REPUBLIC OF TURKIYE
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF BUSINESS

META-REQUIREMENTS FOR ENHANCING USER EXPERIENCE IN CONVERSATIONAL AGENTS: A DESIGN SCIENCE RESEARCH APPROACH

PHD THESIS

Neşe BAZ AKTAŞ
ORCID: 0000-0001-7840-349X

Department: Management Information Systems

Thesis Supervisor: Assoc. Prof. Adem AKBIYIK
ORCID: 0000-0001-7634-4545

JANUARY – 2026

This thesis entitled " Meta-Requırements for Enhancing User Experience in Conversational Agents: A Design Science Research Approach", prepared Neşe BAZ AKTAŞ, was found successful as a result of the Thesis Defense Examination held on 12/12/2026 in accordance with the relevant articles of Sakarya University Graduate Education and Training Regulation and was accepted as a Doctoral Thesis by our jury.

Thesis Supervisor: Assoc. Prof. Adem AKBIYIK

Sakarya University

Jury Members: Assoc. Prof. Naciye Güliz UĞUR

Sakarya University

Dr. Kayhan AYAR

Sakarya University

Prof. Erman COŞKUN

Bahçeşehir University

Prof. Yeliz EKİNCİ

İstanbul Bilgi University

SAKARYA ÜNİVERSİTESİ	**T.C.** **SAKARYA ÜNİVERSİTESİ** **İŞLETME ENSTİTÜSÜ** **TEZ SAVUNULABİLİRLİK VE ORİJİNALLİK BEYAN FORMU**	Sayfa: 1/1

Öğrencinin

Adı Soyadı	:	Neşe Baz Aktaş
Öğrenci Numarası	:	D209054001
Enstitü Anabilim Dalı	:	Yönetim Bilişim Sistemleri
Enstitü Bilim Dalı	:	
Programı	:	
Tezin Başlığı	:	Meta-Requırements for Enhancing User Experience in Conversational Agents: A Design Science Research Approach
Benzerlik Oranı	:	%8

Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen tez çalışmasının benzerlik oranının herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

...... / ...... / 20….
İmza
Öğrenci

Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen öğrenciye ait tez çalışması ile ilgili gerekli düzenleme tarafımca yapılmış olup, yeniden değerlendirlilmek üzere gsbtez@sakarya.edu.tr adresine yüklenmiştir.

Bilgilerinize arz ederim.

...... / ...... / 20….
İmza
Danışman

Uygundur

Danışman
Unvanı / Adı-Soyadı: Doç. Dr. Adem Akbıyık

Tarih: 13 / 11 / 2025

İmza:

☐ Kabul Edilmiştir ☐ Reddedilmiştir EYK Tarih ve No: …… / …… / 20…. …… / ……	**Enstitü Birim Sorumlusu Onayı**

 00.ENS.FR.72

PREFACE

My warmest thanks go to my advisor, Assoc. Prof. Adem Akbıyık, who motivated me many times along the way and guided me with wisdom and care.

I would also like to extend my sincere thanks to my thesis committee, Assoc. Prof. Naciye Güliz Uğur and Dr. Kayhan AYAR, and the defense jury, Prof. Dr. Erman Coşkun and Prof. Dr. Yeliz Ekinci for their valuable feedback and support.

I am especially grateful to Assoc. Prof. Simone BORSCI and Asst. Prof. Gayane SEDRAKYAN, whom I met during this journey. Their guidance and encouragement were invaluable at key moments.

I owe thanks to many colleagues and friends who were part of this process—those who listened patiently to every detail, challenged my thinking, and helped me move my thesis forward. Tuğba KOÇ, Merve ÖZ, Cansu KORKMAZ, Aysun BOZANTA HAKYEMEZ, Meltem MUTLUTÜRK, Yue LI, Eda Nur ÇAPKAN, Sofia MORANDINI, Oğuzhan ARI, Ömer Sezai AYKAÇ, Huseyin HUSEYİNZADE.

I would like to express my deepest gratitude to my family, my mother Seval BAZ, my father Selehattin BAZ, and my sisters Esin BAZ and Bengü BAZ. They have always believed in me, often more than I believed in myself. Their constant support, trust, and encouragement gave me the strength to continue, and I am deeply proud to have achieved this milestone on their behalf.

My deepest gratitude goes to the love of my life and my greatest supporter, my husband, Dr. Bahadır AKTAŞ. None of this would have been possible without you.

And my lovely daughter, Tomris AKTAŞ, you gave me the strength and belief to reach this achievement. Even your gentle movements gave me encouragement and patience throughout this journey. I am truly thankful to have you in my life.

Neşe BAZ AKTAŞ

12.12.2025

ACKNOWLEDGEMENT

This doctoral research was supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) within the scope of the 2214-A International Doctoral Research Fellowship Programme, Application No. 1059B142401168. The financial support provided by TÜBİTAK enabled the author to conduct part of this doctoral research abroad and to benefit from an international academic research environment.

The author gratefully acknowledges TÜBİTAK for its support, which made it possible to carry out advanced research activities, access academic resources, and engage in academic collaboration during the doctoral studies. This support significantly contributed to the academic development of the researcher and to the successful completion of this PhD thesis.

Neşe BAZ AKTAŞ

12.12.2025

TABLE OF CONTENT

ABBREVIATIONS

AIDUA	: Artificial Intelligence Device Acceptance
AOI	: Area of Interest
AT	: Attribution Theory
AVE	: Average Variance Extracted
BCFA	: Bayesian Confirmatory Factor Analysis
BEFA	: Bayesian Exploratory Factor Analysis
BUS-11	: Chatbot Usability Scale
CAs	: Conversational Agents
CASA	: Computers are Social Actors
CAT	: Cognitive Appraisal Theory
CLT	: Cognitive Load Theory
DRT	: Dissimilarity Reputation Theory
DSR	: Design Science Research
ECM	: Expectation Confirmation Model
ECT	: Expectation Confirmation Theory
ECV	: Explained Common Variance
EPTD	: Expected Percentage of True Differences
EVT	: Expectancy Violations Theory
HCI	: Human-Computer Interaction
IAs	: Intelligent Agents
IBT	: Information Behavior Theory
IS	: Information Systems
ISDT	: Information Systems Design Theory
ISO	: International Organization for Standardization
LLMs	: Large Language Models
MIS	: Management Information Systems
ML	: Machine Learning
NLG	: Natural Language Generation
NLP	: Natural Language Processing
NLU	: Natural Language Understanding
PA	: Parallel Analysis
PRAs	: Product Recommender Agents
PRMSE	: Proportional Reduction of Mean Squared Error

PT	: Proxemic Theory
SAT	: Social Attraction Theory
SDLC	: Systems Development Life Cycle
SDT	: Self-Determination Theory
SET	: Social Exchange Theory
SFT	: Social Facilitation Theory
SIT	: Social Identity Theory
SLR	: Systematic Literature Review
SPT	: Social Presence Theory
SRT	: Social Response Theory
ST	: Schema Theory
TAM	: Technology Acceptance Model
TTFF	: Time to First Fixation
TRA	: Theory of Reasoned Action
UMUX	: Usability Metric for User Experience
URT	: Uncertainty Reduction Theory
UTAUT	: Unified Theory of Acceptance and Use of Technology

TABLES

FIGURES

FIGURES CONTINUED

ABSTRACT

Baz Aktas, N. (2025). *Meta-requirements for enhancing user experience in conversational agents: a design science research approach* (Unpublished doctoral thesis). Sakarya University.

Conversational Agents (CAs), such as chatbots, have evolved from rule-based systems that followed predefined scripts into AI-driven agents capable of understanding, generating, and adapting natural language interactions. These advancements have enabled more natural and context-aware communication between humans and agents. However, achieving such experience depends on how these agents are designed. Therefore, identifying which design features and principles most strongly influence user experience has become a central research challenge.

To address this challenge, this dissertation adopts the Design Science Research (DSR) approach, which integrates theory-driven reasoning with iterative design and empirical evaluation. It is particularly effective for developing and validating artifacts that bridge theory and practice in three cycles: relevance, design, and rigor. Firstly, a systematic literature review (SLR) of 97 empirical studies from Human–Computer Interaction (HCI) and Information Systems (IS) was conducted to identify how users perceive and evaluate CAs with varying characteristics. This synthesis produced a structured framework of design dimensions, including anthropomorphic design, agent characteristics, and agent competency, that describe how design decisions shape user experience. Second, insights from the SLR informed the formulation of meta-requirements, representing high-level design needs that connect user expectations with CA design practices. To complement these literature-based insights, a laboratory study with 87 participants was conducted to explore user requirements not identified in the literature. Concurrently, the study aimed to identify a valid and integrative approach to assessing user experience with conversational agents. Therefore, a new overall user experience scale was developed to capture the combined effect of core user perception dimensions, usability, social presence, trust, usefulness, and enjoyment. The Bayesian Exploratory Analysis showed that these factors are not distinct dimensions. They collectively represent a single underlying construct of overall user experience, supporting the validity of a unified measurement approach. Lastly, the meta-requirements were instantiated by analyzing real service chatbots, mapping their design dimensions and elements against the established framework. A series of empirical evaluations followed, including an eye-tracking experiment and an online study with 363 observations, to validate the unidimensional scale across contexts. The results confirmed the robustness of the unidimensional scale. They provided convergent evidence that varying design configurations partially satisfied the meta-requirements, demonstrating consistency between the theoretical framework and user evaluations.

This dissertation contributes theoretically by integrating CA design research into a unified, empirically validated framework that bridges the perspectives of HCI and IS. Practically, it delivers (1) a validated user experience measurement instrument, and (2) a set of meta-requirements and design guidelines that inform the creation of socially engaging and functionally competent conversational agents. These outcomes advance methodological rigor and provide evidence-based guidance for next-generation CAs.

Keywords: Conversational Agents, Design, User Experience, Meta-Requirements, Design Science Research

ÖZET

Baz Aktas, N. (2025). *Konuşma temelli sistemlerde kullanıcı deneyimini geliştirmeye yönelik meta-gereklilikler: bir tasarım bilimi araştırma yaklaşımı* (Yayımlanmamış doktora tezi). Sakarya Üniversitesi.

Diyalog bazlı sistemler veya konuşma aracıları (KA), örneğin sohbet botları, önceden tanımlanmış komut dizilerini izleyen kural tabanlı sistemlerden, doğal dil etkileşimlerini anlama, üretme ve uyarlama yeteneğine sahip yapay zekâ destekli aracılara evrilmiştir. Bu gelişmeler, insanlar ve aracılar arasında daha doğal ve bağlama duyarlı iletişimi mümkün kılmıştır. Ancak, böyle bir deneyimin elde edilmesi bu aracılarının nasıl tasarlandığına bağlıdır. Bu nedenle, hangi tasarım özellikleri ve ilkelerinin kullanıcı deneyimini en güçlü şekilde etkilediğini belirlemek temel bir araştırma konusu haline gelmiştir.

Bu zorluğu ele almak amacıyla, bu tez teoriye dayalı akıl yürütmeyi yinelemeli tasarım ve ampirik değerlendirme süreçleriyle bütünleştiren Tasarım Bilimi Araştırması yaklaşımını benimsemektedir. Bu yaklaşım, teori ile pratiği üç döngüde uygunluk, tasarım ve titizlik üzerinden bir araya getiren yapıtların geliştirilmesi ve doğrulanmasında özellikle etkilidir. Bu tez ilk olarak, İnsan-Bilgisayar Etkileşimi (İBE) ve Bilgi Sistemleri (BS) alanlarından 97 ampirik çalışmanın sistematik bir literatür taraması (SLT) gerçekleştirilmiş, kullanıcıların farklı özelliklere sahip KA'ları nasıl algıladıkları ve değerlendirdikleri belirlenmiştir. Bu sentez, kullanıcı deneyimini şekillendiren tasarım kararlarını tanımlayan antropomorfik tasarım, aracı özellikleri ve aracı yetkinliği dâhil olmak üzere yapılandırılmış bir tasarım boyutları çerçevesi üretmiştir. İkinci olarak, SLT'den elde edilen içgörüler, kullanıcı beklentilerini KA tasarım uygulamalarıyla ilişkilendiren yüksek düzeyli tasarım gereksinimlerini temsil eden meta-gereksinimlerin oluşturulmasını bilgilendirmiştir. Bu literatür temelli içgörüleri tamamlamak için, literatürde tanımlanmamış kullanıcı gereksinimlerini keşfetmek amacıyla 87 katılımcı ile bir laboratuvar çalışması yürütülmüştür. Aynı zamanda, çalışma, konuşma araçlarıyla kullanıcı deneyimini değerlendirmek için geçerli ve daha kapsayıcı bir yaklaşım belirlemeyi amaçlamıştır. Bu nedenle, temel kullanıcı algısı boyutları, kullanılabilirlik, kullanışlılık, sosyal varlık, güven, fayda ve keyif faktörlerinin birleşik etkisini yakalamak için yeni bir genel kullanıcı deneyimi ölçeği geliştirilmiştir. Bayesyen yaklaşımı ile Keşifsel Analiz gerçekleştirilmiştir. Bu yaklaşım KA etkileşiminde bu faktörlerin ayrı boyutlar olmadığını göstermiştir. Bunlar birlikte genel kullanıcı deneyiminin tek bir temel yapısını temsil etmekte, birleşik bir ölçüm yaklaşımının geçerliliğini desteklemektedir. Son aşamada, belirlenen meta-gereksinimler ilgili tasarım boyutlarıyla ilişkilendirilmiş ve her bir boyuta yönelik uygulama unsurları keşfedilmiştir. Bu doğrultuda, gerçek hizmet sohbet botları üzerinde yapılan analizlerle meta-gereksinimlerin bu tasarım boyutlarına nasıl yansıdığı incelenmiştir. Bir dizi ampirik çalışma aracılığıyla hem bu yansımaların kullanıcı deneyimi üzerindeki etkisi hem de geliştirilen konuşma aracı kullanıcı deneyimi ölçeğinin geçerliliği ve güvenilirliği değerlendirilmiştir. Elde edilen bulgular, ölçeğin tek boyutlu yapısının sağlamlığını doğrulamış ve farklı tasarım yapılandırmalarının kullanıcı deneyimi üzerindeki etkilerini ortaya koymuştur. Böylece, meta-gereksinimlerin yansımaları üzerinden kuramsal çerçeve ile kullanıcı değerlendirmeleri arasındaki tutarlılık gösterilmiştir.

Bu tez, İBE ve BS bakış açılarını birleştirerek KA tasarım araştırmasını bütünleşik ve ampirik olarak doğrulanmış bir çerçeveye entegre etmek suretiyle kuramsal bir katkı sunmaktadır. Uygulamada ise (1) doğrulanmış bir kullanıcı deneyimi ölçüm aracı ve (2) sosyal açıdan etkileşimli ve işlevsel olarak yetkin konuşma araçlarının oluşturulmasına rehberlik eden meta-gereksinimler ve tasarım yönergeleri seti sağlamaktadır. Bu çıktılar yöntemsel titizliği ilerletmekte ve yeni nesil KA'lar için kanıta dayalı rehberlik sunmaktadır.

Anahtar Kelimeler: Diyalog Bazlı Sistem, Sohbet Robotu, Tasarım Bilimi Araştırması, Kullanıcı Deneyimi, Meta Gereksinimler

INTRODUCTION

The first section of this dissertation presents the motivation for this study endeavor (section I.1). Then, the research position and gaps, as well as the research questions, are presented (I.2). Following the research importance (I.3), method (I.4), and the scope and limitations are provided (I.5).

Motivation

"People will forget what you said, but they will never forget how you made them feel."
(Angelou, n.d)

Conversational agents have become a prominent interface for human–computer interaction, transforming how people access services, information, and support. Success stories can be seen in customer service, education, healthcare, and other domains, where CAs provide immediate responses, reduce workload, and expand accessibility (Gnewuch et al., 2017; Jo et al., 2023). Using natural language in CAs has made digital systems more accessible and effortless (Gnewuch et al., 2017). The introduction of ChatGPT in 2022 has changed how people use these technologies in their daily lives (Skjuve et al., 2024). With other companies introducing new tools such as DeepSeek or Groq, CAs have become increasingly integrated into daily routines (Jin et al., 2025). As a result, users now interact with various agents and develop preferences. Although many successful implementations exist, users face new challenges when engaging with these systems (Følstad & Brandtzaeg, 2020; Jo et al., 2023; Luria et al., 2019; Złotowski et al., 2015). For instance, research by Joe et al. (2023) shows that chatbots provide immediate responses and support user goals. However, maintaining meaningful conversations remains difficult due to limited long-term memory and a lack of personalization. In addition, users report that repetitive dialogues, misinterpretations, and irrelevant responses make interaction challenging (Følstad & Brandtzaeg, 2020). Furthermore, task-oriented or keyword-based chatbots provide limited functionality (Haugeland et al., 2022).

