental health has become a critical global public health concern. Recent estimates reveal that over 970 million individuals worldwide were living with mental health disorders in 2019, with depressive and anxiety disorders being the most prevalent (Moitra et al. 2023). In Germany, epidemiological data show that annually about 27.8% of adults experience a mental health condition, translating to approximately 17.8 million individuals (Jacobi et al. 2016).

The COVID-19 pandemic has additionally impacted global mental health, with increased rates of anxiety, depression, and stress, particularly among vulnerable groups such as adolescents, healthcare workers, and individuals with pre-existing psychiatric conditions. Social isolation exacerbated these issues, leading to higher levels of loneliness and emotional distress, especially in younger populations. Healthcare professionals faced substantial psychological strain due to infection fears and high workloads, contributing to elevated burnout and trauma-related disorders. While some individuals demonstrated resilience over time, evidence suggests that long-term psychiatric consequences are still emerging. Moreover, COVID-19 infection itself may heighten the risk of subsequent mental health disorders, underscoring the need for ongoing psychosocial support and targeted interventions to mitigate the pandemic's lasting psychological effects (Manchia et al. 2022).

Globally, the demand for mental health services far exceeds the available resources. Low treatment coverage remains a significant barrier, especially in low- and middle-income countries, where only 8% of individuals with major depressive disorders receive adequate treatment (Moitra et al. 2023). But even in high-income nations like Germany, waiting times for psychotherapy can extend beyond several months, particularly for children and adolescents, where the demand outpaces the available slots (Bachmann et al. 2023). The reason for this is a shortage of mental health professionals, which is a multifaceted issue driven by systemic and demographic factors. According to the German Federal Ministry of Labour and Social Affairs medium-term workforce monitoring forecasts for 2022–2026, demand for healthcare professionals is projected to outstrip supply in 87 out of 140 occupational groups. In addition to fields such as medicine, nursing, and emergency medical services, this also includes psychotherapy and other mental health professions (BMAS 2022). Long-term projections up to 2040 indicate that the healthcare sector will face the largest increase in labor demand across all industries due to demographic changes. By 2040, it is estimated that 16.2% of the workforce in Germany will need to be employed in healthcare to meet the anticipated demand, compared to 13% today (BMAS 2023). The COVID-19 pandemic has further intensified workforce shortages in the mental health sector, exacerbating existing challenges. Research has highlighted increased burnout, job dissatisfaction, and emotional distress among mental health professionals due to heightened workloads, the transition to telehealth, and the psychological toll of treating patients during a crisis (Bae et al. 2022; Crocker et al. 2023).

Recruitment challenges for healthcare employers are expected to increase as vacancies take longer to fill. Contributing factors include an aging population, which not only increases demand for health services but also leads to a shrinking workforce as more professionals retire. Furthermore, the profession’s high stress levels, relatively low pay, and limited opportunities for advancement exacerbate recruitment and retention difficulties (Adams et al. 2019) These challenges are particularly acute in rural and underserved areas, where healthcare infrastructure is already strained (BMAS 2022).

These different and severe challenges associated with mental health necessitate the search for innovative ways to improve psychological well-being as a preventive factor for mental health. As the search for innovative approaches to enhancing mental health continues, gratitude has emerged as a promising yet underutilized tool (Portocarrero et al. 2020) In the next chapter, a deeper exploration of the different facets of gratitude will be undertaken, highlighting its multidimensional nature and how it manifests in daily life. Both the psychological and behavioral components of gratitude will be examined to establish a clearer understanding of its connection to mental health. This discussion will provide insights into how gratitude can be cultivated and harnessed to foster well-being and resilience in diverse populations.

2.2 Gratitude

Gratitude has emerged as a powerful intervention in the domain of mental health, demonstrating substantial benefits in fostering emotional resilience, alleviating stress, and mitigating symptoms of depression and anxiety. A growing body of research underscores its role in enhancing psychological well-being, strengthening social bonds, and contributing to overall life satisfaction (Portocarrero et al. 2020)

The word gratitude is derived from the Latin gratia, meaning favor, and gratus, meaning pleasing (Emmons and Stern 2013). Gratitude is a multifaceted psychological construct that has been conceptualized in various ways, including as an emotion, a virtue, an attitude, a habit, and even as a coping mechanism. Broadly defined, gratitude is the appreciation of what is valuable and meaningful to oneself, reflecting a general state of thankfulness and recognition of positive aspects in one’s environment (Sansone and Sansone 2010). This definition allows for gratitude to be understood both as a transient emotional state and as a more enduring personality trait.

Gratitude can be categorized into two distinct forms: "gratitude for" and "gratitude to." This distinction aligns with the theoretical framework differentiating between benefit-triggered gratitude, or "grateful to," and generalized gratitude, or "grateful for" (Lambert et al. 2009). The former refers to the gratitude experienced when receiving a specific benefit from another person, reinforcing interpersonal relationships and social cohesion. In contrast, generalized gratitude arises from an awareness and appreciation of broader life aspects, such as nature, health, or existence itself. Research indicates that individuals with a disposition toward generalized gratitude tend to have a lower threshold for experiencing gratitude, which fosters a more sustained sense of well-being (McCullough et al. 2002).

Gratitude can be conceptualized on three distinct levels: as a trait, an emotion, and a mood. These classifications reflect varying temporal and psychological dimensions of gratitude, each contributing uniquely to individual well-being and social functioning (Rosenberg 1998).

Gratitude as an emotion represents a transient, acute, and often intense psychophysiological response triggered by receiving a perceived benefit from another person or external circumstance. Unlike trait gratitude, which reflects a general disposition, emotional gratitude is situationally bound and arises spontaneously in response to specific events. This emotional experience not only enhances momentary well-being but also serves a critical social function by strengthening interpersonal relationships through reciprocity and acknowledgment of kindness. In contrast, gratitude as a mood operates on a broader temporal scale, exerting a subtle yet pervasive influence on an individual’s emotional state. Unlike an isolated emotional response, a grateful mood extends over time and shapes one’s overall affective experience. A sustained grateful mood has been associated with reduced susceptibility to negative emotions, increased resilience, and greater life satisfaction. The presence of a grateful mood suggests that gratitude can function as a self-reinforcing mechanism that contributes to long-term psychological stability and adaptive coping strategies (McCullough et al. 2004).

Trait gratitude refers to a stable dispositional characteristic that influences an individual’s tendency to experience and express gratitude across various situations. Individuals with high trait gratitude possess a lower threshold for recognizing and appreciating positive contributions from others. This disposition has been linked to enhanced psychological well-being, stronger social bonds, and a generally more optimistic outlook on life. Research suggests that trait gratitude fosters a habitual cognitive orientation toward recognizing and valuing benevolence, thereby reinforcing positive emotions and prosocial behavior over time (Emmons and Mishra 2011).

As an affective trait, gratitude comprises several facets, or aspects, which are distinct yet interconnected dimensions of how gratitude is experienced and expressed. Facets help to break down complex traits into measurable components, providing a deeper understanding of how individuals engage with gratitude. The first facet is intensity, referring to how deeply a person feels gratitude when encountering positive experiences. Those with higher intensity experience stronger emotional responses to acts of kindness. The second facet, frequency, describes how often gratitude is felt in daily life, ranging from occasional moments to a continuous awareness of positive influences. Another important facet is span, which indicates the breadth of life domains for which one feels grateful. A broad gratitude span means acknowledging a variety of aspects, such as relationships, health, or opportunities. Lastly, density reflects how many people are recognized for a single positive outcome. Some individuals attribute success to a wide network of supporters, while others focus on just a few key contributors (McCullough et al. 2002).

Understanding these facets provides a foundation for developing effective gratitude interventions. These structured exercises, classified as a form of positive activity intervention (PAI), are designed to enhance well-being by encouraging individuals to focus on and express gratitude in meaningful ways. These interventions are simple, low-cost cognitive behavioral strategies that mirror the thoughts and behaviors of naturally happy individuals, such as expressing thankfulness, optimism, or kindness, to improve psychological and emotional well-being (Layous and Lyubomirsky 2014). Research suggests that gratitude interventions significantly enhance positive affect while simultaneously reducing negative emotions. Practicing daily gratitude over two weeks has been shown to increase subjective happiness, enhance life satisfaction, and reduce depressive symptoms, highlighting its potential in promoting long-term emotional well-being (Cunha et al. 2019).

Similarly, synthesized findings from multiple studies show that gratitude interventions not only improve psychological outcomes, such as life satisfaction and mood, but also exhibit promising effects on physical health, including enhanced sleep quality and reduced markers of inflammation (Boggiss et al. 2020). Moreover, gratitude is closely linked to improved cardiovascular responses and stress regulation. A study explored the psychophysiological mechanisms underpinning gratitude expression, finding that individuals who engaged in gratitude texting exhibited favorable cardiovascular responses, including reduced total peripheral resistance, which is indicative of lower stress levels. The study further highlighted that individuals with high levels of depression and low trait gratitude were less likely to engage in gratitude interventions, suggesting that individual differences in personality traits influence the effectiveness of such practices (Enko et al. 2021).

Gratitude interventions include structured practices that help cultivate appreciation and enhance psychological resilience. Different forms of these interventions have been identified, each serving distinct functions. These interventions can generally be categorized into three primary types: gratitude journaling, gratitude letters, and behavioral expressions of gratitude (Wood et al. 2010).

Gratitude journaling, one of the most researched interventions, involves regularly writing down things one is grateful for. Studies indicate that this practice contributes to increased positive emotions, higher life satisfaction, and reduced depressive symptoms. Participants in such studies have shown long-term improvements in emotional well-being and stronger social connections. By focusing on positive life events, journaling reinforces thought patterns that support happiness and emotional stability. This form of gratitude falls under generalized gratitude ("grateful for"), as it promotes appreciation for broader life experiences rather than gratitude directed at a specific person (Emmons and McCullough 2003).

Another widely used intervention is writing gratitude letters, in which individuals express appreciation for others through written messages. Research has shown that this practice can lead to significant improvements in mental health, with benefits lasting for weeks after the intervention. Unlike journaling, which encourages personal reflection, writing gratitude letters strengthens relationships by fostering social bonds and promoting positive interactions. Expressing gratitude in this way has been associated with greater trust, empathy, and overall psychological resilience, all of which contribute to enhanced emotional well-being (Wong et al. 2018). This intervention aligns with benefit-triggered gratitude ("grateful to"), as it involves expressing thankfulness to a specific person who has provided a benefit.

Behavioral expressions of gratitude, such as verbally expressing thanks or engaging in acts of kindness, further reinforce gratitude’s social dimension. These interventions emphasize active engagement with others and can lead to reciprocal positive interactions, thereby deepening social connections. Studies have found that individuals who consistently practice gratitude through verbal or behavioral expressions experience greater levels of subjective well-being and social support (Wood et al. 2010). Like gratitude letters, this intervention is classified as benefit-triggered gratitude ("grateful to"), since it directly acknowledges a specific benefactor and fosters social reciprocity.

The effectiveness of gratitude interventions, however, is influenced by the phenomenon of hedonic adaptation. Hedonic adaptation refers to the psychological phenomenon whereby individuals tend to return to a baseline level of subjective well-being following positive or negative life changes. It is described as the tendency of affective responses to diminish over time in response to persistent changes in life circumstances. This adaptation occurs due to cognitive and affective mechanisms, such as shifts in attention, expectation adjustments, and the re-normalization of experiences. Consequently, individuals often experience only transient changes in happiness after major life events, as they gradually habituate to both positive and negative circumstances (Frederick and Loewenstein 1999).

Individuals often become habituated to gratitude exercises, leading to a decline in their impact over time. This hedonic adaptation suggests that, to sustain benefits, gratitude practices must incorporate variety and novelty. Alternating between different gratitude exercises, increasing the depth of reflection, or integrating new gratitude-focused activities can help maintain engagement and prolong the psychological benefits of these interventions. Given the potential for adaptation, researchers propose several strategies to enhance the long-term efficacy of gratitude interventions. The Hedonic Adaptation Prevention Model suggests that introducing elements of surprise, personalization, and social reinforcement can counteract the diminishing returns of repetitive gratitude exercises. For instance, gratitude interventions embedded within digital platforms, such as AI-driven chatbots, can leverage adaptive algorithms to present novel gratitude exercises tailored to individual user preferences and engagement patterns (Fritz et al. 2017).

While gratitude is often regarded as a positive and socially beneficial emotion, several critical aspects warrant scrutiny. One concern is its potential to reinforce power imbalances in relationships. When expressed within hierarchical structures, such as employer-employee or caregiver-recipient dynamics, it can perpetuate dependence and inhibit autonomy. Individuals may feel obligated to express gratitude, even when support stems from expected duties or social contracts. Moreover, gratitude can function as a mechanism for conformity, pressuring individuals to accept unfavorable circumstances. For example, those in unjust situations may feel compelled to show gratitude despite inequity, discouraging critique or advocacy. This aligns with gratitude’s potential role in fostering complacency rather than empowerment. Additionally, gratitude can contribute to emotional burden. Some individuals experience “gratitude fatigue,” where the expectation to continually express thankfulness becomes psychologically draining. This is especially evident in caregiving contexts, where recipients may feel guilty for not showing enough appreciation despite a legitimate need for support (Buck 2004).

Gratitude, as a key component of positive psychology, has been shown to enhance emotional resilience and foster psychological well-being through structured interventions such as reflective journaling and gratitude lists. While these practices have demonstrated efficacy in improving mental health, their long-term impact and scalability remain challenges. To address these limitations, digital health solutions offer promising avenues for expanding the reach and effectiveness of gratitude-based interventions. Emerging technologies are enabling new ways to integrate gratitude into everyday routines, making mental health support more accessible and engaging.

2.3 Digital Health Solutions

Over the past decades, digital technologies have profoundly transformed various aspects of human activity, including healthcare. These developments have sparked significant debates concerning their advantages and challenges. The integration of mechanical and digital systems for recording, analyzing, and utilizing medical data has laid the foundation for revolutionary advancements in individual patient care, population-wide health strategies, and real-time generation of medical insights. Collectively, these advancements are categorized under the umbrella term of digital health.

Digital health refers to the integration of digital technologies into healthcare to enhance medical outcomes, streamline healthcare delivery, and empower individuals in managing their health. It encompasses a wide range of innovations, including mobile health (mHealth), telemedicine, artificial intelligence, wearable devices, and health informatics. These technologies facilitate real-time data collection, predictive analytics, and personalized medicine, optimizing both preventive and therapeutic interventions (Fatehi et al. 2020). The WHO defines digital health as the use of digital tools to improve public health and healthcare accessibility, particularly in underserved populations (WHO 2020). Beyond clinical care, digital health also includes public health surveillance and behavioral interventions.

The early uses of technology in health can be traced back to the 1960s, when computers first facilitated information processing in medical settings. During this period, biofeedback technology emerged, allowing individuals to monitor physiological functions such as heart rate and muscle tension, thereby enabling better management of stress and certain medical conditions (Blanchard 1982). By the 1980s and 1990s, telephone-based health interventions became a pivotal development, providing remote access to healthcare services. Tobacco quitlines, for instance, leveraged telephone counseling to support individuals seeking to stop smoking, demonstrating the scalability of remote healthcare solutions (Cummins et al. 2007). As mobile phone adoption increased in the early 2000s, text messaging became a widely used tool for health interventions. Programs such as Text4baby, which provided health education to pregnant women via SMS, exemplified the potential of mobile communication in improving public health outcomes (Evans et al. 2012). The proliferation of smartphones in the late 2000s marked a turning point in digital health, with mHealth applications becoming increasingly prevalent. These apps enabled real-time health monitoring, personalized interventions, and interactive health coaching. Health apps incorporating evidence-based behavioral strategies, such as self-monitoring and goal-setting, can effectively promote healthier lifestyles (Payne et al. 2015). In parallel with the rise of mHealth applications, social media platforms emerged as influential tools for health promotion and engagement. By the 2010s, health organizations and researchers increasingly utilized platforms such as Facebook, Twitter, and Instagram to disseminate health information, foster peer support networks, and conduct digital health interventions (Pagoto and Waring 2016). Online health communities provided individuals with chronic conditions the opportunity to connect, share experiences, and receive emotional support. However, the spread of misinformation and privacy concerns remain significant challenges in this space (Wicks et al. 2018).

Wearable technology has further expanded the possibilities of digital health, offering continuous health tracking through devices like Fitbit, Apple Watch, and other smart wearables. These devices enable users to monitor physical activity, sleep patterns, heart rate, and other health metrics, providing real-time feedback and motivation for maintaining healthy behaviors (Evenson et al. 2015).

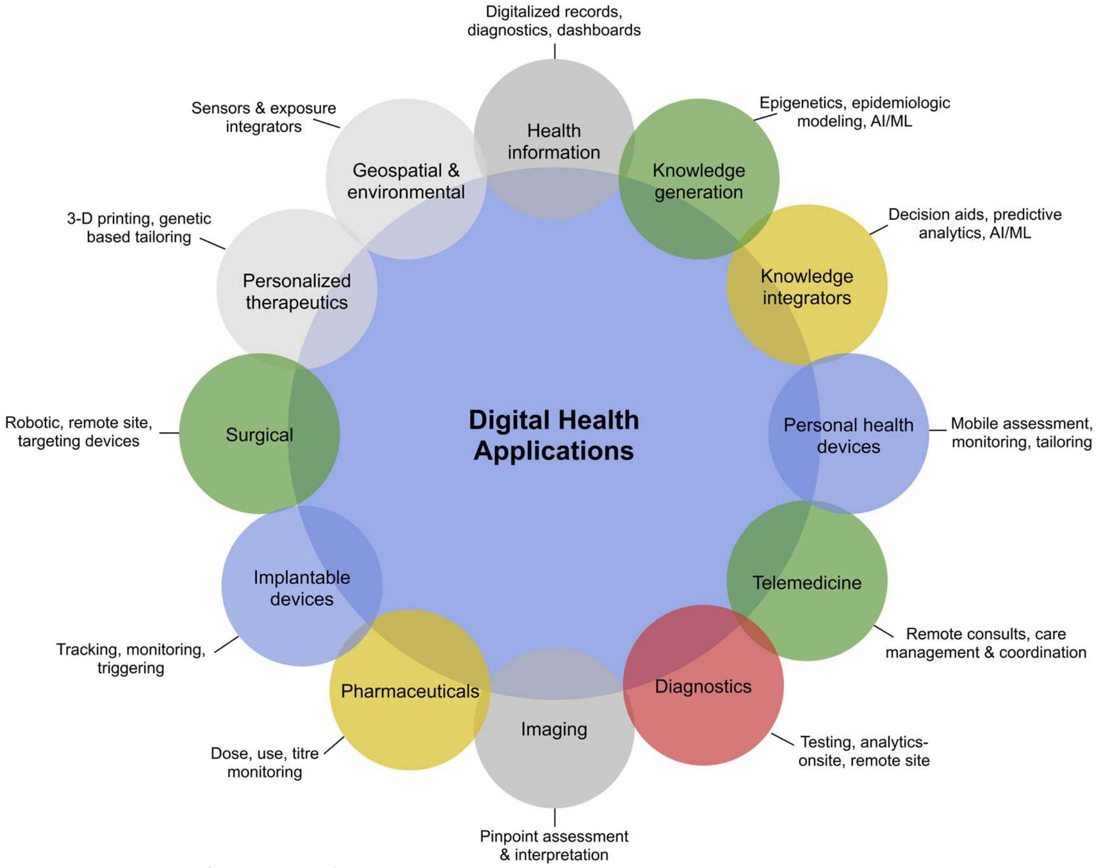

Figure 1: Overview of Digital Health Applications (based on: Abernethy et al. 2022, p. 3)

Figure 1 illustrates the broad scope of digital health applications, categorizing digital tools into twelve distinct areas of healthcare and public health. The figure emphasizes how digital technologies have permeated every aspect of health and healthcare, facilitating innovations in diagnosis, treatment, and patient management. These digital advancements contribute to ensuring continuity of care, enabling remote patient management through telemedicine, and empowering individuals to take an active role in their health. Additionally, digital health tools support system-wide improvements by reducing medical errors and inefficiencies in healthcare delivery (Abernethy et al. 2022). The diagram also highlights digital solutions aimed at improving population health and addressing upstream determinants of health, such as geospatial and environmental sensors, personal health devices, and knowledge integrators. These applications extend beyond traditional medical care, integrating data from diverse sources to enhance health outcomes and preventive care strategies. Notably, the importance of digital health tools in integrating social services into healthcare delivery became evident during the COVID-19 pandemic, when disparities in healthcare access and outcomes were exacerbated for marginalized and economically disadvantaged communities (Isasi et al. 2021). By leveraging digital technologies, healthcare systems can better identify and mitigate these structural disparities, advancing equity in healthcare provision. When implemented effectively, digital health technologies foster long-term partnerships between individuals and healthcare providers, reinforcing positive health behaviors and improving patient engagement. Additionally, if thoughtfully designed and equitably deployed, these tools have the potential to reduce disparities in healthcare access and quality (Craig et al. 2020).

Digital health technologies offer substantial benefits, particularly in enhancing accessibility, efficiency, and patient empowerment within healthcare systems. One of the most significant advantages is the improved access to healthcare services, particularly for older adults and individuals with mobility limitations. Digital health platforms enable remote consultations, telemedicine services, and mobile health applications, reducing the need for physical visits and thereby increasing convenience and timely access to medical care. Furthermore, digital solutions facilitate patient self-management by providing health monitoring tools, medication reminders, and real-time data tracking, which support proactive disease prevention and chronic condition management (Kainiemi et al. 2023)

Additionally, digital health technologies contribute to cost-effectiveness by optimizing resource allocation, minimizing administrative burdens, and reducing hospital admissions. They also foster greater patient engagement through personalized health information, facilitating informed decision-making and promoting healthier lifestyles. From a broader perspective, digital health services enable healthcare providers to leverage big data analytics for early disease detection, trend analysis, and personalized treatment strategies (Kainiemi et al. 2023).

Digital health technologies offer significant advancements in healthcare but also introduce substantial privacy risks. One major concern is the exposure of sensitive personal health information through data breaches, unauthorized access, or improper handling. Digital health ecosystems rely on interconnected devices, including mobile apps, wearable sensors, and cloud-based platforms, which create multiple points of vulnerability. Cyber threats such as hacking, phishing, and malware attacks have become primary causes of security breaches, often leading to medical identity theft, fraud, or unauthorized data sharing. Additionally, the increasing use of mobile health applications poses risks, as many apps collect and transmit personal health data without adequate encryption or user consent, sometimes sharing information with third parties, including advertisers. The aggregation of health data with financial, behavioral, or genomic information further amplifies the risk of re-identification, even when data is anonymized (Filkins et al. 2016).

Another significant risk of digital health applications is the potential exclusion of individuals who either lack access to or do not wish to engage with these technologies. Vulnerable populations, including older adults, migrants, and individuals with mental health conditions, frequently encounter barriers to digital health services due to insufficient digital literacy, limited access to suitable devices, or inadequate local language skills. Digital health services are often perceived as impersonal or ineffective, leading to reluctance in adoption. Additionally, digital alternatives are not always available or sufficiently promoted, leaving certain populations unaware of their potential benefits. Concerns regarding privacy and data security further discourage engagement with digital health solutions. To ensure equitable healthcare access, the maintenance of traditional face-to-face services alongside digital options remains essential, addressing disparities in digital health adoption (Kaihlanen et al. 2022).

There are also challenges related to validation and regulatory oversight. In particular, many commercial health apps lack rigorous scientific validation, raising concerns about their efficacy and adherence to clinical guidelines (Jake-Schoffman et al. 2017). The absence of standardized evaluation criteria makes it difficult for patients, clinicians, and payers to distinguish effective digital health solutions from those with unproven benefits. While some digital health applications undergo rigorous clinical validation, many rely on limited evidence or fail to demonstrate meaningful clinical impact. Furthermore, the rapid pace of development in the digital health sector often leads to gaps in regulatory oversight, allowing some products to enter the market before undergoing comprehensive safety and efficacy evaluations. To address these issues, a structured validation framework to ensure digital health solutions meet predefined standards is needed (Mathews et al. 2019).

Germany has pioneered a proactive approach to this challenge and integrated digital solutions into healthcare with the introduction of Digital Health Applications (DiGAs) under the Digital Healthcare Act in 2019. DiGAs are certified medical applications that can be prescribed by physicians and reimbursed by statutory health insurance providers. These digital health applications play a crucial role in disease management, rehabilitation, and preventive care, empowering patients with tools for monitoring and self-management (Sauermann et al. 2022).

A particularly significant area within digital health is mental health support, as psychological well-being is deeply interconnected with overall health. Digital Mental Health Interventions (DMHIs) play a central role in mental health-focused e-health applications, providing scalable, cost-effective, and personalized care solutions. DMHIs refer to structured interventions that utilize digital technologies such as mobile applications, websites, text messaging, and wearable devices to promote health behavior change, offer remote therapeutic support, and enhance self-management strategies. These interventions target a broad spectrum of health issues, including chronic disease management, lifestyle modifications, and mental health conditions. For mental health promotion, DMHIs are typically complex, multi-component systems integrating various evidence-based strategies. These may include psychoeducation, cognitive-behavioral therapy (CBT), mindfulness training, and interactive self-assessment tools. By enabling individuals to monitor their health states, receive personalized feedback, and engage in behavior change processes, DMHIs support users through automated algorithms or human guidance. Additionally, these interventions facilitate improved communication between patients and healthcare providers, providing structured decision-making aids and data-driven insights to optimize treatment outcomes. The effectiveness of DMHIs depends on their accessibility, usability, and engagement, necessitating iterative adaptation and rigorous evaluation to ensure sustained impact (Murray et al. 2016). The effectiveness of DMHIs in mental health has been extensively studied, demonstrating positive outcomes in mental health promotion, prevention, and treatment. Systematic reviews indicate that eHealth and telemedical interventions improve stress management, well-being, and symptom reduction in anxiety and depression. Furthermore, these interventions have shown efficacy across different populations, including adolescents, adults, and high-risk groups, with effect sizes ranging from small to medium. Blended-care approaches, combining digital and conventional therapeutic components, have exhibited superior outcomes compared to standalone digital interventions, particularly for anxiety and depressive disorders (Rauschenberg et al. 2021).

DMHIs leverage a diverse range of technologies, reflecting broader advancements in digital health. These technologies include synchronous and asynchronous communication methods such as videoconferencing, telephone consultations, and messaging-based interactions (Philippe et al. 2022). Furthermore, web-based and mobile therapy programs offer structured psychological interventions, either in guided or self-administered formats (Andersson et al. 2019). Virtual reality (VR) applications have demonstrated efficacy in exposure therapy and immersive cognitive-behavioral interventions, while cognitive training programs delivered via web-based platforms and mobile applications are increasingly employed to address neurodevelopmental and neurocognitive disorders (Botella et al. 2017). Additionally, peer support networks facilitated through social media and online forums provide community-based interaction and psychoeducational resources (Mendes-Santos et al. 2020). However, the effectiveness and implementation of these technologies vary considerably across different mental health conditions. Synchronous digital interventions, such as video-based psychotherapy and telepsychiatry, have demonstrated effectiveness comparable to traditional face-to-face therapy, particularly for anxiety and depressive disorders (Torous et al. 2020). Conversely, asynchronous modalities, including SMS-based reminders and email-based interventions, are primarily employed to support adherence and symptom monitoring rather than as standalone therapeutic tools. Evidence suggests that their efficacy in acute treatment phases is limited, underscoring the need for practitioner involvement (Philippe et al. 2022).

VR-based exposure therapy has proven particularly effective for anxiety disorders, PTSD, and phobia treatments by offering immersive environments that enable controlled exposure to stress-inducing stimuli (Botella et al. 2017). Cognitive training programs targeting executive function and memory have been successfully utilized in neurocognitive and psychiatric disorders, including schizophrenia and dementia (Torous et al. 2020). Web-based and mobile applications have emerged as widely accessible mental health care tools, particularly for anxiety, mood, and trauma-related disorders (Andersson et al. 2019). Digital therapy programs have been successfully implemented for treating anxiety and depressive disorders (Mendes-Santos et al. 2020).

Several factors contribute to the effectiveness of DMHIs, including the quality of therapeutic content, accessibility, personalization, and user engagement. Among these, engagement has emerged as a critical determinant of intervention success, influencing both adherence and clinical outcomes. Studies have consistently shown that users who actively engage with DMHIs experience greater improvements in mental health symptoms, highlighting the need to optimize engagement strategies to maximize intervention efficacy. Higher engagement levels in DMHIs are associated with significant reductions in symptoms of depression, anxiety, and psychological distress. However, low engagement remains a major challenge, with many users failing to complete intervention modules or dropping out early. Research indicates that over 70% of users do not complete all treatment modules, and more than half disengage before reaching the intervention’s midpoint. These findings suggest that while DMHIs hold great potential, their real-world effectiveness depends on strategies that foster sustained engagement (Gan et al. 2021). Various engagement strategies have been identified to address these challenges. Personalized feedback, reminders, and gamification elements, such as rewards and progress tracking, have been shown to improve user adherence. Social support mechanisms, including peer forums and e-coaching, further enhance engagement by fostering a sense of community and accountability. Additionally, interventions that offer flexibility and intuitive design increase retention by integrating seamlessly into users’ daily routines. These strategies collectively contribute to higher adherence and improved intervention outcomes, underscoring the need for engagement-focused design principles in DMHIs (Saleem et al. 2021).