In addition to enhancing the agent's capabilities, various design strategies have been employed to mitigate or eliminate negative interactions. Scholars applied Social Response Theory (SRT) to design conversational agents that appear more humanlike and compensate for their limited capabilities (Brendel et al., 2020). Similarly, the Anthropomorphism Theory has been applied to reduce uncertainty (Seeger et al., 2017) and to support long-term usage (Magyar et al., 2019). The Computers Are Social Actors (CASA) perspective has been widely adopted to trigger social engagement with agents through the incorporation of social cues in design (Chattaraman et al., 2012; Kim & Sundar, 2012; Pelau et al., 2021). Qiu and Benbasat (2008) employed Social Presence Theory to design a recommender agent by integrating different voices and texts. Finally, studies on the Uncanny Valley effect have primarily examined situations where users perceive CA design negatively and how specific design choices shape these perceptions (Brendel et al., 2020; Chung et al., 2023).

Building on these theoretical perspectives, it becomes important to understand user preferences by analyzing their experiences with CAs. Today, users decide whether to adopt technology and then evaluate their experience with the technology to continue using it in the post-acceptance stage (Bhattacherjee, 2001; Davis et al., 1989; Hsiao & Chen, 2021). Identifying factors that shape this stage is essential to uncover meaningful design considerations for the CAs (Nguyen, 2023). For example, Zhang et al. (2023) reported that perceived anthropomorphism, social influence, and performance expectancy are antecedents of continued chatbot use in tourism (Li et al., 2021). In the finance domain, the chatbot's satisfaction and perceived competence were key drivers of continued use (Nguyen et al., 2021a). In healthcare, perceived usefulness, satisfaction, and expectation confirmation shaped users' decisions to maintain chatbot use (Li et al., 2022). Other studies have highlighted additional determinants of continuance, including social presence, human likeness, enjoyment, reliability, understandability, trust, and effort expectancy (Balakrishnan et al., 2022; Prakash et al., 2023; Zhou et al., 2022).

When the design of CAs is considered carefully, it can foster a satisfactory and trustworthy experience, which leads to greater continuance intention and sustained use, as long as the interaction corresponds to users' expectations. Hence, this raises a problem for CA designers and developers on how they can make design choices in CA that consistently support a positive user experience.

This thesis contributes by investigating CA design and its relationship to user experience. The goal is to generate knowledge for both the academic community and practitioners by adapting the Design Science Research (DSR) approach, which enables the deployment of a creative solution to a problem by creating an artifact and evaluating its practicality in real-world settings. Examining this design problem intersects with three different research streams (Figure 1). Conversational agent design represents the domain-specific focus, addressing how agents are structured, what design elements are implemented, and how these elements contribute to user experience. HCI contributes the theories, terminology, and interaction aspects needed to study user experience, emphasizing how people perceive and engage with technology. Information Systems design and development offer methodological tools, particularly DSR, which support the formulation of meta-requirements and design knowledge for CA design, as well as methods to frame adoption, post-acceptance evaluation, and continuance of technology use.

Research Gaps and Questions

This dissertation addresses the design choices and their effect on user experience. To do so, it explores CA design while considering user evaluations. The following section outlines the research questions and the gaps this study seeks to address.

Over the past two decades, the design of conversational agents has developed into a multidisciplinary research area spanning HCI, IS, and domain-specific applications (e.g., health, tourism, etc). Within HCI, research has examined how design features influence user experience dimensions such as trust, enjoyment, and social presence. From an IS perspective, scholars have examined adoption and continuance by applying established models. Domain-specific studies further illustrate the diversity of CA applications. In healthcare, agents have supported patient education and chronic illness management (Kocaballi et al., 2020). They have acted as tutors or learning companions in education, offering guidance and reflection support (Fryer et al., 2017). In customer service, chatbots handle repetitive queries to improve efficiency (Følstad & Brandtzaeg, 2020), while in tourism, agents serve as interactive recommenders that enhance trip planning and customer experience (Tussyadiah & Miller, 2019).

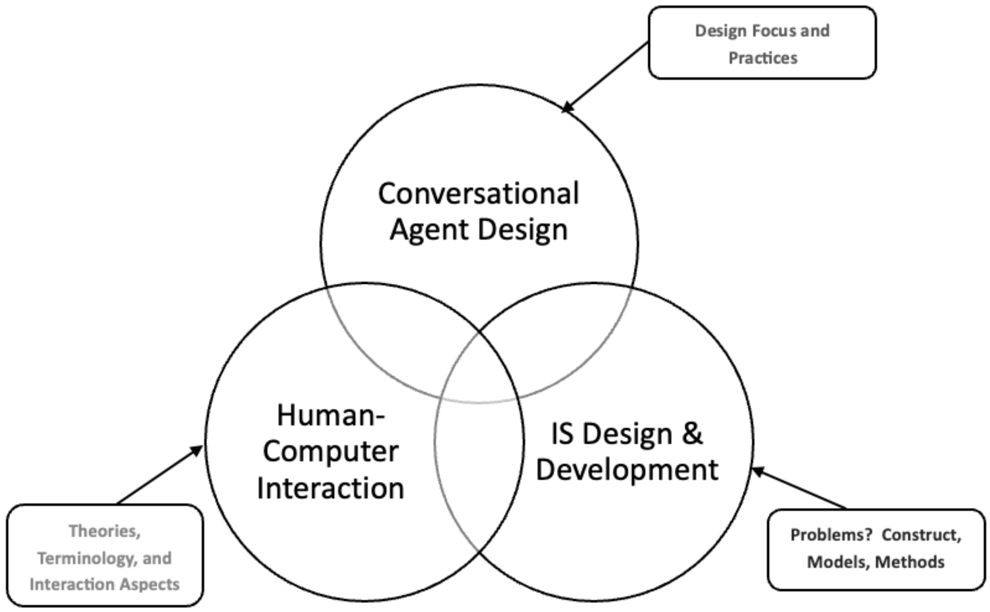
Much of the design incorporates human-like characteristics such as appearance, personality, or social behaviors. Early work by Nass and colleagues showed that even when users are aware they are interacting with a machine, they tend to respond socially, a phenomenon explained through anthropomorphism and formalized in the CASA theory (Nass et al., 1994; Nass & Moon, 2000). Research building on CASA demonstrates that anthropomorphic cues, such as polite behaviors, gestures, or human-like names, can enhance perceptions of reliability and trust (Chattaraman et al., 2012; Pelau et al., 2021). In e-commerce, anthropomorphic design in product recommender agents has been shown to strengthen social presence and credibility (Qiu & Benbasat, 2008), while studies on service robots indicate that anthropomorphic features increase interaction time and engagement (Belanche et al., 2020). In healthcare, older adults are more willing to adopt virtual assistants with anthropomorphic design, supporting long-term use (Magyar et al., 2019). Furthermore, anthropomorphism has been linked to trust by reducing uncertainty in interactions with unfamiliar systems (Seeger et al., 2017; Troshani et al., 2021). At the same time, scholars caution that human likeness must be applied carefully. The Uncanny Valley theory highlights that designs that are too human-like may evoke discomfort (Li & Suh, 2021; Mori, 2012). Other research emphasizes that CAs also possess non-human attributes such as re-embodiment or multi-presence, which should be acknowledged in their design (Luria et al., 2019). More recently, Chandra et al. (2022) argue that CAs should be designed with cognitive, relational, and emotional competencies, moving beyond superficial cues toward deeper integration of human-like capabilities.

Second, several reviews have attempted to address this fragmentation, yet they, too, reveal limitations. For instance, some focus on narrow domains, such as tutorial agents Huang et al., 2022; Zawacki-Richter et al., 2019) or text-based chatbots (Rapp et al., 2021). Like Feine et al. (2019) offer detailed taxonomies of social cues but restrict their scope to interpersonal communication behaviors. Broader mappings, such as Elshan et al. (2022) connect design elements with acceptance constructs but remain thematic rather than implementation-oriented, while Rhue et al. (2021) concentrate solely on trust-building features.

This diversity of perspectives emphasizes the difficulty of forming cumulative knowledge about how design choices shape user experience. To address this, the first research question investigates the key user considerations that influence interactions with CAs across different studies and application domains. The second research question examines the design practices that have been most prominent in shaping user experience.

Figure 1

Dissertation Research Streams



Source: Created by the author.

RQ1: What are the key user considerations when interacting with conversational agents (CAs)?

RQ2: What design practices are used to improve user experience, and which elements are most prominent in CA design?

The research questions are first directed at examining the current landscape of CA design research and systematizing existing knowledge on user experience (RQ1). Second, they explore successful design practices and choices reported across studies (RQ2). Together, these investigations provide an understanding of the overarching design problem. Building on these insights, the final step is synthesizing findings from prior research questions with established knowledge from the CA design knowledge base to formulate design-oriented meta-requirements. This leads to the third research question:

RQ3: What meta-requirements in CA design can be identified to enhance user experience?

Importance of Research

This thesis will contribute to advancing the design of CAs by identifying the meta-requirements in CA design and design elements that influence user experience, particularly during the post-acceptance phase. Drawing from findings in HCI and IS, the study synthesizes empirical insights into requirements that guide the development of CAs that are usable, trustworthy, and satisfying to interact with. In doing so, it moves beyond focusing on technical performance to address how users engage with conversational systems over time.

The growing adoption of AI-powered conversational tools in domains such as education, healthcare, and customer service makes their usability and effectiveness increasingly critical (Hwang & Chang, 2023; Kuhail, et al., 2023b). This thesis contributes by organizing empirical findings into meta-requirements that provide structured guidance for CA design. Rather than prescribing rigid design rules, the results highlight principles that can inform intentional and strategic design decisions, ensuring that CAs are technically proficient, socially acceptable, and emotionally engaging. The empirical studies conducted in this research emphasize the importance of user-based evaluation for understanding CA performance. By grounding the findings in behavioral and subjective measures of post-acceptance use, the study delivers concrete implications for improving interaction quality and sustaining long-term engagement.

Finally, developing a validated theoretical model strengthens academic and practical contributions. For researchers, it provides a foundation to build cumulative knowledge on CA design; for practitioners, it offers actionable requirements that can be adapted across different domains, user groups, and CA types. As a result, this contributes to the ongoing effort to ensure that users continue to use and benefit from these technologies in the long term.

Research Methodology

This PhD thesis presents design-oriented requirements for developing a CA to enhance user experience. To achieve this, the study adopts a Design Science Research approach to identify the problem, explore the solution, and design an artifact that guides designers and developers in making strategic design decisions. The DSR approach enables the deployment of a new solution to a problem by creating an artifact and evaluating its practicality in real-world settings. Moreover, this approach ensures that theory and practice inform the design process (Hevner et al., 2004a). In this thesis, the artifact is a set of design requirements that help build conversational agents that function well, meet user expectations, and support sustainable interaction. Through this process, the thesis demonstrates how DSR can bridge the gap between design theory and practice in human-agent interaction.

The methodology follows a structured sequence of steps consistent with the three-cycle view of DSR (Hevner, 2007). DSR provides a process in which the relevance, design, and rigor cycles interact continuously rather than progressing in isolation (Hevner & Chatterjee, 2010; Wieringa, 2014). Each research activity in this thesis contributes to more than one cycle, ensuring both practical relevance and theoretical grounding.

The Systematic Literature Review (SLR) illustrates this interconnection. As a structured investigation of existing CA studies, the SLR served the relevance cycle by identifying fragmented practices and highlighting the need for consolidated design knowledge. At the same time, it contributed to the design cycle by treating literature as a dataset from which design dimensions and elements were derived and linking these with reported user evaluations. Similarly, the scale development process spanned across cycles. The need for an appropriate measurement instrument was identified in the relevance cycle. The scale was designed and adjusted during the design cycle to capture user experience with CAs. While it was empirically validated through user studies within the design cycle, its application also fulfilled the rigor cycle by testing and grounding the findings in established theoretical models. The empirical studies further illustrate the cyclical interplay. They supported the relevance cycle by uncovering user considerations and expectations, informed the design cycle by guiding the formulation of design-oriented requirements, and contributed to the rigor cycle through empirical validation of the theoretical model of user experience (grounded in the Expectation Confirmation Model (Bhattacherjee, 2001))

By structuring the research this way, the thesis ensures that problem definition, artifact development, and theoretical validation are not separate stages but interdependent activities. This iterative process enables the formulation of meta-requirements that are theoretically robust, empirically validated, and practically helpful in guiding CA design.

Research Limitations

In academic exploration, this thesis acknowledges its inherent limitations and carefully delineates its specific scope, like every scholarly pursuit. This thesis is situated within HCI and IS, focusing specifically on design issues that aim to enhance user experience. The scope deliberately excludes the technical development of conversational agents, as the research centers on human-centered design considerations. HCI and IS were chosen as the foundational domains because they place the user at the core of their research.

A further limitation of this study concerns the type of system used. Due to resource constraints and the complexity involved in developing advanced conversational systems, the research was conducted using a chatbot that was feasible to implement and modify. While the findings provide valuable insights that can inform the design of a broad range of conversational agents, some design aspects may not be generalizable to more complex or embodied systems.

CHAPTER 1. BACKGROUND ON CONVERSATIONAL AGENTS

This chapter lays the theoretical groundwork for understanding the design and evaluation of conversational agents. It first defines what CAs are and introduces their functional architecture. It then categorizes different types of CAs based on their purpose and capabilities. By distinguishing between these types, the chapter aims to clarify the varied roles CAs can play in user interaction. The chapter concludes with a comparative overview and reflection on design considerations common across agent types.

1.1. Conversational Agents: Definitions and Architecture

Conversational Agents, commonly referred to as chatbots or bots, are systems designed to emulate human conversation through natural language interfaces. They interact with users through text or voice interfaces, making interactions with digital systems feel more intuitive and humanlike (Følstad & Brandtzaeg, 2020; Shevat, 2017). These agents rely on technologies like natural language processing (NLP), natural language generation (NLG), and machine learning (ML), enabling them to understand, interpret, and respond to user inquiries in meaningful ways (Dale, 2016; Sarikaya, 2017). Examples of CAs include Amazon's digital assistant (Alexa, Echo) for shopping, social bots (Facebook Messenger, Kuki) for interactive conversations, Microsoft's remote assistant for service delivery, and Apple's Siri for checking the weather and playing music at home. Also, chatbots in customer service, popular social robots in travel and hospitality (Spencer, Connie). Additionally, chatbots are widely used in customer service domains, while social robots such as Spencer and Connie have been deployed in travel and hospitality contexts (Shevat, 2017). More recently, systems powered by large language models, such as ChatGPT, have expanded the capabilities of general-purpose conversational agents by supporting open-domain dialogue across tasks, including information retrieval, creative writing, and problem-solving (Rajamaran, 2023). These developments signal a shift toward replacing or augmenting frontline roles with AI-driven systems capable of sustained interaction and adaptive support (Chaves et al., 2022; Kuhail, et al., 2023b).

CAs are involved in service encounters in tourism (Chaves et al., 2022), voice response systems in customer service (Taylor et al., 2020), personal assistants, and brand representatives (Silva & Bonetti, 2021) in different industries instead of front-line employees (Balakrishnan & Dwivedi, 2021; Belanche et al., 2020; Kim et al., 2021; Liu & Hung, 2021; Mariani et al., 2023). They are gaining popularity, and the global conversational agent market is expected to reach US$206.6 billion by 2034 (Market.us, 2025). With the use of NLP and NLG techniques and ML tools, many businesses will have more powerful conversational agents (Dale, 2016; Sarikaya, 2017). According to IBM Corporation, CAs powered by AI can help to provide personalized customer service and cut labor costs by as much as 30% in customer service (Papas, 2018). CAs can deliver nonstop service, which is one of the most significant advantages of increasing customer satisfaction and loyalty by always being available (Himanshu, 2021; Papas, 2018).

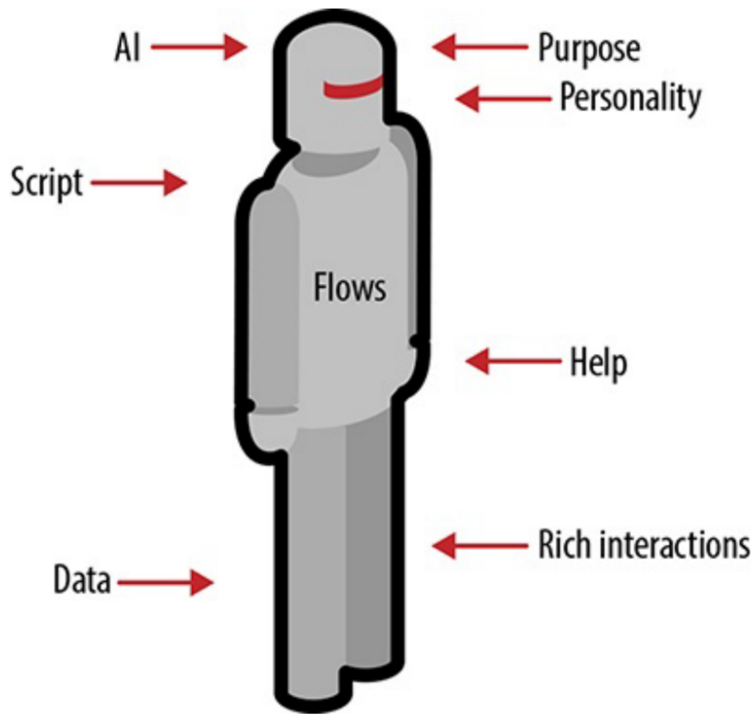
Building CA includes a combination of architectural and experiential layers that define how the agent delivers interaction. According to Shevat (2017), which is illustrated in Figure 2, a CA is "a user experience layer" that exposes a product or service using conversational interaction. With the rise of conversational interfaces, traditional graphical interfaces are moving toward more dialog-based experiences that prioritize flexibility and accessibility. CAs have multiple interconnected components to deliver coherent and effective user experience.

A well-structured CA comprises multiple interdependent components that support adaptive, user-centered communication. The first foundational element in conversational agent design is its purpose, which delineates its core functionality. Clearly articulating what the agent is designed to do is essential. Designers often incorporate features such as buttons or quick replies to maintain user interactions within the agent's functional boundaries (Liu & Martens, 2024). Some also include self-introduction mechanisms to clarify the agent's role and guide user expectations (Ben Mimoun et al., 2017). Ultimately, the effectiveness of a CA depends on the clarity and relevance of its services. A poorly defined purpose or limited perceived usefulness often leads to decreased user engagement and eventual abandonment.

The second key component in conversational agent design is personality, which influences the agent's tone and communication style, and plays a role in brand representation (Shevat, 2017). Designers create personality through visual elements such as logos, avatars (animal or human), or images of real people. Naming is also a critical strategy, as it helps make the agent more recognizable and memorable. In addition, the tone of voice used by the agent shapes user perception. For example, using a humorous or overly casual tone in sensitive contexts, such as purchasing health insurance, can reduce trust and disengage users (Mathies et al., 2016). Maintaining a consistent personality across different platforms is essential to ensure coherence and user trust (Shevat, 2017).

Figure 2

CAs Architectural Layers



Source: Shevat, (2017)

The third essential component in conversational agent design is the integration of artificial intelligence, which enhances the agent's capacity to manage complex interactions and contributes to its ability to learn and adapt over time. AI supports key functionalities such as natural language understanding (NLU), dialogue management, intent recognition, prediction, and sentiment analysis (Shevat, 2017). In particular, using NLU enables more fluid and coherent interactions, especially in complex or domain-specific contexts, allowing for smoother text-based communication. Traditionally, many conversational agents were rule-based and did not incorporate AI, primarily due to the high development costs and the need for extensive training data. However, recent advancements in large language models and the availability of accessible APIs have enabled designers to create more natural and flexible conversational experiences (Lin et al., 2023). Despite these benefits, the use of AI introduces new challenges, most notably, the issue of AI hallucinations, which can reduce the agent's reliability and negatively impact user satisfaction (Alkaissi & McFarlane, 2023; Cheng & Jiang, 2020).

The fourth component is scripting, which defines the conversational structure and flow of interactions between the agent and the user. Scripting organizes the dialog logic and determines how the agent responds to various user inputs across different pathways (Shevat, 2017). It includes ideal interaction paths, often called “happy paths”, and fallback or error-handling sequences, ensuring the conversation remains coherent even when user input deviates from expectations. Rather than just writing replies, scripting involves designing clear, structured, and reusable conversation patterns. It ensures that the agent communicates in a way that feels natural while guiding users toward successful task completion (Følstad & Brandtzæg, 2020). Well-crafted scripts contribute to a smooth and coherent experience. In contrast, poorly structured scripts often confuse users and cause frustration or early abandonment. Shortly, scripting is related to

- What the bot says first.
- How it guides the user.
- What happens when things go wrong?
- How each response connects to the next step (Shevat, 2017).

In addition, help mechanisms are embedded to provide contextual assistance and error recovery, enhancing usability and reliability.

The fifth component in conversational agent design is the help mechanism, which ensures users receive guidance when confused, make errors, or lose track of the conversation. According to Shevat (2017), a well-designed help feature contributes significantly to user retention and task success by providing clear options when the user is “stuck.” This can take the form of onboarding flows, FAQ menus, embedded help commands, or persistent buttons that offer access to further support. Effective help mechanisms should be proactive as well as reactive. Proactive support may include welcome messages to explain the bot’s capabilities or suggest example queries, while reactive help responds to unclear input or signs of user frustration. Integrating help options into the core dialogue structure ultimately reinforces usability and enhances the overall user experience, especially in complex or unfamiliar interaction scenarios.

The six components in CA design include rich interaction capabilities. Importantly, it refers to using enhanced interface elements, such as visual elements, quick replies, or voice features, to extend the agent's communication beyond plain text, supporting more intuitive engagement. These features are often employed to make interactions feel more natural and humanlike. In many studies, designers incorporated visual or multimodal cues (e.g., emoticons, avatars, or cards) to enhance user experience (Riquel et al., 2021b). Rich interactions serve to reduce cognitive load by visually organizing information, allowing users to quickly scan and act without relying on text. However, their use should be carefully aligned with the agent's purpose and user context. Overuse or poor integration can create inconsistency or an unexpected user experience (Luger & Sellen, 2016; Rapp et al., 2024). When applied thoughtfully, rich interaction elements support conversational continuity, reinforce intended actions, and contribute to a more efficient and satisfying user experience (Shevat, 2017).

The final component in conversational agent design is data, which plays a foundational role in enabling personalization, learning, and continuous improvement. As emphasized by Shevat (2017), data collected through user interactions, such as preferences, usage patterns, and conversational history, can be leveraged to improve responses, adapt content, and improve overall system performance. Access to structured and meaningful data allows the agent to deliver more relevant, context-aware support over time. However, while data handling and learning mechanisms are central to the long-term effectiveness of CAs, this component lies outside the scope of the present thesis, which focuses on design-oriented and user-facing dimensions. Still, it is important to acknowledge that without robust data strategies, even well-designed conversational agents may fail to evolve in ways that sustain user engagement and utility. Together, these components form the basis of an effective CA design that is purposeful, coherent, and capable of supporting diverse user needs.

1.2. Categorizing Conversational Agents

Conversational agents can be grouped according to the roles they are meant to play. This makes it easier to understand what each type of agent is good at and where it is most useful. For instance, some agents are designed to answer generic inquiries to accomplish domain-specific tasks. On the other hand, some are intended to handle general-purpose inquiries that address frequent questions across many domains, such as education, health, or customer service, as a tutor, sales representative, or caregiver (Kuhail et al., 2023b; Kuhail et al., 2024; Silva & Bonetti, 2021; Taylor et al., 2020). The complexity and flexibility of these agents often vary based on whether they follow scripted rule-based patterns or incorporate AI-driven capabilities (Stryker, 2025). In line with this distinction, Shevat (2017) emphasizes that understanding the role and scope of a bot early in the design process is essential for setting the right expectations, shaping the bot's personality, and planning conversation flows effectively.

CAs can also be classified according to how they communicate. For instance, text-based agents (web-based chat windows or SMS) use text, voice-based agents (smart speakers) employ audio features, and multimodal agents that combine text, voice, and visual components (Luria et al., 2019). This diversity in purpose and form also requires thoughtful design choices. Consequently, various agent types have been encountered in recent literature, each tailored to different use cases and user expectations. The medium through which a bot communicates has a significant impact on how interactions are scripted, what kinds of responses are feasible, and how users perceive the agent's effectiveness (Shevat, 2017).

1.3. Types of Conversational Agents

1.3.1. Chatbots

The most common form of CAs is chatbots. They are designed to respond to structured or semi-structured user inquiries. Chatbots are widely used in customer service (e.g., onboarding or FAQ scenarios), where the scope of interaction is well-defined (Dooley, 2025), and often designed as a rule-based or one-size-fits-all design (Diederich et al., 2022). One example is the KLM Royal Dutch Airlines chatbot, which assists customers with booking a ticket via platforms like Facebook Messenger, providing dialogue flows in a structured way (KLM, 2025).

Nowadays, ChatGPT (OpenAI), Gemini (Google DeepMind), and Grok (xAI) are a new generation of chatbots, and they can talk about any topic, including travel planning, content creation, language learning, etc (Dilmegani, 2025). These systems are leveraged by large language models (LLMs), allowing them to generate accurate, contextually relevant, and adaptive responses. These chatbots are used explicitly while generating code, searching information, and planning and scheduling (Rajamaran, 2023; Rudolph et al., 2023; Skjuve et al., 2024).

Building on this, these systems require careful design and deployment strategies. Scholars have suggested offering a personalized user experience, eliminating CA hallucinations, and maintaining alignment with ethical norms to enhance user experience (Alkaissi & McFarlane, 2023; Dwivedi et al., 2023; Park et al., 2021). It is noteworthy that when incorporating AI or rule-based chatbots, their usage contexts and strengths demand tailored approaches to usability, trust, and engagement (Borsci et al., 2022a; Seeger & Heinzl, 2021; Skjuve et al., 2024).

1.3.2. Conversational Recommender Agents

Recommender agents (RA) (or decision aids) are designed to generate personalized suggestions considering user preferences or behavior patterns (Xiao & Benbasat, 2007). These systems are frequently applied in e-commerce, healthcare, finance, and education with conversational capabilities (Dascalu et al., 2015; Katzman et al., 2018; Sepideh Ebrahimi & Benbasat, 2022). Users get help through the decision-making process with multi-turn dialogue (Hernandez-Bocanegra & Ziegler, 2023). When effectively designed, these agents contribute to improved decision quality and reduce users' cognitive effort by providing tailored feedback (Xiao & Benbasat, 2007).

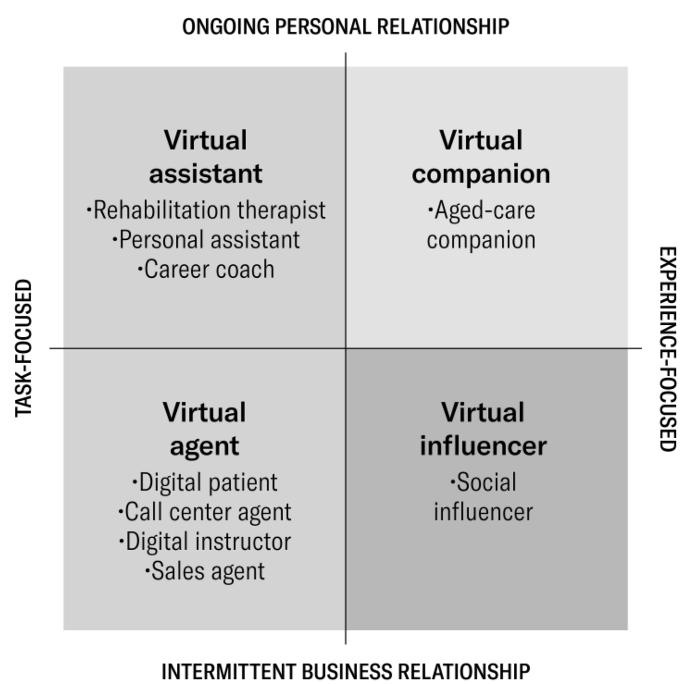
Recent research on recommender agents adopted anthropomorphic or social design elements to improve trust and user satisfaction (Herse et al., 2023; Qiu & Benbasat, 2008). Similarly, Al-Natour et al. (2022) emphasized the importance of features like "perceived caring" and "informativeness" in RA design. They found that these features are important in users' competence and integrity perceptions. Since these agents are developed to be more conversational and user-oriented, their design must integrate the algorithmic rationale behind recommendations and communication strategies that ensure users can understand, trust, and act on the suggestions (Al-Natour et al., 2022; Hernandez-Bocanegra & Ziegler, 2023).

1.3.3. Virtual Humans

Virtual humans represent a more complex form of conversational agent, combining conversational abilities and interactive/embodied representation of a human (e.g., image, avatar). Specifically, while building a virtual human, the software builds a lifelike image of a person (Lucas et al., 2014; Seymour et al., 2018). Virtual humans, sometimes called digital humans, have been classified into four types based on the interaction type, such as task-focused or experience-focused (Figure 3). They provide more interactive communication than chatbots, intelligent agents, and virtual assistants (HBR, 2023).

Figure 3

Virtual Human Categories



Source: HBR, (2023)

Virtual humans are often visualized as animated avatars or 3D-rendered characters that engage with users through speech, facial expressions, gestures, and gaze, which results in emotional engagement (Spadoni et al., 2023). Their main goal is to enhance social presence and relational engagement, especially in domains that require emotional sensitivity or sustained interpersonal connection (Kim et al., 2016; Morina et al., 2014). A notable example is SIMSensei, a virtual human developed to conduct mental health interviews. SIMSensei uses real-time speech analysis and facial behavior tracking to adjust its responses and maintain empathetic engagement, simulating behaviors such as nodding, eye contact, and mirroring user expressions (DeVault et al., 2014). This type of agent has been used in healthcare, training, education, and counseling, where relational credibility and humanlike interaction are essential (Lahav et al., 2020; Morina et al., 2014).

The design of virtual humans involves challenges in multimodal coordination, ensuring that verbal responses are aligned with facial expressions, body language, and gaze direction (non-verbal responses). Another critical design consideration is the degree of anthropomorphism. Highly realistic agents may invoke user expectations of humanlike understanding, which can lead to underestimating the system's capabilities (the so-called “uncanny valley” effect). To avoid giving users false expectations, designers often balance realism and simplicity, using expressive but stylized human features instead of aiming for complete realism (HBR, 2023; Li, 2024; Seymour et al., 2018).

1.3.4. Virtual Agents

Virtual agents have been described differently in the literature. For instance, Huang et al. (2023) described virtual agents as an umbrella term encompassing various CAs, such as chatbots and intelligent agents. On the other hand, Harvard Business Review (HBR, 2023) defines them as task-focused virtual humans designed to assist users in completing specific, often one-time tasks. These agents leverage AI capabilities to enable context-aware, multi-turn dialogue, setting them apart from rule-based chatbots.

These agents may appear as embodied digital humans or be purely text- or voice-based. Common examples include virtual doctors (Robb et al., 2014), nurses (Bickmore et al., 2009), or a pedagogical agent (Goldberg & Cannon-Bowers, 2015). Their functionality is limited to assisting with immediate tasks, such as answering queries, providing instructions, or simulating training scenarios (HBR, 2023; Lucas et al., 2014).

In design, emphasis is placed on accuracy and task completion. However, the role of virtual agents brings consideration to emotional and social aspects (Muniady et al., 2020; Wang et al., 2020). In addition, scholars have benefited from social cues to enhance engagement (Lawson-Guidigbe et al., 2023; Zehnder et al., 2021).

1.3.5. Intelligent Agents

Intelligent agents (IAs) are advanced AI-powered systems designed. IAs perceive their environment, respond in real time, and interact with humans or other agents. They can exhibit a high degree of autonomy and adaptability, while communicating naturally and performing tasks across various domains. These agents are particularly valuable where they support users through human-like interaction. They shift human-computer interaction by introducing more responsive, context-aware, and personalized experiences (Elshan et al., 2022; Rudowsky, 2004).

Their applications extend decision support in finance and logistics to hands-free assistance in clinical or industrial settings (Cui et al., 2012; Laumer et al., 2019). In particular, their ability enables them to interact in natural language and adapt their responses based on prior interactions or learned behavior (Maedche et al., 2019). This capability enables a more humanlike and flexible communication style, which is increasingly critical in their design.

1.3.6. Digital Assistant

Digital assistants are devices we use daily, such as smartphones, smart speakers, and home appliances, for navigation, calendars, reminders, controlling IoT devices, and asking questions (SAP, 2024). Siri from Apple, Google Assistant, Amazon' alexa, and IBM Watson Assistant are the most popular examples, and they are increasingly integrated into consumer and business applications. These systems utilize natural language processing with task automation and continuous learning based on user behavior, as well as integration with smart devices and apps.

Digital assistants are primarily task-oriented; however, their ability to build long-term user profiles supports more proactive and personalized engagement. Compared to intelligent agents, digital assistants are often more domain-general and optimized for rather than specialized decision-making. Their design must include multimodal input (text, voice), continuity across devices, and smooth integration with third-party platforms, making them a central interface for collaborative human-AI interaction in both personal and professional contexts (Brill et al., 2019; Maedche et al., 2019; Marikyan et al., 2022).