Building on the importance of user engagement in digital mental health interventions, Conversational Agents have emerged as a promising solution. These AI-powered tools enhance accessibility, offer personalized support, and maintain user involvement, addressing many of the challenges traditional DMHIs face (Boucher et al. 2021).The following section explores the potential of Conversational Agents in revolutionizing mental health care delivery.

2.4 Conversational Agents

Conversational agents (CAs) are software systems that interact with users through natural language, either via text or speech. These agents can process user inputs, generate responses, and, in some cases, perform specific actions in virtual or physical environments. The term "Conversational Agent" is often used interchangeably with "Chatbot," but they have distinct characteristics. A Conversational Agent is a broad category encompassing any system that engages in dialogue with users. These systems may use advanced technologies such as deep learning and natural language understanding (NLU) to provide contextual and adaptive interactions. They can be text-based, voice-based, or embodied as virtual or physical agents (Allouch et al. 2021).

A Chatbot, on the other hand, is a subset of conversational agents, typically designed to simulate human-like conversations with predefined replies or response models. Chatbots often employ rule-based or retrieval-based techniques, whereas more advanced conversational agents may integrate generative models and deep learning for richer interactions (Allouch et al. 2021).

The distinction between the two lies in their functionality and complexity: while all chatbots are conversational agents, not all conversational agents are chatbots. Some conversational agents perform task-specific actions, such as virtual assistants that help with scheduling or customer service bots that assist users in e-commerce platforms. Others, such as social bots, focus on maintaining open-domain dialogues for engagement and entertainment (Hussain et al. 2019). For the purpose of this study, we will focus on chatbots, as they represent a fundamental and widely used category of conversational agents.

The history of chatbots can be traced back to Alan Turing's famous 1950 question of whether a machine could exhibit human-like conversation, known as the Turing Test (Turing 1950). One of the first notable chatbots was ELIZA, developed in 1966, which simulated a psychotherapist through pattern-matching techniques. Although it lacked deep understanding, it demonstrated that simple rule-based interactions could create the illusion of conversation (Weizenbaum 1966). In 1972, PARRY introduced a more sophisticated conversational model by simulating a patient with schizophrenia. Unlike ELIZA, PARRY implemented a system of assumptions and emotional responses, making it the first chatbot to incorporate personality traits and mimic human mental states to a limited degree (Colby et al. 1971). A major milestone was ALICE, introduced in 1995, which leveraged the Artificial Intelligence Markup Language (AIML) for structured chatbot responses. ALICE improved upon ELIZA's rule-based approach by expanding its knowledge base and making responses more contextually relevant, though it still lacked true understanding and learning capabilities (Wallace 2009). The 2001 release of SmarterChild on platforms like AOL Messenger showcased chatbots as practical tools for retrieving real-world information. Unlike ALICE, SmarterChild could fetch dynamic data such as sports scores, weather updates, and stock prices, demonstrating the potential for chatbots as digital assistants rather than just conversational partners. The era of virtual assistants began with Apple Siri (2010), which introduced voice recognition and integrated with various apps to provide users with practical assistance. IBM Watson (2011) took chatbot capabilities even further by utilizing deep learning and vast knowledge bases, famously defeating human champions in the quiz show Jeopardy! Watson’s ability to process and analyze large amounts of unstructured data made it revolutionary in AI-driven responses. Google Assistant (2016), Amazon Alexa, and Microsoft Cortana followed, focusing on seamless voice interaction, smart home integration, and personalization. These assistants marked a shift from purely text-based interfaces to voice-activated AI, enabling more natural interactions with users (Adamopoulou and Moussiades 2020a).

In the recent years, Large Language Models (LLMs) have revolutionized the field of artificial intelligence, demonstrating unprecedented capabilities in natural language processing and generation. These models, built on transformer architectures, are pre-trained on vast amounts of textual data, enabling them to understand, generate, and adapt language in a highly contextual manner. A key innovation in LLMs is in-context learning, which allows models to interpret and generate responses dynamically without requiring explicit retraining for new tasks. This has significantly enhanced their versatility across diverse applications, from conversational AI to code generation and content creation. Another fundamental advancement in LLMs is the integration of reinforcement learning from human feedback, which aligns model outputs with human preferences, improving response coherence, factual accuracy, and ethical considerations (Wu et al. 2023).

Chatbots can be classified into several categories based on their functionality and purpose as visualized in Figure 2. One primary classification criterion is their knowledge domain: generic chatbots can handle a wide range of topics, whereas domain-specific chatbots focus on particular areas, such as customer support or healthcare (Nimavat and Champaneria 2017).

Another classification is based on the chatbot's goal. Informative chatbots provide users with specific data, such as weather updates or news. Conversational chatbots aim to engage users in free-flowing dialogue, whereas task-oriented chatbots assist users in performing actions like booking a hotel or ordering food (Kucherbaev et al. 2018). Additionally, chatbots can be classified based on their level of autonomy: fully autonomous chatbots operate without human intervention, while human-mediated chatbots require occasional human oversight (Zumstein and Hundertmark 2017).

Lastly, chatbots differ based on their interaction mode, including text-based chatbots, voice-based assistants, and multimodal chatbots that incorporate text, voice, and even images for interaction (Shum et al. 2018).

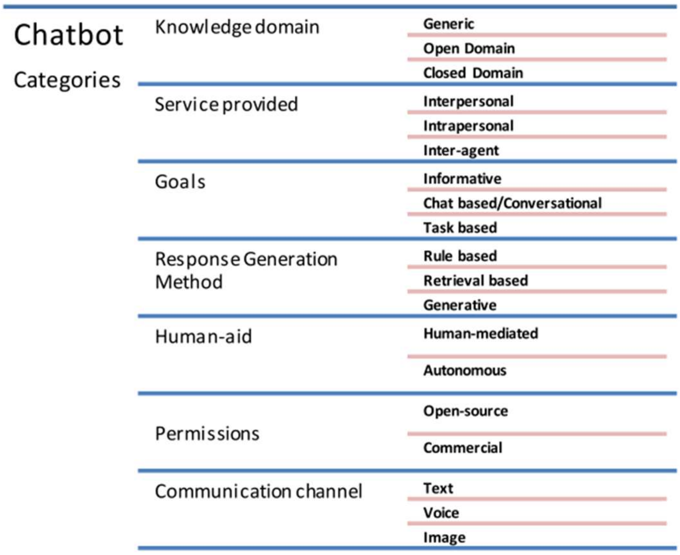

Figure 2: Overview of Chatbot Categories (Adamopoulo and Moussiades 2020a, p. 5)

The development of chatbots follows two primary approaches: pattern matching and machine learning. Pattern matching chatbots use predefined rules to match user input with stored responses. Early chatbots like ELIZA and ALICE relied on this approach, using structured templates and scripts to generate replies (Wallace 2009). While pattern matching ensures quick responses, it lacks flexibility and cannot handle complex conversations effectively (Marietto et al. 2013). In contrast, machine learning-based chatbots leverage NLP and neural networks to generate responses dynamically. These chatbots can be further divided into retrieval-based models, which select the most appropriate response from a predefined database, and generative models, which construct responses from scratch using deep learning techniques like recurrent neural networks and transformers (Xu et al. 2017).

The popularity of chatbots stems from their widespread applications in fields such as education, healthcare, and customer service. Unlike traditional interfaces, chatbots provide instant interaction without requiring users to switch between applications. Their platform-independent nature and ability to integrate seamlessly into messaging apps make them particularly appealing (Klopfenstein et al. 2017). Additionally, they offer functionalities such as automated notifications, payment integration, and multi-threaded conversations (Adamopoulou and Moussiades 2020b).

Given their potential, Chatbots have also increasingly gained significance as digital health interventions due to their scalability, efficiency, and user convenience (Xue et al. 2023). Their anonymous interaction capability makes them especially valuable for underserved or stigmatized populations, who might otherwise avoid seeking support (Abd-alrazaq et al. 2019; Boucher et al. 2021).

The application fields of chatbots in digital health are diverse and encompass diagnostic and screening processes, particularly in mental health conditions such as depression and anxiety (Boucher et al., 2021). They also facilitate therapeutic interventions and counseling, employing approaches like CBT, mindfulness practices, positive psychology, and motivational interviewing (Abd-Alrazaq et al., 2019; Boucher et al., 2021). Specifically, gratitude interventions delivered via chatbots have been demonstrated to significantly enhance users' gratitude and overall positive emotions while reducing negative emotions, highlighting the complex and beneficial emotional experience associated with gratitude (Lee et al., 2024). Additionally, chatbots support symptom and behavioral management, including sleep tracking, physical activity monitoring, and mood tracking (Boucher et al., 2021). They contribute to self-management and health monitoring, assisting users in maintaining healthy behaviors and routines. Furthermore, chatbots serve educational and training purposes, aiding in developing social skills and vocational competencies, and play an essential role in prevention and relapse prevention strategies (Abd-Alrazaq et al., 2019; Boucher et al., 2021).

The effectiveness of chatbots is documented through small to medium effects on mental health symptoms such as anxiety and depression (Boucher et al., 2021). Generally, users demonstrate a positive attitude and high acceptance of chatbots, though critical perspectives regarding trustworthiness and emotional depth also exist (Boucher et al., 2021). Enhanced user engagement is achieved through regular interaction points and personalized experiences. Nevertheless, significant challenges remain, notably the lack of robust evidence from rigorous research and safety concerns, especially in crisis situations involving suicidal ideation (Xue et al., 2023; Boucher et al., 2021).

Key factors enhancing chatbot effectiveness include their ability to establish relational capacity through social dialogues, empathy, humor, self-disclosure, and meta-relational communication (regular check-ins on conversation quality) (Xue et al., 2023). Personalizing interactions to users' needs further improves user experiences and engagement (Xue et al., 2023). Moreover, strategies such as frequent check-ins, personalized reminders, and adaptive feedback are effective in maintaining user participation (Xue et al., 2023; Boucher et al., 2021).

Critical ethical considerations include the chatbot's response capability in crisis situations, such as suicidal thoughts, highlighting the importance of safety measures (Xue et al., 2023; Abd-Alrazaq et al., 2019). Ensuring data protection and security is essential to build and maintain user trust (Xue et al., 2023; Boucher et al., 2021). There is a pressing need for rigorous study designs, standardized evaluation methods, and comprehensive analyses of real-world user experiences to further establish chatbot efficacy (Xue et al., 2023; Boucher et al., 2021).

Looking forward, the advancement of AI and NLP will enhance realistic interactions and facilitate hyper-personalization of chatbot interactions (Boucher et al., 2021). Comprehensive empirical validation remains crucial to substantiate the effectiveness of chatbots compared to traditional digital interventions (Boucher et al., 2021).

Understanding how users interact with and respond to chatbots involves exploring broader theoretical frameworks. One such perspective is the Computers Are Social Actors (CASA) framework, which provides insight into user perceptions and behaviors toward technology-based interactions, setting the stage for a deeper investigation into human-chatbot interactions.

2.5 Computers Are Social Actors

The Computers Are Social Actors (CASA) framework, originating from the "Media Equation" theory, posits that humans tend to interact with machines and digital media as if they were social entities. It suggests that individuals apply social scripts and norms, typically used in human-human interactions, to their interactions with media agents, even when they are fully aware that these are non-human entities (Reeves and Nass 1996). CASA was first conceptualized by Nass & Moon in 2000, building on earlier work by Nass et al. 1994. It highlights that humans mindlessly apply learned social behaviors to computers when they exhibit social cues. These cues trigger automatic social responses, which facilitate interaction with technology in ways analogous to interpersonal communication. This paradigm is widely utilized in fields such as human-computer interaction (HCI), human-agent interaction (HAI), and human-robot interaction (HRI) (Gambino et al. 2020).

According to the CASA paradigm, humans focus on specific cues from computers (e.g., a female avatar) that lead them to categorize the computer as a social entity (e.g., perceiving the computer as female) while ignoring its non-human nature (Nass and Moon 2000). As a result, people instinctively apply social rules, expectations, and communication scripts to computers, such as associating gender stereotypes with them (Nass et al. 1994). It is argued that human brains evolved in an environment where only humans exhibited social behavior (Nass and Moon 2000). To efficiently navigate social interactions, the brain developed automatic social responses, which are triggered by any entity displaying social-like behavior. This process occurs subconsciously, leading individuals to engage with computers in a social manner without realizing it (Fogg 2002).

The CASA paradigm provides a theoretical framework for understanding how users interact with technology in a manner similar to human-human interactions. One of the key constructs associated with CASA is social presence, which refers to the psychological perception of a medium or an artificial entity as a real social actor. Social presence plays a crucial role in explaining users' reactions to emerging technologies, particularly those designed with human-like characteristics such as social robots, chatbots, and voice assistants (Xu et al. 2022).

2.6 Social Presence

Social presence, defined as the subjective experience of being with a "real" person and having access to their thoughts and emotions (Biocca 1997), is a crucial factor in understanding human interactions in virtual environments. It is fundamentally linked to the degree to which a virtual entity, whether controlled by a human or an artificial agent, is perceived as socially real and capable of meaningful interaction (Lee et al. 2006).

To contextualize social presence, it is essential to first understand the broader constructs of immersion, telepresence, and self-presence. Immersion is a technological property that describes the extent to which a medium can generate perceptually realistic experiences, thus detaching users from their physical reality (Slater and Wilbur 1997). The level of immersion is determined by the quality of sensory input, including visual fidelity, spatialized audio, and haptic feedback (Cummings and Bailenson 2016; Welch et al. 1996). Telepresence, or spatial presence, refers to the sensation of being physically situated within a mediated environment rather than in one's immediate surroundings (Steuer 1992). When telepresence is strong, users often experience a perceptual shift, losing awareness of their actual physical space and becoming fully engaged with the virtual one (Lombard and Ditton 2006). Self-presence concerns the degree to which users identify with their virtual self or avatar, experiencing it as an extension of their real-world identity (Ratan and Hasler 2009).

In contrast to these dimensions, social presence is inherently relational; it requires an entity that is perceived as socially and cognitively responsive (Biocca et al. 2003). The core of social presence lies in the perception of a mediated other as a sentient and socially relevant counterpart, making it a fundamental element in virtual environments that facilitate social interactions (Lee et al. 2006).

The concept of social presence was originally developed by Short et al. in 1976 in the context of communication media and was initially defined as the degree of salience of an interlocutor in a mediated interaction. Short et al. (1976) emphasized two primary factors shaping social presence: intimacy, or the emotional closeness experienced in an interaction, and immediacy, referring to the psychological proximity between communicators. These attributes were thought to be influenced primarily by the richness of a communication medium, with more immersive and interactive technologies facilitating higher levels of social presence. This early view of social presence as a medium-dependent construct was reinforced by media richness theory (Daft and Lengel 1986), which argued that different communication technologies vary in their ability to transmit nuanced social cues such as facial expressions, gestures, and vocal tone. This perspective led to the assumption that face-to-face interactions inherently offer the highest level of social presence, while text-based communication is significantly limited in this regard (Gunawardena and Zittle 1997). However, subsequent research challenged this medium-deterministic approach, highlighting the role of adaptive human behavior in shaping social presence (Walther 1992). According to Social Information Processing Theory, users are capable of compensating for the lack of nonverbal cues in text-based communication by developing alternative strategies, such as extended response times and richer textual expressions (Walther 1996). This adaptive capacity suggests that social presence is not solely dictated by technological affordances but is also shaped by user behavior and contextual factors (Ramirez et al. 2002).

Social presence plays a critical role in determining the effectiveness and impact of virtual interactions. One of its most significant functions is its influence on social influence and persuasion (Fogg and Tseng 1999). When users perceive a high level of social presence in virtual environments, they tend to exhibit stronger emotional responses and greater susceptibility to persuasion. For instance, Lee et al. (2006) found that individuals who experienced higher social presence in an interaction with a virtual agent reported increased attraction and engagement. Beyond persuasion, social presence has been shown to enhance trust and user experience in online interactions. Hassanein and Head (2007) demonstrated that heightened social presence in an online shopping context led to greater consumer trust, increased enjoyment, and stronger purchase intentions. Similarly, in educational and collaborative virtual settings, increased social presence has been linked to greater engagement and cooperation among users (Biocca et al. 2003). Moreover, social presence is a key determinant in the effectiveness of virtual agents and avatars. Whether interacting with a humanoid robot, a virtual instructor, or a digital assistant, users are more likely to engage meaningfully when they perceive the entity as socially and emotionally responsive (Lee et al. 2006). This underscores the importance of designing virtual environments that optimize the factors contributing to social presence, such as immersive sensory input, naturalistic avatar behavior, and real-time interactivity (Cummings and Bailenson 2016).

Having established the significance of social presence, this thesis explores two approaches to enhancing it in digital interactions. Given its impact on user experience, researchers have investigated multiple strategies to reinforce this perception within human-computer interactions. In the next section, one of the most prominent strategies for enhancing social presence will be examined: the use of anthropomorphic cues. These design elements, which include human-like visual features, conversational styles, and behavioral adaptations, have been widely studied for their role in improving user engagement and perceived relational depth in digital interactions (Oh et al. 2018).

2.7 Anthropromorphic Cues

Social cues refer to design elements in human-computer interaction that facilitate social responses by users, particularly in interactions with conversational agents (CAs). These cues trigger human-like reactions by leveraging verbal, visual, auditory, and invisible communication channels. The CASA paradigm suggests that people tend to subconsciously apply social norms to computers when these cues are present, fostering a perception of social presence and relational depth in digital interactions. Various terminologies have been used to describe social cues across different research domains. In human-computer interaction and social signal processing, they are often referred to as social cues, social signals, anthropomorphic features, or human-like characteristics (Feine et al. 2019).

Social cues can be divided into several distinct categories, shown in Figure 3, each influencing user interaction differently. These categories include verbal, visual, auditory, and invisible cues, each of which contributes uniquely to the perception of social presence and emotional engagement in human-computer interactions.

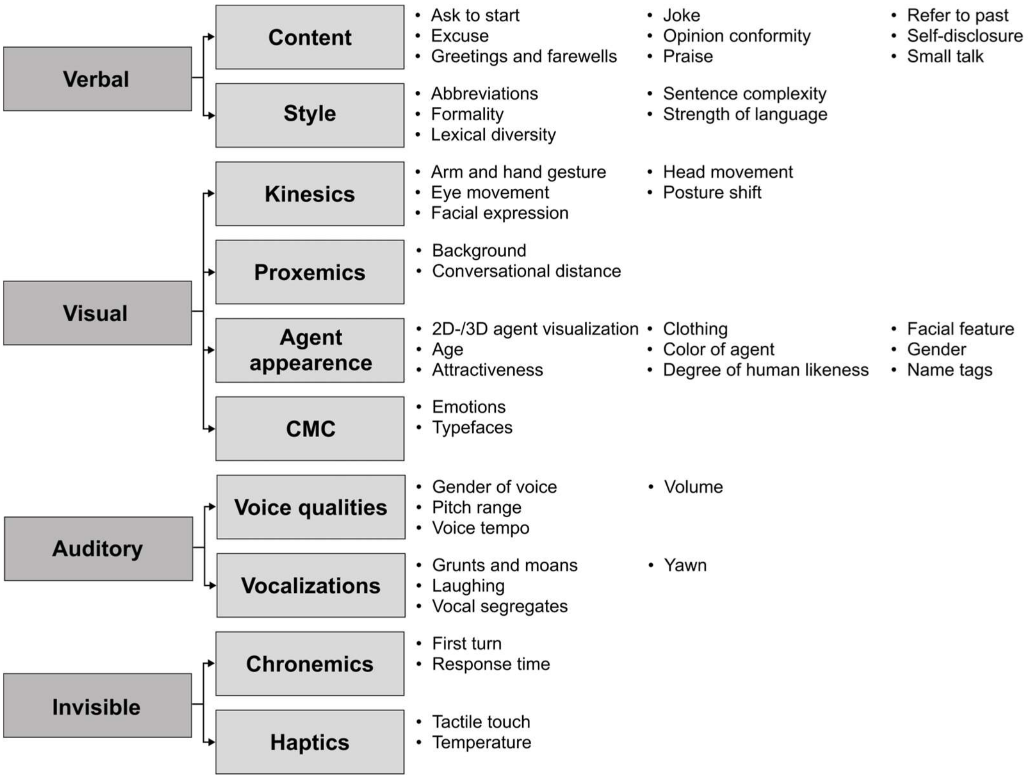

Figure 3: Overview of Social Cues (based on: Feine et. al. 2019, p. 31)

Verbal cues encompass all social signals conveyed through language. Everything what people write or say belongs to the discourse of an interaction, which represents "social action made visible in language" (Antaki 2008, p.2). Language use in human-computer interaction can be analyzed through multiple frameworks, including syntax, semantics, and pragmatics (Titscher et al. 2000). Additionally, verbal cues can be divided into content and style elements (Collier 2014). Content cues refer to the strict meaning of the words used, while style cues pertain to how language is expressed, including aspects like conversational behavior, lexicon, and speech modulation (Mairesse et al. 2007). The deployment of verbal cues in chatbots enables the establishment of social presence by fostering engagement and perceived human-likeness (Walther 2006).

Visual cues encompass all nonverbal elements that are visually perceptible. These cues include kinesics (body language and gestures), proxemics (use of space and distance), and artifacts (appearance, clothing, and accessories) (Eaves and Leathers 2017). For conversational agents, agent appearance plays a crucial role in enhancing social presence, as avatars and expressive facial features can simulate human-like interactions. Moreover, digital elements such as typefaces and emoticons serve as additional visual cues, expanding the interpretability of textual messages in computer-mediated communication (CMC) (Candello et al. 2017).

Auditory cues, often referred to as vocalics or prosody, include nonverbal sounds and voice-related features (Burgoon 1994). These cues are categorized into voice set, voice qualities, and vocalizations (Trager 1958). Voice set refers to quasi-permanent characteristics such as gender and age, while voice qualities involve dynamic features like pitch variation and speech tempo (Nöth 1995). Vocalizations, including laughter, sighs, and hesitation markers ("uh-huh," "mhm"), further contribute to the perception of authenticity and engagement in conversation (James 2017).

Invisible cues, or "silent language" elements, include chronemics, haptics, and olfactory communication (Eaves and Leathers 2017). Chronemics refers to temporal elements of communication, such as response latency and pacing, which influence perceptions of immediacy and attentiveness. Haptic cues involve tactile interactions, such as simulated touch in virtual environments, which play a role in emotional engagement (Burgoon 1994). Olfactory communication, while relevant in face-to-face interactions, is largely absent in digital communication and is thus excluded from chatbot design considerations (Eaves and Leathers 2017).

Social cues can also be classified based on the cognitive processing they trigger. This distinction introduces the concepts of mindful and mindless cues, which determine how users perceive and interpret these interactions. Mindful cues require deliberate cognitive processing and conscious attribution of human-like qualities to a system. This process is characterized by users actively recognizing and evaluating the system's human-like features, such as a virtual assistant's facial expressions or the use of natural language in conversations. These cues engage systematic information processing, where users critically assess the system's behavior before forming judgments about its credibility or social presence. In contrast, mindless cues trigger automatic and unconscious social responses to computers. These responses arise due to prolonged exposure to human-like digital entities or interactive interfaces that mimic human communication patterns. Mindless anthropomorphism is often driven by heuristic processing, where individuals rely on simple cues, such as the presence of a human-like avatar or a conversational tone, to form immediate impressions without deep analysis. Their study revealed that even subtle design choices, like the gendered voice of a virtual assistant, could elicit stereotypical social reactions from users, demonstrating the unconscious nature of mindless cues. (Kim and Sundar 2012)

Social cues play a crucial role in shaping user perceptions and interactions with conversational agents (CAs). These cues, which include verbal, visual, auditory, and invisible elements, have been shown to significantly impact various psychological and behavioral responses (Araujo 2018; Nass and Moon 2000). Research suggests that social cues can enhance user engagement, trust, satisfaction, and perceived social presence while also presenting potential drawbacks if not implemented appropriately (Verhagen et al. 2014; Von Der Pütten et al. 2010).

One of the most widely recognized benefits of social cues in CAs is their ability to increase the perception of social presence (Araujo 2018). Users tend to attribute human-like qualities to CAs when they exhibit social behaviors, leading to a more immersive and engaging interaction (Gnewuch et al. 2018). This increased social presence fosters a stronger connection between the user and the CA, making the interaction more natural and enjoyable (De Visser et al. 2016). Additionally, social cues contribute to the establishment of trust in CAs (Bickmore and Picard 2005).

When CAs employ human-like gestures, emotional expressions, or conversational strategies such as small talk or politeness, users are more likely to perceive them as credible and reliable sources of information (Verhagen et al. 2014). This trust is a crucial factor in user acceptance and long-term engagement with CAs (Demeure et al. 2011).

Another important effect of social cues is their impact on user satisfaction (Verhagen et al. 2014). The presence of well-designed social cues enhances the overall user experience by making interactions more fluid and intuitive (Pelachaud and Bilvi 2003). For instance, nonverbal cues such as appropriate response timing, tone variation, and facial expressions can help simulate human-like conversational flow, thereby improving user comfort and reducing frustration (De Carolis et al. 2004). Furthermore, social cues can influence user attitudes toward CAs by making them appear more competent and persuasive (Fogg 2002). When users perceive CAs as socially competent, they are more likely to respond positively to recommendations, advice, or assistance provided by the system (Araujo 2018).

Despite these advantages, social cues can also have negative effects when poorly designed or mismatched with user expectations (Brandtzaeg and Følstad 2018). Overuse or inauthentic application of social cues may lead to discomfort, confusion, or even rejection of the CA (Ghazali et al. 2018). For example, an exaggerated or unnatural use of gestures, facial expressions, or vocal variations may trigger the uncanny valley effect (Wallis and Norling 2005). The "uncanny valley" hypothesis suggests that when artificial entities appear almost but not entirely human, they can evoke discomfort and rejection (Mori et al. 2012). It has been demonstrated that an overabundance of social cues or an incongruence between verbal and nonverbal cues can reduce the perceived authenticity of an agent (Seeger et al. 2018). Additionally, inappropriate social cues, such as excessive familiarity or overuse of humor, may be perceived as intrusive or unprofessional, leading to reduced user trust and engagement (Fogg 2002).

Overall, social cues are powerful design elements that can enhance user experience and interaction quality with CAs (Bickmore and Picard 2005). When applied effectively, they contribute to increased social presence, trust, satisfaction, and engagement (De Visser et al. 2016). However, their misuse can lead to negative perceptions, resistance, and decreased usability (Brandtzaeg and Følstad 2018). Therefore, designers must adopt a balanced approach, ensuring that social cues align with user expectations, cultural norms, and contextual appropriateness to maximize their positive impact.

Beyond immediate verbal, visual, auditory, and invisible signals, conversational memory enables a deeper, more personalized user experience by allowing conversational agents to retain contextual knowledge across interactions. This capacity enhances continuity, improves user satisfaction, and fosters a sense of familiarity, ultimately strengthening the relational depth between the user and the system. The following chapter delves into the concept of conversational memory, examining its mechanisms, implications, and challenges in maintaining coherence and user trust within digital dialogues.

2.8 Conversational Memory

In human dialogue, conversational memory refers to the cognitive ability to store and retrieve information from interpersonal dialogues. This form of memory is essential for social interactions as it ensures the coherence and efficiency of conversations. Studies highlight that successful conversational memory depends on multiple cognitive processes that work in tandem to enable effective communication One crucial aspect is the ability to generate relevant topics from previous interactions, allowing individuals to build on prior exchanges and maintain coherence in conversations.

This process is closely linked to the retrieval of both general and specific semantic schemas, which provide the structural foundation for understanding and integrating information within a given dialogue (Stafford and Daly 1984).

These schemas serve as cognitive frameworks that help individuals categorize and interpret conversational content, enabling them to respond appropriately based on past experiences and contextual cues. Moreover, the application of social interaction rules within specific conversational contexts plays a significant role in conversational memory. These rules, often implicit, guide individuals in navigating turn-taking, maintaining politeness, and adhering to cultural and situational norms. The ability to integrate new information into pre-existing memory structures further ensures that conversations remain dynamic and adaptive. This integration allows for the seamless assimilation of novel insights and perspectives, thereby enriching the ongoing discourse and enhancing interpersonal understanding. In addition to these mechanisms, conversational memory involves anticipating probable conversational trajectories. This predictive ability enables individuals to foresee potential directions a conversation might take, allowing them to adjust their responses and contributions accordingly. By doing so, they can maintain engagement, preempt misunderstandings, and align their communication strategies with the expectations and intentions of their conversational partners. Finally, the capacity to store and structure memories for future interactions ensures continuity and relational depth in communication. By retaining relevant conversational elements, individuals can establish stronger connections, personalize future interactions, and demonstrate attentiveness, all of which contribute to the development of meaningful and effective communication over time (Stafford and Daly 1984).

Conversational memory in chatbots is structured around multiple memory architectures that allow the system to retain, retrieve, and utilize past interactions in a coherent and contextually appropriate manner. Chatbot memory can be categorized into short-term memory (STM), long-term memory (LTM), and episodic memory, each serving a specific function in enhancing chatbot intelligence and interaction quality. STM in chatbots functions as a transient storage system that holds conversation data relevant to the current interaction. It enables the chatbot to maintain dialogue coherence within a single session, allowing for follow-up responses based on previous user inputs. This type of memory typically persists only for the duration of the conversation and is cleared once the session ends. The primary purpose of STM is to facilitate smooth, contextually aware exchanges without requiring the system to store every interaction permanently. Long-term memory (LTM) extends beyond a single session and enables chatbots to remember user preferences, past discussions, and recurring queries over time. Unlike STM, LTM stores structured information, often in databases or knowledge graphs, that allows the chatbot to personalize interactions and offer more relevant recommendations. LTM also supports user profiling, where stored data about user behaviors and preferences enhance the chatbot's ability to provide tailored responses over multiple interactions (Elvir et al. 2017).