Table 1

Conversational Agent Types and Their Features

Type	Embodied	Support Task Completion	Support General or Detailed Conversation	Personalized	Long-term Relation	Example Use Case
Chatbot Rule-based	Avatar/Image	✓	X	X	X	Booking hotel room via website chat
AI-based	Avatar/Image	✓	✓	✓	✓	Assisting user understanding complex topics
Conversational Recommender Agent	Avatar/Image/Virtual Interactive	✓	X	✓	Possibly	Product recommendation in online shopping
Virtual human	Virtual Interactive	✓	✓	✓	✓	Conducting Interviews
Virtual Agent	Avatar/Image/Virtual Interactive	✓	✓	✓	X	Education patients using virtual nurses
Intelligent Agent	Avatar/Image/Virtual Interactive	✓	✓	✓	✓	Supporting diagnosis of disease
Digital Assitant	Not necessarily embodied (multimodal)	✓	✓	✓	✓	Control doors, lights or

Source: Created by the author.

1.4. Comparative Summary of Agent Types

Literature has revealed that CAs vary in their technical structure, design priorities, and interaction goals. As shown in Table 1, all aim to support communication and task completion; their ability to build personalized interaction, maintain long-term engagement, or operate across multiple modalities differs based on their type and usage purpose. Designing these systems, therefore, requires careful attention to the specific role an agent is expected to play. For this reason, scholars benefit from theories to improve CAs' design, as well as experience with them.

CHAPTER 2. THEORETICAL BACKGROUND ON CONVERSATIONAL AGENT DESIGN

The design of conversational agents has increasingly drawn on theories from psychology, communication, and human-computer interaction to improve user experience. Scholars have applied a wide range of theoretical approaches to understand how users perceive, evaluate, and respond to different agent behaviors and appearances. These theories serve not only to explain user reactions but also to inform design strategies that make agents more engaging, trustworthy, and effective. This section establishes the theoretical background necessary to understand CAs' design and its implications for user experience. Section 2.1 introduces the theoretical foundations that inform CA design. Section 2.2 links CAs to the fields of HCI and IS, highlighting how each discipline approaches design and evaluation. Section 2.3 presents technology adoption and continuance models (e.g., TAM, UTAUT, ECM) to understand what factors drive long-term use. Section 2.4 extends this discussion by identifying the key predictors of continuance intention that have emerged as critical in CA interaction research. Finally, Section 2.5 synthesizes these insights into a theoretical evaluation model that serves as the conceptual foundation for the empirical studies later in this thesis.

2.1. Theoretical Foundations of Conversational Agent Design

The user experience with conversational agents has frequently been examined using HCI theories such as the Anthropomorphism Theory (Seeger & Heinzl, 2017), CASA (Pelau et al., 2021), and Social Response Theory (SRT) (Diederich et al., 2020a). Among these theories, the main concern has been creating interactions with CAs that resemble human interaction. Anthropomorphism, SRT, and CASA all focus on agents' human likeness and its effects on interaction to enhance positive user experience. At the same time, the Uncanny Valley Theory seeks to find a balance between humanlike and computerlike design. Together, these perspectives provide essential insights for CA design. This section will address the theoretical foundations of Anthropomorphism Theory, Social Response Theory, CASA, Social Presence Theory, Uncanny Valley Theory, and other related theories.

2.1.1. Anthropomorphism Theory

To begin, anthropomorphism theory provides one of the most fundamental lenses for explaining why users perceive CAs as social entities. Former studies focused on the human likeness of the CAs because of the phenomenon of anthropomorphism (Nass et al., 1996). Psychological studies state that two motivational forces cause this phenomenon. These are needs (1) to establish social relationships and (2) to understand and control the environment (Epley et al., 2007). In essence, the anthropomorphism theory results from an inductive reasoning process (Epley et al., 2008). Importantly, anthropomorphism occurs when characteristics of the object cause the user to interpret it through a human lens (Waytz et al., 2010). Besides, it is stated that concepts of social presence and familiarity cause nonhumans to anthropomorphize (Gefen & Straub,2004). Previous studies found that conversational agents' interaction through natural language leads to assigning anthropomorphic aspects to them (Diederich et al., 2020a; Seeger et al., 2018).

Moreover, research shows that anthropomorphism is positively related to trust, as it helps reduce uncertainty about newly interacted partners' behavior (Seeger et al., 2017; Troshani et al., 2021). The anthropomorphic design of product recommendation agents (PRA) in e-commerce platforms (especially anthropomorphized voices) positively affects social presence. Also, it helps High Tech Product Recommender Agents (PRAs) to lead to trust and credibility (Qiu & Benbasat, 2008). A study on robot design, customer characteristics, and service encounter characteristics in service robots expresses that anthropomorphism enables human-robot interaction and encourages customers to spend more time with robots (Belanche et al., 2020). Another study in health argues that older people will accept virtual assistants with anthropomorphic designs and can use applications for a long time (Magyar et al., 2019). Furthermore, anthropomorphism is the most crucial antecedent in digital assistants generating positive attitudes and purchase intention (Balakrishnan & Dwivedi, 2021).

Another study describes consumers' willingness to accept AI devices. There is a three-stage acceptance process for consumer service interaction when deciding whether to accept AI devices during service interactions. Positive emotions determine consumers' willingness to accept the use of AI devices. The perceived anthropomorphism of an AI device significantly increases customers' perceived effort expectancy of using it (Gursoy et al., 2019). Although the human-human trust perspective suggests that anthropomorphic design is helpful for the trustworthiness of agents, the human-machine trust perspective points to minimizing anthropomorphic design to make agents more trustworthy (Seeger & Heinzl, 2021). In contrast, Pelau et al. (2021) revealed that the human characteristics of service robots are not enough to be reliable. It must also act like a human and behave according to social norms.

2.1.2. Computers are Social Actors (CASA)

Building on anthropomorphism, the CASA paradigm deepens the explanation by showing that people interact with computers as if they were social beings, even with minimal cues (Nass et al., 1994). CASA argues that the interaction between humans and computers is social, and people see computers as entities separately from their designers (Nass & Moon, 2000). Therefore, scholars have benefited from the CASA paradigm in understanding how people perceive computers, AI, and CAs in the context of HCI.

Regarding CAs, research that adopted the CASA perspective provides evidence that anthropomorphic design enhances perceptions of the reliability of computer agents through the use of social cues (names, appearance) or human behavior (politeness, gestures) (Chattaraman et al., 2012; Pelau et al., 2021). Scholars revealed that people trust systems more because systems are more rational and objective than people (Kim & Sundar, 2012). At the same time, CASA highlights that even minimal social cues can trigger social responses from users, such as politeness or turn-taking, regardless of whether the user consciously perceives the agent as human (Bührke et al., 2021). This automatic and unconscious social attribution explains why even simple design choices, such as using a human name or applying a polite tone, can significantly shape the user's experience and attitude toward the system (Diederich et al., 2021). In this way, CASA continues to provide a foundational lens for explaining how and why users respond socially to conversational agents.

2.1.3. Social Response Theory (SRT)

Closely related to CASA, Social Response Theory emphasizes that users unconsciously apply human social norms to their interactions with CAs (Nass & Moon, 2000; Reevess & Nass, 1996). The theory is grounded in the observation that human social behaviors and norms are used in human-to-human interaction when interacting with technology. CA researchers have studied this theory since they know that CA imitates human behaviors to achieve human roles. Therefore, applying social norms in human-human interaction to human-agent interaction is natural. Similar to CASA, researchers adopted the SRT to make CA designs more human-like. For example, Morana et al. (2020) experimented with making the design more human-like using images of humans and robots to investigate how users perceive the changes. Diederich et al. (2020) benefit from SRT when CA has limited capabilities by making their design more humanlike. They used apologizing verbs and social cues during the failures to align the agent with social norms, improving service satisfaction. Huang & Lee (2022) conducted their CA design around the social interactivity cues contributing to the continuance intention behavior.

2.1.4. Social Presence Theory (SPT)

While CASA and SRT explain why people respond socially, Social Presence Theory focuses on how users feel another social entity is present during the interaction (Lee, 2004). SPT originates from communication studies and describes the degree to which the feeling of another's presence is experienced through a communication medium (Biocca, 1997; Lee, 2004). "The other" may refer to a human being, a computer, or an AI-based agent (Lombard & Ditton, 1997). In the context of CAs, SPT highlights how the perception of social presence is shaped by design elements such as verbal cues and identity, revealing that these factors significantly influence user trust and engagement (Hess et al., 2009; Janson, 2023).

Scholars have employed various design strategies to enhance the perception of social presence. For example, Qiu and Benbasat (2008) examined how humanoid embodiment and output modality (voice or text) influence the social presence of a product recommendation agent. Their findings indicate that using human voices induced a stronger perception of social presence, positively affecting trust, perceived usefulness, and enjoyment. Similarly, Schlesener et al. (2025) found that the age and gender of virtual humans can influence users' perceived presence, demonstrating that identity characteristics modulate social perception in agent design.

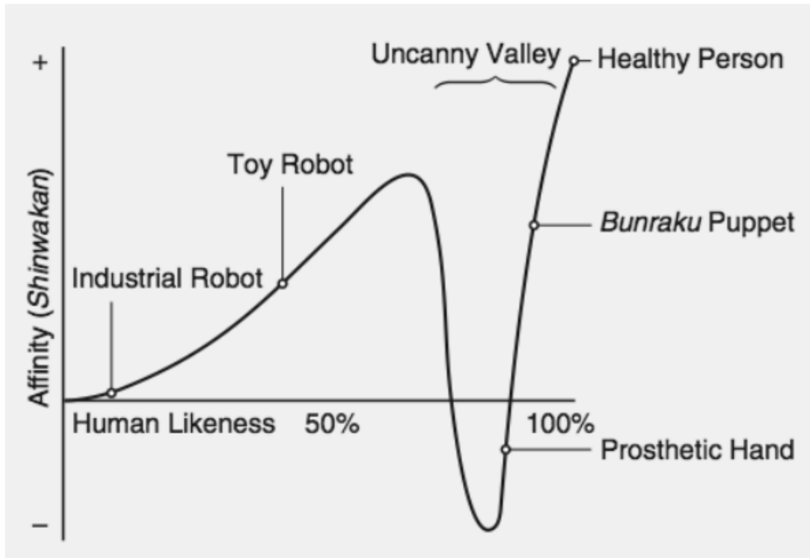
As a result, SPT provides a valuable theoretical lens for understanding how specific design elements can simulate social presence and, in doing so, impact users' emotional and cognitive evaluations of conversational agents.

2.1.5. Uncanny Valley Theory

At the same time, being more human-like is not always better. The Uncanny Valley Theory introduces a cautionary note, showing that poorly calibrated anthropomorphism can evoke discomfort and distrust (Mori, 2012). Luria et al. (2019) argued that CAs possess unique characteristics unrelated to humans. Therefore, designers should recognize these non-humanized attributes when developing CAs. In addition, the Uncanny Valley theory explains the adverse effects caused by anthropomorphism since more humanoid technology causes people to say they experience eeriness (Li & Suh, 2021; Mori et al., 2012). Research on Uncanny Valley Theory explored the implications of such design choices, which do not invariably lead to positive outcomes (Mathur & Reichling, 2016; Saygin et al., 2012; Seymour et al., 2017). Uncanny Valley theory specifically informs researchers that an artificial agent attempting to mimic human traits but failing to achieve a lifelike presence fully can trigger discomfort. The reason for this is that people may perceive a disconnect between the agent's non-human and human-like features, which can prevent them from feeling at ease during the interaction (Mori, 2012). In this context, Chung et al. (2023) discovered that agents designed with voice interfaces to appear more human-like intensified users' concerns about privacy. Moreover, Brendel et al. (2020) found that an agent's inability to respond appropriately can evoke uncanny feelings. In designing CA with humanlike features, the Uncanny Valley serves as a critical reminder that more humanlike is not always required (Figure 4), and careful calibration is needed when designing visual, vocal, or social cues for agents intended to simulate social presence.

Figure 4

Uncanny Valley



Source: Mori, (2012)

2.1.6. Other Theories

In the sphere of conversational agent design research, various theories have been applied across different studies to explore a range of interactions and effects. Al-Natour et al. (2021) have adapted the Social Exchange Theory (SET), which initially was an interpersonal context, to human-agent interaction, drawing attention to using explanations of virtual advisors' design during communication. Yang & Kankanhalli (2023) examined how making a CA's responses humorous for service recovery could satisfy people more. Qiu & Benbasat (2010) and Zhao et al. (2019) utilized Similarity Attraction Theory (SAT) and Dissimilarity-Repulsion Theory (DRT) to explore the effects of demographic similarities between users and agents. Cafaro et al. (2016) applied Proxemics Theory (PT) to investigate the perception of realism of the agent's nonverbal behaviors, as well as expressive design. Cai et al. (2023) utilized the Self-Determination Theory (SDT) to explore the adapting CA design in response to user feedback to enable self-awareness and expression. Ceha & Law (2022) have adopted the Cognitive Load Theory (CLT) and suggested that it is essential for these agents to show emotions to build a better connection with users when dealing with the problem of teaching agents. Further theories and concepts that researchers have applied to the study include

Attribution Theory (AT) (Wang & Benbasat, 2016), Information Behavior Theory (IBT) (Sin & Munteanu, 2020), Frustration-Aggression Theory (FAT) (Riquel et al., 2021a), Social Identity Theory (SIT) (Liu & Yao, 2023; Pietrantoni et al., 2022), Uncertainty Reduction Theory (URT) (Kang & Gratch, 2014), Social Facilitation Theory (SFT) (Hwang & Won, 2021), Schema Theory (ST) (Haas & Moussawi, 2020) and Expectancy Violations Theory (EVT) (Gnewuch et al., 2022; Go & Sundar, 2019).

2.1.7. Summary

Theoretical perspectives on CA design consistently highlight human-agent interaction's social and perceptual nature. Scholars have shown that users respond to CAs as social actors, applying norms and expectations commonly seen in human communication. This occurs through design elements such as embodiment, verbal and non-verbal cues, and identity signaling, which trigger mechanisms like anthropomorphism, social presence, and trust attribution.

Across the reviewed theories, Anthropomorphism Theory, CASA, SRT, SPT, and Uncanny Valley Theory, it becomes clear that design decisions are not merely technical but psychological interventions. These choices shape how users interpret the agent's role, credibility, and emotional presence. While much research focuses on making agents more humanlike to enhance engagement and trust, recent studies caution against over-anthropomorphizing when it creates discomfort or unmet expectations. Therefore, calibration of humanlikeness appears to be critical. Other theoretical lenses, such as Similarity Attraction Theory, Self-Determination Theory, and Cognitive Load Theory, help expand the scope by explaining how agent design can impact perceived relevance, user autonomy, or learning outcomes.

These theoretical models inform how users form expectations, evaluate trustworthiness, and sustain engagement. They emphasize the importance of design intentionality, where social responses are elicited not by accident, but by strategic use of communicative and relational cues. These insights provide a foundational rationale for the design dimensions later explored in this thesis and motivate the empirical investigation into how such choices impact user perceptions in practice.

2.2. Disciplinary Foundations: HCI and IS Perspectives

2.2.1. Human-Computer Interaction Perspective

Human-Computer Interaction is rooted in an interdisciplinary tradition that draws from computer science, cognitive psychology, design studies, and social sciences to examine how people interact with technological systems. Early HCI research, particularly in the 1980s and 1990s, focused on optimizing usability and task performance, drawing heavily on cognitive models such as the GOMS model (Goals, Operators, Methods, and Selection Rules) and theories of mental workload (Card et al., 2008). During this time, HCI was primarily concerned with efficiency, error reduction, and learnability, a concept Norman (1986) referred to as "cognitive ergonomics." However, as computing technologies have become more ubiquitous, social, and embedded in everyday life, the scope of HCI has expanded significantly. Contemporary HCI now engages with emotional, affective, and social dimensions of interaction, including trust, engagement, enjoyment, and long-term user satisfaction (Bødker, 2006; Lazar et al., 2017).

This expansion redefined HCI's focus from merely facilitating human use of machines to fostering meaningful and sustained interaction experiences. Modern HCI distinguishes between first-wave HCI, which emphasized human factors and usability engineering, and second-wave HCI, which turned toward social contexts and situated interaction, as well as third-wave HCI, which explores affect, embodiment, and values in design (Bødker, 2006; Harrison et al., 2007). This evolution is especially relevant in the context of CAs, where users' experiences are shaped by the system's technical accuracy and how well the agent performs socially and emotionally. Agents that mimic conversational norms, express personality, or adapt to user behavior exemplify how HCI's third-wave concerns directly influence system design.

Researchers within HCI have also taken varied epistemological stances. While some adopt a behavioral science approach, using controlled experiments to assess variables such as usability and satisfaction (Hornbæk, 2006), others favor design-based and ethnographic methods, exploring how users make sense of interactive technologies in real-world contexts (Dourish, 2001; Rogers et al., 2011). This pluralism enables HCI to address design questions, what should be built, and evaluation questions, how do users experience what has been built Gerlach and Kuo (1991) emphasized that effective information systems must reflect both perspectives: the technical affordances of a system and the behavioral context of its use. In the case of conversational agents, this dual lens helps explain why certain design elements (e.g., verbal politeness, avatar embodiment, error handling strategies) are not merely stylistic but crucial to how users interpret the agent's competence, trustworthiness, and intent.

2.2.2. Information Systems Perspective

A key distinction within the HCI domain lies between approaches grounded in management information systems (MIS) and computer science. While computer science–oriented HCI often prioritizes practical interface design and iterative user testing, MIS-oriented HCI emphasizes theoretical rigor and model-driven inquiry.

In the MIS tradition, HCI is treated as a subdomain concerned with broader organizational, cognitive, and behavioral factors that influence the success of information systems. Theoretical models such as the TAM, UTAUT, and ECM are commonly employed to examine how users form attitudes, adopt technologies, and maintain use over time (Bhattacherjee, 2001; Davis, 1989; Venkatesh et al., 2003, 2011). These models have become central in how the MIS-HCI community investigates user interaction with systems and informs the design of user-centered technologies (Gerlach & Kuo, 1991; Lazar et al., 2017).

The HCI and IS perspectives provide complementary foundations for understanding and designing user-centered systems. HCI contributes insights into interaction dynamics, while IS offers explanatory models for understanding technology use in context.

2.3. Technology Adoption and Continuance Models

From HCI and IS perspectives, understanding the user experience that determines acceptance and continuous usage is essential. This is particularly relevant in the MIS, where user attitudes, expectations, and intention to use are systematically examined through theoretical models. The MIS researcher incorporated conventional acceptance models, including the TAM, RAT, DOI, and UTAUT. These models help explain the psychological and contextual factors shaping user behavior in both initial (pre) and long-term (post) use. Drawing from this tradition, the following subsections explore how pre-adoption and post-adoption processes have been conceptualized in the literature.

2.3.1. Pre-Adoption Model

IS research has contributed notably to understanding the antecedents of technology usage and acceptance. Specifically, IS aims to understand why people resist new technologies and how to overcome those obstacles (Venkatesh et al., 2003). These models state several antecedents that can influence the adoption of the technology, such as perceived usefulness, perceived ease of use, social influence, facilitating conditions, and behavioral beliefs and intentions (Davis et al., 1989; Venkatesh et al., 2003).

Table 2

Previous studies on CA Acceptance

Acceptance Model	Constructs	Sample Research For CAs	CAs Application	Relationship
TAM	Perceived Usefulness (PU), Perceived Ease of Use (PEOU) Subjective norms		Travel planning chatbot, enterprise collaboration chatbots	PEOU→AIN (β= 0.328) PU→AIN (β= 0.266), PU→BI (f^2 = 0.142) PEOU→BI (f^2 =0.0084)
RAT	Attitudes Toward Behavior, Subjective norms	(Maar et al., 2023)	Restaurant or dentist chatbot for scheduling service	Chatbot-Related Attitude → Chatbot-Related Usage Intention (β= 0.853)
DOI (Rogers, 1995)	Interactivity (I), Compatibility (C), Complexity (Co), Observability (O), Trialability (T)	(Hari et al., 2022)	chatbots for banking services	I→CBE (β=0.194) C→CBE (β=0.314) T→CBE (β=0.055)
(UTAUT) (Venkatesh et al., 2003)	Performance expectancy (PE), Effort expectancy (EE), Social influence (SI), Facilitating conditions (FC), Perceived Risk	(N. Terblanche & Kidd, 2022)	AI coach chatbot (goal-attainment)	PE→BI (β= 0.50) SI→BI (β= 0.30) FC→BI (β= 0.25)

AIN: (Adoption Intention), CBE: (Customer Brand Engagement), "β" represents the standardized path coefficient, "f2" represents the effect size, indicating the proportion of variance explained by the predictor variable(s)

Source: Created by the author.

Individual differences, such as age and technical expertise, can also be important in accepting technology (Venkatesh et al., 2003). These models offer a different perspective on the factors influencing technology acceptance and use, specifically within office automation, software development, and business application tools (Legris et al., 2003).

Researchers have applied conventional models, such as the TAM (Rietz et al., 2019) and the UTAUT (van Bussel et al., 2022), to study factors influencing the acceptance of CAs (Table 2). Additionally, researchers focused on CASA (Nass et al., 1995), SPT (Adam et al., 2021), and Uncanny Valley Theory (Ho & MacDorman, 2010; Thaler et al., 2020) to outline the influencing factors when interacting with CAs. Ling et al. (2021)conducted a systematic literature review and explored existing theories and models to determine the factors influencing the intention to use/adopt CAs. They revealed that agent-characteristic (anthropomorphism, social attractiveness, etc.), user-related factors (demographic and personal innovativeness), usage-related factors (perceived usefulness, utilitarian benefit, hedonic benefit, etc.), and other factors (social influence, facilitating conditions) are key to understanding acceptance and willingness to use CAs. Mariani et al. (2023) classified the drivers that lead to the usage and adoption of the CA, including the design of CAs (anthropomorphic design, personalized design), user-related features (perceived enjoyment, perceived usefulness, etc.), business-related outcomes (word of mouth, branding, etc.), and customer satisfaction. However, these studies did not consider emotions to be one of the factors affecting users' willingness to use CAs. Literature supports that the emotions we experience in response to technology are essential antecedents to using that technology, such as enjoyment, happiness, playfulness, anger, anxiety, irritation, and fear (Beaudry & Pinsonneault, 2010; Cenfetelli, 2004).

Only recently, the Artificial Intelligence Device Acceptance (AIDUA) model, composed by Gursoy et al. (2019), regards emotions in the context of Cognitive Appraisal Theory (CAT) (Lazarus et al., 1980). The model suggests a three-stage appraisal process to understand the willingness to accept using CAT-based AI devices. The first stage involves evaluating the significance and relevance of AI devices (anthropomorphism, social influence, and hedonic motivation), and the second stage focuses on benefits and expectancy (performance and effort expectancy), which form the emotions toward using AI devices. In addition to their applicability in various domains, existing technology acceptance models can be applied to accepting AI devices. In general, the factors that influence technology acceptance, such as perceived usefulness, perceived ease of use, and social influence, are likely to be predictors in the acceptance of AI devices as well (Liu & Tao, 2022).

Finally, emerging NeuroIS research further explores the role of unconscious cognitive and emotional processes in technology interaction. Studies using neuroimaging (e.g., fMRI, EEG) and physiological measures have linked trust, perceived usefulness, and satisfaction to neural activity (Dimoka et al., 2010; Riedl & Léger, 2016). Such findings open new pathways for understanding affective and subconscious drivers of CA adoption and user experience.

2.3.2. Post-Adoption Model

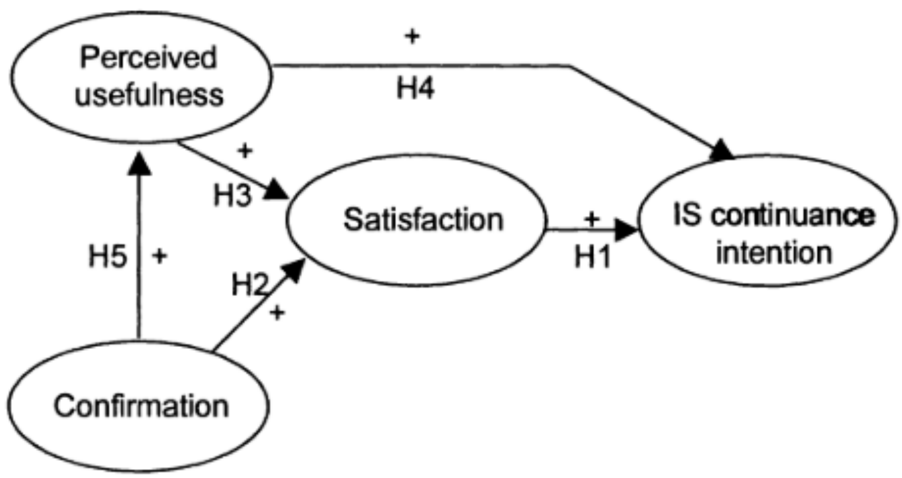
Conventional models for technology acceptance explain the adoption of IS usage, factors influencing resistance to technology, and users' intention and willingness to adopt the technology (Davis et al., 1989; Venkatesh et al., 2003). These models shape the attitudinal beliefs of users in the pre-usage stage, which is the foundation of users' cognitive biases, initial expectations, and judgments. (Davis et al., 1989, 1992; Venkatesh et al., 2003). However, continuance intention focuses on post-acceptance or post-stage usage behavior, where users possess concrete knowledge of the IS. Upon using the product or service or after the pre-acceptance stage, users assess its performance based on their experiences (Bhattacherjee, 2001; Oliver, 1980). This means that users will use a technology based on their experience (Bhattacherjee, 2001; Hayashi et al., 2004; Karahanna et al., 1999).

Regarding this, Bhattacherjee (2001) developed IS's Expectation Confirmation Model (ECM) (Figure 5) based on the Expectation Confirmation Theory (ECT), which essentially describes consumer behavior after purchasing a product or service. According to ECT, consumer behavior is characterized by the intention to repurchase a product or service when satisfaction occurs and expectations are confirmed (Oliver, 1980). The ECM incorporates the concept of perceived usefulness, which represents the expectations formed after consumption (Bhattacherjee, 2001; Nguyen, Chiu, et al., 2021). Perceived usefulness is one of the key indicators of satisfaction. Additionally, ECM also considers confirmation of expectations, comparing users' expectations with the real performance of technology (Bhattacherjee, 2001). Confirmation indicates the realization of the anticipated benefits from IS, which is also positively associated with satisfaction (Bhattacherjee, 2001; Venkatesh et al., 2011). Furthermore, the level of satisfaction will influence the intention to continue using (Bhattacherjee, 2001).

ECM highlights users' assessment of the real performance of IS and how those assessments impact their satisfaction and intention to continue using the IS. Bhattacherjee (2001) built the foundation for IS user continuance intention, and the model successfully explained users' behavior. Moreover, ECM is widely applied in various contexts, such as online course environments (M. Chen et al., 2021) and e-learning systems (Hayashi et al., 2004), mobile apps (Oghuma et al., 2015), mobile internet (Zhou, 2011), and chatbots (Ashfag et al., 2020; Nguyen, Chiu, et al., 2021).

Figure 5

Expectation Confirmation Model



Source: Bhattacherjee (2001)

2.3.3. Post Adoption Model Applications in CA Studies

Recent literature focused on the classical IS models of technology adoption and continuance for conversational agents (Table 3). In particular, scholars have extended these models by integrating constructs that reflect the unique nature of chatbot interaction, such as personalization, perceived intelligence, trust, and anthropomorphism. For example, Lin et al.

(2021) extended the post-adoption model of IS continuance by incorporating chatbot-specific quality dimensions, including understandability, reliability, responsiveness, and interactivity. Their study in the tourism and hospitality domain demonstrated that these dimensions were moderated by technology anxiety and significantly influenced confirmation, satisfaction, and use continuance.

In a different approach, Jin and Youn (2023) examined the influence of anthropomorphism, social presence, imagery processing, and psychological ownership on chatbot continuance intention. Their findings showed that social presence impacts forming users' behavioral intentions (See Table 3 for additional studies linking social presence to continuance intention). Nguyen et al. (2021b) combined the ECM with DeLone and McLean's IS Success Model to explore continuance intention in banking chatbots. As a result, their model highlighted the influence of information quality, service quality, trust, and confirmation on user satisfaction and continuance intention. Furthermore, Zhu et al. (2022) examined how personalization, enjoyment, learning, and contextual conditions affect satisfaction and continuance intention in mental health chatbots. Similarly, Hsiao and Chen (2021) applied the Theory of Reasoned Action (TRA) and SERVQUAL with ECM. They investigated food-ordering chatbots' continuance intention drivers, including problem-solving capacity, anthropomorphism, and trust.

In tourism, Zhang et al. (2023) integrated UTAUT2 and the Theory of Perceived Risk to investigate continuance intention, including constructs such as performance expectancy, social influence, anthropomorphism, and privacy risk. They found that hedonic and risk-related factors are important in long-term user engagement. In another example, Balakrishnan et al. (2022)extended the meta-UTAUT framework by introducing perceived intelligence, anthropomorphism, and social self-efficacy to explain attitude and continuance intention.

Critically, these studies demonstrate that traditional adoption and continuance models remain central but are frequently extended with constructs tailored to the socio-technical context of CAs. These extensions allow researchers to capture both functional and experiential dimensions of user interaction, reflecting the shift toward more contextualized and user-centered models in studying CA long-term usage.

Table 3

Summary of Studies Employing Adoption and Continuance Models in the Context of Conversational Agents

Research Title	Topic	Theory	Research Area	Research Method	Constructs	Results (Hypothesis Supported)	Results (Hypothesis Not Supported)
What makes you continuously use chatbot services? Evidence from Chinese Online travel agencies (Li et al., 2021)	The study incorporated the chatbot quality dimensions by extending the post-acceptance model of IS continuance, aiming to understand better chatbot users and their adoption of services within the tourism and hospitality industry.	Extended post-acceptance model of IS continuance.	Tourism/ Travel Agency Chatbots	Survey	Chatbot Quality Dimensions: Understandability, Reliability, Responsiveness, Assurance, Interactivity, Technology Anxiety, Confirmation, Satisfaction, Use Continuance	Understandability ➞ Confirmation Reliability ➞ Confirmation Assurance ➞ Confirmation Interactivity ➞ Confirmation Understandability x TA ➞ Confirmation Reliability x TA➞ Confirmation Assurance x TA ➞ Confirmation Interactivity x TA ➞ Confirmation Confirmation ➞ Satisfaction Satisfaction ➞ Use Continuance	Responsiveness x TA ➞ Confirmation Responsiveness ➞ Confirmation
Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction (Jin & Youn, 2023)	This study explores the association between anthropomorphism, social presence, imagery processing, and psychological ownership as a predictor of continuance intention in Human-AI-Interaction.	Theory of cognitive development and human intelligence	Fashion and Hospitality	Online survey. Participants viewed screenshots about a conversation with customer and chatbot for hotel room offer or shoes.	Anthropomorphism: Human-likeness, Animacy Intelligence. Social presence, Imagery processing, Psychological ownership (it is 'MINE), Chatbot continuance intention.	Human-likeness ➞ Social presence Human-likeness ➞ Imagery processing Intelligence ➞ Imagery processing Social presence ➞ Imagery processing Imagery processing ➞ Psychological ownership	Social presence ➞ Psychological ownership Animacy ➞ Social presence Intelligence ➞ Social presence Animacy ➞ Imagery processing x

Table Continued

						Social presence ➞ Continuance intention Imagery processing ➞ Continuance intention	Psychological ownership ➞v
Determinants of Continuance Intention towards Banks' Chatbot Services in Vietnam: A Necessity for Sustainable Development (Nguyen et al., 2021a)	The study explores the factors influencing customers' continuance intention to use Bank Chatbot in Vietnam in order to improve customer experience in the banking industry.	Expectation-Confirmation Model, DeLone and McLean's information systems success (D&M ISS) model, Trust	Bank/Bank Chatbot	Survey with Bank customers based on their previous experience.	Information quality, System quality, Service quality, Trust, Confirmation of expectations, Perceived Usefulness, Satisfaction, Continuance intention	Information Quality ➞ Trust Information Quality ➞ Satisfaction System Quality ➞ Trust Service Quality ➞ Trust Service Quality ➞ Satisfaction Confirmation of Expectations ➞ Trust Confirmation of Expectations ➞ Satisfaction Satisfaction ➞ Continuance Intention Perceived Usefulness -➞ Satisfaction Perceived Usefulness -➞ Continuance Intention Trust -➞ Continuance Intention	System Quality ➞ Satisfaction
I am chatbot, your virtual mental health adviser." What drives citizens'	The theory of Consumption Value has been employed to explain the determinants behind users' satisfaction and	Theory of Consumption Values	Health / AI-based mental health chatbot	Cross-sectional study with Online survey for mental	Functional Value: Personalization Voice Interaction Emotional Value: Enjoyment	Personality-➞ Satisfaction Personality -➞ Continuance	Voice Interaction-➞ Satisfaction Voice