Chatbot memory plays a crucial role in fostering user engagement and self-disclosure by enabling the chatbot to recall past interactions and demonstrate continuity in conversations. General advantages of conversational memory include heightened trust, improved user engagement, and a stronger sense of social presence, making interactions feel more personalized and empathetic (Bickmore and Picard 2005; Richards and Bransky 2014). Users typically perceive chatbots employing conversational memory as more intelligent, attentive, and empathetic, leading to longer interaction durations and increased willingness to disclose personal information (Fu et al. 2021; Jo et al. 2024).

In the specific context of mental health, conversational memory offers substantial benefits due to its ability to mimic key aspects of therapeutic relationships. Trust-building and the formation of a therapeutic alliance are particularly enhanced when chatbots remember and appropriately reference previous emotional disclosures or ongoing personal challenges (Siddals et al. 2024). Users interacting with mental health chatbots equipped with conversational memory report feeling emotionally supported, experiencing reduced loneliness, and demonstrating greater openness in sharing sensitive mental health information, crucial for effective therapeutic interventions (Jo et al. 2024; Siddals et al. 2024)

Furthermore, chatbots with conversational memory have shown promise in supporting users in managing emotional stress, trauma, and anxiety, largely due to their perceived empathy and continuous emotional presence. For instance, conversational memory allows these digital assistants to recall specific user situations, fostering a deeper emotional connection and providing tailored emotional reassurance or encouragement (Siddals et al. 2024). Such personalized interactions not only enhance user engagement and adherence to therapeutic recommendations but also help sustain long-term user involvement, thereby potentially improving overall mental health outcomes (Cox et al. 2023).

Nevertheless, the application of conversational memory in mental health contexts must be approached cautiously, considering privacy and the sensitivity of health-related information. An overly detailed or verbatim recall of past user statements may raise privacy concerns, potentially undermining trust. Hence, implementing conversational memory with nuance, such as referencing user experiences contextually rather than literally, is essential to balance personalization and privacy effectively (Cox et al. 2023). In conclusion, conversational memory significantly enhances chatbot interactions, particularly within mental health applications, by fostering trust, empathy, engagement, and therapeutic alliance. When sensitively implemented, it can greatly enrich user experience, promote emotional well-being, and support long-term mental health interventions.

2.9 Research Gap

The growing global burden of mental health challenges, exacerbated by systemic issues such as professional shortages and long wait times, highlights the urgent need for preventive and low-threshold mental health interventions (Jacobi et al., 2016). In recent years, gratitude has emerged as an effective strategy for promoting psychological well-being. Studies indicate that gratitude interventions can foster resilience, reduce symptoms of depression and anxiety, and improve overall life satisfaction (Portocarrero et al., 2020).

DMHIs have shown particular promise in increasing the scalability and accessibility of such approaches (Murray et al., 2016). Within this context, chatbots have become a compelling medium for delivering psychological support. Their ability to simulate human-like conversation enables sustained engagement, especially when they are enhanced with anthropomorphic design features (Boucher et al., 2021). HCI research suggests that such features, including social cues and conversational memory, significantly influence how users perceive and emotionally respond to digital agents (Nass & Moon, 2000).

Key concepts from the CASA paradigm and social presence theory underline the importance of perceived social presence in chatbot-user interactions. When users perceive chatbots as socially present, they tend to engage more deeply and report greater satisfaction and trust (Lee et al., 2006; Gnewuch et al., 2018). Despite this knowledge, existing research has not yet examined how these social dynamics shape the effectiveness of gratitude interventions.

This research gap points to the need for further investigation into how social cues and perceived social presence impact gratitude expression and psychological outcomes. The following section introduces the research model developed to address this gap and to explore how chatbot design can optimize the delivery of gratitude interventions.

2.10 Research Model

This study seeks to examine the role of chatbot-based gratitude interventions and how anthropomorphic design features, such as social cues and conversational memory, shape perceived social presence, and well-being outcomes. Specifically, it aims to assess the impact of social cues, including conversational memory and human-like responses, on perceived social presence. Furthermore, it investigates the relationship between social presence and gratitude expression, exploring whether a chatbot perceived as socially present can facilitate more meaningful gratitude exercises. Additionally, the study explores the influence of social presence on well-being outcomes, assessing whether the emotional depth created by human-like chatbot interactions leads to improved psychological states. By addressing these objectives, the study contributes to the understanding of how digital mental health interventions can be optimized through human-like chatbot interactions.

Building upon prior research in digital health interventions and social presence theory, the proposed conceptual model suggests a pathway in which anthropomorphic chatbot features enhance social presence, which, in turn, fosters gratitude expression and improves well-being outcomes. Social cues and conversational memory are critical components in this process, as they enhance users' perception of social presence by making interactions with the chatbot feel more authentic and engaging. Social presence is expected to strengthen emotional engagement within gratitude interventions. Gratitude, serving as a central outcome of the intervention, is anticipated to mediate the relationship between social presence and well-being, reinforcing the idea that greater emotional engagement leads to more profound psychological benefits. Finally, well-being is measured as an indicator of intervention effectiveness, reflecting the improved emotional states resulting from chatbot-facilitated gratitude exercises.

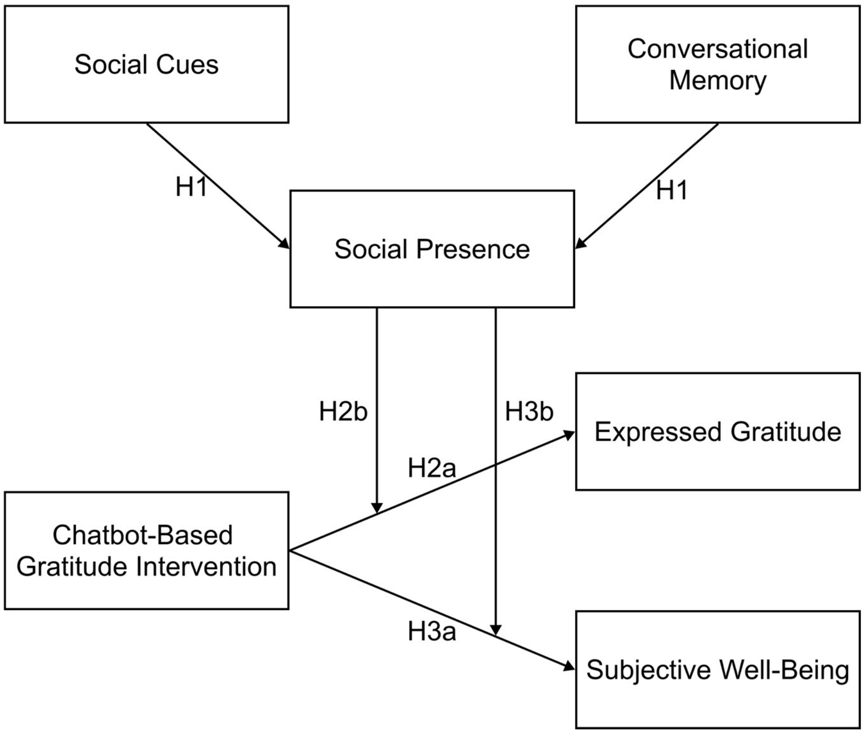

Figure 4: Conceptual Research Model

Based on the conceptual model visualized in Figure 4, this study formulates key hypotheses that explore the relationships between chatbot design, social presence, gratitude expression, and well-being. These hypotheses aim to empirically assess whether anthropomorphic chatbot features can enhance user engagement and emotional outcomes in gratitude interventions.

H1: Social cues and conversational memory enhance perceived social presence.

Anthropomorphic chatbot features, such as conversational memory, personalized responses, and natural dialogue interactions, are expected to increase users' perception of social presence. Prior research suggests that human-like attributes in chatbots contribute to greater user trust and engagement, thereby making interactions feel more meaningful and fostering a stronger sense of connection.

H2a: Chatbot-based gratitude interventions lead to an increase in gratitude expression.

Users engaging in gratitude interventions are expected to experience a rise in gratitude expression, aligning with previous research that demonstrates the efficacy of structured gratitude exercises.

H2b: A chatbot with higher social presence leads to a greater increase in gratitude expression than a chatbot with lower social presence.

Furthermore, users who perceive a chatbot as socially present are more likely to engage in gratitude exercises in a more profound and meaningful manner. This aligns with findings from human-computer interaction research, which indicate that social presence fosters emotional self-disclosure and deeper engagement with digital interventions.

H3a: Chatbot-based gratitude interventions lead to improved well-being outcomes.

Higher levels of gratitude expression resulting from chatbot-facilitated interventions are expected to correlate positively with well-being improvements. Research suggests that individuals who experience meaningful social interactions, even with digital agents, demonstrate greater emotional resilience and psychological well-being.

H3b: A chatbot with higher social presence leads to greater improvements in well-being outcomes than a chatbot with lower social presence.

By creating a more engaging and socially immersive gratitude intervention, a chatbot with high social presence is expected to enhance users' overall emotional well-being to a greater extent than one with low social presence.

This study extends the findings of Lee et al. (2024), who demonstrated that chatbot-based gratitude interventions can increase users' gratitude and positive emotions while reducing negative emotions. Since the effectiveness of gratitude chatbots has already been measured, this study does not formulate an additional hypothesis in this regard. While Lee et al. primarily examined the effectiveness of gratitude-focused chatbot interactions, it did not explicitly investigate the role of anthropomorphic cues such as conversational memory and human-like responses. This research fills this gap by exploring how anthropomorphic chatbot design influences perceived social presence and, consequently, the depth of gratitude expression and well-being outcomes.

3 Method

3.1 Experimental Setup

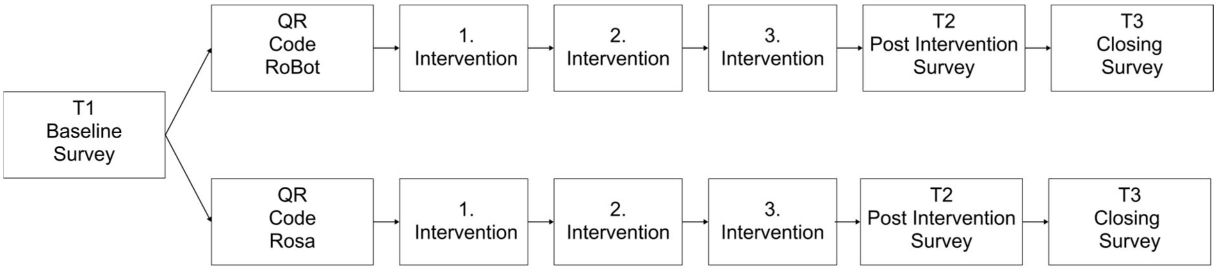

Figure 5: Experimental Setup

The study employed a randomized controlled trial with a quantitative longitudinal design to examine the effects of chatbot anthropomorphism on gratitude interventions, which is shown in Figure 5. The experimental setup largely mirrored that of the study conducted by Lee et al. (2023) to ensure comparability of results and maintain similar methodological conditions. By aligning with an established framework, the study aimed to replicate and extend findings related to chatbot-mediated gratitude interventions. The first stage of data collection at T1 commenced with a baseline survey that assessed demographic characteristics, prior exposure to gratitude interventions, and initial emotional states. Upon accessing the survey via a provided link, participants completed validated psychological instruments to assess their baseline levels of gratitude expression and emotional well-being. Upon completion of these measures, the system automatically generated a randomized QR code corresponding to the chatbot condition to which the participant had been randomly assigned. Scanning this QR code provided immediate access to the respective chatbot on the designated messaging platform, ensuring a seamless transition from survey completion to intervention participation.

The core intervention phase spanned three consecutive days and comprised a structured gratitude exercise each day. On the first day, participants engaged in reflective gratitude by identifying and contemplating aspects of their present life for which they were grateful. The second day encouraged them to recall a gratitude experience from the past, deepening their emotional engagement by linking the gratitude practice to a meaningful personal memory. The final day required participants to compose a gratitude message directed toward a specific person in their life, fostering an externalized expression of gratitude. This sequence was intentionally designed to counteract the phenomenon of hedonic adaptation. The inclusion of varied exercises aimed to sustain engagement and maximize the emotional impact of the intervention. Upon completion of the intervention phase, participants engaged in a post-intervention survey to evaluate immediate changes in gratitude levels and emotional states.

The post intervention survey at T2 replicated the questions from the baseline assessment and measured the extent to which participants perceived their interaction with the chatbot as socially present. This assessment was crucial in determining whether the anthropomorphic design elements of the chatbot enhanced the perceived relational depth of the experience.

To assess the retention of gratitude-related benefits over time, a follow-up measurement at T3 was conducted one-week post-intervention. Participants were asked to complete the same gratitude and emotional well-being scales used in previous assessments regarding gratitude expression and emotional well-being. This final data collection point provided insights into the longevity of the gratitude effects and the potential long-term influence of anthropomorphic chatbot interactions on emotional well-being.

Incorporating a randomized assignment, repeated measurements, and a structured intervention approach, enabled a comprehensive examination of the impact of chatbot anthropomorphism on gratitude interventions. Careful consideration was also given to making the study as convenient and accessible as possible for participants to minimize drop-out rates. The surveys were designed to be concise yet comprehensive, ensuring that they did not place excessive cognitive burden on respondents. Additionally, the use of QR codes for chatbot access streamlined the process, allowing participants to transition smoothly between survey completion on a desktop computer and chatbot interactions on their mobile devices.

3.2 Measures

The collected demographic data included age, gender, education, and employment status. Age distribution was analyzed for sample representativeness and was categorized into two groups: under 40 and over 40. Gender identities helped assess the balance among male, female, and non-binary participants. Educational background was based on participants' highest qualifications and was classified into academic and non-academic categories. Employment status identified whether individuals were students, employed, or in other situations, offering insights into factors influencing engagement and responses.

The present study employs three well-established measurement instruments to assess gratitude levels, emotional responses, and perceived social presence, ensuring methodological rigor and reliability. To evaluate *gratitude*, the study utilizes the Gratitude Questionnaire-Five Item Form (GQ-5-G), an adapted version of the original GQ-6 scale. The GQ-5-G has been validated to enhance model fit while maintaining robust psychometric properties. This five-item scale measures multiple aspects of gratitude, including its intensity, frequency, span, and density. Responses are recorded on a seven-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) (Example item: *"There is so much in my life for which I feel grateful."*). The total score ranges from 5 (minimum) to 35 (maximum) (Hudecek et al. 2020). The GQ-5-G has demonstrated strong internal consistency in this study (α = 0.808). Previous research has confirmed its construct and criterion validity, with higher gratitude scores being positively associated with well-being, life satisfaction, and prosocial behaviors, while inversely correlating with anxiety and depressive symptoms (McCullough et al. 2002). *Emotional well-being* is assessed using the Scale of Positive and Negative Experience (SPANE), a 12-item instrument that evaluates the frequency of positive and negative emotions experienced. SPANE consists of two subscales: SPANE-P, measuring positive affect (e.g., *"I felt joyful"*), and SPANE-N, measuring negative affect (e.g., *"I felt sad"*). Participants rate their emotional experiences on a five-point Likert scale ranging from 1 (*never or very rarely*) to 5 (*very often or always*). The SPANE enables the calculation of an affect balance score (SPANE-B) by subtracting the total negative affect score from the total positive affect score, yielding a range from -24 (maximum negative affect) to +24 (maximum positive affect). Both the positive affect and negative affect subscales consist of 6 items each, resulting in subscale scores ranging from 6 to 30. This scoring method allows for a nuanced assessment of overall emotional well-being (Diener et al. 2010). The SPANE has demonstrated robust psychometric properties (α = 0.936 for positive affect; α = 0.87 for negative affect).

The third instrument employed is the Social Presence Scale (SPRES), a 14-item measure that evaluates users' perception of *social presence* and the realness of interactions in a mediated communication context. The SPRES assesses key dimensions of user interaction, including perceived immediacy, comfort, and the sense of community established during chatbot interactions. Responses are provided on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) (Example item: *"I feel comfortable opening up to the chatbot.").* The total score ranges from 14 (minimum) to 70 (maximum) (Gunawardena and Zittle 1997). The SPRES showed strong internal constistency in this study (α = 0.881).

3.3 Chatbot Implementation

To evaluate the influence of anthropomorphic design features in chatbot-assisted gratitude interventions, two chatbot variants were developed: RoBot, a neutral chatbot, and Rosa, an anthropomorphic chatbot. These chatbots were designed to assess the potential impact of human-like attributes on user engagement and the perceived effectiveness of gratitude exercises. Both chatbots were implemented using pattern matching and followed a rule-based system, as outlined in the theoretical background. RoBot was structured to deliver gratitude interventions in a functional and impartial manner. It followed a pre-established script, guiding users through gratitude exercises without incorporating adaptive responses or human-like characteristics. The chatbot script was based on the framework proposed by Lee et. al (2024). By adapting elements from their intervention design, both Chatbots followed a structured gratitude intervention approach to optimize user engagement and emotional responses.

Rosa was designed to integrate human-like features aimed at enhancing social presence and fostering a stronger emotional connection with users. Specifically, this chatbot was characterized by several mindless anthropomorphic attributes: it was assigned a female gender identity, given a unique name, and equipped with a visual avatar in Telegram featuring a real female profile picture, which was not AI-generated to avoid the Uncanny Valley effect. By avoiding AI-generated avatars, which often fall into this valley due to subtle but unsettling inconsistencies in facial expressions and realism, Rosa's design aimed to enhance user trust and emotional engagement (Mori et al. 2012). Additionally, it incorporated a mindful element in the form of a conversational memory, allowing it to recall and reference prior interactions. In each intervention, Rosa inquired about the user's well-being and stored the response in a database as well as the name of the user. During the subsequent intervention, this stored response was retrieved and processed by a language model, which generated a contextually appropriate follow-up statement, which was additionally personalised by mentioning the name. This approach enabled Rosa to acknowledge past interactions and tailor responses accordingly, fostering a more dynamic and engaging conversational experience. While Rosa included human-like elements such as a name and avatar, its responses were still constrained by predefined rules, meaning that it did not possess true adaptive learning capabilities or contextual memory. Instead, any perceived personalization resulted from scripted interactions designed to enhance social presence rather than genuine adaptivity. By ensuring that both chatbots adhered to a pattern-matching and rule-based approach, this study isolates the effect of anthropomorphic design elements without introducing confounding variables related to adaptive AI behavior.

Both chatbot versions were deployed on Telegram, a widely used messaging application, selected for its accessibility and ease of use, particularly on mobile devices. Telegram's intuitive interface facilitated seamless interaction with the chatbots, minimizing technical barriers and thereby reducing the likelihood of participant dropout. By leveraging an existing communication platform, the study ensured a user-friendly experience, making it easier for participants to integrate the gratitude interventions into their daily routines. The chatbot was developed using Botpress, an open-source chatbot development platform known for its flexibility and ease of integration with various messaging applications. Botpress enables the creation of conversational agents with advanced natural language understanding (NLU) capabilities, rule-based decision-making, and dynamic response generation. The platform supports modular architecture, allowing developers to implement custom features, including API integrations and conversational memory. To manage intervention schedules and automate interactions, the chatbot processed intervention data via code execution within Botpress. These data were then transmitted via webhook to Make.com, an automation and integration platform that facilitates seamless workflows between various applications. Make.com was chosen due to its user-friendly interface and powerful automation capabilities, enabling the chatbot to interact with external databases and scheduling mechanisms effectively.

For automation, Make.com was utilized to store intervention data and trigger the chatbot's responses at predefined intervals. Specifically, a database within Make.com maintained user-specific intervention schedules. A cron job executed periodic HTTP requests containing conversation IDs, ensuring that the chatbot engaged users at appropriate moments. This approach allowed the chatbot to deliver gratitude interventions dynamically and consistently, ensuring adherence to intervention timelines while reducing manual administrative workload.

3.4 Intervention procedure

In the following, the differences between the gratitude interventions of the respective chatbots will be discussed. While both chatbots adhered to a structured and consistent intervention framework to ensure comparability, their design differences were intended to examine the role of anthropomorphic cues and social presence in digital gratitude interventions. By maintaining a uniform structure, wording, and tone, the study controlled for potential confounding variables while isolating the effect of minor variations in chatbot design.

At the outset of the first conversation, participants were required to provide their unique Survey ID. This procedural step was crucial for ensuring consistency in data collection and facilitating potential follow-up interviews. As pictured in Figure 6, RoBot initiated the intervention in a neutral and functional manner, directly greeting participants without requesting their name. This approach maintained an impersonal and pragmatic interaction style, focusing on efficiency rather than personalization. In contrast, Rosa first asked participants for their name before addressing them personally throughout the conversation. This personalized greeting aimed to establish a sense of familiarity and engagement. Furthermore, Rosa continued to reference the participant's name throughout the conversation, reinforcing social presence and fostering a relational depth that mimicked human-like interaction.

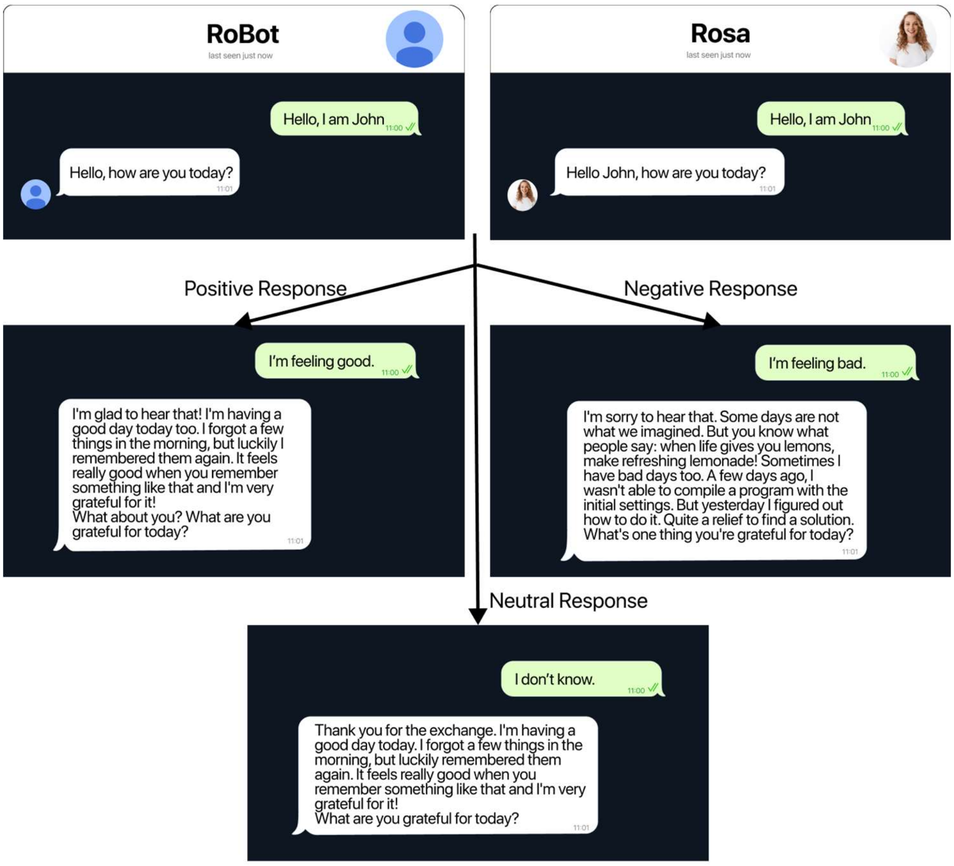

Figure 6: Exemplary First Intervention

Both chatbots were programmed to respond to participants' emotional states in an identical manner to maintain consistency in the intervention. When participants expressed a positive emotional state, both RoBot and Rosa acknowledged their response with an affirming statement, shared a minor anecdote related to gratitude, and then prompted the user to reflect on what they were grateful for. Conversely, when participants reported a negative emotional state, the chatbots acknowledged their response in a neutral and supportive manner before gently encouraging them to engage in gratitude reflection. The intervention design ensured that neither chatbot displayed empathetic variance in wording or tone, mitigating potential discrepancies in user experience. In cases where participants indicated uncertainty in their responses, both chatbots responded identically by reiterating the gratitude prompt and encouraging further reflection. No additional guidance or suggestions were provided, preserving the intervention's uniformity across both chatbot conditions.

Subsequent interactions followed a structured progression as Figure 7 shows. RoBot maintained a consistent and neutral re-engagement pattern by beginning each session with a standard greeting, ensuring a uniform and impersonalized approach. Rosa, however, incorporated continuity into subsequent interactions by referring to the participant's previously expressed emotions in a generalized manner. By acknowledging past interactions, the chatbot aimed to create a more personalized and engaging user experience, fostering a sense of continuity and attentiveness to the participant's emotional journey.

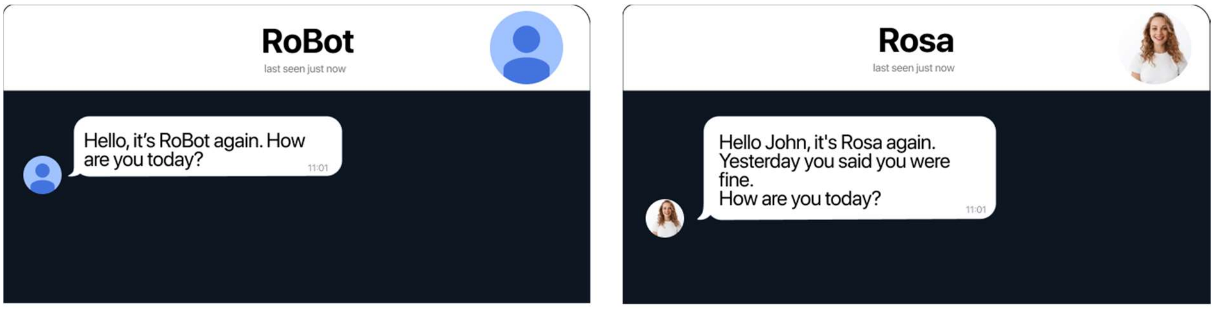

Figure 7: Exemplary Second Intervention

3.5 Sample Procedure

The data collection for this study was conducted between January 20, 2025, and March 4, 2025. This period allowed for sufficient time to recruit participants and ensure the completion of all measurement phases.

Participants were recruited via social media platforms such as WhatsApp and LinkedIn. Special efforts were made to involve master's students from the Information Management program at the University of Applied Sciences Neu-Ulm. Additionally, recruitment was supplemented through Prolific, an online research participant pool, to ensure a broader and more diverse sample. Prolific was chosen as a recruitment platform due to its well-documented advantages in providing high-quality participant data while ensuring diversity in demographics. The platform enables researchers to access a large and varied participant pool, allowing for a more representative sample beyond the specific academic network. Moreover, Prolific participants are pre-screened for various demographic and behavioral attributes to ensure that the study reached individuals who met the necessary criteria (Palan and Schitter 2018). For this study, the pre-screening included fluent German speakers residing in Germany. This criterion was established to ensure linguistic proficiency in chatbot interactions, preventing misunderstandings or misinterpretations of gratitude prompts. Additionally, selecting participants within Germany allowed for cultural homogeneity, reducing confounding variables related to differences in gratitude expression across cultures.

3.6 Statistical Analysis

To analyze the data collected in this study, statistical computations were conducted using RStudio version 2024.12.1 with R 4.4.0.

Descriptive statistics, including mean, standard deviation and median were calculated to summarize the distribution characteristics of key variables across experimental conditions (RoBot and Rosa) and measurement time points (T1, T2, T3). Additionally, frequency distributions were assessed to detect patterns and identify potential anomalies such as outliers or missing data, thereby ensuring dataset validity.

Demographic characteristics were analyzed to assess representativeness and potential confounding variables. Proportions evaluated gender distribution, age was stratified into categories for subgroup analysis, and educational and employment data were visualized via bar charts to identify disparities possibly affecting outcomes.

To confirm the suitability of statistical models, assumption tests were conducted. Normality was tested using the Shapiro-Wilk test across dependent variables, sphericity was evaluated through Mauchly's test, and Levene's test assessed variance homogeneity, identifying possible distortions of parametric results. Although the initial plan involved mixed ANOVAs for within- and between-subject effects, assumption violations regarding normality, sphericity, and homogeneity led to the use of Linear Mixed Models (LMMs).

These models account for within-subject correlations, variance heterogeneity, and repeated measures, offering robust inference for longitudinal data (Liu et al. 2012). Hypotheses were tested via LMMs to assess effects of time and group on Gratitude Score, Affect Balance, Positive Affect, and Negative Affect. A two-sample t-test compared perceived social presence between chatbot conditions at T2 to evaluate anthropomorphic design effects. LMMs ensured appropriate handling of repeated measurements and data variability.