Table Continued

satisfaction and continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China (Zhu et al., 2022)	their intention to continue using mental health chatbots during the COVID-19 pandemic.			health chatbot (Xiaolv) users	Epistemic Value: Learning Condition Value: Condition User Satisfaction Continuance Intention	Intention Enjoyment -→ Satisfaction Enjoyment-→ Continuance Intention Learning -→ Satisfaction Learning-→ Continuance Intention Condition -→ Satisfaction Condition-→ Continuance Intention Satisfaction → Continuance Intention	Interaction-→ Continuance
What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction (Hsiao & Chen, 2021)	The study employed behavioral belief and outcome evaluation to predict users' intention to continue using a food-ordering chatbot.	Theory of reasoned action (TRA), SERVQUAL	Food Service/ LINE chatbot of Kentucky Fried Chicken (KFC)	Online survey for food ordering bot service users. The screenshots were utilized.	Anthropomorphism, Service Quality: Problem-Solving User Interface, Trust, Satisfaction, Continuous Using Intention	Anthropomorphism -→ Satisfaction (not very strong) Anthropomorphism -→ Trust (not very strong) Problem-Solving -→ Satisfaction Problem-Solving -→ Trust User Interface -→ Satisfaction Trust → Continuance Intention Satisfaction → Continuance Intention Trust → Satisfaction	User Interface -→ Trust

Table Continued

Investigating Patients' Continuance Intention Toward Conversational Agents in Outpatient Departments: Cross-sectional Field Survey (Li et al., 2022)	This study explored the fundamental elements that impact patients' continuance intention to use CAs.	Expectation-Confirmation Model	Health	Cross-sectional field survey with patients	Perceived Usefulness, Perceived Ease of Use, Confirmation, Satisfaction, Continuance Intention	Perceived usefulness → continuance intention Perceived usefulness → satisfaction Confirmation → Perceived Usefulness Confirmation → satisfaction Confirmation → Perceived ease of use Satisfaction → continuance intention	Perceived ease of use → continuance intention Perceived ease of use → satisfaction Perceived ease of use → Perceived Usefulness
Hey chatbot, why do you treat me like other people? The role of uniqueness neglect in human-chatbot interactions (Kallel et al., 2023)	The study focuses on resistance behavior to chatbots and understanding customer reactions to interactions with bank chatbots	Warmth, Competent, Uniqueness Neglect	Bank/Bank Chatbot	Survey based on previous experience in France	Uniqueness neglect Chatbot competence Chatbot warmth Satisfaction Recommendation intention Continuance intention	Satisfaction → Continuance Intention Satisfaction → Recommendation Intention Competence → Satisfaction Competence X Uniqueness Neglect → Satisfaction	Competence X Warmth → Satisfaction
"I Am Here to Assist Your Tourism": Predicting Continuance Intention to Use AI-based Chatbots for Tourism. Does Gender Really Matter?	The study determined the predictors behind the customer's continuous intention to use tourism chatbots by including UTAUT2, TPR, anthropomorphism, and personalization.	Unified Theory of Adoption and Use of Technology 2 (UTAUT2), the Theory of Perceived Risk (TPR), Anthropomorphism, and Personalization	Tourism/online Tourism market chatbots	Survey with tourism company customers who have experience with tourism chatbots in China.	TPR: Time Risk, Privacy Risk UTAUT2: Performance Expectancy, Effort Expectancy, Social Influence, Hedonic Motivation,	Performance Expectancy ➞ Continuance Intention Social Influence ➞ Continuance Intention Habit ➞ Continuance Intention Privacy-Risk ➞	Effort Expectancy ➞ Continuance Intention Facilitating Conditions ➞ Continuance Intention Hedonic Motivation ➞

Table Continued

(Zhang et al., 2023)					Habit Supplementary Antecedents: Anthropomorphism, Personalization Continuance Intention	Continuance Intention (negative) Time-Risk ➞ Continuance Intention (negative) Anthropomorphism ➞ Continuance Intention(positive) Personalization ➞ Continuance Intention	Continuance Intention
Determinants and consequences of trust in AI-based customer service chatbots (Prakash et al., 2023)	The study focuses on trust due to its importance in accepting technology to determine how trust influences the continuance intention to use chatbots.	Trust in Technology Model (TTM), TAM, Social Presence	General	Online survey based on previous experience of the participants in India	Conversational Ques: Perceived Contingency, Perceived Interactivity. TAM: Perceived Usefulness, Perceived Ease of Use. Social Presence Trust in Technology Model: Propensity to Trust Technology, Trusting Belief, Privacy Risk. Trust Intention. Continuance Intention	Perceived Contingency → Perceived Usefulness Perceived Contingency → Perceived Ease of Use Perceived Contingency → Social Presence Perceived Interactivity → Perceived Usefulness Perceived Interactivity → Perceived Ease of Use Perceived Interactivity → Social Presence Perceived Usefulness → Trusting Belief Perceived Ease of Use→ Trusting Belief Perceived Usefulness →	Privacy Risk → Trust Intention

Table Continued

						Continuance Intention Perceived Usefulness → Trust Intention Trusting Belief → Continuance Intention Trusting Belief → Trust Intention	
The role of meta-UTAUT factors, perceived anthropomorphism, perceived intelligence, and social self-efficacy in chatbot-based services? (Balakrishnan et al., 2022)	The study explored the important attributes of chatbot user continuance. This was achieved by expanding the meta-UTAUT framework and incorporating self-efficacy, using perceived intelligence and anthropomorphism as additional factors.	meta-UTAUT	General: based on previous experience engaging with chatbots	A single cross-sectional design with data collected using a survey	Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Perceived Intelligence, Perceived Anthropomorphism, Social Self-Efficacy, Attitude toward chatbot, Continuing Intention	Performance Expectancy → Attitude Effort Expectancy → Attitude Facilitating Conditions → Attitude Social Influence → Attitude Perceived Intelligence → Attitude Perceived Anthropomorphism → Attitude Performance Expectancy → Continuing Intention Effort Expectancy → Continuing Intention Facilitating Conditions → Continuing Intention Perceived Intelligence → Continuing Intention Perceived	Performance Expectancy x Social Self-Efficacy → Continuing Intention Effort Expectancy x Social Self-Efficacy → Continuing Intention Facilitating Conditions x Social Self-Efficacy → Continuing Intention Social Influence x Social Self-Efficacy → Continuing Intention Social Influence → Continuing Intention

Table Continued

						Anthropomorphism ➞ Continuing Intention Attitude ➞ Continuing Intention Perceived Intelligence x Social Self-Efficacy ➞ Continuing Intention Perceived Anthropomorphism x Social Self-Efficacy ➞ Continuing Intention	
Exploring chatbot trust: Antecedents and behavioral outcomes (Alagarsamy & Mehrolia, 2023)	The study has been conducted with an emphasis on trust, aiming to determine antecedents and outcomes of chatbot trust.	TAM, Unified Theory of Acceptance and Use of Technology (UTAUT), Diffusion of Innovation theory (DOI), D&M success model, Trust	Bank/Bank Chatbot	Online Survey with Bank Customers in India	Perceived Ease of Use, Perceived Enjoyment, Perceived Usefulness, Service Quality, Information Quality, Interface and Design, Perceived Risk, Structural Assurances, Privacy and Security Concerns, Disposition to Trust, Technology Fear, Ubiquity, Chatbot Trust, Attitude, Behavioral Intention, User Satisfaction	Perceived Ease of Use → Chatbot Trust Perceived Enjoyment → Chatbot Trust Perceived Usefulness → Chatbot Trust Service Quality → Chatbot Trust Information Quality → Chatbot Trust Chatbot Trust → User Satisfaction Social Presence → Trusting Belief Perceived Risk → Chatbot Trust Privacy and Security Concerns → Chatbot Trust Disposition to Trust → Chatbot Trust Social Presence → Trust Intention Propensity to Trust	Interface and Design → Chatbot Trust Perceived Ease of Use→ Continuance Intention Perceived Ease of Use → Trust Intention Structural Assurances → Chatbot Trust Technology Fear → Chatbot Trust Social Presence → Continuance Intention Propensity to Trust Technology → Trust Intention Privacy Risk → Trusting Belief Privacy Risk

Table Continued

						Technology → Trusting Belief Propensity to Trust Technology → Continuance Intention Ubiquity → Chatbot Trust Chatbot Trust → Attitude Chatbot Trust → Behavioral Intention	→ Continuance Intention
I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents (Ashfag et al., 2020)	The study extended the ECM to examine the factors influencing users' satisfaction and their intention to continue using chatbot-based customer service.	Expectation-confirmation model (ECM), Information system success (ISS) model, TAM, and the need for interaction with a service employee (NFI-SE)	General: text-based customer service chatbots	Survey	Information quality Service quality Perceived Enjoyment Perceived Usefulness Perceived ease of use Need for interaction with a service employee (NFI-SE) Satisfaction Continuance intention.	Information Quality → Satisfaction Service Quality → Satisfaction Satisfaction → Continuance Intention Perceived Enjoyment → Satisfaction Perceived Usefulness -→ Satisfaction Perceived Enjoyment → Continuance Intention Perceived Usefulness -→ Continuance Intention Perceived Ease of Use → Continuance Intention NFI-SE → satisfaction Perceived Ease of Use × NFI-SE →	Perceived Enjoyment × NFI-SE → satisfaction Perceived Ease of Use → Satisfaction

Table Continued

	satisfaction Perceived Usefulness × NFI-SE ➞ satisfaction

Source: Created by the author.

2.4. Key Predictors of Continuance Intention in CA Interaction

Regarding CAs, recent literature emphasized that these systems differ from traditional systems due to their capabilities and the replacement of human roles. Extending conventional acceptance models related to agent characteristics can highlight on how users experience them. Prior research highlights several psychological and interaction-related constructs that affect the user evaluations of the CAs. This section discusses five key predictors of continuance intention as established in prior IS and HCI literature. The following five constructs are identified and discussed: perceived social presence, perceived usability, perceived usefulness, perceived enjoyment, and trust.

2.4.1. Perceived Social Presence

As a starting point for understanding social presence, we must grasp the meaning of presence itself. We can observe this sense of presence when individuals watch a television program, as they feel like the program's world is being brought right to them. A notable example from the early days of cinema illustrates this sensation, where specific theater audiences panicked and hurried towards the exits when a black and white film of an approaching locomotive was screened (Schoen, 1976). Lombard (1995) explained why this occurred, as media users could not distinguish between the images and the referents. As a result, they reacted as if what they saw in the film was present in their physical environment. Fundamentally, the sense of presence lies at the heart of these mediated experiences (Biocca, 1997). Emerging technologies like video conference tools, virtual reality simulations, and computers provide mediated experiences that strongly evoke a sense of presence (Aitamurto et al., 2018; Lombard & Ditton, 1997; Sirkin et al., 2011). Consequently, presence is essential in traditional media (TV, radio) and human-computer interfaces (Biocca, 1997; Biocca et al., 2003).

Due to the significant interest in this concept, some scholars have attempted to conceptualize the definition of presence. Biocca (1997) defined presence as the illusion of "being there," regardless of whether that "there" exists in physical space. This definition connects presence and mediated experience, referring to the "being there" sensation within a virtual or mediated environment (Biocca et al., 2003; Lee, 2004). However, Lombard and Ditton (1997) defined presence as “the perceptual illusion of nonmediation.” This definition illustrates how someone might not notice the presence of a medium and react as if the tool is not there. According to Biocca's (1997) definition, it is about experiencing a virtual world as if you were there, while Lombard and Ditton's (1997) definition highlights not realizing that you are interconnected with technology. Meanwhile, scholars have proposed different terms for presence, such as telepresence, mediated, physical, social, and virtual presence, all referring to the same concept (Barfield & Furness, 1995).

Lee (2004) aimed to define presence to address the aforementioned interpretations. Firstly, his attempt sheds light on presence research related to virtual experiences. This approach focuses on virtual and actual objects to emphasize their connections and similarities. As a result, the concept of presence is distinguished from experiences with actual objects (real experience) or imaginary objects (hallucination). This distinction leads to a primary focus in presence research: the investigation of psychological similarities between virtual and actual objects. Additionally, Lee’s (2004) approach compromises para-authentic (representations of real objects) and artificial objects (objects that do not exist in the physical world), suggesting that when people use virtual objects, they focus on the psychological similarities with real ones. Hence, presence should be operationalized as a psychological construct. Analogously, scholars stated that the feeling of presence occurs in the psychology of the users in virtual environments (Barfield & Furness, 1995; Felton & Jackson, 2022).

On the other hand, social presence occurs when technology users do not consciously recognize or think about nonhumans in virtual environments (Lee, 2004). Consequently, interacting with other entities in the virtual environment feels like engaging with real social actors (Biocca et al., 2003; Lee, 2004; Lombard & Ditton, 1997). Social presence is one of the dimensions of presence that conveys a "sense of being with others" or "interacting with another social being"(Biocca et al., 2003; Lombard & Ditton, 1997). These "others" can be another person, artificial intelligence, humanoid, or animal-like agents (Lee, 2004; Lombard & Ditton, 1997).

CASA theory tells us that individuals apply social rules, which arise in interpersonal relationships, to computers (Nass et al., 1996; Nass & Moon, 2000). HCI studies support the rationale for applying social rules specifically to agents or chatbots due to users perceiving them as socially present (Gefen & Straub, 2004; Lee et al., 2022; Munnukka et al., 2022; Nass & Moon, 2000; Prakash et al., 2023). According to Lee (2004), individuals presume the social presence of artificial beings due to their characteristics in sensory or nonsensory ways. A former study revealed that people easily comprehended the humanlike behaviors of agents (Luria et al., 2019). Moreover, human-likeness positively influences the perception of social presence (Munnukka et al., 2022). From a socio-relational perspective, the social presence of the CAs conveys a sense of human contact, warmth, and friendliness, further aligning with individuals' goals (Hess et al., 2009).

Luria et al. (2019) studied social presence with different design configurations of CAs and conducted experimental research. They found that perceiving social presence during communication is like interacting with another entity, and participants felt more comfortable with the feeling of social presence from CAs. In addition, they stated that agents also have non-humanlike design attributes, such as a singular social presence that can hop from one body to another (e.g., Siri, Alexa). Therefore, it should be noted that human and nonhuman likeness attributes are essential in feeling a sense of social presence from agents. Designers should focus on human likeness and know that agents are complex social presences encompassing humanlike and non-humanized attributes.

Social presence has been shown to influence how users perceive and interact with technology. For instance, having a higher sense of social presence can increase trust and enjoyment of the technology (Gefen & Straub, 2004; Mishra et al., 2021) In the context of retail websites, Chattaraman et al. (2012) employed virtual assistants to support older people when interacting on websites. Social presence in retail websites through virtual agents has aided elderly users in perceiving higher levels of social support; this, in turn, enriches their trust in online retail stores. Furthermore, a higher sense of social presence also positively influences website satisfaction and purchase intentions (Lu et al., 2016). Scholars in HCI have demonstrated that the social presence exhibited by agents is a key predictor for perceived usefulness, trust, and enjoyment, leading to adoption intentions and willingness to use (De Cicco et al., 2020; Jin & Youn, 2023; Lee et al., 2022; Munnukka et al., 2022; Tan & Liew, 2020; Toader et al., 2020).

2.4.2. Perceived Usability

Since the growing attention on CAs, it is crucial to understand factors that influence user experience when interacting with technological artifacts (Borsci et al., 2015; Kujala et al., 2011; Nicolescu & Tudorache, 2022). International Organization for Standardization (ISO) 9241-11 is one of the standardized frameworks that guides the user experience by evaluating usability and interaction quality. It summarizes principles and methods for conducting usability evaluations, including effectiveness, efficiency, and user satisfaction (ISO, 2018). Finstad (2010) suggested a Usability Metric for User Experience (UMUX) organized around ISO 9241-11 to capture a product's user experience. Moreover, the System Usability Scale (SUS) is another standardized assessment tool to capture users' subjective perceptions of the usability of a system (Brooke, 1996). However, these tools cannot be suitable for gathering valuable insights into conversational aspects of CAs because they are centered around traditional interface usability and user experience metrics (Borsci, et al., 2022a; Chandra et al., 2022; Lewis et al., 2018; Sugisaki & Bleiker, 2020).

Sugisaki and Bleiker (2020) attempted to cover conversational aspects to offer practical and usable interactions with Conversational User Interfaces (CUI). However, their endeavors have been limited to identifying CUI requirements guidelines. Eventually, Borsci et al. (2022b) developed the Chatbot Usability Scale (BUS-11) by adapting the quality attributes of Radwill and Benton (2017) in response to the need for a reliable and comparable instrument to assess the quality of conversational interactions. The scale aims to measure end-user perception of usability after using a chatbot. Besides, the scale highlights the conversational ability of chatbots, such as engaging in communication and maintaining efficient and effective conversational exchanges.

BUS-11 follows the usability concept of ISO 9241 (ISO, 2018), concentrating on satisfaction. Essentially, it evaluates chatbot interaction quality across four dimensions: (1) Perceived accessibility to chatbot functions, (2) Functional interactive conversations, (3) Perceived privacy and security, and (4) Time response (Borsci & Schmettow, 2024). Scholars noted that the scale does not encompass specific attributes, such as personality and enjoyment, despite their frequent endorsement in the literature, because personality and enjoyment attributes might have less impact on short-term interactions but could be more influential in determining long-term interactions (Borsci, Schmettow, et al., 2022). Moreover, BUS-11 has adapted to different languages and demonstrated its effectiveness in explaining the perception of usability in chatbot quality assessment.

2.4.3. Trust

Trust is associated with expectations shared between the trustor and the trustee. It represents a psychological state where the trustor has positive expectations regarding the intentions or behavior of the trustee (Mayer et al., 1995; Simpson, 2007). The definition of trust has been made based on interpersonal trust. In organizational studies, Mayer et al. (1995) defined trust as a trustee's ability, benevolence, and integrity. Analogously, for virtual teams, Jarvenpaa et al. (1997) defined trust with the aforementioned terms. Accordingly, the ability is about the trustee's perceived competence in the specific domain (Jarvenpaa et al., 1997; Mayer et al., 1995). Benevolence is perceived as having genuine care and concern for others and being willing to help them (Jarvenpaa et al., 1997). Moreover, benevolence refers to loyalty, openness, caring, and availability (Mayer et al., 1995). Integrity refers to the trustee's actions that align with ethical principles, values, or standards while encompassing attributes like consistency, reliability, congruence, and fairness (Jarvenpaa et al., 1997; Mayer et al., 1995).

Schneiderman (2000) stated that trust is a social construction based on interpersonal relationships. Trust conceptualization based on interpersonal relationships by Mayer et al. (1995) is widely accepted by researchers, including recommender agents (Wang & Benbasat, 2007), electronic commerce (Bhattacherjee, 2001), chatbots (Seeger et al., 2017; Seeger & Heinzl, 2021), mobile payment (Franque et al., 2023; Talwar et al., 2020), and e-government technology (Venkatesh et al., 2011).

In the context of IS, trust is recognized as a critical predictor in the acceptance of technology (Gefen, 2000, p. 200; Komiak & Benbasat, 2006; Qiu & Benbasat, 2008; Shneiderman, 2000). The current literature suggests that humans evaluate computer systems based on rationality, objectivity, and reliability. Moreover, the systems are considered more trustworthy than humans (Devitt, 2018). On the other hand, Schneiderman (2000) has separated trust in person from trust in technology and stated that trust in technology is associated with confidence in its performance and consistency. Trust in interpersonal relations may not directly apply to human-to-computer trust (Rheu et al., 2021). However, regarding conversational agents, Seeger et al. (2017) used a three-dimensional conceptualization of interpersonal trust and distinguished trust as qualification-based and goodwill-based trustworthiness for CAs. Accordingly, qualification-based trust expresses CA's ability to perform the anticipated task related to the necessary skills and qualifications (competence). Meanwhile, benevolence and integrity are categorized as goodwill-based trust, which is concerned with intentions and the ethical aspect of behavior (Seeger et al., 2017; Seeger & Heinzl, 2021). Seeger et al. (2021) made this distinction, stating that "computers do not have volitional control over intentions and motives, but anthropomorphizing nonhuman objects by definition assigns such human characteristics to them."

In the context of CAs, researchers widely employed trust in HCI. More specifically, trust is essential in determining user adoption of chatbots (Guo et al., 2022). Especially, prior studies found that trust is one of the critical predictors of user satisfaction and the intention to continue using a CA (Alagarsamy & Mehrolia, 2023; Chung et al., 2020; Eren, 2021; Hsiao & Chen, 2021; Kallel et al., 2023; Prakash et al., 2023) (Table 3). Moreover, if users trust the CAs and receive reliable service, it motivates them to continue using them (Nguyen et al., 2021a). A recent study involving interviews with customer service chatbot users has revealed that users tend to get information from chatbots instead of employees in the company if the chatbot is trustworthy (Følstad et al., 2018). Another study demonstrated that the trust of chatbot users enhances the usage intention of chatbots and encourages customer engagement (Mostafa & Kasamani, 2022). Chandra et al. (2022) utilized a mixed-method approach involving three distinct human-like chatbot interactions to investigate the role of trust. Importantly, they discovered that trust is crucial in keeping users engaged during exchanges with chatbots. In their study, participants also stated that if the chatbot responds successfully to their questions, they will trust it. Lastly, their analysis revealed that the chatbot's competencies are primarily a builder of trust rather than human-like attributes.

2.4.4. Perceived Enjoyment

Using technology is an intrinsically motivated activity wherein individuals become entirely absorbed in the activity, committed to it, and derive enjoyment, regardless of the expected outcome of the usage (Davis et al., 1992; Deci, 1975; Venkatesh, 2000). This enjoyment in system usage can be seen as an inherent motivation for adopting technology (Song & Han, 2009; Venkatesh, 2000). Van der Heijden (2004) found that perceived enjoyment is central to user acceptance in systems designed for pleasure and entertainment. However, in task-oriented systems, perceived enjoyment indirectly influences the intention to use them (Venkatesh et al., 2003). Recent research, however, provides insight into the importance of considering both utilitarian (task-oriented) and hedonic (pleasure-driven) aspects in technology acceptance and customer engagement (Davis et al., 1992; Japutra et al., 2022; Koufaris, 2002; Song & Han, 2009; van der Heijden, 2004). Essentially, enjoyment substantially influences intentions when users perceive the computer system as applicable (Davis et al., 1992).

In computer-mediated environments, intrinsic enjoyment and pleasure levels result in positive attitudes (H. Kim et al., 2013). More specifically, in the context of CAs, perceiving the CA as enjoyable is essential; its absence can evoke negative emotions, users' assessment of CA affects their usage, and they consider it effortful (Mariani et al., 2023; Venkatesh, 2000). Furthermore, enjoyment contributes positively to customer loyalty and satisfaction across various domains, including online shopping customers (Jarvenpaa et al., 1997), mobile phone users (Kujala et al., 2011), mobile application users (Song & Han, 2009), and chatbot users (Ashfag et al., 2020; Zhu et al., 2022).

Moreover, research has highlighted the importance of enjoyment in influencing the intention to use social shopping websites (Shen, 2012). Similarly, Song and Han (2009) investigated perceived enjoyment as one of the key factors driving the intention to use mobile applications. Hew et al. (2018) provide support for the role of enjoyment in mobile social tourism shopping intention. Recently, some research has examined the positive influence of perceived enjoyment on user satisfaction in various contexts, such as chatbot interactions (Ashfag et al., 2020), movie recommender systems (Lee & Choi, 2017), and mental health chatbots (Zhu et al., 2022). According to Qui and Benbasat (2008), "product recommender agents with the humanoid embodiments and human voice invoked stronger perceptions of social presence, which elevates perceived enjoyment." Consequently, this situation resulted in a higher usage intention. Therefore, CA developers should incorporate entertainment elements into their designs to enhance user experience (Wei et al., 2016).

2.4.5. Perceived Usefulness

Perceived usefulness refers to how individuals can enhance their job performance using a specific information system (Davis, 1989; Venkatesh & Davis, 2000). TAM addresses perceived usefulness as a cognitive belief related to the intention and behavior of using technology (Davis, 1989). In several studies, perceived usefulness has been accepted as a fundamental construct in decisions to use information technology (Davis et al., 1989; Venkatesh & Davis, 2000). For instance, Davis et al. (1989) discovered a substantial (.76) association between perceived usefulness and usage intention for a word-processing program.

Davis (1989) observed that individuals primarily adopt technology due to the functions it performs for them. On the other hand, TAM2 posits that individuals assess the system's usefulness by considering its output quality, particularly in the context of their job requirements (Venkatesh & Davis, 2000). Furthermore, Davis et al. (1992) define perceived usefulness as extrinsic motivation. This highlights that perceived usefulness focuses on extrinsic rewards or consequences instead of oneself.

Perceived usefulness extends its influence beyond the pre-acceptance stage and impacts subsequent continuance decisions (Bhattacherjee, 2001) and grasps users' cognitive expectations regarding the system's performance in the post-usage stage (Venkatesh et al., 2011). Unlike the pre-acceptance stage, individuals gain first-hand experience in the post-acceptance stage, which makes perceived usefulness an unbiased and more realistic construct (Bhattacherjee, 2001; Venkatesh et al., 2011). Bhattacherjee (2001) reported that perceived usefulness is the most significant past expectation influencing satisfaction after the hands-on experience based on the Expectation Confirmation Model.

Perceived usefulness is a widely applied construct for various systems, such as business graphics programs (Davis et al., 1992), business software(Agarwal & Prasad, 1999), e-government technologies (Venkatesh et al., 2011), digital library and agile Web portal (Hong et al., 2014), operating system (Steelman et al., 2014), travel website (Liu & Park, 2015). Legris et al. (2003) critically reviewed the technology acceptance model for different types of software (office automation tools, software development tools, and business application tools). They revealed that perceived usefulness is a significant factor in user acceptance of all types of software.

In CAs, HCI researchers have leveraged the perceived usefulness construct. Ashfag et al. (2020) observed a significant correlation between perceived usefulness, satisfaction, and continuance intention for customer service chatbots. Kasilingam (2020) demonstrated the considerable impact of perceived usefulness on attitudes toward e-commerce chatbots on Facebook. Pillai and Sivathanu (2020) observed that the perceived usefulness of AI-powered chatbots in hospitality and tourism impacts chatbot adoption intention. Belanche et al. (2019) reported that perceived usefulness positively affects attitudes toward financial robo-advisors. Analogously, Heerink et al. (2010) determined that the intention to use assistive social agents is driven by perceived usefulness. On the other hand, former research found that the interaction style of agents affects perceived usefulness (Chattaraman et al., 2019; Prakash et al., 2023). Surprisingly, no other attempt has been made to identify the predictors that influence the perceived usefulness of CAs. This topic remains unexplored in prior literature, with no clear explanation provided.

2.5. Summary for Design Evaluations

This chapter demonstrated that user experience with conversational agents is shaped by system performance and design choices embedded within the agent's interaction model. Drawing from both pre-adoption (e.g., TAM, UTAUT) and post-adoption (e.g., ECM) perspectives, a consistent set of predictors, including perceived usefulness, perceived enjoyment, trust, perceived usability, and perceived social presence, emerges across empirical studies as central to user evaluations and continuance intentions.

Importantly, these predictors are not isolated constructs but are dynamically influenced by how design elements are operationalized. For instance, features such as interaction style, verbal framing, or social cues can affect users' trust, while usability and enjoyment are shaped by response structure and transparency. Across studies, findings show that the same design choice may yield varying effects depending on the interaction context (e.g., education vs. tourism), user characteristics (e.g., prior experience, expectations), or agent role (e.g., task-oriented vs. relational). This reinforces the need for context-aware design strategies that align design features with the intended use case.

A notable distinction emerges between the HCI and IS traditions. While the HCI perspective foregrounds usability and user satisfaction as critical to interaction quality, IS studies often focus on satisfaction from a post-adoption perspective without directly incorporating usability-specific measures. The HCI literature suggests that experiences perceived as unsatisfactory can significantly alter user behavior and technology preference (Hornbæk, 2006). From this view, the quality of interaction becomes a key determinant of both initial impressions and continued use. Therefore, when designing conversational agents, it is crucial to ensure that design choices serve not only the functional goals of the agent but also support positive, satisfying, and trustworthy user experiences. The theoretical and empirical insights presented here highlight the need for purposeful, context-aware design strategies that reinforce long-term engagement through their influence on key psychological predictors.

Building on these insights, this thesis proposes an integrated evaluation model that consolidates the identified predictors, perceived social presence, trust, enjoyment, usefulness, and perceived usability as determinants of continuance intention. This model serves as the conceptual foundation for the thesis's empirical work. It provides a structured lens to examine how specific design decisions influence psychological predictors and long-term user engagement. The method chapter explains these theoretical insights to develop a design and evaluation framework, which guides the artifact development and empirical study. The goal is to examine how design decisions affect user evaluations.

CHAPTER 3. RESEARCH METHODOLOGY

This section outlines the dissertation's research methodology, problem, and goals, which adopts the Design Science Research approach, the broader methodological framework guiding this dissertation through three interconnected cycles: relevance, design, and rigor. First, the conceptual overview of the DSR methodology was presented, followed by foundational principles and rationale for use in CA’s design. It then presents the research design, explaining how the three DSR cycles are interwoven across the stages of problem investigation, requirement derivation, artifact development, and evaluation. To understand the design problem and inform the development of CA design guidelines, a systematic literature review was conducted. These activities supported eliciting user experience needs and formulated initial design requirements. In parallel, a measurement instrument was developed to assess overall user experience with CAs. Empirical studies were then conducted to validate the proposed model, as well as evaluate CA interactions. Finally, the results of these investigations were synthesized into a set of meta-requirements and mapped to actionable design elements.

3.1. Research Approach: Design Science Research

Design Science Research is a prominent research paradigm within Information Systems that focuses on developing creative artifacts to address real-world problems while generating theoretical knowledge. It contrasts with behavioral science, which aims to explain and predict phenomena by instead seeking to build solutions through purposeful design. Articulated initially by Simon (1996) and Nunamaker et al. (1991), and later formalized within IS by Hevner et al. (2004), DSR positions artifact creation as a central activity of research. It moves beyond traditional empirical inquiry by embedding design as a method of investigation, emphasizing the construction, use, and evaluation of artifacts that serve practical and theoretical purposes (Hevner et al., 2024b; Hevner & Chatterjee, 2010).

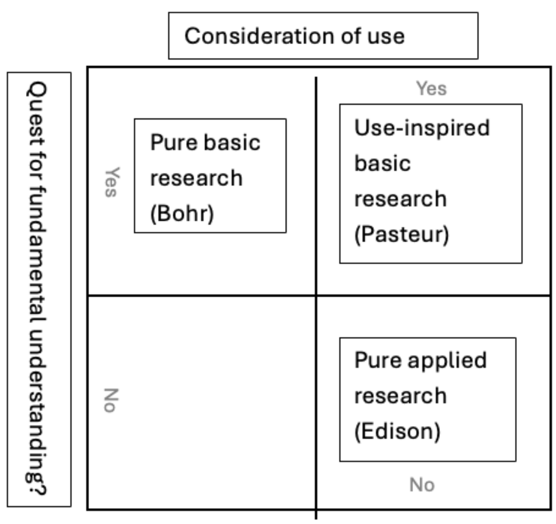
3.1.1. Foundation and Purpose of DSR

DSR is a problem-solving paradigm in which researchers aim to produce innovative, evaluated artifacts that respond to real-world human and organizational needs. As Hevner and Chatterjee (2010) explain, the DSR process does not merely involve building systems but also generating theoretical contributions through reflective evaluation and generalization of design outcomes. By focusing on utility, effectiveness, and theoretical insight, DSR bridges the longstanding relevance gap in IS research: the disconnect between academic rigor and practical utility (Hevner, 2007).

DSR can be further situated within Stokes' (1997) quadrant of research motivations, which classifies research according to its pursuit of fundamental understanding and practical utility (Hoppe de Sousa et al., 2009). The matrix (Figure 6) illustrates that DSR reflects Pasteur's quadrant, representing theoretically motivated and practically relevant studies. Unlike Edisonian experiments (which are helpful but not theory-oriented) or Bohr's principles (which are theoretical but not immediately useful), DSR contributes simultaneously to scientific understanding and real-world impact. Thus, DSR aims to answer practical questions while contributing to the theoretical base of the IS discipline.

Figure 6

Quadrant Model of Scientific Research



Source: Stokes (1997)

In this light, IS research is often understood to comprise two core paradigms: behavioral science, which focuses on explaining phenomena, and design science, which aims at building artifacts (Hevner et al., 2024b). The design paradigm addresses problem-solving through creating, justifying, and evaluating artifacts.

DSR paradigm defines artifacts as purposeful creations. These creations embody solutions to identified problems. These solutions can simultaneously act as tools for intervention and as vehicles for knowledge discovery. Constructs represent the vocabulary necessary to express design problems and their components; models provide abstractions that map relationships between problems and solutions (Weber, 2006); methods guide the development or implementation of artifacts; instantiations demonstrate the operationalization of these elements in functioning systems. At a higher level of abstraction, design theories offer generalizable principles for creating and evaluating future artifacts (Gregor & Hevner, 2013). Moreover, DSR spans engineering and theoretical domains, producing various outputs in form and abstraction. According to Hevner and Chatterjee (2010), valid artifacts in DSR include the (1) Constructs, (2) Models, (3) Methods, (4) Instantiations, and (5)Design Theories. They must be grounded in existing theory and iteratively tested to ensure functionality and contribution to knowledge.