Covariate analyses explored influences of age, gender, and education. LMMs with covariates assessed their independent and interaction effects, identifying moderating roles on group-time-dependent variable relationships. Exploratory analyses further examined covariate effects through subgroup analyses. These included the role of gender in gratitude expression, education as a predictor of affective well-being, and age-specific group interactions. Subgroups with low baseline gratitude were also analyzed to assess differential intervention impacts and better understand effectiveness across participant profiles.

4 Results

4.1 Sample Description

A total of 123 participants were recruited for this study, all of whom completed the initial measurement phase (T1). Of these, 98 participants continued to the second measurement phase (T2), ensuring the inclusion of social presence data. A total of 91 participants successfully completed all three measurement phases (T1, T2, and T3), resulting in a completion rate of approx. 74%. The dropout rate throughout the study was approx. 26%, reflecting participant attrition across the different phases. Regarding group assignment, participants were divided into two conditions: the RoBot group and the Rosa group. A total of 49 (50%) participants were assigned to the RoBot chatbot, while 49 (50%)participants interacted with Rosa. This division ensured that both experimental conditions had sufficient sample sizes for meaningful statistical analysis.

The following sections will provide a detailed statistical analysis of these demographic characteristics, referring specifically to participants who completed at least T1 and T2. This subset was chosen because social presence was measured at T2, ensuring that the collected data remains valid for analysis. For the evaluation of other measures, only the 91 participants who completed all three time points (T1, T2, and T3) will be considered.

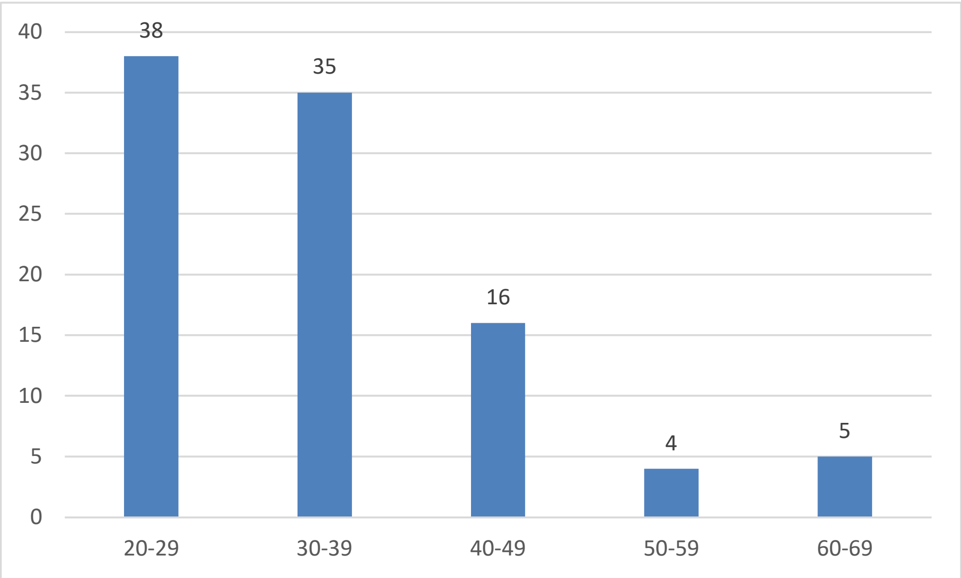

Figure 8: Age Distribution

Figure 8 shows the distribution of the sample in age groups. The sample consisted of participants across five age groups. The largest groups were 20–29 years old (38 participants; 38.8%) and 30–39 years old (35 participants; 35.7%). The 40–49 age group included 16 participants (16.3%), while the 50–59 and 60–69 age groups comprised 4 (4.1%) and 5 participants (5.1%), respectively.

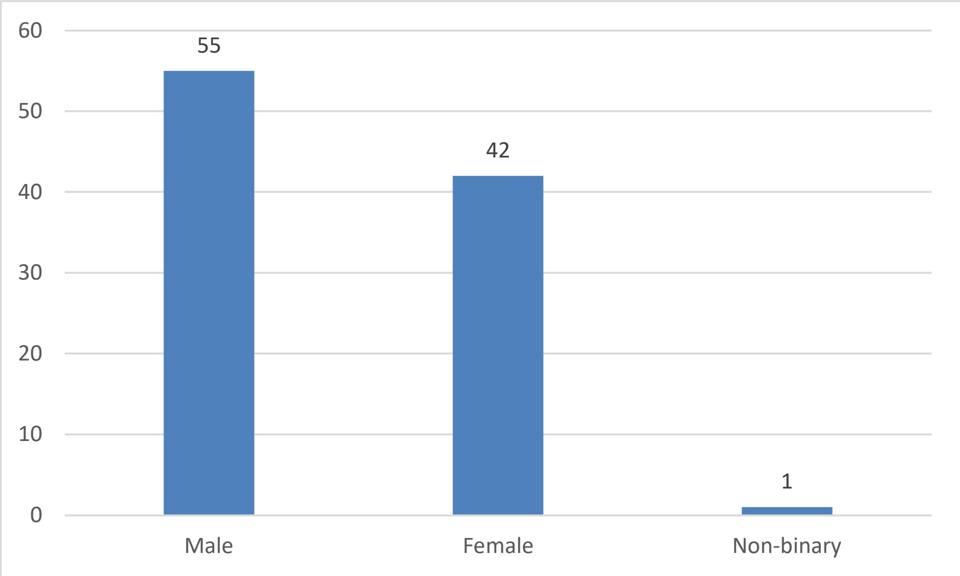

Figure 9: Gender Distribution

The distribution by gender is shown in Figure 9. The sample included 55 (56%) male and 42 (43%) female participants. Additionally, one participant identified as non-binary.

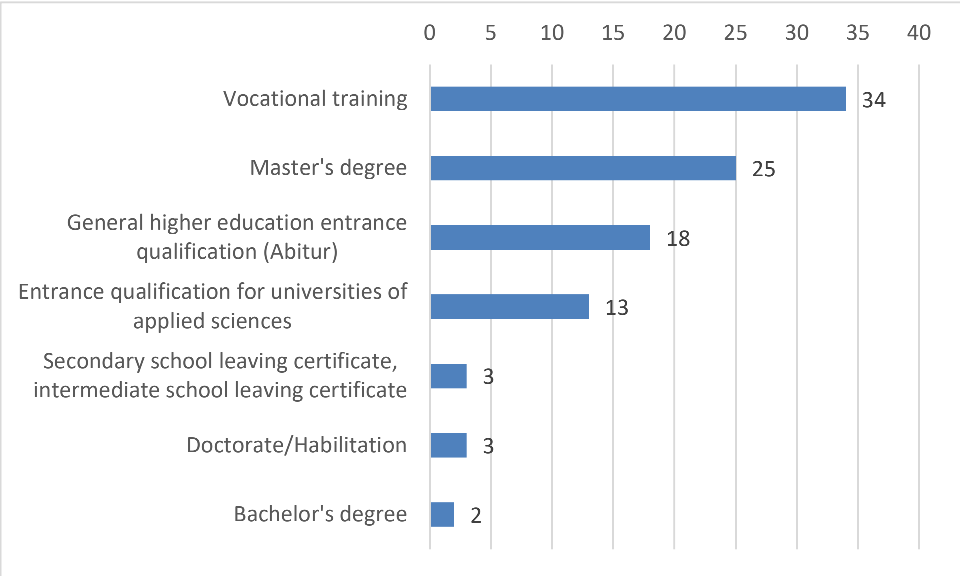

Figure 10: Educational Background

Figure 10 depicts the distribution of the educational background of the sample. Among the participants, 34 had completed vocational training (34.7%), and 25 held a master's degree (25.5%). A total of 18 participants had obtained a general higher education entrance qualification (Abitur; 18.4%), while 13 held an entrance qualification for universities of applied sciences (13.3%). Three participants had completed a doctorate or habilitation (3.1%), and another three had a secondary school leaving certificate (3.1%). Two participants held a bachelor's degree (2.0%).

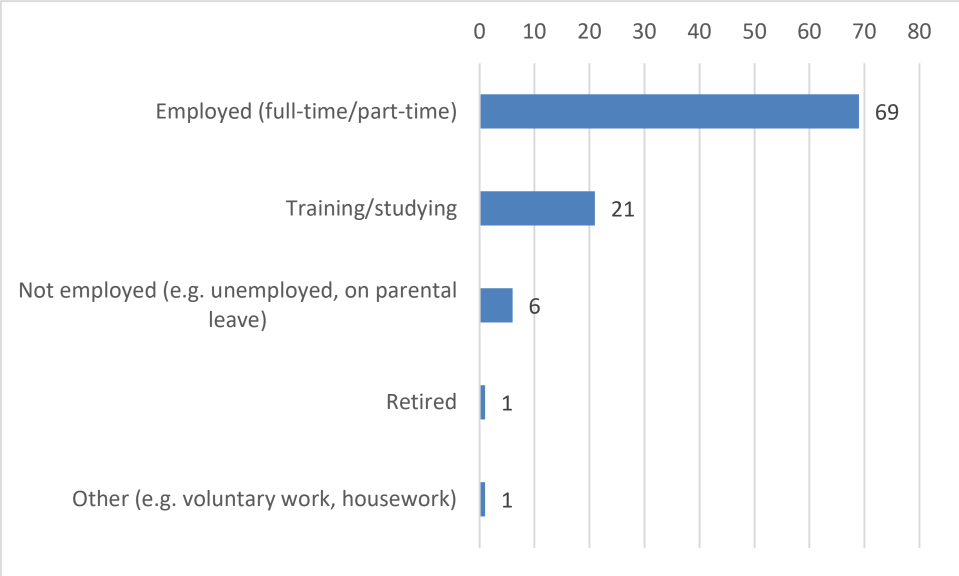

Figure 11: Employment Status

The employment status distribution is visualized in Figure 11. The majority of participants (69; 70.4%) were employed either full-time or part-time. A total of 21 participants (21.4%) were engaged in training or studying. Six participants (6.1%) were not employed, including those on parental leave. One participant (1.0%) was retired, and one (1.0%) was engaged in other activities such as voluntary work or household responsibilities.

4.2 Descriptive Statistics

The descriptive statistics provide an overview of the key variables measured at different time points (T1, T2, T3) and across experimental groups. The primary variables analyzed include Gratitude Score, Affect Balance, Positive Affect and Negative Affect.

The Social Presence Score (SPRES) was measured exclusively at T2 and varied between the two experimental conditions. In RoBot, the mean SPRES score was 29.04 (SD = 7.22), with a median of 29 and a range of 11 to 43. In Rosa, the mean SPRES score was 30.22 (SD = 7.48), with a median of 31 and a range of 13 to 49. The distributions of SPRES scores across groups were relatively similar, with RoBot showing a slight negative skew (-0.12) and Rosa displaying a slightly positive skew (0.25). The standard deviations suggest a moderate variability in perceived social presence across participants.

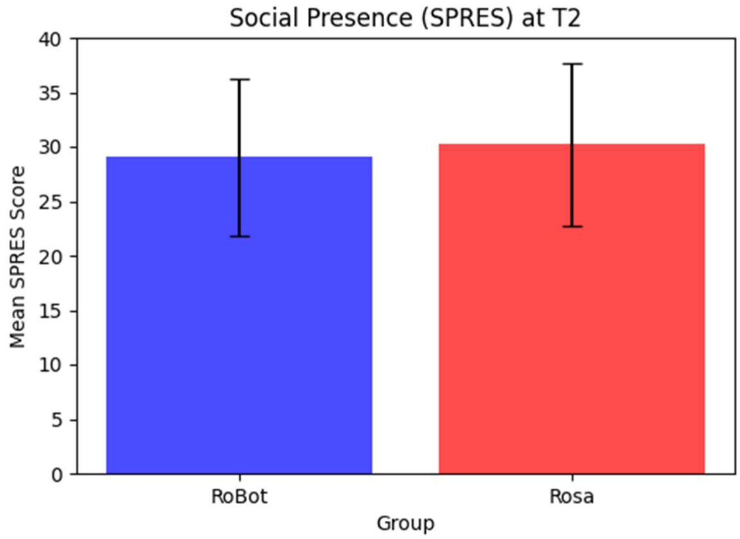

Figure 12: Social Presence Score by Group

The Gratitude Score was assessed across both experimental groups at three time points. In RoBot, the mean gratitude score at T1 was 23.82 (SD = 5.41), with a median of 25 and a range between 12 and 33 as depicted in Figure 12. At T2, the mean increased slightly to 24.20 (SD = 5.44), with a median of 25 and a range of 11 to 33. At T3, the mean was 23.93 (SD = 5.57), with a median of 24 and a range of 10 to 33.

In Rosa, the Gratitude Score at T1 was slightly higher, with a mean of 25.43 (SD = 4.08) and a median of 25. At T2, the mean gratitude score decreased slightly to 24.84 (SD = 4.68), and at T3, it was recorded at 24.64 (SD = 4.58). The range of gratitude scores across all time points remained relatively stable, spanning from a minimum of 9 to 33, suggesting a moderate to high level of gratitude expression within both groups. This progression can be seen in Figure 13.

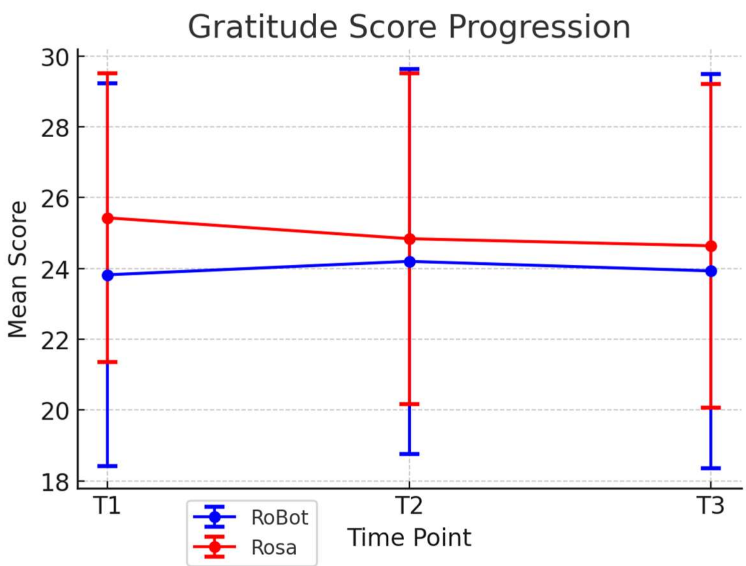

Figure 13: Gratitude Score Progression

The Affect Balance Score, which captures the difference between Positive Affect and Negative Affect, varied across time points and groups as shown in Figure 14. In RoBot, the mean Affect Balance Score at T1 was 6.35 (SD = 6.88), with a median of 7. At T2, the mean decreased to 5.52 (SD = 7.45), and at T3, it remained stable at 5.50 (SD = 6.49).

In contrast, Rosa started with a slightly lower mean Affect Balance score of 5.67 (SD = 5.22) at T1, which increased to 6.88 (SD = 6.43) at T2, before decreasing again to 5.53 (SD = 6.89) at T3. Across all time points, the minimum and maximum values ranged widely from -12 to 22, indicating considerable variability in affect balance among participants.

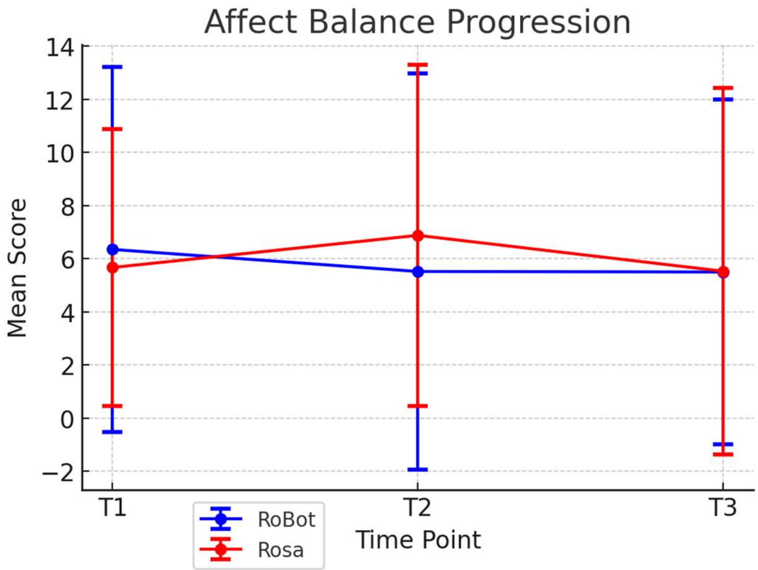

Figure 14: Affect Balance Progression

Positive Affect and Negative Affect were assessed separately using the corresponding sub-scales. In RoBot, Positive Affect remained relatively stable, with a mean of 21.80 (SD = 4.73) at T1, 21.80 (SD = 4.84) at T2, and a slight decrease to 20.50 (SD = 4.77) at T3. In Rosa, the mean Positive Affect scores were 21.71 (SD = 3.47) at T1, 21.86 (SD = 3.65) at T2, and 21.30 (SD = 3.83) at T3 as visualized in Figure 15.

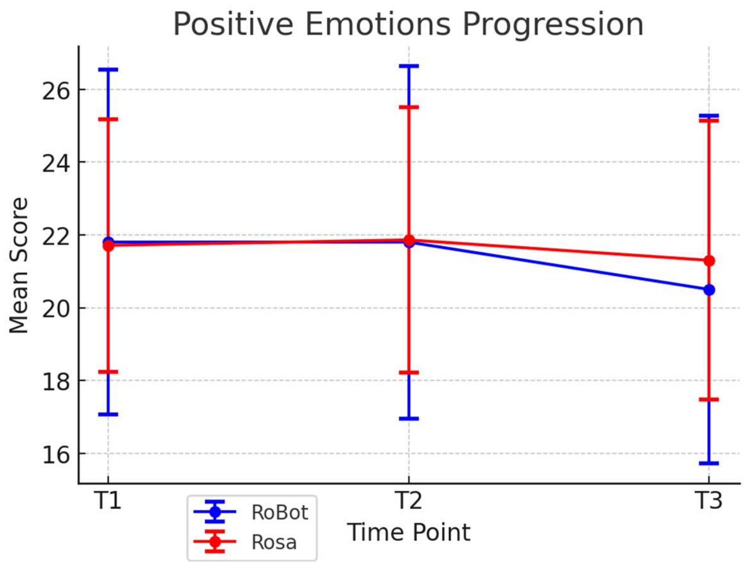

Figure 15: Positive Emotions Progression

For Negative Affect, RoBot exhibited a decrease from 16.92 (SD = 4.12) at T1 to 15.73 (SD = 4.72) at T2, with a slight increase to 16.09 (SD = 4.05) at T3. Similarly, Rosa showed a decrease in Negative Affect from 15.76 (SD = 4.81) at T1 to 15.24 (SD = 4.99) at T2, and further down to 14.91 (SD = 4.51) at T3. This progression is depicted in Figure 16.

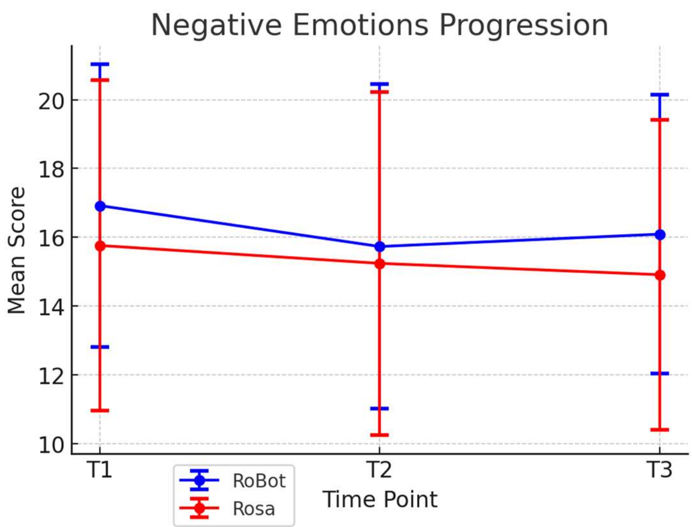

Figure 16: Negative Emotions Progression

4.3 Requirements

4.3.1 Normality

To ensure the validity of the mixed ANOVA, normality tests were conducted for each experimental group and each time point for the Dependent Variables Gratitude Score, Affect Balance, Positive and Negative Affect.

Dependent Variable	***Group***	***W***	***p-value***
Gratitude Score	RoBot	0.951 ***	<.001 ***
	Rosa	0.969	.00233
Affect Balance	RoBot	0.984	.0942
	Rosa	0.992	.544
Positive Affect	RoBot	0.953 ***	<.001
	Rosa	0.959 ***	<.001
Negative Affect	RoBot	0.976	.0121
	Rosa	0.957 ***	<.001

Table 1: Results of the Shapiro-Wilk Test by Group

A Shapiro-Wilk test was conducted to assess the normality of the distributions for the Dependent Variables across the two groups (RoBot and Rosa). The results of the Shapiro-Wilk test (listed in Table 1) for normality indicate that the assumption of normality is violated in several cases. Specifically, for Gratitude Score, both groups (RoBot and Rosa) exhibit significantly low p-values, suggesting a substantial deviation from a normal distribution. Similarly, Positive Affect shows strong evidence of non-normality in both groups.

For Affect Balance, the test results differ between groups: while the RoBot group exhibits a p-value close to the threshold, indicating a potential deviation from normality, the Rosa group does not show significant evidence of non-normality. Regarding Negative Affect, the RoBot group's p-value suggests a violation of normality, while the Rosa group's result indicates a stronger deviation.

Dependent Variable	***Time-Point***	***W***	***p-value***
Gratitude Score	T1	0.945 ***	<0.001
	T2	0.963	.007
	T3	0.963 **	.010
Affect Balance	T1	0.991	.732
	T2	0.992	.859
	T3	0.989	.654
Positive Affect	T1	0.971 *	.028
	T2	0.952 ***	.001
	T3	0.950 ***	.001
Negative Affect	T1	0.969 *	.022
	T2	0.959 **	.004
	T3	0.987	.480

Table 2: Results of the Shapiro-Wilk Test by Time-Point

A Shapiro-Wilk test was conducted to assess the normality of the distributions for Gratitude Score, Affect Balance, Positive Affect, and Negative Affect across three time points (T1, T2, and T3), the result of which are shown in Table 2. This test evaluates whether the sample data significantly deviates from a normal distribution. The results indicate that the assumption of normality is violated for Gratitude Score at all three time points, as the p-values are consistently low, suggesting a significant deviation from normality. For Affect Balance, the test results show no significant deviation from normality at any time point. In the case of Positive Affect, all time points show significant deviations from normality, with low p-values indicating non-normal distributions. Regarding Negative Affect, the assumption of normality is violated at T1 and T2, but T3 does not show significant evidence of non-normality, suggesting a potential shift in distribution over time.

Overall, the results indicate that the assumption of normality is violated in multiple cases, particularly for Gratitude Score, Positive Affect, and Negative Affect across different groups and time points.

4.3.2 Sphericity

A Mauchly’s test of sphericity was conducted to assess whether the assumption of sphericity was met for the repeated-measures ANOVA across different dependent variables. The test examines whether the variances of differences between repeated measures are equal, which is a key assumption for standard ANOVA procedures, the results are shown in Table 3.

The results indicate that the assumption of sphericity was not violated for Gratitude Score, Affect Balance, and Negative Affect, as the p-values were above the conventional significance threshold. This suggests that the variances of differences between repeated measures are approximately equal, allowing for the use of standard ANOVA procedures without the need for corrections.

Consequently, the interpretation of effects in these variables can proceed without concerns regarding violations of sphericity. However, the test revealed that the assumption of sphericity was violated for Positive Affect, as indicated by a statistically significant result. This implies that the differences between time points do not have equal variances, which could introduce bias in the standard repeated-measures ANOVA framework. In such cases, appropriate adjustments should be applied to ensure the validity of statistical conclusions.

Overall, while most dependent variables met the assumption of sphericity, the significant violation for Positive Affect highlights the necessity of employing appropriate statistical adjustments.

Dependent Variable	***Effect***	***W***	***p-value***
Gratitude Score	Time	0.963	.191
	Time:Group	0.963	.191
Affect Balance	Time	0.998	.902
	Time:Group	0.998	.902
Positive Affect	Time	0.925 *	.033
	Time:Group	0.925 *	.033
Negative Affect	Time	0.967	.233
	Time:Group	0.967	.233

Table 3: Results of the Sphericity Test

4.3.3 Homogeneity of Variances

Measure	***Df (group)***	***Df (residual)***	***F-value***	***p-value***
Gratitude Score	1	287	5.282 *	.022
Balance Affect	1	285	0.932	.335
Positive Affect	1	287	5.312 *	.022
Negative Affect	1	287	4.456 *	.036

Table 4: Results of the Homogeneity of Variances Test

To ensure that the assumption of homogeneity of variance is met, Levene's test was conducted for each dependent variable across the groups (results in Table 4). The results for the Gratitude Score revealed a statistically significant difference in variances between the groups, suggesting a violation of the homogeneity of variance assumption for this variable.

For the Affect Balance Score, the test result suggests that the homogeneity of variance assumption is met. In contrast, the Positive Affect variable showed a significant difference in variance between the groups, indicating a violation of the homogeneity assumption. Similarly, the Negative Affect measure also demonstrated a statistically significant difference in variances, suggesting another violation of the homogeneity assumption.

Taken together, the results of Levene's test suggest that for the Gratitude Score, Positive Affect, and Negative Affect, the assumption of homogeneity of variance is violated. This violation implies that variance differs significantly across groups for these variables. However, for the Affect Balance Score, homogeneity of variance is retained. Given these findings, statistical adjustments may be necessary, such as using Welch's t-test instead of a standard t-test or applying robust statistical methods to account for heterogeneity in variance where violations occur. Additionally, transformation techniques or non-parametric alternatives could be considered to mitigate the impact of unequal variances on subsequent analyses.

Due to the violation of sphericity and the homogeneity of variance assumption in multiple variables, mixed ANOVAs were not deemed appropriate. Instead, Linear Mixed Models (LMMs) were employed, as they effectively handle violations of these assumptions by explicitly modeling random effects and accounting for within-subject variability and heterogeneity of variances across groups. Unlike traditional ANOVA approaches, LMMs do not rely on strict assumptions of homogeneity of variance or sphericity, thus providing more robust and unbiased estimates, especially when dealing with longitudinal data or repeated measurements. This methodological choice ensures that statistical inferences remain reliable and are not compromised by variance heterogeneity (West et al. 2022).

4.4 Testing the Hypotheses

4.4.1 Hypothesis 1

H1: Social cues and conversational memory enhance perceived social presence.

A two-sample t-test was conducted to examine whether the mean perceived social presence (SPRES) differed significantly between the two chatbot groups (Rosa vs. RoBot) at post-intervention measurement (T2). The t-test revealed no statistically significant difference in perceived social presence between the two groups ($t(94.98) = -0.79$, $p = .430$).

The mean SPRES score for Rosa was 29.04, whereas RoBot had a mean SPRES score of 30.22. The 95% confidence interval for the mean difference ranged from -4.15 to 1.78, further suggesting that the observed difference is not statistically robust.

4.4.2 Hypothesis 2

Effect	***df (num)***	***df (denom)***	***F-value***	***p-value***
(Intercept)	1	185	2642.208 ***	<.0001
Group	1	98	0.598	.441
Time	2	185	0.059	.942
Group:Time	2	185	1.900	.153

Table 5: Results of the LMM for Gratitude Expression

A LMM was employed to assess the effects of time and group differences on gratitude expression, accounting for within-subject variability and variance heterogeneity. The results are listed in Table 5.

H2a: Chatbot-based gratitude interventions lead to an increase in gratitude expression.

The analysis within the LMM considered gratitude scores as a function of time points. Results indicated that overall gratitude scores were significantly different from zero, but the main effect of time was not statistically significant. This suggests that gratitude levels remained stable across the measurement points throughout the intervention period, with no significant changes observed.

H2b: A chatbot with higher social presence leads to a greater increase in gratitude expression than a chatbot with lower social presence.

The LMM further examined the interaction between group and time to determine whether group assignment affected changes in gratitude expression over the intervention duration. Results demonstrated that while overall gratitude scores were significantly different from zero, neither the main effect of group nor the interaction between group and time reached statistical significance. This indicates stable gratitude levels across all conditions, with no statistically significant differences between the different chatbot groups observed over time.

4.4.3 Hypothesis 3

Dependent Variable	***Effect***	***df (num)***	***df (denom)***	***F-value***	***p-value***
Affect Balance	(Intercept)	1	185	57.786 ***	<.001
	Group	1	98	0.225	.637
	Time	2	185	3.085 *	.048
	Group:Time	2	185	1.429	.242
Positive Affect	(Intercept)	1	185	2756.067 ***	<.001
	Group	1	98	0.001	.977
	Time	2	185	4.256 *	.016
	Group:Time	2	185	2.242	.109
Negative Affect	(Intercept)	1	185	1306.666 ***	<.001
	Group	1	98	0.674	.414
	Time	2	185	9.934 ***	<.001
	Group:Time	2	185	0.902	.408

Table 6: Results of the LMM for Well-Being

LMMs were again utilized to investigate the effects of time and group differences on affect balance, positive affect, and negative affect, effectively handling within-subject variability and variance heterogeneity, their results are shown in Table 6.

H3a: Chatbot-based gratitude interventions lead to improved well-being outcomes.

The analysis of affect balance indicated a statistically significant main effect of time, suggesting meaningful changes in affect balance across the measured time points. Similarly, the analysis of positive affect also demonstrated a significant main effect of time, indicating notable variations in positive affect throughout the intervention period. Regarding negative affect, results similarly showed a significant effect of time, suggesting that negative affect levels changed significantly across measurement points.

To explore this time effect in more detail, post-hoc pairwise comparisons were computed using estimated marginal means (EMMs). These model-derived means are adjusted for the effects of other factors in the model and are therefore more robust than raw means when group sizes or distributions are unequal. The pairwise contrasts were corrected for multiple testing using the Bonferroni method. The post-hoc results indicated a statistically significant increase in Affect Balance from T1 to T2 (estimate = -0.918, SE = 0.371, t(185) = -2.48, p = .0425), reflecting a rise in overall positive emotional state following the first day of the gratitude intervention. No significant differences were observed between T1 and T3 (estimate = -0.542, SE = 0.381, t(185) = -1.42, p = .4707) or between T2 and T3 (estimate = 0.376, SE = 0.381, t(185) = 0.99, p = .9741).

Overall, these results provide evidence for a short-term emotional benefit of the gratitude intervention, particularly after the first day, but raise questions regarding the sustainability of this effect in the absence of continued novelty or interpersonal reinforcement.

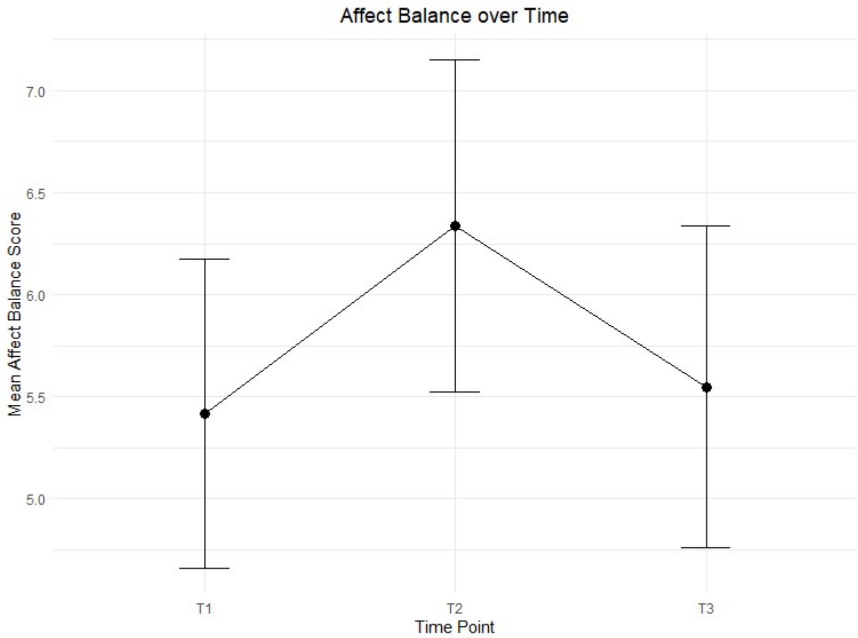

Figure 17: Affect Balance over Time

H3b: A chatbot with higher social presence leads to greater improvements in well-being outcomes than a chatbot with lower social presence.

The interaction between group and time was evaluated to assess potential differences in affect trajectories between the groups. Results revealed that the interaction effect for affect balance was not statistically significant, suggesting comparable changes over time across both groups. Similarly, the interaction effect for positive affect was also non-significant, indicating that changes in positive affect levels over time were parallel between the two groups. Additionally, the interaction effect for negative affect did not reach statistical significance, implying consistent patterns of negative affect across both groups over the intervention period.

Collectively, these findings indicate that group assignment did not significantly moderate the temporal changes in affect balance, positive affect, or negative affect.

4.5 Inclusion of covariances

4.5.1 Influence of covariates on gratitude expression

Effect	*df (num)*	*df (denom)*	*F-Value*	*p-value*
(Intercept)	1	177	2897.154 ***	<.001
Time	2	177	0.096	.908
Group	1	89	0.612	.436
Age	1	89	0.579	.449
Gender	1	89	3.479	.065
Educational Background	1	89	0.399	.529
Time:Group	2	177	1.800	.168
Time:Age	2	177	0.216	.806
Time:Gender	2	177	2.214	.112
Time:Educational Background	2	177	0.705	.496
Group:Age	1	89	0.666	.417
Group:Gender	1	89	0.039	.843
Group:Educational Background	1	89	1.751	.189

Table 7: Results of the LMM for Gratitude Expression including the Covariate

A LMM was conducted to examine the effects of time, group differences, and covariates (age, gender, and education level) on gratitude expression, accounting for within-subject variability and variance heterogeneity. Unlike the previous LMM that only assessed the effects of time and group, this model incorporates covariates to control for potential confounding influences.The results shown in Table 7 indicate that the overall model intercept was statistically significant, confirming meaningful variation in gratitude scores.

However, the main effects of time and group were not statistically significant, suggesting that gratitude expression remained stable across measurement points and did not significantly differ between groups. Similarly, the interaction effect between time and group was not significant, indicating that changes in gratitude expression over time were comparable across groups.

Regarding covariates, gender showed a marginally significant main effect on gratitude expression, suggesting a potential influence of gender on gratitude levels. However, age and education level did not show significant main effects, indicating that these factors were not independently associated with gratitude expression.

The interaction effects between time and the covariates (age, gender, and education) were all non-significant, meaning that the trajectory of gratitude expression over time was not moderated by these demographic factors. Additionally, interactions between group and age, gender, or education level did not yield significant effects, suggesting that group differences in gratitude expression were consistent across these demographic variables.

Overall, the findings suggest that gratitude expression remained relatively stable over time and was not significantly influenced by group assignment or the included covariates, apart from a marginally significant gender effect. These results highlight the relative stability of gratitude expression across different demographic factors during the intervention period.

4.5.1 Influence of covariates on affect balance

Effect	***df (num)***	***df (denom)***	***F-Value***	***p-value***
(Intercept)	1	177	77.101 ***	<.001
Time	2	177	3.017	.052
Group	1	89	0.286	.594
Age	1	89	0.559	.457
Gender	1	89	0.198	.658
Educational Background	1	89	7.126 **	.009
Time:Group	2	177	1.633	.198
Time:Age	2	177	0.411	.664
Time:Gender	2	177	0.788	.456
Time:Educational Background	2	177	0.678	.509
Group:Age	2	89	7.731 **	.007
Group:Gender	1	89	0.032	.859
Group:Educational Background	1	89	0.084	.773

Table 8: Results of the LMM for Affect Balance including the Covariate

A LMM was conducted to examine the effects of time, group differences, and covariates (age, gender, and education level) on affect balance accounting for within-subject variability and variance heterogeneity. Unlike the previous LMM that only assessed the effects of time and group, this model incorporates covariates to control for potential confounding influences.

The results shown in Table 8 indicated that the overall model intercept was statistically significant, confirming meaningful variation in affect balance scores. The main effect of time approached statistical significance, suggesting marginal changes in affect balance across measurement points. However, the main effect of group was not statistically significant, indicating no systematic differences between groups. Similarly, the interaction between time and group was not significant, indicating comparable trajectories in affect balance across both groups over time.

Regarding covariates, education level showed a significant main effect on affect balance, indicating that education significantly influenced affect balance scores. In contrast, age and gender did not have significant main effects, suggesting these variables did not independently impact affect balance.

The interaction effects between time and the covariates (age, gender, and education) were all non-significant, suggesting that changes in affect balance over time were not moderated by these demographic factors. Additionally, interactions between group and gender, as well as group and education, were also not statistically significant, suggesting consistent group effects regardless of these demographic factors.

However, a significant interaction effect emerged between group and age, highlighting that age moderated the relationship between group assignment and affect balance. This finding indicates the importance of age in understanding group differences in affect balance. Overall, the results suggest relative stability of affect balance over time, influenced predominantly by education level and notably moderated by age in the context of group differences.

4.5.2 Influence of covariates on positive affect

Effect	***df (num)***	***df (denom)***	***F-Value***	***p-value***
(Intercept)	1	177	3258.473 ***	<.001
Time	2	177	3.770 *	.025
Group	1	89	0.003	.957
Age	1	89	0.148	.701
Gender	1	89	1.472	.228
Educational Background	1	89	4.748 *	.032
Time:Group	2	177	2.560	.080
Time:Age	2	177	0.306	.737
Time:Gender	2	177	0.862	.424
Time:Educational Background	2	177	0.041	.960
Group:Age	2	89	2.545	.114
Group:Gender	1	89	0.005	.946
Group:Educational Background	1	89	0.354	.553

Table 9: Results of the LMM for Positive Affect including the Covariate

A LMM was conducted to examine the effects of time, group differences, and covariates (age, gender, and education level) on positive affect, while accounting for variance heterogeneity across groups. This approach allows for more accurate modeling of repeated measures data by incorporating random effects.

The results listed in Table 9 indicate that the overall model intercept was statistically significant, confirming that the dependent variable exhibited meaningful variation. The main effect of time was statistically significant, suggesting that positive affect varied across measurement points. However, the interaction between time and group was not significant, indicating that changes in positive affect over time did not differ systematically between groups.

Regarding covariates, neither age nor gender had a significant effect on positive affect. However, education level exhibited a statistically significant main effect, indicating that individuals from different educational backgrounds experienced variations in their affective states.

The interaction effects between time and covariates (age, gender, and education) were all non-significant, suggesting that the trajectory of positive affect over time was not moderated by these demographic factors. Additionally, the interactions between group and covariates (age, gender, and education) were all non-significant, indicating that group-related differences in positive affect were not significantly influenced by these demographic variables.

Overall, the findings suggest that while positive affect fluctuated across time points, it was not significantly influenced by group assignment. Furthermore, the only demographic factor that played a significant role was education level, which had a meaningful impact on positive affect. The absence of significant interaction effects underscores the relatively stable nature of positive affect across different demographic and group-related contexts.

4.5.3 Influence of covariates on negative affect

Effect	*df (num)*	*df (denom)*	*F-Value*	*p-value*
(Intercept)	1	177	1491.986 ***	<.001
Time	2	177	9.721 ***	<.001
Group	1	89	0.773	.382
Age	1	89	2.831	.096
Gender	1	89	0.131	.718
Educational Background	1	89	6.526 *	.012
Time:Group	2	177	0.964	.384
Time:Age	2	177	1.125	.327
Time:Gender	2	177	0.607	.546
Time:Educational Background	2	177	1.396	.250
Group:Age	2	89	10.738 **	.002
Group:Gender	1	89	0.053	.818
Group:Educational Background	1	89	1.117	.294

Table 10: Results of the LMM model for Negative Affect including the Covariate

A LMM was conducted to examine the effects of time, group differences, and covariates (age, gender, and education level) on negative affect, while accounting for within-subject correlations. This approach allows for more accurate modeling of repeated measures data by incorporating random effects.

The results shown in Table 10 indicate that the overall model intercept was statistically significant, confirming that the dependent variable exhibited meaningful variation. The main effect of time was statistically significant, suggesting that negative affect varied across measurement points. However, the interaction between time and group was not significant, indicating that changes in negative affect over time did not differ systematically between groups.

Regarding covariates, age did not have a significant main effect on negative affect, whereas education level exhibited a statistically significant effect. This finding suggests that individuals from different educational backgrounds reported varying levels of negative affect. Conversely, gender did not have a significant impact, indicating no meaningful differences in negative affect levels between genders.

The interaction effects between time and the covariates (age, gender, and education) were all non-significant, suggesting that the trajectory of negative affect over time was not moderated by these demographic factors. Similarly, the interactions between group and gender, as well as group and education, were not statistically significant, indicating that group-related differences in negative affect were not influenced by these covariates.

However, a significant interaction effect was found between group and age, suggesting that age played a role in how negative affect differed between groups. This indicates that the relationship between group assignment and negative affect may be moderated by age, warranting further investigation.

4.6 Explorative Analyzes

The exploratory analyses aimed to examine potential covariates influencing gratitude expression and affect balance. Specifically, the relationships between gender and gratitude score, gratitude score and the interaction of group and educational background, affect balance and educational background, and affect balance and the interaction between age and group were investigated.

While various covariates were explored, further analysis focused solely on affect balance due to similar patterns emerging in Positive Affect and Negative Affect. Since Affect Balance is derived from these two components, it is reasonable to assume that the observed effects would be consistent across all three measures. This approach ensures a streamlined interpretation of the results while maintaining focus on the most relevant covariates.

4.6.1 Gender as a Predictor of Gratitude

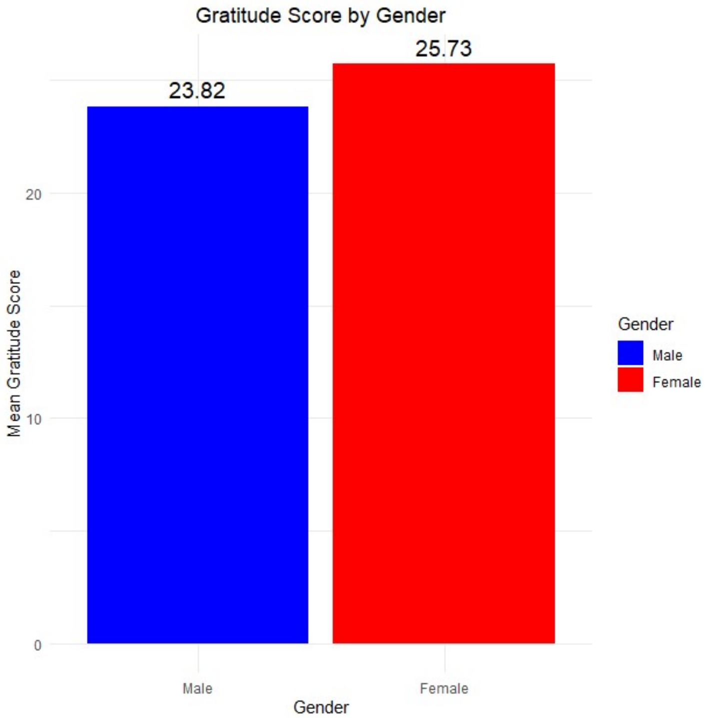

Figure 18: Differences in Gratitude Expression by Gender

The analysis using the LMM identified gender as a significant covariate, indicating that gender influences gratitude expression and should be accounted for in the interpretation of gender differences. To further investigate this effect, a Tukey HSD post-hoc test was conducted. The results revealed significant differences between male and female participants. Specifically, female participants reported significantly higher gratitude scores than male participants with a mean difference of 1.90. Figure 18 shows the descriptive difference in mean Gratitude Score by gender.

4.6.2 Educational Background as a Predictor of Well-Being

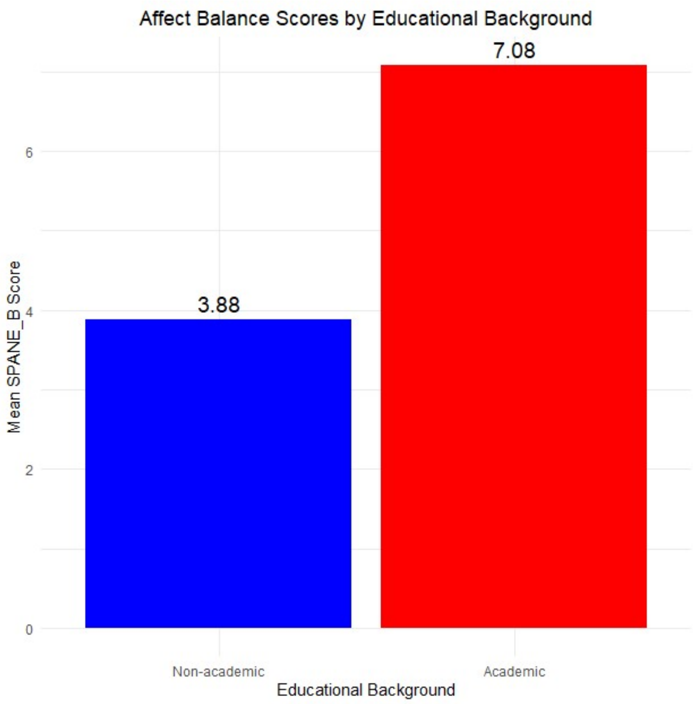

Figure 19: Differences in Affect Balance by Educational Background

The LMM identified educational background as a significant covariate of gratitude scores, indicating that differences in educational attainment may influence gratitude expression.

The results indicate that individuals with an academic background report significantly higher affect balance scores compared to those with a non-academic background as in Figure 19 depicted. This suggests that educational attainment may be associated with differences in affective well-being.

4.6.3 Age as a Predicor of Well-Being

To further investigate the significant interaction between age and group in predicting affect balance, a LMM was separately conducted for participants over and under 40 years old. The results are listed in Table 11.

Age Group	***Effect***	***df (num)***	***df (denom)***	***F-value***	***p-value***
Over 40	(Intercept)	1	45	20.958212	<.001 ***
	Group	1	23	7.114953	.014 *
	Time	2	45	0.081082	.922
	Group:Time	2	45	0.498274	.611
Under 40	(Intercept)	1	134	52.90187	<.001 ***
	Group	1	70	0.63933	.427
	Time	2	134	4.10179	.019 *
	Group:Time	2	134	3.69787	.027 *

Table 11: Results of the LMM on Affect Balance separated separated in Age Group

For individuals older than 40, the analysis revealed a significant main effect of group, indicating that group membership was associated with differences in affect balance. However, neither the main effect of time nor the interaction between group and time reached statistical significance. This suggests that affect balance remained stable across measurement points and that group differences did not evolve over time in this age group.

For individuals younger than 40, the main effect of group was not statistically significant, suggesting that group membership did not strongly influence affect balance levels. However, the main effect of time was significant, indicating that affect balance fluctuated across measurement points. Additionally, a significant interaction between group and time suggests that group differences in affect balance evolved over time in this younger cohort.

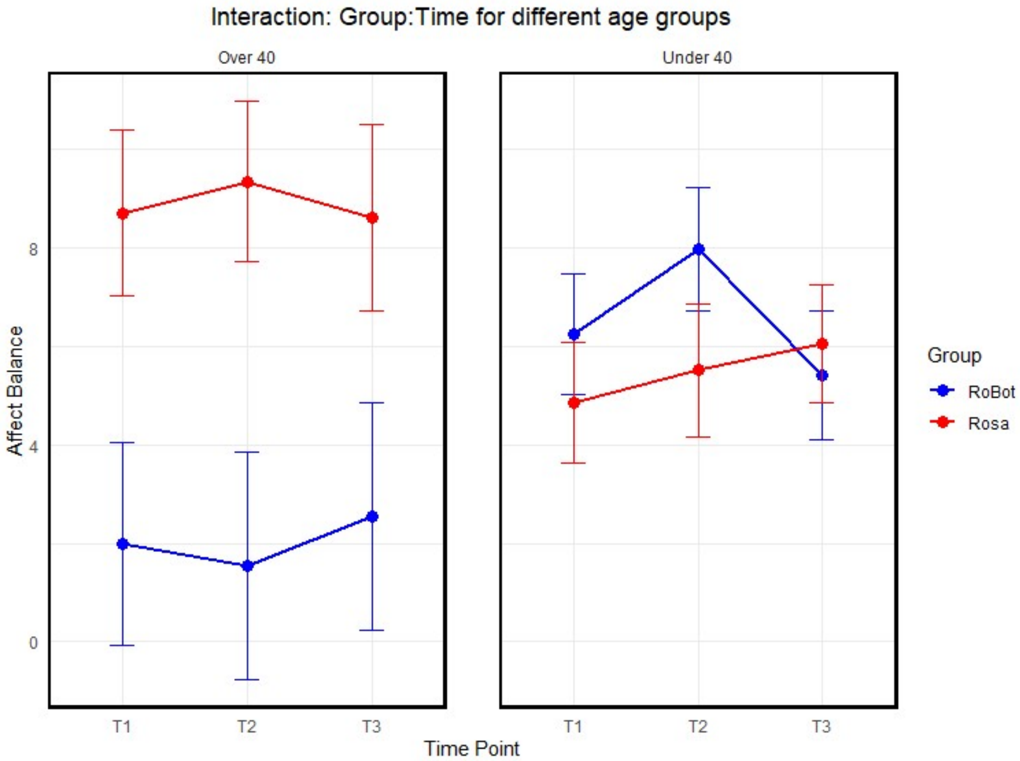

Figure 20: Difference in the course of the Affect Balance of the age groups

Figure 20 illustrates the interaction between group assignment (RoBot vs. Rosa) and measurement time points (T1, T2, and T3) across two distinct age groups (Over 40 vs. Under 40). The dependent variable, Affect Balance is plotted as a function of time, with error bars representing the standard error of the mean.

In the Over 40 group, participants in the Rosa group exhibit a consistently higher Affect Balance across all time points compared to the RoBot group. The trend for the Rosa group remains relatively stable, with a slight peak at T2 before returning to baseline at T3. In contrast, the RoBot group demonstrates lower Affect Balance scores throughout, maintaining a relatively stable pattern over time with minor fluctuations.

In the Under 40 group, the Rosa group shows a more linear pattern, with an increase from T1 to T2 to T3. The RoBot group exhibits a pronounced peak at T2, with a subsequent decline at T3, converging towards the values of the Rosa group. The fluctuations within the RoBot group suggest greater variability in Affect Balance over time compared to the relatively stable Rosa group.

These results indicate that group differences in Affect Balance are more pronounced in the Over 40 group, where the Rosa group consistently reports higher scores than the RoBot group. In the Under 40 group, the RoBot group initially experiences an increase in Affect Balance, surpassing the Rosa group at T2 before declining at T3, suggesting a more dynamic response pattern.

4.6.4 Principal Component Analysis of GQ-6

The objective of this analysis was to identify underlying latent structures in the gratitude scale and assess whether a smaller set of components could adequately explain the variability observed in the data.

The summary output of the PCA indicated that the first principal component (PC1) accounted for a substantial proportion of the variance, suggesting that a single component might sufficiently represent the gratitude construct. However, the second component (PC2) exhibited a unique loading pattern, particularly with respect to the Third Item (original statement: "*When I look at the world, I do not see much for which I am grateful.*"), indicating potential divergence from the primary latent construct. This item was reverse-coded in the original scale.

An examination of the component loadings showed that the Third Item was strongly associated with PC2, while the other items were primarily associated with PC1. This distinction indicated that the Third Item captured a different aspect of gratitude than the rest of the scale. This discrepancy suggested that the Third Item might not be measuring the same underlying construct as the other items. Consequently, the Third Item was excluded from the subsequent score calculation to enhance construct validity and internal consistency.

The exclusion of the Third Item was further supported by an internal consistency analysis using Cronbach's alpha. The initial scale (including the Third Item) yielded a reliability coefficient of 0.808, whereas the modified scale (excluding the Third Item) demonstrated an improved alpha of 0.863. This increase in internal consistency reinforced the decision to remove the Third Item, ensuring that the final gratitude score was more reliable and internally coherent.

Effect	***df (num)***	***df (denom)***	***F-value***	***p-value***
(Intercept)	1	185	2516.8532	<.001 ***
Group	1	98	0.1229	.727
Time	2	185	0.1084	.897
Group:Time	2	185	4.3863	.014 *

Table 12: Results of the LMM on Gratitude Expression after excluding the Third Item

A comparison of the LMM results (shown in Table 12) before and after the application of PCA highlights substantial changes in the statistical significance of key effects. The primary observation is the emergence of a significant interaction effect between Group and Timepoint after the refinement of the Gratitude Score. This effect was previously non-significant when using the original Gratitude Score. The newfound significance suggests that differences in gratitude scores between groups evolved distinctly over time, a pattern that may have been obscured by the inclusion of the divergent third item.

The main effects of Group and Timepoint, however, remained non-significant before and after the PCA refinement. The Group effect and the Timepoint effect exhibited no meaningful changes, indicating that the primary shift in statistical outcomes resulted from a clarified interaction effect rather than from overarching changes in group means.

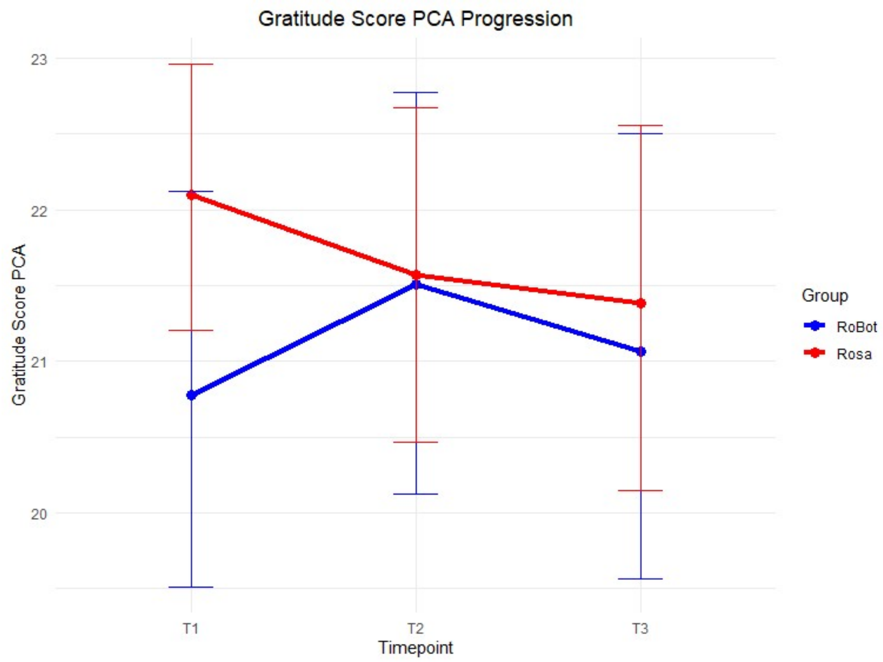

Figure 21: Progression of the Gratitude Score excluding the Third Item

4.6.5 Intervention Effect in Participants with Low Baseline Gratitude

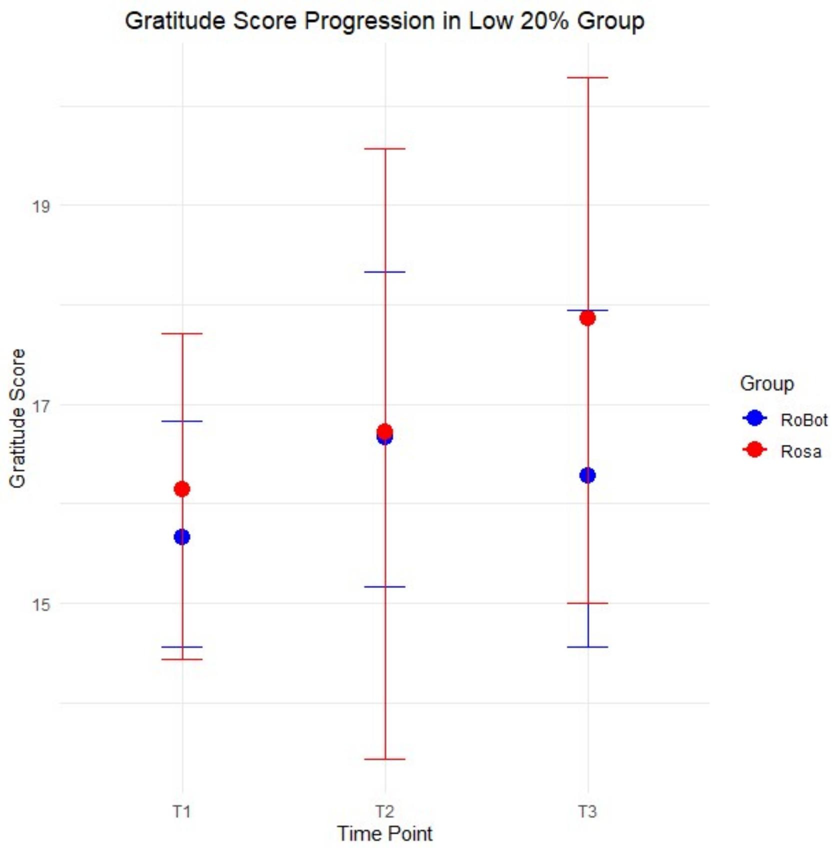

Figure 22: Progression of Gratitude Expression of the Low 20% Group by Group

Descriptive analyses of the dataset revealed that the overall gratitude scores at baseline (T1) were already relatively high, indicating a potential ceiling effect. Given this, it was essential to explore whether the intervention had a different impact on individuals who initially reported lower gratitude levels. To achieve this, an exploratory subgroup analysis was conducted, focusing on participants whose gratitude scores at T1 fell within the lowest 20% of the distribution. To identify participants with particularly low initial gratitude, the 20th percentile of the T1 gratitude scores was calculated. Participants with scores at or below this threshold were classified as the "Low 20%" group. This subset was then analyzed separately to examine whether the intervention led to distinct patterns of gratitude development over time. A LMM was applied to investigate the effects of time using the Gratitude Score without the third item, group assignment ("RoBot" vs. "Rosa"), and their interaction within this specific subset.

The results of the LMM indicated that the interaction between time and group was not statistically significant ($F(2, 46) = 1.22$, $p = .3051$), suggesting that the two groups did not develop differently over time. Additionally, the main effect of time approached marginal significance ($F(2, 46) = 2.62$, $p = .0839$), indicating a possible trend toward increased gratitude scores across all participants in the Low 20% group, irrespective of their group assignment. However, the main effect of group was non-significant ($F(1, 23) = 0.24$, $p = .6273$), showing that "RoBot" and "Rosa" participants did not exhibit substantial differences throughout the study period. Figure 22 visualizes the progression of gratitude scores over time within the Low 20% group, separated by chatbot condition. While the statistical analysis did not reveal a significant interaction effect, the Rosa group shows a noticeable upward trend in gratitude scores across time points.

5 Discussion

5.1 Key Findings

The present study aimed to examine the effects of chatbot interaction on gratitude expression, affect balance, and social presence. Through a series of statistical analyses, several key findings emerged.

The results did not indicate a significant difference in perceived social presence between the two chatbot groups (Rosa vs. RoBot) at the post-intervention measurement (T2). Participants experienced similar levels of social presence regardless of the chatbot condition, contradicting the assumption that one chatbot might foster a greater sense of presence than the other.

Regarding gratitude expression, LMM analyses revealed that gratitude levels remained stable throughout the study period, with no statistically significant main effects of time or group differences observed. Additionally, the interaction between group assignment and time was also non-significant, indicating that chatbot exposure did not differentially influence gratitude expression across measurement points.

Analyses of affect balance, positive affect, and negative affect demonstrated statistically significant time-based effects, indicating meaningful fluctuations in participants' affective states throughout the study period. Notably, affect balance increased significantly after the initial interaction (from T1 to T2), though this improvement did not persist at T3. However, no significant interaction effects between group and time were observed, suggesting that chatbot exposure did not differentially affect these measures.

Exploratory analyses highlighted educational background and age as important predictors of affective well-being. Participants with an academic background demonstrated higher positive affect and affect balance scores compared to those with non-academic backgrounds. Additionally, older participants in the Rosa condition exhibited significantly higher affect balance scores compared to their counterparts in the RoBot condition, suggesting that chatbot interaction effects are moderated by age-related factors.

In conclusion, while chatbot interactions did not significantly alter gratitude expression, they did influence affective states in a temporally specific manner. Demographic factors, such as educational background and age, played a crucial role in shaping psychological outcomes. The implications of these results will now be discussed in detail.

5.2 Discussion of the Results

H1: Social cues and conversational memory enhance perceived social presence.

The theoretical foundation of this study was based on the hypothesis that social cues and conversational memory enhance perceived social presence in human-chatbot interactions. This expectation was grounded in established theories of mediated communication, human-computer interaction, and social cognition.

Social Presence Theory (Short et al. 1976) posited that the perceived "realness" of an interaction partner in a mediated environment would influence the quality and depth of user engagement. According to this perspective, digital agents that exhibit social cues, such as human-like expressions, personalized responses, and relational memory, were expected to evoke a sense of co-presence, thereby increasing user immersion and trust. This aligned with the Computers Are Social Actors (CASA) paradigm (Reeves and Nass 1996), which suggested that humans instinctively apply social norms to computers and other digital entities when they exhibit social behaviors. The CASA framework provided a strong rationale for the assumption that chatbots simulating human conversational patterns would be perceived as more socially present than those that did not.

Based on these theoretical underpinnings, it was hypothesized that Rosa, a chatbot designed with conversational memory and social cues, would be perceived as more socially present than RoBot, a chatbot lacking these features. This expectation stemmed from empirical evidence suggesting that chatbots with anthropomorphic characteristics, such as human-like avatars, the ability to recall user information, and adaptive dialogue, elicit stronger social responses (Pereira et al. 2014; Van Der Goot 2022). The anticipated effect of these design elements was an enhanced sense of social presence, which in turn was presumed to facilitate deeper emotional engagement and greater perceived support from the chatbot.

Contrary to these theoretical expectations, the results of the study did not provide significant evidence supporting the hypothesis. While Rosa's Social Presence score was marginally higher than that of RoBot, this difference was not statistically significant.

These findings were unexpected and diverge from the initial assumptions that a more human-like chatbot would elicit higher perceptions of social presence. Nonetheless, the findings align with prior research that suggests the effects of social presence mechanisms can be highly context-dependent. For instance, Haugeland et al. (2022) found that while topic-led conversations increased perceived anthropomorphism and hedonic quality, they did not significantly enhance perceived social presence. Similarly, Araujo (2018) reported no substantial differences in social presence between human-like and neutral chatbots unless participants were explicitly primed to perceive the chatbot as intelligent. These results suggest that social presence perceptions are not solely dictated by chatbot design but are also influenced by users' expectations and situational context.

Additionally, some studies (Gnewuch et al. 2018) indicate that social cues enhance perceived social presence in customer service interactions, where relational engagement is more relevant. However, the effectiveness of anthropomorphic features varies significantly across applications. Go and Sundar (2019) highlight the delicate balance required in chatbot design: excessive anthropomorphism can lead to skepticism and reduced engagement, while insufficient anthropomorphism may fail to create the intended social connection. This balance may explain why Rosa's design did not significantly impact social presence perceptions in the current study.

To better understand the lack of significant findings, this section first examines possible methodological explanations before exploring alternative theoretical perspectives.

Assessing social presence through explicit self-report measures presents significant challenges, as highlighted by Biocca et al. (2001). Social presence is inherently subjective and multifaceted, making it difficult to capture through standardized questionnaires alone. Participants may struggle to articulate their sense of social presence, as it often operates at a subconscious level rather than being a fully conscious experience (Nass & Moon, 2000). This discrepancy between perception and measurement may have led to an underestimation of the actual social presence experienced during interactions with the chatbots. The characteristics of the participant sample may also have influenced the results. If the sample consisted of predominantly tech-savvy individuals, they might have been more critical of anthropomorphic design features and less inclined to perceive the chatbot as socially present. Prior experience with AI-based conversational agents could have led to heightened scrutiny of Rosa's social cues and conversational memory, making these features less impactful than anticipated.

Moreover, cognitive load during the interactions may have affected participants' attention to social presence cues. According to Lee and Nass (2003), when users are focused on task completion, such as engaging with gratitude exercises, their cognitive resources may be primarily allocated to the task itself rather than to evaluating the chatbot's social characteristics. This shift in focus could have reduced the salience of social presence cues, diminishing their overall effect.

Another possible explanation for the lack of significant findings is the expectation discrepancy created by Rosa's human-like traits. If Rosa's social cues and conversational memory raised users' expectations for highly natural, human-like interactions, but the chatbot ultimately failed to meet these expectations, perceived social presence could have been negatively impacted. Instead of fostering a sense of engagement, these discrepancies may have made users more aware of the artificial nature of the interaction, reducing their overall sense of connection. Presence would have been weakened.

These findings are consistent with Rheu et al. (2024), who demonstrated that a mismatch between a chatbot's social role cues (e.g., expert label) and its actual interactional behavior (e.g., providing generic instead of contingent feedback) can result in negative expectancy violations. Chatbots labeled as experts received lower ratings of trust, sincerity, caring, and understanding when they failed to provide contingent responses, compared to non-expert chatbots. This effect was reflected in lower ratings of trust, sincerity, caring, and understanding. Emotional validation was also reduced under these expectancy violation conditions. Such findings suggest that users develop expectations based on a chatbot's persona and, when these are unmet, the resulting discrepancy can significantly impair perceived social presence and communication outcomes (Rheu et al. 2024).

Another crucial factor could be the authenticity of Rosa's social behaviors. If the chatbot's responses appeared repetitive or overly scripted, users might have perceived the interaction as formulaic rather than genuinely engaging (Gnewuch et al. 2018). In human conversations, variation and spontaneity play an essential role in building relational depth. If Rosa's social behaviors did not dynamically adapt to user input, its social cues may have been dismissed as superficial or insincere. This would explain why Rosa's enhanced features failed to translate into a significantly higher perception of social presence. The design and implementation of Rosa's conversational memory may also have influenced the results. Rosa primarily retained trivial user inputs, such as names, greetings, or simple preferences, without integrating this information into deeper, context-aware responses, the memory function might not have meaningfully enhanced the interaction (Luger and Sellen 2016).

H2a: Gratitude interventions lead to an increase in gratitude expression.

H2b: A chatbot with higher social presence leads to a greater increase in gratitude expression than a chatbot with lower social presence.

H2a posits that gratitude interventions lead to a measurable increase in gratitude expression. Grounded in positive psychology, this expectation aligns with prior research demonstrating that structured gratitude exercises can enhance an individual's capacity to notice and articulate the positive aspects of daily life (Sin and Lyubomirsky 2009). The underlying rationale is that systematic reflection, whether through journaling or guided chatbot prompts, broadens one's perspective on existing blessings and fosters more frequent instances of grateful thinking. Over time, this persistent awareness is hypothesized to manifest as heightened expressions of gratitude.

H2b extends this idea by suggesting that a chatbot with higher social presence catalyzes a greater increase in gratitude expression than a chatbot characterized by comparatively lower social presence. When people perceive an interaction as more authentic and interactive, hallmarks of heightened social presence, they tend to respond more fully on both emotional and behavioral levels (Tsai et al. 2021). The anthropomorphic features of chatbots, including personalized dialogue, memory of previous user inputs, or human-like communication styles, are known to strengthen relational depth (Bae Brandtzæg et al. 2021). In the context of gratitude interventions, such features may amplify users' motivation to engage with the exercises and share deeper reflections, thereby leading to an even stronger boost in expressed gratitude.

However, the empirical findings of this study do not support these theoretical expectations. Contrary to H2a, participants who engaged with gratitude interventions did not exhibit a statistically significant increase in gratitude expression over time. Similarly, H2b was not confirmed, as the chatbot's social presence did not yield a differential effect on gratitude expression. Importantly, the foundational assumption underlying H1, that the implementation of social cues would result in higher perceived social presenc, was not supported. This lack of effect likely influenced the outcomes related to H2b, as the hypothesized mechanism linking social presence to increased gratitude expression could not unfold in the absence of a clear difference in perceived social presence between the chatbot conditions.

In particular, the results of Hypothesis 2a are incongruent with those of the study by Lee et al. (2023), which will be reexamined from a different perspective. One of the most notable concerns in Lee et al.'s study is that the significance of their results on gratitude, measured by GQ-6, may not stem from a meaningful increase in the chatbot condition but rather from the decline observed in the control group. While the gratitude scores in their chatbot condition showed a small improvement, the control group exhibited a decline in gratitude over time. This raises the possibility that the statistical significance of their findings is driven by a downward trend in the control group rather than a robust effect of the chatbot intervention.

In contrast, the present study implemented a control group that followed the same structure as Lee et al.'s experimental group, a chatbot without anthropomorphic features but still delivering gratitude interventions. Notably, in this study, gratitude levels in the control group did not decline, remaining relatively stable over time. This finding suggests that the control group's decrease in gratitude in Lee et al.'s study may not be a generalizable effect but rather a context-dependent artifact. If their control group had remained stable, the statistical significance of their results might have been diminished, raising concerns about whether the chatbot intervention was genuinely effective or whether the observed significance was an artifact of control group fluctuations.

The lack of significance of both Hypotheses 2a and 2b can be attributed to a number of factors. When participants begin an intervention with already elevated levels of gratitude, the capacity for further improvement is inherently constrained, a phenomenon widely known as the ceiling effect (Bohlmeijer et al. 2021). In such cases, there is insufficient room to observe marked changes in gratitude because the baseline is too high to accommodate further growth. This limitation is not merely a statistical artifact but also a theoretical consideration: individuals with consistently high trait gratitude may have already integrated reflective or appreciative practices into their daily cognitive and emotional routines, thereby blunting the potential impact of structured gratitude interventions (Sin and Lyubomirsky 2009). In the present study, descriptive analyses indicated that gratitude levels at baseline were already relatively high across participants, suggesting that a ceiling effect may have influenced the results. This raises important considerations regarding the differential impact of gratitude interventions depending on participants' initial levels of gratitude. To explore this issue further, an analysis was conducted on individuals whose gratitude scores at baseline fell within the lowest 20% of the distribution. This subgroup, theoretically less affected by the ceiling effect, showed slight but inconsistent increases in gratitude over time. However, the absence of statistically significant differences between the two chatbot conditions suggests that while individuals with initially lower gratitude levels may benefit more from gratitude interventions, the specific design and implementation of the intervention in this study were not sufficient to produce substantial or systematic improvements.

Empirical findings further support this dynamic. A school-based study, for instance, showed that learners who initially displayed high gratitude levels demonstrated only minimal improvement from an intervention, whereas those with moderate or low baseline scores reported substantially greater gains (Watkins 2014). This discrepancy underscores the importance of baseline variability in determining the effectiveness of gratitude-focused activities. Moreover, elevated initial scores decrease the statistical power to detect meaningful differences over time, as score fluctuations among high-gratitude individuals tend to be smaller and less stable (Seligman et al. 2005). Consequently, the intervention's measurable impact is diminished, potentially leading to misleading conclusions about its efficacy.

Another potential explanation for the lack of observed effects in gratitude expression is the short duration of the intervention. A three-day gratitude intervention may be insufficient to produce meaningful and lasting changes in gratitude levels. Prior research on gratitude interventions consistently suggests that longer intervention periods, typically spanning several weeks to months, are necessary to induce measurable improvements in gratitude and well-being (Bohlmeijer et al. 2021). A meta-analysis of gratitude interventions found that programs lasting four or more weeks led to more pronounced increases in gratitude and psychological well-being compared to shorter interventions (Sin and Lyubomirsky 2009). This aligns with findings from experimental studies, which indicate that the development of habitual gratitude requires sustained practice over time. Significant benefits of gratitude interventions only begin to emerge after at least four weeks, with no measurable effects found in interventions lasting merely two weeks (Bohlmeijer et al. 2021). These findings suggest that repeated engagement with gratitude exercises over an extended period is necessary to foster a stable shift in gratitude expression.

Moreover, studies on digital gratitude exercises further support the notion that a short intervention period may be inadequate. The effects of the widely used "Three Good Things" gratitude exercise were found to be small and non-significant when the intervention was conducted over a short duration, suggesting that longer exposure and continuous reinforcement are necessary to consolidate cognitive and emotional changes associated with gratitude practice (Seligman et al. 2005). A three-day intervention, as implemented in this study, may not have provided participants with enough opportunities to internalize gratitude as a habitual response or to undergo the cognitive restructuring required for meaningful shifts in gratitude expression.

Furthermore, gratitude is a multifaceted construct that encompasses emotional, cognitive, and behavioral dimensions, which may not be fully captured by standardized quantitative measures (Watkins 2014). The measurement tools employed in this study, such as the GQ-6 scale, are primarily designed to capture stable, long-term changes in gratitude and emotional well-being (Diener et al. 2010; McCullough et al. 2002). Given that these instruments are better suited to detecting gradual changes over weeks or months, the short intervention period and the relatively brief follow-up measurement window (one week) may have limited the ability to statistically capture any potential gratitude gains. Instead, observed variations in gratitude expression may have been influenced by transient mood fluctuations, measurement errors, or participants' daily experiences rather than reflecting true intervention-induced effects.

H3a: Gratitude interventions lead to improved well-being outcomes.

H3b: A chatbot with higher social presence leads to greater improvements in well-being outcomes than a chatbot with lower social presence.

H3a posits that gratitude interventions lead to improved well-being outcomes. Prior research indicates that expressing gratitude fosters positive emotions, strengthens social bonds, and mitigates stress-related psychological burdens. Given that chatbots facilitate structured gratitude interventions, their implementation is expected to yield similar benefits by prompting users to reflect on and articulate grateful thoughts. The underlying mechanisms align with cognitive theories of emotion regulation, wherein the focus on positive aspects of life counters negative affective states. Studies suggest that individuals who engage in gratitude exercises exhibit increased emotional resilience and life satisfaction, effects that persist over time (Cunha et al. 2019). Furthermore, the accessibility and adaptability of chatbot-based gratitude interventions can enhance adherence and engagement, leading to sustained well-being improvements. The social and affective reinforcement provided by the chatbot interaction may further amplify these benefits by making the gratitude practice feel more immersive and meaningful (Boucher et al. 2021). Thus, it is anticipated that chatbot-facilitated gratitude interventions will lead to significant improvements in emotional well-being, reinforcing H3a that gratitude expression enhances psychological health.

H3b states that a chatbot with higher social presence leads to greater improvements in well-being outcomes than a chatbot with lower social presence. The effectiveness of chatbot-based gratitude interventions may be further moderated by the chatbot's perceived social presence. Social presence has been shown to enhance user engagement and emotional connection. A chatbot with high social presence, characterized by anthropomorphic cues, conversational memory, and personalized responses, is expected to create a more engaging and relationally immersive experience compared to a chatbot with lower social presence. Research suggests that individuals respond to socially engaging chatbots similarly to human interactions, applying social norms and emotional expectations to the interaction. Consequently, a chatbot with heightened social presence can elicit stronger affective responses, making gratitude interventions more emotionally impactful (Oh et al. 2018).

The CASA (Computers Are Social Actors) paradigm supports this assertion, positing that individuals unconsciously attribute social characteristics to chatbots that exhibit human-like behaviors. By fostering a sense of companionship and interactional warmth, chatbots with higher social presence may enhance users' motivation to engage deeply with gratitude exercises, thereby leading to more pronounced improvements in well-being. In contrast, chatbots with minimal social presence might still facilitate gratitude interventions effectively but lack the relational depth required to maximize emotional benefits. Thus, it is expected that users interacting with a chatbot featuring higher social presence will report greater improvements in well-being compared to those engaging with a chatbot exhibiting lower social presence, supporting H3b.

The results of the present study partially supported these theoretical expectations. Contrary to the anticipated improvements, chatbot-facilitated gratitude interventions did not lead to significant long-term changes in emotional resilience. However, there was evidence of short-term improvements in affect balance immediately following the initial intervention phase, indicating a temporary positive emotional shift. This finding aligns with research suggesting that gratitude interventions can generate short-term emotional benefits, though sustaining these improvements over time may require additional interpersonal reinforcement or ongoing interaction.

Regarding H3b, the expected advantage of a chatbot with higher social presence over one with lower social presence was not observed. While participants interacting with the socially engaging chatbot reported slightly higher levels of perceived social presence, this did not translate into significant differences in well-being outcomes. These results indicate that while anthropomorphic cues and conversational memory may enhance user interaction experiences, their direct impact on emotional health outcomes remains inconclusive and possibly influenced by additional moderating variables such as user characteristics, baseline mental health, and individual differences in digital interaction preferences.

The findings of the present study align with a growing body of research indicating that fully automated chatbot interventions generally yield only minimal effects on psychological well-being (Andersson and Titov 2014). Other findings support this trend. For example, moderate short-term effects of AI-based chatbots on depression and anxiety were observed, but these diminished substantially by the three-month follow-up, indicating limited long-term benefits (Zhong et al. 2024). Similarly, small initial improvements in positive emotions following mobile positive-psychological interventions were not maintained over time (Tomczyk and Ewert 2025). A review of 19 studies on digital self-help interventions also revealed a small standardized mean effect of only 0.19 on well-being (Li et al. 2023). Together, these meta-analytic findings underscore the limited and typically transient nature of emotional and psychological improvements resulting from chatbot-based interventions.

The variability in affective outcomes, particularly in measures such as positive and negative affect assessed through instruments akin to SPANE, further underscores the inconsistency of chatbot interventions. Some studies have reported improvements, while others have found no significant change, suggesting that individual differences, context, and methodological variations might account for these discrepancies. The present study’s findings, namely, the absence of significant SPANE changes and the lack of a notable Group × Time interaction, thus fit well within the broader evidence base that questions the robustness of chatbot-driven well-being enhancements (Li et al., 2023).

Meta-analyses examining broader digital mental health interventions confirm this pattern. Chatbot-based interventions may provide minor relief in specific conditions such as stress, anxiety, and depression, but their influence on general psychological well-being has been found to remain minimal. Additionally, digital self-help interventions have been shown to provide only slight benefits when compared to no intervention, with these benefits disappearing when an active control condition is introduced (Groot et al. 2023). These findings suggest that the chatbot itself may not provide a unique advantage over other self-help formats.

Several theoretical and practical explanations can account for these results, aligning with previous research on digital mental health interventions. The following sections outline key factors that may have contributed to the observed outcomes.

One central explanation concerns the length and intensity of the chatbot intervention. The role of intervention length was already discussed in the context of Hypothesis 2, however, here it is further examined within the broader framework of mental health interventions. A meta-analysis of AI-based chatbot interventions for depression and anxiety indicated that short interventions typically yield only modest, short-term effects, whereas longer interventions appear necessary to achieve enduring psychological change. The analysis found that the strongest effects emerged after approximately eight weeks of use, with benefits diminishing by the three-month follow-up (Zhong et. al., 2024). Moreover, digital self-help tools frequently suffer from premature disengagement, with studies indicating that users complete, on average, only 40% of the provided content (Winkler and Soellner 2018). These findings suggest that insufficient exposure and limited adherence may explain why chatbot-based gratitude interventions fail to elicit substantial changes in global well-being measures such as SPANE.

Another important factor could be the role of social support in fostering well-being. Psychological research consistently highlights the necessity of authentic social reassurance in mental health interventions. While chatbots can simulate supportive interactions, they do not offer genuine human connection, which can reduce adherence and limit their effectiveness. Social-cognitive models of human-computer interaction emphasize that trust and perceived presence are key determinants of an intervention's success. Users must not only recognize the chatbot as a competent partner but also trust its ability to provide relevant support over time. However, evidence suggests that chatbot interactions often lack the sustained trust and engagement needed to facilitate long-term well-being improvements (Vaidyam et al. 2019).

Overall, the results of this study underscore the fundamental limitations of chatbots in delivering psychological support and fostering well-being improvements. Despite advancements in artificial intelligence, chatbots remain constrained by their inability to form genuine empathetic connections with users. In psychological interventions, the therapeutic alliance, characterized by trust, empathy, and a deep interpersonal bond, is widely recognized as a key determinant of efficacy. However, chatbots lack the nonverbal communication cues, nuanced emotional responsiveness, and adaptive social intelligence necessary to cultivate such an alliance (Vaidyam et al. 2019). As a result, users may perceive chatbot interactions as impersonal or superficial, limiting their willingness to engage deeply or disclose emotions fully.

Another significant limitation lies in the standardized nature of chatbot interactions. Unlike human therapists, who tailor their responses to the unique needs of individuals, chatbots operate within pre-programmed frameworks that restrict their adaptability. While AI-driven systems can provide structured guidance, they often fail to address complex psychological concerns that require nuanced, context-sensitive intervention strategies (Vaidyam et al. 2019). This lack of personalization may contribute to the absence of measurable improvements in well-being, as interventions remain surface-level and fail to generate the depth of cognitive and emotional processing necessary for sustained psychological change.

These limitations also extend to gratitude interventions, which rely on emotional engagement and reflective processing to enhance well-being. While gratitude exercises have demonstrated efficacy in traditional settings, their impact may be diminished when facilitated by chatbots that lack human warmth and dynamic responsiveness.

The emotional depth required to fully internalize gratitude may not be achieved in automated exchanges, potentially explaining the lack of significant well-being improvements observed in this study. Taken together, these findings reinforce the growing recognition that chatbots, while offering scalable and accessible mental health support, are unlikely to replace human-delivered interventions in their current form. Their inability to provide deep emotional resonance, tailored psychological insights, and sustained therapeutic relationships constrains their effectiveness in fostering well-being. Future advancements should focus on integrating more adaptive, emotionally intelligent systems and hybrid mode

5.3 Discussion of the Explorative Analyses

While the primary hypotheses of this study focus on predefined relationships between gratitude, affect balance, and chatbot-based interventions, it remains crucial to examine the influence of additional covariates. These exploratory analyses offer valuable insights into demographic factors, such as gender, educational background and age, that may impact the effectiveness of gratitude interventions. Although these covariates were not explicitly incorporated into the primary hypotheses, their potential influence on intervention outcomes underscores the importance of contextual factors in digital mental health interventions.

The analysis revealed a significant gender effect on gratitude scores, with female participants reporting higher levels of gratitude compared to their male counterparts. This finding is consistent with prior psychological research suggesting that women tend to exhibit greater emotional expressiveness and prosocial behavior. Socialization processes may play a central role in shaping these gender differences, as women are often encouraged from an early age to express gratitude and appreciation more openly than men. Additionally, emotional regulation strategies may differ between genders, with women potentially utilizing gratitude as a coping mechanism to a greater extent than men (Kashdan et al. 2009). These findings align with studies indicating that women score higher on measures of positive affect and interpersonal sensitivity, which may facilitate the cultivation and expression of gratitude. Future research should explore whether cultural variations influence these gender-based differences and whether targeted gratitude interventions can be designed to engage both men and women effectively.

The results demonstrated that educational background interacts with group assignment to influence gratitude scores. Specifically, while gratitude levels in the Rosa group remained stable across different education levels, participants with an academic background in the RoBot group exhibited slightly higher gratitude scores. One possible explanation for this trend is that individuals with a higher education level may possess greater cognitive flexibility, enabling them to engage more deeply with gratitude interventions. Supporting this explanation, previous research indicates that higher education enhances cognitive functioning broadly, including executive functioning and reasoning skills, with particular improvements noted in tasks requiring cognitive flexibility and complex reasoning. These cognitive enhancements may equip individuals with the necessary skills to process gratitude interventions more effectively (Guerra-Carrillo et al. 2017). Thus, the observed increase in gratitude among academically educated individuals in the RoBot group aligns with findings that cognitive flexibility, strengthened through educational experiences, facilitates deeper engagement and responsiveness to psychological interventions.

Moreover, individuals with advanced education may have greater exposure to psychological self-regulation strategies, allowing them to reflect more effectively on gratitude-related experiences. Research supports this notion, indicating that higher educational achievement correlates with an enhanced ability to engage in self-regulatory behaviors. Specifically, adults with higher education levels demonstrated improved impulse control, decision-making capabilities, and goal orientation. These self-regulation skills facilitate reflective practices and effective cognitive engagement, further supporting the capability of highly educated individuals to derive greater benefit from structured interventions like gratitude exercises. Consequently, the educational context provides individuals with not only the cognitive but also the psychological tools necessary for effective self-regulation and deeper experiential reflection (Vaculíková 2024). However, it remains unclear why this effect was observed predominantly in the RoBot condition. Future studies should investigate whether tailored educational interventions could enhance gratitude expression among non-academic populations by incorporating different pedagogical approaches or varying the complexity of gratitude exercises.

Educational background also emerged as a significant predictor of affect balance, with individuals possessing an academic background reporting higher affect balance scores than those without. This finding supports previous research linking higher education levels to enhanced emotional regulation and psychological resilience. One explanation for this relationship is that education equips individuals with problem-solving skills and coping strategies that mitigate negative emotional experiences, thereby fostering greater emotional stability. Furthermore, those with higher education levels may have increased access to mental health resources and social support networks, which can contribute to overall well-being (Cutler and Lleras-Muney 2010). Given that affect balance serves as a comprehensive indicator of emotional well-being, these results underscore the importance of educational attainment in fostering psychological resilience. Future research should examine whether specific educational experiences, such as exposure to emotional intelligence training or mindfulness practices, contribute to these observed differences. Additionally, targeted interventions could be designed to support individuals with lower education levels in cultivating affect balance through structured gratitude exercises or psychoeducational programs.

Age was found to interact significantly with group assignment in predicting affect balance. Notably, among participants over the age of 40, those assigned to the Rosa group exhibited substantially higher affect balance scores than their counterparts in the RoBot group. Conversely, among participants younger than 40, the RoBot group demonstrated slightly higher affect balance scores relative to the Rosa group. These findings suggest that older participants may respond more positively to chatbot interventions that incorporate socially engaging design elements, such as anthropomorphic cues and conversational memory. In contrast, younger participants, who may be more accustomed to interacting with digital interfaces, may not require these human-like elements to benefit from chatbot-based gratitude interventions (Rodríguez-Martínez et al. 2023). The observed differences highlight the importance of designing chatbot interventions that are tailored to the preferences and expectations of different age groups. Future studies should investigate whether user-specific customization, such as adaptive conversational styles or varying levels of anthropomorphism, enhances intervention efficacy across different age demographics. Moreover, research could explore whether older participants derive greater psychological benefits from digital interventions that explicitly mimic human social interaction, as opposed to younger users who may prioritize efficiency and functionality over social presence.

The findings underscore the need for personalized chatbot interventions that consider demographic variables to maximize engagement and effectiveness. Specifically, developing chatbot interactions that adjust based on a user's age, gender, or educational background could significantly enhance the user experience and optimize the effectiveness of digital gratitude interventions.

6 Strengths & Limitations

This study addresses a pressing global mental health challenge by exploring the role of anthropomorphic chatbots in fostering gratitude-based interventions. Given the increasing prevalence of anxiety and depression, particularly in the aftermath of the COVID-19 pandemic (Manchia et al. 2022), innovative and scalable solutions for mental health support are urgently needed. By integrating positive psychology principles with artificial intelligence, this research contributes to an emerging interdisciplinary field at the intersection of psychology, human-computer interaction, and digital health technologies. While previous research has demonstrated the effectiveness of gratitude interventions, few studies have examined the role of anthropomorphic chatbot features in enhancing user experience and emotional impact. This study directly addresses this gap by evaluating whether human-like chatbot attributes improve the depth and effectiveness of gratitude interventions.

The study employs a randomized quantitative longitudinal design, allowing for the assessment of both short-term and sustained effects of chatbot-mediated gratitude interventions. Unlike cross-sectional studies, which provide only a snapshot of user responses, this methodological approach enables the tracking of emotional and cognitive changes over multiple time points. This increases the internal validity of the findings and offers a more nuanced understanding of gratitude development over time.

A key strength of the study lies in its strong theoretical foundation, drawing upon multiple well-established frameworks such as Social Presence Theory, the CASA paradigm, and the Hedonic Adaptation Prevention Model. By grounding the study in these theories, the research enhances its conceptual rigor and contributes to theory-building in AI-mediated interventions.

The research provides practical insights into how chatbot design influences user engagement and emotional well-being. This has direct implications for developers of digital mental health interventions, suggesting that personalized conversational memory can enhance user engagement, anthropomorphic cues such as avatars and personalized responses may increase social presence, and structured gratitude interventions in digital formats can provide scalable mental health support. These findings are valuable for healthcare practitioners, AI designers, and policymakers, particularly in the realm of Digital Health Applications in Germany.

The study leverages a widely used messaging platform to deliver chatbot interventions, ensuring high ecological validity. Many chatbot studies rely on controlled laboratory settings, which do not accurately reflect real-world usage patterns. By embedding the chatbot in an existing communication platform, this study mirrors real-life engagement scenarios, increasing the generalizability of the findings.

In summary, this study excels in its rigorous design, theoretical depth and practical applicability. By integrating AI-driven gratitude interventions with established psychological theories, it makes a valuable contribution to both research and practice. The findings not only advance scientific understanding of chatbot-mediated mental health interventions but also provide actionable insights for the design of future digital mental health tools. The study stands as a strong foundation for further exploration of AI-human interactions in the context of psychological well-being.

While this study presents several strengths, including its robust theoretical foundation, rigorous methodological design, and valuable practical implications, it is also essential to acknowledge its limitations. These constraints provide critical insights into areas where future research can refine and expand upon the findings, ensuring that chatbot-mediated gratitude interventions continue to evolve and improve in effectiveness.

One of the primary limitations of this study is that both chatbots utilized in the intervention were rule-based rather than adaptive. This design choice constrained the flexibility and responsiveness of the chatbot interactions, limiting the ability to generate truly personalized and context-sensitive dialogue. While conversational memory was integrated into one chatbot, its implementation remained scripted, meaning that responses were not dynamically generated based on nuanced user input. As a result, user engagement and perceived social presence may have been lower than if a more sophisticated machine learning-based chatbot had been used. The lack of adaptive learning capabilities may also have impacted the chatbot's ability to foster deeper emotional engagement, as human-like interactions often depend on contextual adaptation and evolving conversational depth.

The relatively short duration of the intervention is another limitation that may have contributed to the lack of significant findings. The chatbot-facilitated gratitude exercises spanned only three consecutive days, which is a much shorter duration than many previous studies on gratitude interventions. Existing research indicates that substantial and lasting changes in gratitude and emotional well-being often require interventions lasting several weeks or even months (Zhong et al. 2024). A three-day period may not have provided participants with sufficient time to internalize gratitude practices, form new cognitive patterns, or experience meaningful shifts in well-being measures. Additionally, the short follow-up period limited the ability to assess long-term retention effects, leaving open the question of whether any delayed benefits might have emerged with continued engagement.

Another limitation concerns the external validity of the study, as the participant sample primarily consisted of students and individuals with relatively high levels of digital literacy. This demographic composition may not fully represent the broader population that could benefit from chatbot-facilitated gratitude interventions. Individuals with lower levels of technological proficiency, older adults, or those from different cultural backgrounds may have distinct interaction patterns with chatbots, affecting the generalizability of the findings. Future research should aim to recruit a more diverse sample to ensure that results are applicable to a wider range of users.

The study also relied on self-report measures to assess key dependent variables, including gratitude expression, social presence, and affect balance. While validated scales were used, self-reported data are inherently susceptible to biases such as social desirability and subjective interpretation. Participants may have responded in ways they perceived as favorable rather than providing entirely accurate reflections of their emotional experiences. Future studies could benefit from incorporating behavioral or physiological measures, such as sentiment analysis of chatbot interactions or biometric indicators of emotional engagement, to complement self-reported outcomes and provide a more objective assessment of intervention effectiveness.

Additionally, the study's findings suggest that the implementation of conversational memory may not have been sufficiently deep to create a strong social connection. While the chatbot was programmed to recall user-provided information, the extent of personalization was limited to retrieving basic details rather than adapting responses in a meaningful way based on past interactions. More advanced implementations of conversational memory, such as the ability to reference user history dynamically and tailor interventions based on prior conversations, could enhance the perceived relational depth and increase engagement.

7 Further Research

Building on the limitations identified in this study, several directions for further research emerge, which could enhance the understanding of anthropomorphic chatbot-based gratitude interventions and improve their effectiveness in digital mental health applications. Addressing these aspects in future studies will not only refine chatbot design but also contribute to a more comprehensive understanding of how AI-driven interventions influence psychological well-being.

One of the most pressing areas for further research is the transition from rule-based to adaptive chatbot models. While the current study employed rule-based chatbots, which provided structured but limited conversational flexibility, future research should explore the implementation of machine learning-based chatbots that can dynamically adjust responses based on user input. Integrating NLP and deep learning algorithms could enhance the chatbot's ability to interpret user sentiment and tailor responses accordingly. Adaptive chatbots could create a more personalized and engaging experience, increasing user retention and fostering a stronger sense of connection. Studies comparing rule-based and AI-driven chatbots in gratitude interventions could provide insights into the extent to which adaptability influences user engagement and psychological outcomes.

Another key avenue for future investigation is extending the intervention duration. The current study's three-day intervention period may have been insufficient for producing measurable long-term changes in gratitude expression and well-being. Previous research suggests that sustained engagement over several weeks is necessary for gratitude practices to become habitual and to yield enduring psychological benefits. Future studies should implement interventions lasting several weeks or even months to assess whether prolonged chatbot interactions result in significant improvements in gratitude and emotional resilience. Longitudinal studies with follow-ups at multiple time points, such as one month, three months, and six months post-intervention, could provide valuable insights into the long-term efficacy of chatbot-based gratitude interventions.

Given the observed ceiling effect in gratitude expression, further research should explore the impact of chatbot interventions on individuals with lower baseline gratitude levels. The current study's findings suggest that participants with already high gratitude scores had limited room for improvement, potentially masking the chatbot's effectiveness. Future research should consider targeting populations with lower initial gratitude levels. By tailoring chatbot interventions to those who may benefit most, studies could yield more pronounced effects and offer practical recommendations for clinical and digital health applications.

Expanding the participant pool to enhance external validity is another important direction for future research. The current study predominantly involved digitally literate individuals, potentially limiting the generalizability of the findings. Future studies should recruit a more diverse sample, including older adults, individuals with limited technological proficiency, and those from different cultural backgrounds. Investigating how demographic factors such as age, education, and digital familiarity influence chatbot interaction and intervention efficacy could provide a more holistic understanding of chatbot-based gratitude practices. Additionally, cross-cultural studies examining how different populations perceive and engage with anthropomorphic chatbots could reveal important variations in user preferences and intervention effectiveness.

Addressing methodological improvements, future studies should incorporate a broader range of objective behavioral and physiological measures in addition to self-reported data. While self-report instruments remain valuable for assessing subjective experiences, they are susceptible to biases such as social desirability and participant interpretation. Future research should explore alternative assessment methods, such as sentiment analysis of chatbot interactions, passive biometric monitoring (e.g., heart rate variability as an indicator of emotional arousal), and linguistic analysis of user responses. These measures could provide a more comprehensive and unbiased evaluation of chatbot effectiveness in fostering gratitude and emotional well-being.

Further exploration of chatbot conversational memory is also warranted. While this study included basic conversational memory features, the chatbot's ability to retain and reference user-specific information remained limited. Future research should examine whether a deeper implementation of conversational memory, such as recalling past interactions, adapting responses over time, and maintaining a user history, enhances social presence and engagement. Investigating whether a chatbot's ability to provide continuity and personalized responses leads to stronger emotional connections and greater intervention efficacy would contribute valuable insights to chatbot development.

Finally, future research should consider integrating chatbot-based gratitude interventions into blended digital mental health care models. While stand-alone chatbot interventions provide scalability, their effectiveness may be enhanced when combined with other therapeutic approaches, such as human counseling or digital cognitive-behavioral therapy (CBT). Future studies should investigate hybrid models where chatbots serve as complementary tools alongside professional mental health support. Evaluating the synergistic effects of chatbot-facilitated gratitude interventions within broader mental health frameworks could reveal their potential role in improving mental well-being at scale.

By addressing these research directions, future studies can refine chatbot-based gratitude interventions, optimize user engagement, and enhance their psychological benefits. Leveraging technological advancements, expanding demographic inclusivity, and implementing rigorous methodological improvements will contribute to the continued evolution of AI-driven mental health solutions. Through these efforts, chatbot-based gratitude interventions can become more effective, accessible, and adaptable to diverse populations, ultimately improving emotional resilience and mental well-being.

8 Conclusion

This study examined the influence of anthropomorphic chatbot design features, specifically social cues and conversational memory, on perceived social presence, gratitude expression, and emotional well-being within digital gratitude interventions. Contrary to the initial hypothesis, the findings indicated that anthropomorphic features did not significantly enhance users' perceived social presence compared to a neutral chatbot. Moreover, perceived social presence did not demonstrate a significant impact on participants' gratitude expression or their emotional well-being.

Further analysis revealed that, overall, the gratitude interventions, regardless of chatbot design, did not significantly improve gratitude expression or emotional well-being measures, including affect balance, positive emotions, or negative emotions. However, exploratory analyses identified that participants with lower baseline gratitude levels benefited more notably from the interventions than those starting with higher gratitude levels. Additionally, demographic factors such as gender, age, and educational background were partially linked to variations in gratitude expression and emotional well-being outcomes, suggesting the relevance of these variables in understanding the effectiveness of digital gratitude interventions.

From a theoretical perspective, these findings challenge specific assumptions within the CASA framework, particularly concerning the effectiveness of anthropomorphic features in short-term, text-based interventions. Practically, the results indicate that incorporating advanced anthropomorphic features in chatbot-based gratitude interventions may not yield proportionate enhancements in user engagement or emotional outcomes. Consequently, simpler chatbot designs might be equally effective while also being more cost-efficient and scalable. Additionally, future interventions could specifically focus on individuals with lower baseline gratitude, given their apparent greater receptivity and benefit from gratitude-based interventions.

Several limitations of the study should be acknowledged. First, the short duration of the intervention might have limited the observable effects of anthropomorphic chatbot features. Second, the characteristics of the study sample restrict the generalizability of these findings to broader populations. Third, reliance on self-reported measures introduces potential methodological biases, including social desirability and subjective reporting inaccuracies.

Future research should explore longer-term interventions to determine if prolonged interaction enhances the effectiveness of anthropomorphic chatbot features. Investigating multimodal anthropomorphic elements, such as auditory and visual cues, could further clarify their potential to improve social presence and emotional outcomes. Additionally, subsequent studies should consider individual differences, such as personality traits and digital affinity, to gain deeper insights into the factors affecting the success of chatbot-based gratitude interventions.

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., and Househ, M. 2019. “An Overview of the Features of Chatbots in Mental Health: A Scoping Review,” *International Journal of Medical Informatics* (132), p. 103978. (https://doi.org/10.1016/j.ijmedinf.2019.103978).

Abernethy, Adams, L., National Academy of Medicine, Barrett, M., ResMed, Bechtel, C., X4 Health, Brennan, P., National Library of Medicine, Butte, A., University of California, San Francisco, Faulkner, J., Epic Systems, Fontaine, E., National Academy of Medicine, Friedhoff, S., Anthem, Inc., Halamka, J., Mayo Clinic, Howell, M., Google Health, Johnson, K., University of Pennsylvania, Long, P., Blue Shield of California, McGraw, D., Ciitizen Corporation, Miller, R., Johns Hopkins Hospital, Lee, P., Microsoft Corporation, Perlin, J., The Joint Commission, Rucker, D., 1up Health, Sandy, L., UnitedHealth Group, Savage, L., Omada Health, Inc., Stump, L., Yale New Haven Health System and Yale School of Medicine, Tang, P., Stanford University, Topol, E., The Scripps Research Institute, Tuckson, R., Tuckson Health Connections, LLC, Valdes, K., and b.well Connected Health. 2022. “The Promise of Digital Health: Then, Now, and the Future,” *NAM Perspectives* (6:22). (https://doi.org/10.31478/202206e).

Adamopoulou, E., and Moussiades, L. 2020a. “Chatbots: History, Technology, and Applications,” *Machine Learning with Applications* (2), p. 100006. (https://doi.org/10.1016/j.mlwa.2020.100006).

Adamopoulou, E., and Moussiades, L. 2020b. “An Overview of Chatbot Technology,” in *Artificial Intelligence Applications and Innovations* (Vol. 584), IFIP Advances in Information and Communication Technology, I. Maglogiannis, L. Iliadis, and E. Pimenidis (eds.), Cham: Springer International Publishing, pp. 373–383. (https://doi.org/10.1007/978-3-030-49186-4_31).

Adams, D. R., Williams, N. J., Becker-Haimes, E. M., Skriner, L., Shaffer, L., DeWitt, K., Neimark, G., Jones, D. T., and Beidas, R. S. 2019. “Therapist Financial Strain and Turnover: Interactions with System-Level Implementation of Evidence-Based Practices,” *Administration and Policy in Mental Health and Mental Health Services Research* (46:6), pp. 713–723. (https://doi.org/10.1007/s10488-019-00949-8).

Allouch, M., Azaria, A., and Azoulay, R. 2021. “Conversational Agents: Goals, Technologies, Vision and Challenges,” *Sensors* (21:24), p. 8448. (https://doi.org/10.3390/s21248448).

Andersson, G., Carlbring, P., Titov, N., and Lindefors, N. 2019. “Internet Interventions for Adults with Anxiety and Mood Disorders: A Narrative Umbrella Review of Recent Meta-Analyses,” *The Canadian Journal of Psychiatry* (64:7), pp. 465–470. (https://doi.org/10.1177/0706743719839381).

Andersson, G., and Titov, N. 2014. “Advantages and Limitations of Internet-Based Interventions for Common Mental Disorders,” *World Psychiatry* (13:1), pp. 4–11. (https://doi.org/10.1002/wps.20083).

Antaki, C. 2008. *Discourse Analysis and Conversation Analysis*. (https://repository.lboro.ac.uk/articles/chapter/Discourse_analysis_and_conversation_analysis/9478202).

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), pp. 183–189. (https://doi.org/10.1016/j.chb.2018.03.051).

Bachmann, C. J., Plener, P. L., Mechels, M., and Lempp, T. 2023. “Wie kann auch zukünftig eine gute stationäre Versorgung im Fachgebiet Kinder- und Jugendpsychiatrie gewährleistet werden? Ein Diskussionsbeitrag aus deutscher Perspektive,” *neuropsychiatrie* (37:1), pp. 39–46. (https://doi.org/10.1007/s40211-023-00458-z).

Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., and Følstad, A. 2021. “When the Social Becomes Non-Human: Young People’s Perception of Social Support in Chatbots,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 6, pp. 1–13. (https://doi.org/10.1145/3411764.3445318).

Bae, H., Kim, J., Lee, S. H., Lee, J., Lee, J.-Y., Park, H. Y., Kim, Y., and Kwon, K. T. 2022. “Difficulties of Healthcare Workers Encountered under Cohort Isolation in a Psychiatric Hospital during the COVID-19 Pandemic: A Qualitative Study,” *Psychiatry Research* (313), p. 114615. (https://doi.org/10.1016/j.psychres.2022.114615).

Bickmore, T. W., and Picard, R. W. 2005. “Establishing and Maintaining Long-Term Human-Computer Relationships,” *ACM Transactions on Computer-Human Interaction* (12:2), pp. 293–327. (https://doi.org/10.1145/1067860.1067867).

Biocca, F. 1997. “The Cyborg’s Dilemma: Progressive Embodiment in Virtual Environments,” *Journal of Computer-Mediated Communication* (3:2), pp. 0–0. (https://doi.org/10.1111/j.1083-6101.1997.tb00070.x).

Biocca, F., Harms, C., and Burgoon, J. K. 2003. “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria,” *Presence: Teleoperators and Virtual Environments* (12:5), pp. 456–480. (https://doi.org/10.1162/105474603322761270).

Blanchard, E. B. 1982. “Behavioral Medicine: Past, Present, and Future.,” *Journal of Consulting and Clinical Psychology* (50:6), pp. 795–796. (https://doi.org/10.1037/h0077935).

BMAS. 2022. “Fachkräftemonitoring Für Das BMAS. Mittelfristprognose Bis 2026.,” BMAS (Bundesministerium für Arbeit und Soziales).

BMAS. 2023. “Langfristprojektion Des Fachkräftebedarfs in Deutschland, 2021-2040. Szenario ‘Fortschrittliche Arbeitswelt’ (Annahmesetzung Nach Dem Koalitionsvertrag von 2021).,” BMAS (Bundesministerium für Arbeit und Soziales).

Boggiss, A. L., Consedine, N. S., Brenton-Peters, J. M., Hofman, P. L., and Serlachius, A. S. 2020. “A Systematic Review of Gratitude Interventions: Effects on Physical Health and Health Behaviors,” *Journal of Psychosomatic Research* (135), p. 110165. (https://doi.org/10.1016/j.jpsychores.2020.110165).

Bohlmeijer, E. T., Kraiss, J. T., Watkins, P., and Schotanus-Dijkstra, M. 2021. “Promoting Gratitude as a Resource for Sustainable Mental Health: Results of a 3-Armed Randomized Controlled Trial up to 6 Months Follow-Up,” *Journal of Happiness Studies* (22:3), pp. 1011–1032. (https://doi.org/10.1007/s10902-020-00261-5).

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., and Baños, R. 2017. “Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review,” *Current Psychiatry Reports* (19:7), p. 42. (https://doi.org/10.1007/s11920-017-0788-4).

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., and Zilca, R. 2021. “Artificially Intelligent Chatbots in Digital Mental Health Interventions: A Review,” *Expert Review of Medical Devices* (18:sup1), pp. 37–49. (https://doi.org/10.1080/17434440.2021.2013200).

Brandtzaeg, P. B., and Følstad, A. 2018. “Chatbots: Changing User Needs and Motivations,” *Interactions* (25:5), pp. 38–43. (https://doi.org/10.1145/3236669).

Buck, R. 2004. “The Gratitude of Exchange and the Gratitude of Caring,” in *The Psychology of Gratitude*, R. A. Emmons and M. E. McCullough (eds.), Oxford University Press, pp. 100–122. (https://doi.org/10.1093/acprof:oso/9780195150100.003.0006).

Burgoon, J. 1994. “Nonverbal Signals,” in *Handbook of Interpersonal Communication*, pp. 229–285.

Candello, H., Pinhanez, C., and Figueiredo, F. 2017. “Typefaces and the Perception of Humanness in Natural Language Chatbots,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2, pp. 3476–3487. (https://doi.org/10.1145/3025453.3025919).

Colby, K. M., Weber, S., and Hilf, F. D. 1971. “Artificial Paranoia,” *Artificial Intelligence* (2:1), pp. 1–25. (https://doi.org/10.1016/0004-3702(71)90002-6).

Collier, G. J. 2014. *Emotional Expression*, (0 ed.), Psychology Press. (https://doi.org/10.4324/9781315802411).

Cox, S. R., Lee, Y.-C., and Ooi, W. T. 2023. *Comparing How a Chatbot References User Utterances from Previous Chatting Sessions: An Investigation of Users’ Privacy Concerns and Perceptions*, arXiv. (https://doi.org/10.48550/ARXIV.2308.04879).

Craig, K. J. T., Fusco, N., Lindsley, K., Snowdon, J. L., Willis, V. C., Arriaga, Y. E., and Dankwa-Mullan, I. 2020. “Rapid Review: Identification of Digital Health Interventions in Atherosclerotic-Related Cardiovascular Disease Populations to Address Racial, Ethnic, and Socioeconomic Health Disparities,” *Cardiovascular Digital Health Journal* (1:3), pp. 139–148. (https://doi.org/10.1016/j.cvdhj.2020.11.001).

Crocker, K. M., Gnatt, I., Haywood, D., Butterfield, I., Bhat, R., Lalitha, A. R. N., Jenkins, Z. M., and Castle, D. J. 2023. “The Impact of COVID-19 on the Mental Health Workforce: A Rapid Review,” *International Journal of Mental Health Nursing* (32:2), pp. 420–445. (https://doi.org/10.1111/inm.13097).

Cummings, J. J., and Bailenson, J. N. 2016. “How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence,” *Media Psychology* (19:2), pp. 272–309. (https://doi.org/10.1080/15213269.2015.1015740).

Cummins, S. E., Bailey, L., Campbell, S., Koon-Kirby, C., and Zhu, S.-H. 2007. “Tobacco Cessation Quitlines in North America: A Descriptive Study,” *Tobacco Control* (16:Suppl 1), pp. i9–i15. (https://doi.org/10.1136/tc.2007.020370).

Cunha, L. F., Pellanda, L. C., and Reppold, C. T. 2019. “Positive Psychology and Gratitude Interventions: A Randomized Clinical Trial,” *Frontiers in Psychology* (10), p. 584. (https://doi.org/10.3389/fpsyg.2019.00584).

Cutler, D. M., and Lleras-Muney, A. 2010. “Understanding Differences in Health Behaviors by Education,” *Journal of Health Economics* (29:1), pp. 1–28. (https://doi.org/10.1016/j.jhealeco.2009.10.003).

Daft, R. L., and Lengel, R. H. 1986. “Organizational Information Requirements, Media Richness and Structural Design,” *Management Science* (32:5), pp. 554–571. (https://doi.org/10.1287/mnsc.32.5.554).

De Carolis, B., Pelachaud, C., Poggi, I., and Steedman, M. 2004. “APML, a Markup Language for Believable Behavior Generation,” in *Life-Like Characters*, Cognitive Technologies, H. Prendinger and M. Ishizuka (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 65–85. (https://doi.org/10.1007/978-3-662-08373-4_4).

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., 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), pp. 331–349. (https://doi.org/10.1037/xap0000092).

Demeure, V., Niewiadomski, R., and Pelachaud, C. 2011. “How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment?,” *Presence: Teleoperators and Virtual Environments* (20:5), pp. 431–448. (https://doi.org/10.1162/PRES_a_00065).

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., and Biswas-Diener, R. 2010. “New Well-Being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings,” *Social Indicators Research* (97:2), pp. 143–156. (https://doi.org/10.1007/s11205-009-9493-y).

Eaves, M., and Leathers, D. G. 2017. *Successful Nonverbal Communication: Principles and Applications*, (Fifth edition.), London: Routledge. (https://doi.org/10.4324/9781315542317).

Elvir, M., Gonzalez, A. J., Walls, C., and Wilder, B. 2017. “Remembering a Conversation – A Conversational Memory Architecture for Embodied Conversational Agents,” *Journal of Intelligent Systems* (26:1), pp. 1–21. (https://doi.org/10.1515/jisys-2015-0094).

Emmons, R. A., and McCullough, M. E. 2003. “Counting Blessings versus Burdens: An Experimental Investigation of Gratitude and Subjective Well-Being in Daily Life.,” *Journal of Personality and Social Psychology* (84:2), pp. 377–389. (https://doi.org/10.1037/0022-3514.84.2.377).

Emmons, R. A., and Mishra, A. 2011. “Why Gratitude Enhances Well-Being: What We Know, What We Need to Know,” in *Designing Positive Psychology: Taking Stock and Moving Forward*, K. M. Sheldon, T. B. Kashdan, and M. F. Steger (eds.), Oxford University Press, pp. 248–262.

Emmons, R. A., and Stern, R. 2013. “Gratitude as a Psychotherapeutic Intervention,” *Journal of Clinical Psychology* (69:8), pp. 846–855. (https://doi.org/10.1002/jclp.22020).

Enko, J., Behnke, M., Dziekan, M., Kosakowski, M., and Kaczmarek, L. D. 2021. “Gratitude Texting Touches the Heart: Challenge/Threat Cardiovascular Responses to Gratitude Expression Predict Self-Initiation of Gratitude Interventions in Daily Life,” *Journal of Happiness Studies* (22:1), pp. 49–69. (https://doi.org/10.1007/s10902-020-00218-8).

European Commission. 2023. “Large Majority of Europeans Consider Mental Health Promotion as Important as Physical Health Promotion.,” , October 9. (https://europa.eu/eurobarometer/surveys/detail/3032).

Evans, W. D., Wallace, J. L., and Snider, J. 2012. “Pilot Evaluation of the Text4baby Mobile Health Program,” *BMC Public Health* (12:1), p. 1031. (https://doi.org/10.1186/1471-2458-12-1031).

Evenson, K. R., Goto, M. M., and Furberg, R. D. 2015. “Systematic Review of the Validity and Reliability of Consumer-Wearable Activity Trackers,” *The International Journal of Behavioral Nutrition and Physical Activity* (12), p. 159. (https://doi.org/10.1186/s12966-015-0314-1).

Fatehi, F., Samadbeik, M., and Kazemi, A. 2020. “What Is Digital Health? Review of Definitions,” in *Studies in Health Technology and Informatics*, A. Värri, J. Delgado, P. Gallos, M. Hägglund, K. Häyrinen, U.-M. Kinnunen, L. B. Pape-Haugaard, L.-M. Peltonen, K. Saranto, and P. Scott (eds.), IOS Press. (https://doi.org/10.3233/SHTI200696).

Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. “A Taxonomy of Social Cues for Conversational Agents,” *International Journal of Human-Computer Studies* (132), pp. 138–161. (https://doi.org/10.1016/j.ijhcs.2019.07.009).

Filkins, B. L., Kim, J. Y., Roberts, B., Armstrong, W., Miller, M. A., Hultner, M. L., Castillo, A. P., Ducom, J.-C., Topol, E. J., and Steinhubl, S. R. 2016. “Privacy and Security in the Era of Digital Health: What Should Translational Researchers Know and Do about It?,” *American Journal of Translational Research* (8:3), pp. 1560–1580.

Fogg, B. J. 2002. "Persuasive Technology: Using Computers to Change What We Think and Do," *Ubiquity* (2002:December), p. 2. (https://doi.org/10.1145/764008.763957).

Fogg, B. J., and Tseng, H. 1999. "The Elements of Computer Credibility," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, Pittsburgh, Pennsylvania, United States: ACM Press, pp. 80–87. (https://doi.org/10.1145/302979.303001).

Frederick, S., and Loewenstein, G. 1999. "Hedonic Adaptation.," in *Well-Being: The Foundations of Hedonic Psychology.*, New York, NY, US: Russell Sage Foundation, pp. 302–329.

Fritz, M. M., Walsh, L. C., and Lyubomirsky, S. 2017. "Staying Happier," in *The Happy Mind: Cognitive Contributions to Well-Being*, M. D. Robinson and M. Eid (eds.), Cham: Springer International Publishing, pp. 95–114. (https://doi.org/10.1007/978-3-319-58763-9_6).

Fu, C., Yoshikawa, Y., Iio, T., and Ishiguro, H. 2021. "Sharing Experiences to Help a Robot Present Its Mind and Sociability," *International Journal of Social Robotics* (13:2), pp. 341–352. (https://doi.org/10.1007/s12369-020-00643-y).

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., and Sartorius, N. 2015. "Toward a New Definition of Mental Health," *World Psychiatry* (14:2), pp. 231–233. (https://doi.org/10.1002/wps.20231).

Gambino, A., Fox, J., and Ratan, R. 2020. "Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm," *Human-Machine Communication* (1), pp. 71–86. (https://doi.org/10.30658/hmc.1.5).

Gan, D. Z. Q., McGillivray, L., Han, J., Christensen, H., and Torok, M. 2021. "Effect of Engagement With Digital Interventions on Mental Health Outcomes: A Systematic Review and Meta-Analysis," *Frontiers in Digital Health* (3), p. 764079. (https://doi.org/10.3389/fdgth.2021.764079).

Ghazali, A. S., Ham, J., Barakova, E., and Markopoulos, P. 2018. "The Influence of Social Cues in Persuasive Social Robots on Psychological Reactance and Compliance," *Computers in Human Behavior* (87), pp. 58–65. (https://doi.org/10.1016/j.chb.2018.05.016).

Gnewuch, U., Morana, S., Heckmann, C., and Maedche, A. 2018. "Designing Conversational Agents for Energy Feedback," in *Designing for a Digital and Globalized World* (Vol. 10844), Lecture Notes in Computer Science, S. Chatterjee, K. Dutta, and R. P. Sundarraj (eds.), Cham: Springer International Publishing, pp. 18–33. (https://doi.org/10.1007/978-3-319-91800-6_2).

Groot, J., MacLellan, A., Butler, M., Todor, E., Zulfiqar, M., Thackrah, T., Clarke, C., Brosnan, M., and Ainsworth, B. 2023. "The Effectiveness of Fully Automated Digital Interventions in Promoting Mental Well-Being in the General Population: Systematic Review and Meta-Analysis," *JMIR Mental Health* (10), p. e44658. (https://doi.org/10.2196/44658).

Guerra-Carrillo, B., Katovich, K., and Bunge, S. A. 2017. "Does Higher Education Hone Cognitive Functioning and Learning Efficacy? Findings from a Large and Diverse Sample," *PLOS ONE* (12:8), (A. R. Dalby, ed.), p. e0182276. (https://doi.org/10.1371/journal.pone.0182276).

Gunawardena, C. N., and Zittle, F. J. 1997. "Social Presence as a Predictor of Satisfaction within a Computer-mediated Conferencing Environment," *American Journal of Distance Education* (11:3), pp. 8–26. (https://doi.org/10.1080/08923649709526970).

Hassanein, K., and Head, M. 2007. "Manipulating Perceived Social Presence through the Web Interface and Its Impact on Attitude towards Online Shopping," *International Journal of Human-Computer Studies* (65:8), pp. 689–708. (https://doi.org/10.1016/j.ijhcs.2006.11.018).

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), p. 102788. (https://doi.org/10.1016/j.ijhcs.2022.102788).

Hudecek, M. F. C., Blabst, N., Morgan, B., and Lermer, E. 2020. “Measuring Gratitude in Germany: Validation Study of the German Version of the Gratitude Questionnaire-Six Item Form (GQ-6-G) and the Multi-Component Gratitude Measure (MCGM-G),” *Frontiers in Psychology* (11), p. 590108. (https://doi.org/10.3389/fpsyg.2020.590108).

Hussain, S., Ameri Sianaki, O., and Ababneh, N. 2019. “A Survey on Conversational Agents/Chatbots Classification and Design Techniques,” in *Web, Artificial Intelligence and Network Applications* (Vol. 927), Advances in Intelligent Systems and Computing, L. Barolli, M. Takizawa, F. Xhafa, and T. Enokido (eds.), Cham: Springer International Publishing, pp. 946–956. (https://doi.org/10.1007/978-3-030-15035-8_93).

Isasi, F., Naylor, M. D., University of Pennsylvania School of Nursing, Skorton, D., Association of American Medical Colleges, Grabowski, D., Harvard Medical School, Hernandez, S., California Health Care Foundation, Montgomery Rice, V., and Morehouse School of Medicine. 2021. “Patients, Families, and Communities COVID-19 Impact Assessment: Lessons Learned and Compelling Needs,” *NAM Perspectives* (11). (https://doi.org/10.31478/202111c).

Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L., Busch, M. A., Maske, U., Hapke, U., Gaebel, W., Maier, W., Wagner, M., Zielasek, J., and Wittchen, H.-U. 2016. “Erratum zu: Psychische Störungen in der Allgemeinbevölkerung. Studie zur Gesundheit Erwachsener in Deutschland und ihr Zusatzmodul „Psychische Gesundheit“ (DEGS1-MH),” *Der Nervenarzt* (87:1), pp. 88–90. (https://doi.org/10.1007/s00115-015-4458-7).

Jake-Schoffman, D. E., Silfee, V. J., Waring, M. E., Boudreaux, E. D., Sadasivam, R. S., Mullen, S. P., Carey, J. L., Hayes, R. B., Ding, E. Y., Bennett, G. G., and Pagoto, S. L. 2017. “Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps,” *JMIR mHealth and uHealth* (5:12), p. e190. (https://doi.org/10.2196/mhealth.8758).

James, A. 2017. “Prosody and Paralanguage in Speech and the Social Media: The Vocal and Graphic Realisation of Affective Meaning,” *Linguistica* (57), p. 137. (https://doi.org/10.4312/linguistica.57.1.137-149).

Jo, E., Jeong, Y., Park, S., Epstein, D. A., and Kim, Y.-H. 2024. “Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 11, pp. 1–21. (https://doi.org/10.1145/3613904.3642420).

Kaihlanen, A.-M., Virtanen, L., Buchert, U., Safarov, N., Valkonen, P., Hietapakka, L., Hörhammer, I., Kujala, S., Kouvonen, A., and Heponiemi, T. 2022. “Towards Digital Health Equity - a Qualitative Study of the Challenges Experienced by Vulnerable Groups in Using Digital Health Services in the COVID-19 Era,” *BMC Health Services Research* (22:1), p. 188. (https://doi.org/10.1186/s12913-022-07584-4).

Kainiemi, E., Saukkonen, P., Virtanen, L., Vehko, T., Kyytsönen, M., Aaltonen, M., and Heponiemi, T. 2023. “Perceived Benefits of Digital Health and Social Services among Older Adults: A Population-Based Cross-Sectional Survey,” *DIGITAL HEALTH* (9), p. 20552076231173559. (https://doi.org/10.1177/20552076231173559).

Kashdan, T. B., Mishra, A., Breen, W. E., and Froh, J. J. 2009. “Gender Differences in Gratitude: Examining Appraisals, Narratives, the Willingness to Express Emotions, and Changes in Psychological Needs,” *Journal of Personality* (77:3), pp. 691–730. (https://doi.org/10.1111/j.1467-6494.2009.00562.x).

Kim, Y., and Sundar, S. S. 2012. “Anthropomorphism of Computers: Is It Mindful or Mindless?,” *Computers in Human Behavior* (28:1), pp. 241–250. (https://doi.org/10.1016/j.chb.2011.09.006).

Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. 2017. “The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms,” in *Proceedings of the 2017 Conference on Designing Interactive Systems*, Edinburgh United Kingdom: ACM, June 10, pp. 555–565. (https://doi.org/10.1145/3064663.3064672).

Kottke, T. E., Stiefel, M., and Pronk, N. P. 2016. “‘Well-Being in All Policies’: Promoting Cross-Sectoral Collaboration to Improve People’s Lives,” *Preventing Chronic Disease* (13), p. 160155. (https://doi.org/10.5888/pcd13.160155).

Krokstad, S., Weiss, D. A., Krokstad, M. A., Rangul, V., Kvaløy, K., Ingul, J. M., Bjerkeset, O., Twenge, J., and Sund, E. R. 2022. “Divergent Decennial Trends in Mental Health According to Age Reveal Poorer Mental Health for Young People: Repeated Cross-Sectional Population-Based Surveys from the HUNT Study, Norway,” *BMJ Open* (12:5), p. e057654. (https://doi.org/10.1136/bmjopen-2021-057654).

Kucherbaev, P., Bozzon, A., and Houben, G.-J. 2018. “Human-Aided Bots,” *IEEE Internet Computing* (22:6), pp. 36–43. (https://doi.org/10.1109/MIC.2018.252095348).

Lambert, N. M., Graham, S. M., and Fincham, F. D. 2009. “A Prototype Analysis of Gratitude: Varieties of Gratitude Experiences,” *Personality and Social Psychology Bulletin* (35:9), pp. 1193–1207. (https://doi.org/10.1177/0146167209338071).

Layous, K., and Lyubomirsky, S. 2014. “The How, Why, What, When, and Who of Happiness,” in *Positive Emotion*, J. Gruber and J. T. Moskowitz (eds.), Oxford University Press, pp. 472–495. (https://doi.org/10.1093/acprof:oso/9780199926725.003.0025).

Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. 2006. “Are Physically Embodied Social Agents Better than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People’s Loneliness in Human–Robot Interaction,” *International Journal of Human-Computer Studies* (64:10), pp. 962–973. (https://doi.org/10.1016/j.ijhcs.2006.05.002).

Lee, M., Contreras Alejandro, J., and IJsselsteijn, W. 2024. “Cultivating Gratitude with a Chatbot,” *International Journal of Human–Computer Interaction* (40:18), pp. 4957–4972. (https://doi.org/10.1080/10447318.2023.2231277).

Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., and Mohr, D. C. 2023. “Systematic Review and Meta-Analysis of AI-Based Conversational Agents for Promoting Mental Health and Well-Being,” *Npj Digital Medicine* (6:1), p. 236. (https://doi.org/10.1038/s41746-023-00979-5).

Liu, S., Rovine, M. J., and Molenaar, P. C. M. 2012. “Selecting a Linear Mixed Model for Longitudinal Data: Repeated Measures Analysis of Variance, Covariance Pattern Model, and Growth Curve Approaches.,” *Psychological Methods* (17:1), pp. 15–30. (https://doi.org/10.1037/a0026971).

Lombard, M., and Ditton, T. 2006. “At the Heart of It All: The Concept of Presence,” *Journal of Computer-Mediated Communication* (3:2), pp. 0–0. (https://doi.org/10.1111/j.1083-6101.1997.tb00072.x).

Luger, E., and Sellen, A. 2016. “‘Like Having a Really Bad PA’: The Gulf between User Expectation and Experience of Conversational Agents,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 7, pp. 5286–5297. (https://doi.org/10.1145/2858036.2858288).

Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. 2007. “Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text,” *Journal of Artificial Intelligence Research* (30), pp. 457–500. (https://doi.org/10.1613/jair.2349).

Manchia, M., Gathier, A. W., Yapici-Eser, H., Schmidt, M. V., De Quervain, D., Van Amelsvoort, T., Bisson, J. I., Cryan, J. F., Howes, O. D., Pinto, L., Van Der Wee, N. J., Domschke, K., Branchi, I., and Vinkers, C. H. 2022. “The Impact of the Prolonged COVID-19 Pandemic on Stress Resilience and Mental Health: A Critical Review across Waves,” *European Neuropsychopharmacology* (55), pp. 22–83. (https://doi.org/10.1016/j.euroneuro.2021.10.864).

Marietto, M. D. G., Aguiar, R. V., Barbosa, G. D. O., Botelho, W. T., Pimentel, E., Franca, R. D. S., and Da Silva, V. L. 2013. “Artificial Intelligence Markup Language: A Brief Tutorial,” *International Journal of Computer Science & Engineering Survey* (4:3), pp. 1–20. (https://doi.org/10.5121/ijcses.2013.4301).

Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., and Cohen, A. B. 2019. “Digital Health: A Path to Validation,” *Npj Digital Medicine* (2:1), p. 38. (https://doi.org/10.1038/s41746-019-0111-3).

Mauz, E., Walther, L., Junker, S., Kersjes, C., Damerow, S., Eicher, S., Hölling, H., Müters, S., Peitz, D., Schnitzer, S., and Thom, J. 2023. “Time Trends in Mental Health Indicators in Germany’s Adult Population before and during the COVID-19 Pandemic,” *Frontiers in Public Health* (11), p. 1065938. (https://doi.org/10.3389/fpubh.2023.1065938).

McCullough, M. E., Emmons, R. A., and Tsang, J.-A. 2002. “The Grateful Disposition: A Conceptual and Empirical Topography.,” *Journal of Personality and Social Psychology* (82:1), pp. 112–127. (https://doi.org/10.1037/0022-3514.82.1.112).

McCullough, M. E., Tsang, J.-A., and Emmons, R. A. 2004. “Gratitude in Intermediate Affective Terrain: Links of Grateful Moods to Individual Differences and Daily Emotional Experience.,” *Journal of Personality and Social Psychology* (86:2), pp. 295–309. (https://doi.org/10.1037/0022-3514.86.2.295).

McGovern, L. 2014. “The Relative Contribution of Multiple Determinants to Health,” Project HOPE, August 21. (https://doi.org/10.1377/hpb20140821.404487).

Mendes-Santos, C., Andersson, G., Weiderpass, E., and Santana, R. 2020. “Mitigating COVID-19 Impact on the Portuguese Population Mental Health: The Opportunity That Lies in Digital Mental Health,” *Frontiers in Public Health* (8), p. 553345. (https://doi.org/10.3389/fpubh.2020.553345).

Moitra, M., Owens, S., Hailemariam, M., Wilson, K. S., Mensa-Kwao, A., Gonese, G., Kamamia, C. K., White, B., Young, D. M., and Collins, P. Y. 2023. “Global Mental Health: Where We Are and Where We Are Going,” *Current Psychiatry Reports* (25:7), pp. 301–311. (https://doi.org/10.1007/s11920-023-01426-8).

Mori, M., MacDorman, K., and Kageki, N. 2012. “The Uncanny Valley [From the Field],” *IEEE Robotics & Automation Magazine* (19:2), pp. 98–100. (https://doi.org/10.1109/MRA.2012.2192811).

Murray, E., Hekler, E. B., Andersson, G., Collins, L. M., Doherty, A., Hollis, C., Rivera, D. E., West, R., and Wyatt, J. C. 2016. “Evaluating Digital Health Interventions,” *American Journal of Preventive Medicine* (51:5), pp. 843–851. (https://doi.org/10.1016/j.amepre.2016.06.008).

Nass, C., and Moon, Y. 2000. “Machines and Mindlessness: Social Responses to Computers,” *Journal of Social Issues* (56:1), pp. 81–103. (https://doi.org/10.1111/0022-4537.00153).

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers Are Social Actors," in *Conference Companion on Human Factors in Computing Systems - CHI '94*, Boston, Massachusetts, United States: ACM Press, p. 204. (https://doi.org/10.1145/259963.260288).

Nimavat, K., and Champaneria, T. 2017. *Chatbots: An Overview. Types, Architecture, Tools and Future Possibilities*, , October.

Nöth, W. 1995. *Handbook of Semiotics*, (First paperback ed., [4. Dr.].), Advances in Semiotics, Bloomington: Indiana Univ. Press.

Oh, C. S., Bailenson, J. N., and Welch, G. F. 2018. "A Systematic Review of Social Presence: Definition, Antecedents, and Implications," *Frontiers in Robotics and AI* (5), p. 114. (https://doi.org/10.3389/frobt.2018.00114).

Pagoto, S., and Waring, M. E. 2016. "A Call for a Science of Engagement: Comment on Rus and Cameron," *Annals of Behavioral Medicine* (50:5), pp. 690–691. (https://doi.org/10.1007/s12160-016-9839-z).

Palan, S., and Schitter, C. 2018. "Prolific.Ac—A Subject Pool for Online Experiments," *Journal of Behavioral and Experimental Finance* (17), pp. 22–27. (https://doi.org/10.1016/j.jbef.2017.12.004).

Payne, H. E., Lister, C., West, J. H., and Bernhardt, J. M. 2015. "Behavioral Functionality of Mobile Apps in Health Interventions: A Systematic Review of the Literature," *JMIR mHealth and uHealth* (3:1), p. e20. (https://doi.org/10.2196/mhealth.3335).

Pelachaud, C., and Bilvi, M. 2003. "Computational Model of Believable Conversational Agents," in *Communication in Multiagent Systems* (Vol. 2650), Lecture Notes in Computer Science, M.-P. Huget (ed.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 300–317. (https://doi.org/10.1007/978-3-540-44972-0_17).

Pereira, A., Prada, R., and Paiva, A. 2014. "Improving Social Presence in Human-Agent Interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada: ACM, April 26, pp. 1449–1458. (https://doi.org/10.1145/2556288.2557180).

Philippe, T. J., Sikder, N., Jackson, A., Koblanski, M. E., Liow, E., Pilarinos, A., and Vasarhelyi, K. 2022. "Digital Health Interventions for Delivery of Mental Health Care: Systematic and Comprehensive Meta-Review," *JMIR Mental Health* (9:5), p. e35159. (https://doi.org/10.2196/35159).

Portocarrero, F. F., Gonzalez, K., and Ekema-Agbaw, M. 2020. "A Meta-Analytic Review of the Relationship between Dispositional Gratitude and Well-Being," *Personality and Individual Differences* (164), p. 110101. (https://doi.org/10.1016/j.paid.2020.110101).

Ramirez, A., Walther, J. B., Burgoon, J. K., and Sunnafrank, M. 2002. "Information-Seeking Strategies, Uncertainty, and Computer-Mediated Communication.: Toward a Conceptual Model," *Human Communication Research* (28:2), pp. 213–228. (https://doi.org/10.1111/j.1468-2958.2002.tb00804.x).

Ratan, R. A., and Hasler, B. 2009. "Self-Presence Standardized: Introducing the Self-Presence Questionnaire (SPQ)," in *Proceedings of the 12th Annual International Workshop on Presence* (Vol. 81).

Rauschenberg, C., Schick, A., Hirjak, D., Seidler, A., Paetzold, I., Apfelbacher, C., Riedel-Heller, S. G., and Reininghaus, U. 2021. "Evidence Synthesis of Digital Interventions to Mitigate the Negative Impact of the COVID-19 Pandemic on Public Mental Health: Rapid Meta-Review," *Journal of Medical Internet Research* (23:3), p. e23365. (https://doi.org/10.2196/23365).

Reeves, B., and Nass, C. 1996. "The Media Equation: How People Treat Computers, Television, and New Media like Real People and Pla," *Bibliovault OAI Repository, the University of Chicago Press*.

Rheu, M. (Mj), Dai, Y. (Nancy), Meng, J., and Peng, W. 2024. "When a Chatbot Disappoints You: Expectancy Violation in Human-Chatbot Interaction in a Social Support Context," *Communication Research* (51:7), pp. 782–814. (https://doi.org/10.1177/00936502231221669).

Richards, D., and Bransky, K. 2014. "ForgetMeNot: What and How Users Expect Intelligent Virtual Agents to Recall and Forget Personal Conversational Content," *International Journal of Human-Computer Studies* (72:5), pp. 460–476. (https://doi.org/10.1016/j.ijhcs.2014.01.005).

Rodríguez-Martínez, A., Amezcua-Aguilar, T., Cortés-Moreno, J., and Jiménez-Delgado, J. J. 2023. "Qualitative Analysis of Conversational Chatbots to Alleviate Loneliness in Older Adults as a Strategy for Emotional Health," *Healthcare* (12:1), p. 62. (https://doi.org/10.3390/healthcare12010062).

Rosenberg, E. L. 1998. "Levels of Analysis and the Organization of Affect," *Review of General Psychology* (2:3), pp. 247–270. (https://doi.org/10.1037/1089-2680.2.3.247).

Saleem, M., Kühne, L., De Santis, K. K., Christianson, L., Brand, T., and Busse, H. 2021. "Understanding Engagement Strategies in Digital Interventions for Mental Health Promotion: Scoping Review," *JMIR Mental Health* (8:12), p. e30000. (https://doi.org/10.2196/30000).

Sansone, R. A., and Sansone, L. A. 2010. "Gratitude and Well Being: The Benefits of Appreciation," *Psychiatry (Edgmont (Pa.: Township))* (7:11), pp. 18–22.

Sauermann, S., Herzberg, J., Burkert, S., and Habetha, S. 2022. "DiGA – A Chance for the German Healthcare System," *Journal of European CME* (11:1), p. 2014047. (https://doi.org/10.1080/21614083.2021.2014047).

Schroeder, S. A. 2007. "We Can Do Better--Improving the Health of the American People," *The New England Journal of Medicine* (357:12), pp. 1221–1228. (https://doi.org/10.1056/NEJMsa073350).

Seeger, A.-M., Pfeiffer, J., and Heinzl, A. 2018. *Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework*.

Seligman, M. E. P., Steen, T. A., Park, N., and Peterson, C. 2005. "Positive Psychology Progress: Empirical Validation of Interventions.," *American Psychologist* (60:5), pp. 410–421. (https://doi.org/10.1037/0003-066X.60.5.410).

Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*, London ; New York: Wiley.

Shum, H.-Y., He, X., and Li, D. 2018. *From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots*, arXiv. (https://doi.org/10.48550/arXiv.1801.01957).

Siddals, S., Torous, J., and Coxon, A. 2024. "'It Happened to Be the Perfect Thing': Experiences of Generative AI Chatbots for Mental Health," *Npj Mental Health Research* (3:1), p. 48. (https://doi.org/10.1038/s44184-024-00097-4).

Sin, N. L., and Lyubomirsky, S. 2009. "Enhancing Well-being and Alleviating Depressive Symptoms with Positive Psychology Interventions: A Practice-friendly Meta-analysis," *Journal of Clinical Psychology* (65:5), pp. 467–487. (https://doi.org/10.1002/jclp.20593).

Slater, M., and Wilbur, S. 1997. "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments," *Presence: Teleoperators and Virtual Environments* (6:6), pp. 603–616. (https://doi.org/10.1162/pres.1997.6.6.603).

Stafford, L., and Daly, J. A. 1984. “Conversational Memory: The Effects of Recall Mode and Memory Expectancies on Remembrances of Natural Conversations,” *Human Communication Research* (10:3), pp. 379–402. (https://doi.org/10.1111/j.1468-2958.1984.tb00024.x).

Steuer, J. 1992. “Defining Virtual Reality: Dimensions Determining Telepresence,” *Journal of Communication* (42:4), pp. 73–93. (https://doi.org/10.1111/j.1460-2466.1992.tb00812.x).

Titscher, S., Meyer, M., Wodak, R., and Vetter, E. 2000. *Methods of Text and Discourse Analysis*, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. (https://doi.org/10.4135/9780857024480).

Tomczyk, S., and Ewert, C. 2025. “Positive Changes in Daily Life? A Meta-analysis of Positive Psychological Ecological Momentary Interventions,” *Applied Psychology: Health and Well-Being* (17:1), p. e70006. (https://doi.org/10.1111/aphw.70006).

Torous, J., Jän Myrick, K., Rauseo-Ricupero, N., and Firth, J. 2020. “Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow,” *JMIR Mental Health* (7:3), p. e18848. (https://doi.org/10.2196/18848).

Trager, G. L. (George L. 1958. “Paralanguage : A First Approximation,” in *Paralanguage : A First Approximation*, Studies in Linguistics ; v.13., Buffalo, N.Y: University of Buffalo, Department of Anthropology and Linguistics.

Tsai, W.-H. S., Liu, Y., and Chuan, C.-H. 2021. “How Chatbots’ Social Presence Communication Enhances Consumer Engagement: The Mediating Role of Parasocial Interaction and Dialogue,” *Journal of Research in Interactive Marketing* (15:3), pp. 460–482. (https://doi.org/10.1108/JRIM-12-2019-0200).

Turing, A. M. 1950. “I.—COMPUTING MACHINERY AND INTELLIGENCE,” *Mind* (LIX:236), pp. 433–460. (https://doi.org/10.1093/mind/LIX.236.433).

Vaculíková, J. 2024. “Comparing Self-Regulation Strategies among Adult Learners from Poland, Serbia, Slovakia, and the Czech Republic,” *Frontiers in Psychology* (15), p. 1382989. (https://doi.org/10.3389/fpsyg.2024.1382989).

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., and Torous, J. B. 2019. “Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape,” *The Canadian Journal of Psychiatry* (64:7), pp. 456–464. (https://doi.org/10.1177/0706743719828977).

Van Der Goot, M. J. 2022. “Source Orientation, Anthropomorphism, and Social Presence in Human-Chatbot Communication: How to Proceed with These Concepts,” *Publizistik* (67:4), pp. 555–578. (https://doi.org/10.1007/s11616-022-00760-w).

Verhagen, T., Van Nes, J., Feldberg, F., and Van Dolen, W. 2014. “Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters,” *Journal of Computer-Mediated Communication* (19:3), pp. 529–545. (https://doi.org/10.1111/jcc4.12066).

Von Der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. 2010. “‘It Doesn’t Matter What You Are!’ Explaining Social Effects of Agents and Avatars,” *Computers in Human Behavior* (26:6), pp. 1641–1650. (https://doi.org/10.1016/j.chb.2010.06.012).

Wallace, R. S. 2009. “The Anatomy of A.L.I.C.E.,” in *Parsing the Turing Test*, R. Epstein, G. Roberts, and G. Beber (eds.), Dordrecht: Springer Netherlands, pp. 181–210. (https://doi.org/10.1007/978-1-4020-6710-5_13).

Wallis, P., and Norling, E. 2005. *The Trouble with Chatbots: Social Skills in a Social World*.

Walther, J. B. 1992. “Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective,” *Communication Research* (19:1), pp. 52–90. (https://doi.org/10.1177/009365092019001003).

Walther, J. B. 1996. “Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction,” *Communication Research* (23:1), pp. 3–43. (https://doi.org/10.1177/009365096023001001).

Walther, J. B. 2006. “Nonverbal Dynamics in Computer-Mediated Communication or: (And the Net: ('S with You, :) And You :) Alone,” in *The SAGE Handbook of Nonverbal Communication*, 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., pp. 461–480. (https://doi.org/10.4135/9781412976152.n24).

Watkins, P. C. 2014. *Gratitude and the Good Life: Toward a Psychology of Appreciation*, Dordrecht: Springer Netherlands. (https://doi.org/10.1007/978-94-007-7253-3).

Weizenbaum, J. 1966. “ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine,” *Communications of the ACM* (9:1), pp. 36–45. (https://doi.org/10.1145/365153.365168).

Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., and Stark, L. W. 1996. “The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence,” *Presence: Teleoperators and Virtual Environments* (5:3), pp. 263–273. (https://doi.org/10.1162/pres.1996.5.3.263).

West, B. T., Welch, K. B., Gałecki, A. T., and Gillespie, B. W. 2022. *Linear Mixed Models: A Practical Guide Using Statistical Software*, (3rd ed.), Boca Raton: Chapman and Hall/CRC. (https://doi.org/10.1201/9781003181064).

WHO. 2020. “Digital Health.” (https://www.who.int/health-topics/digital-health#tab=tab_3, accessed February 14, 2025).

Wicks, P., Mack Thorley, E., Simacek, K., Curran, C., and Emmas, C. 2018. “Scaling PatientsLikeMe via a ‘Generalized Platform’ for Members with Chronic Illness: Web-Based Survey Study of Benefits Arising,” *Journal of Medical Internet Research* (20:5), p. e175. (https://doi.org/10.2196/jmir.9909).

Winkler, R., and Soellner, M. 2018. “Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis,” *Academy of Management Proceedings* (2018:1), p. 15903. (https://doi.org/10.5465/AMBPP.2018.15903abstract).

Wong, Y. J., Owen, J., Gabana, N. T., Brown, J. W., McInnis, S., Toth, P., and Gilman, L. 2018. “Does Gratitude Writing Improve the Mental Health of Psychotherapy Clients? Evidence from a Randomized Controlled Trial,” *Psychotherapy Research* (28:2), pp. 192–202. (https://doi.org/10.1080/10503307.2016.1169332).

Wood, A. M., Froh, J. J., and Geraghty, A. W. A. 2010. “Gratitude and Well-Being: A Review and Theoretical Integration,” *Clinical Psychology Review* (30:7), pp. 890–905. (https://doi.org/10.1016/j.cpr.2010.03.005).

World Health Organization. 2004. “Promoting Mental Health: Concepts, Emerging Evidence, Practice (Summary Report),” Geneva: World Health Organization, World Health Organization.

World Health Organization. 2022. “World Mental Health Report: Transforming Mental Health for All,” World Health Organization.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., and Tang, Y. 2023. “A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development,” *IEEE/CAA Journal of Automatica Sinica* (10:5), pp. 1122–1136. (https://doi.org/10.1109/JAS.2023.123618).

Xu, A., Liu, Z., Guo, Y., Sinha, V., and Akkiraju, R. 2017. “A New Chatbot for Customer Service on Social Media,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2, pp. 3506–3510. (https://doi.org/10.1145/3025453.3025496).

Xu, K., Chen, X., and Huang, L. 2022. “Deep Mind in Social Responses to Technologies: A New Approach to Explaining the Computers Are Social Actors Phenomena,” *Computers in Human Behavior* (134), p. 107321. (https://doi.org/10.1016/j.chb.2022.107321).

Xue, J., Zhang, B., Zhao, Y., Zhang, Q., Zheng, C., Jiang, J., Li, H., Liu, N., Li, Z., Fu, W., Peng, Y., Logan, J., Zhang, J., and Xiang, X. 2023. “Evaluation of the Current State of Chatbots for Digital Health: Scoping Review,” *Journal of Medical Internet Research* (25), p. e47217. (https://doi.org/10.2196/47217).

Zhong, W., Luo, J., and Zhang, H. 2024. “The Therapeutic Effectiveness of Artificial Intelligence-Based Chatbots in Alleviation of Depressive and Anxiety Symptoms in Short-Course Treatments: A Systematic Review and Meta-Analysis,” *Journal of Affective Disorders* (356), pp. 459–469. (https://doi.org/10.1016/j.jad.2024.04.057).

Zumstein, D., and Hundertmark, S. 2017. “Chatbots – an Interactive Technology for Personalized Communication, Transactions and Services,” *IADIS International Journal on WWW/Internet* (15), pp. 96–109.

Appendix A: Declaration on the use of GenAI tools

In the preparation of this paper, I have used following tools based on generative artificial intelligence (GenAI):

1. ChatGPT
2. DeepL Write
3. DeepL Translate

I further declare that

- I have labeled the content taken from the GenAI tools listed above with my details in the table below,
- I have verified that the content generated by the above-mentioned GenAI tools and adapted by me is factually correct,
- I am aware that, as the author of this work, I am responsible for the information and the statements made in it, and
- I am aware that violating the disclosure of the use of generative AI in my work is a deception and leads to an evaluation with an insufficient grade.

I have used the above-mentioned AI systems as indicated below.

Areas of contribution	**AI tool(s) used**	**Description of the manner of use and compliance with good scientific practice (if applicable, please indicate the section of the thesis)**
Development and conception of the research project	None	
Identification of literature	1	I used the ChatGPT “Deep Research” for some literature suggestions that I inspected afterwards
Synthesizing of literature		
Structuring the text		
Formulation of text	2	I used a translator to express phrases in English that I was familiar with in my mother tongue. I also had ChatGPT give me suggestions for phrasing sentences and text modules
Revision of text	2	To improve the readability of the language, I used DeeplWrite and ChatGPT.
Creation of visualizations		
Further contributions		

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Masterarbeit selbständig angefertigt, nicht anderweitig für Prüfungszwecke vorgelegt, alle benutzten Quellen und Hilfsmittel angegeben, sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Neu-Ulm 17.04.25

Ort, Datum

Unterschrift