3.1.2. The Structured Process Model of DSR

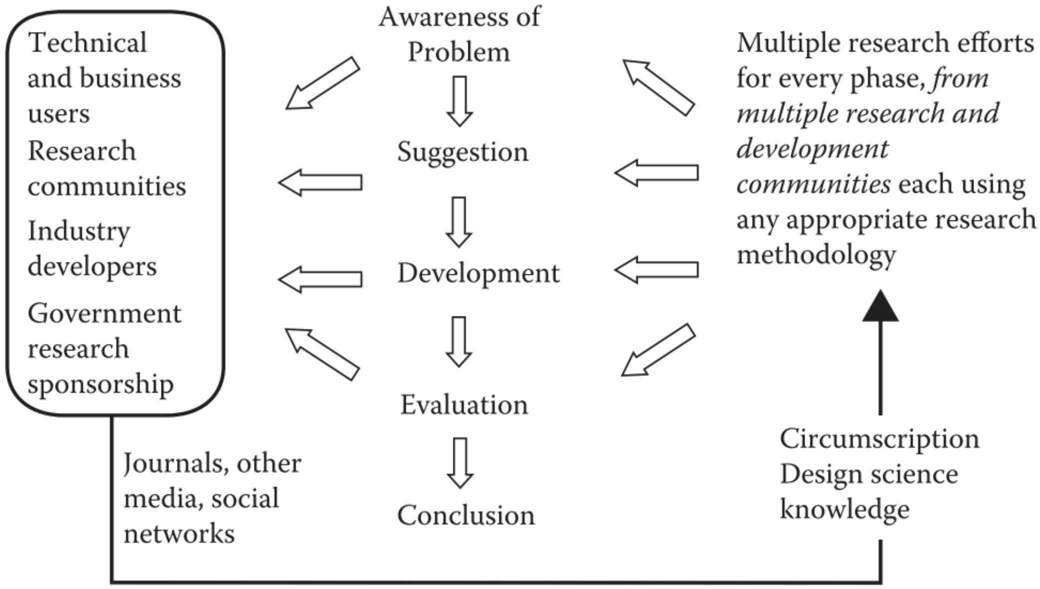
To implement DSR in practice, a systematic approach, which includes a six-step process, was proposed by Peffers et al. (2007):

- Problem Identification and Motivation – Defining the practical and theoretical importance of the problem.
- Definition of the Objectives for a Solution – Translating the problem into clear and design-relevant goals.
- Design and Development – Incorporating appropriate design knowledge, tools, or principles to create artifacts.
- Demonstration – Implementing the artifact in a relevant context.
- Evaluation – Using proper metrics and methods to assess the artifact, including its effectiveness and performance.
- Communication – Disseminating the research outcomes to academic and practitioner audiences.

Besides, the iterative nature of this process is illustrated in the general design cycle proposed by Vaishnavi and Kuechler (2007), shown in Figure 7, which highlights the cyclic movement between problem awareness, suggestion, development, evaluation, and conclusion, all embedded in a broader social and scientific knowledge environment.

Figure 7

General DSR Design Cycle



Source: Vaishnavi and Kuechler, (2007)

These models support reflective iteration, where insights from evaluation and demonstration stages loop back into refinement cycles. Furthermore, evaluation must be appropriate to the context, taking the form of laboratory experiments, field trials, expert assessments, or analytical simulations, depending on the maturity and purpose of the artifact (Hevner et al., 2024b).

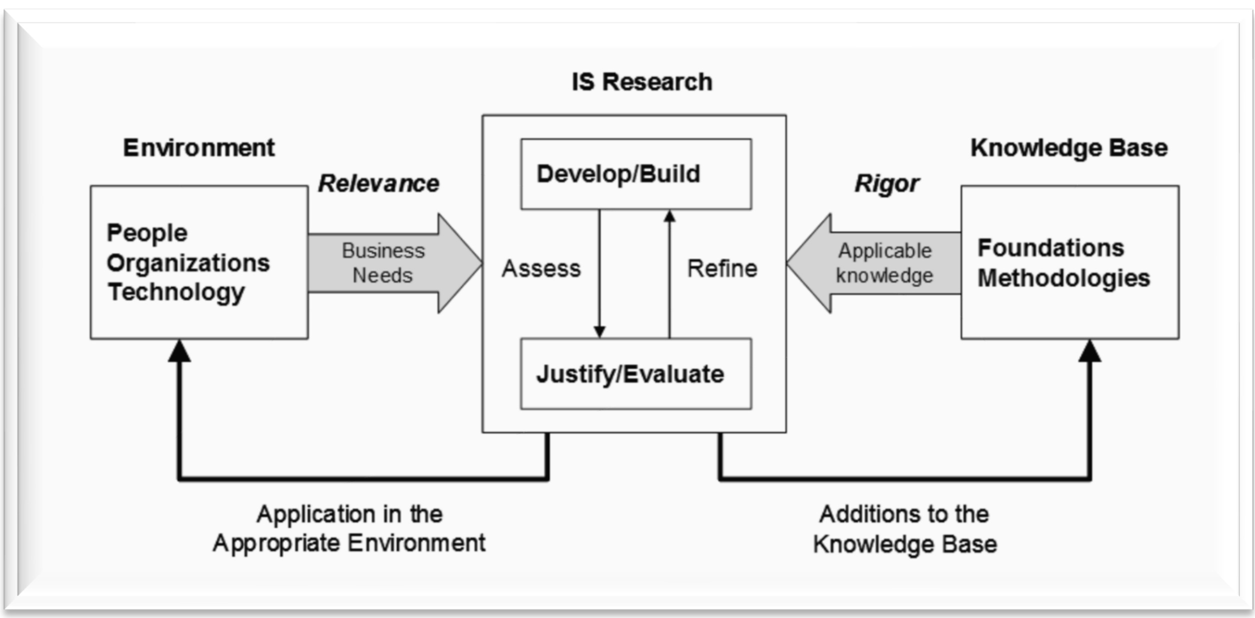
3.1.3. The Three Cycles of DSR

DSR is not isolated but embedded within broader knowledge development and problem-solving processes. It integrates knowledge from reference disciplines and theoretical foundations, responds to contextual needs and requirements, and advances through iterative cycles of design and evaluation. This structured and recursive nature of DSR has led to the establishment of three distinct but interrelated research cycles. Hevner et al. (2004) elaborated the operational model in terms of three interacting cycles (relevance, rigor, and design, that guide DSR activities (Hevner, 2007) (Figure 8):

- The Relevance Cycle connects the research project to the problem/application environment. It brings in the real-world requirements, use cases, and contextual constraints. This cycle assures that the developed artifact addresses a relevant requirement in real-world situations. The cycle includes preliminary problem framing through stakeholder engagement or empirical exploration of existing issues.
- The Design Cycle: This cycle is the core cycle of DSR, which involves the creation and iterative refinement of the artifact, following steps like modeling, development, demonstration, and formative evaluation. Feedback from testing and stakeholder engagement is used to clarify the artifact and its underlying assumptions. This cycle integrates knowledge from the rigor cycle and constraints from the relevance cycle to develop a feasible and valuable artifact.

Figure 8

Design Science Research Cycles



Source: Hevner et al. (2004)

- The Rigor Cycle: This cycle connects the research to the existing knowledge base (Dahanayake & Thalheim, 2011) It involves grounding the research in established theories, models, methods, and empirical findings. Moreover, it supports the theoretical justification of design decisions and enables contributions back to the academic community through extended theories and generalized design principles.

The rigor cycle ensures that the artifact development is not ad hoc but informed by prior knowledge.

These cycles are interdependent and cyclical, enabling DSR to function as both a design method and a knowledge-generating process. They allow researchers to manage the balance between novel solution development and scientific contribution, facilitating a robust approach to solving ill-structured or emerging problems. In sum, researchers can ensure their work's practical impact and scientific contribution by using the three-cycle of DSR and integrating conceptual modeling activities. In this thesis, these cycles structure the activities related to understanding the problem, designing and refining artifacts, and validating outcomes through empirical tests.

3.2. Research Design

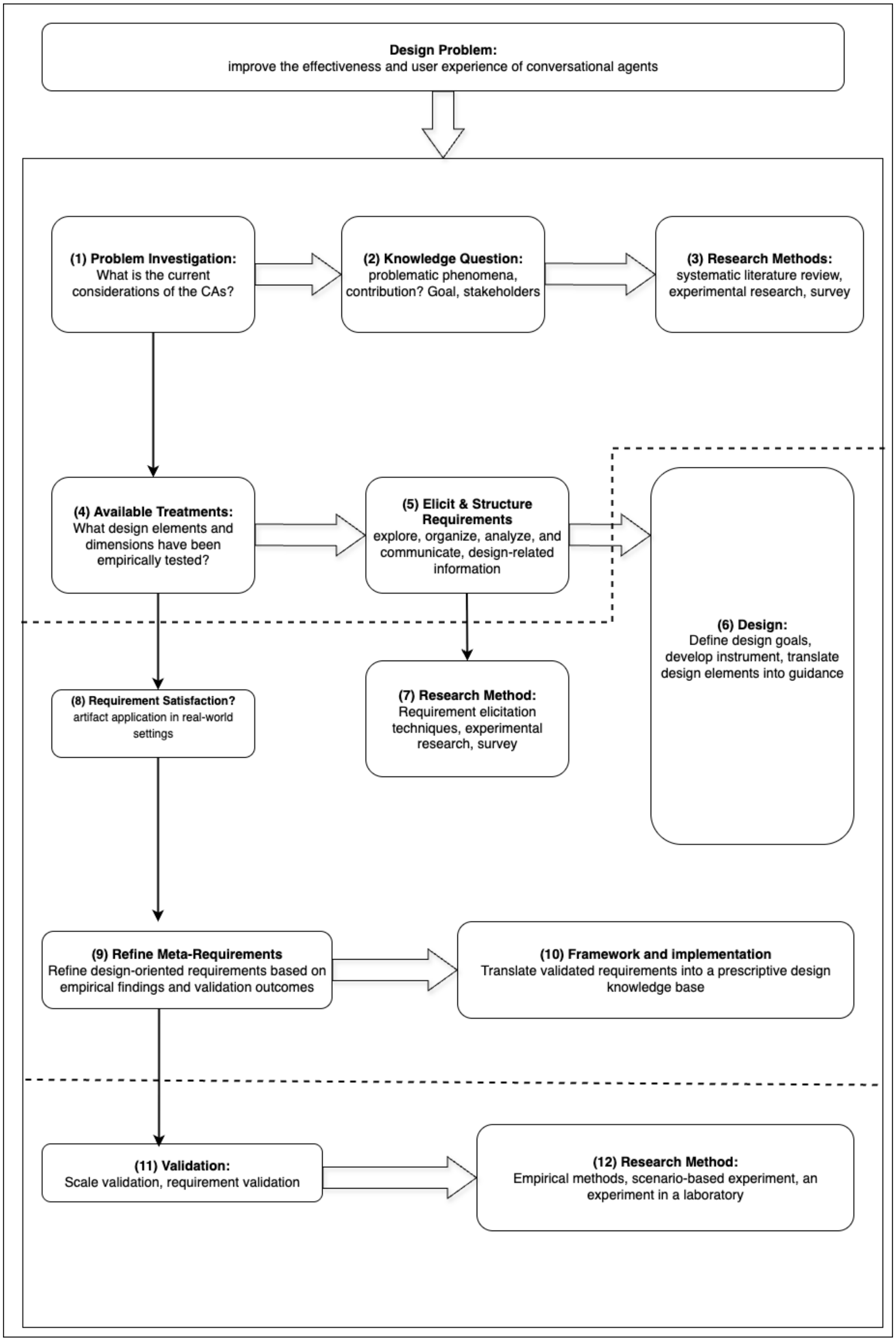
To address the research goal through a deeper understanding of user experience, this thesis adopts a DSR approach conceptualized by Hevner et al. (2004). The three-cycle model provides a structured foundation for the iterative development and evaluation of artifacts (Hevner, 2007). As outlined in Section 3.1, this approach is particularly suitable for guiding the design of human-centered systems in complex environments, such as conversational agents (Diederich et al., 2020a; Gregor & Hevner, 2013).

DSR enables the creation of the artifacts while facilitating theoretical contributions through reflective evaluation and generalization (Wieringa, 2014). In the context of HCI and IS, DSR effectively leads the design and iterative evaluation of artifacts that support user experience (Baskerville et al., 2018; Peffers et al., 2007). This is especially important for CA systems, as design choices frequently have immediate and complex consequences for interface quality, user trust, and adoption (Følstad & Brandtzaeg, 2020; Luger & Sellen, 2016). Moreover, DSR enables the researcher to integrate empirical insights from user evaluations into the design cycle while grounding the work in prior theoretical knowledge (rigor), ensuring scientific robustness and practical relevance (Hevner, 2007). This is consistent with the dual objective of this research: (1) to improve real-world CA design and (2) to contribute to design theory within the IS discipline.

To structure the overall flow and documentation of the research process within the thesis, the methodological framework proposed by Wieringa (2014) was used (see Figure 9). Wieringa's structure complements the DSR paradigm by explicitly aligning research goals, problem investigation, artifact design, and evaluation within a coherent thesis structure. All research activities undertaken throughout this study are within the DSR paradigm and structured according to its core cycles described in Section 3.1. The subsequent sections outline the research process as a coherent sequence of activities. These include the investigation of the problem and elicitation of design-oriented requirements (Sections 3.3 and 3.4), the development and refinement of a theoretically grounded measurement scale (Section 3.5), and a series of empirical studies to evaluate user interaction and validate the measurement scale (Section 3.6). Each phase contributes to solving a practically relevant problem and generating design knowledge suitable for dissemination within the Information Systems discipline.

Figure 9

Research Design



Source: Adapted from Wieringa, (2014)

3.3. Development and Structuring Design-Oriented Requirements

This section explains how design-oriented requirements are conceptualized, structured, and delivered within the Design Science Research paradigm. The approach draws on Information Systems Design Theory (ISDT) to frame requirements as design-relevant knowledge, serving as a foundation for artifact construction and as a contribution to theory and practice.

3.3.1. Information Systems Design Theory as a Structuring Framework

In Information Systems research, the Systems Development Life Cycle (SDLC) has traditionally been a foundational framework for organizing system development activities. It structures the process into sequential or iterative phases, such as requirement definition, system design, development, implementation, and maintenance (Mantei & Teorey, 1989). Within this model, requirements are formally captured during the initial stages, often through stakeholder interviews, document analysis, and process modeling. These requirements are then used to guide subsequent technical design decisions (Aurum & Wohlin, 2005). For example, in enterprise systems research, the requirement phase often involves modeling business processes to identify system functionalities aligned with organizational workflows (Avison et al., 2006). In software engineering-focused studies, use cases or functional specifications are developed to express these requirements in actionable terms for developers (Dennis et al., 2014).

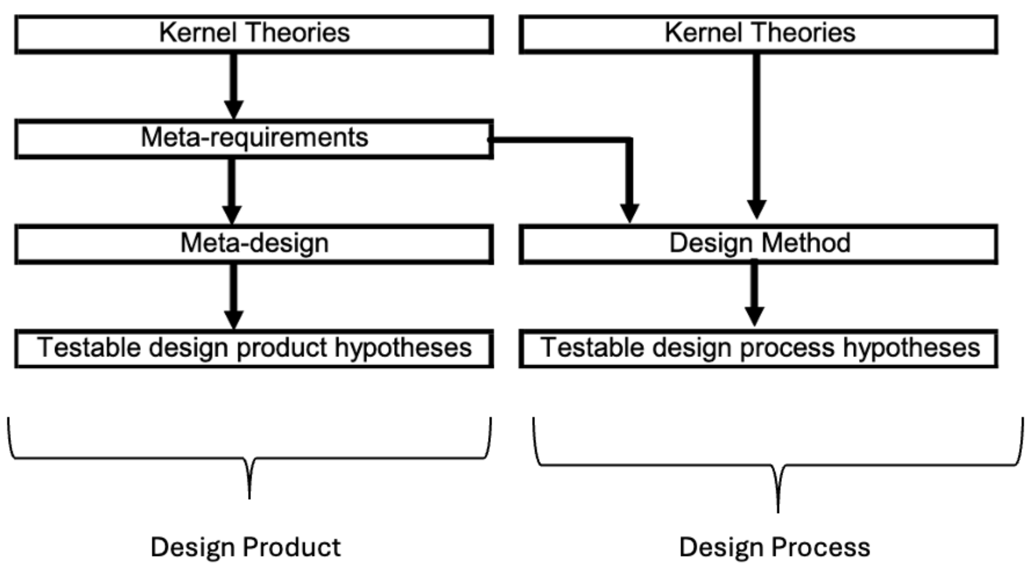
This structured, input-driven view of requirements has also influenced various extensions of the SDLC, including iterative models like the Spiral Model (Boehm, 1989)and agile-inspired hybrid frameworks, where requirement refinement remains central but is distributed across cycles (Dennis et al., 2014). Across these models, requirements are positioned as the basis for aligning stakeholder expectations, reducing design ambiguity, and controlling scope. They are often categorized as functional or non-functional, depending on whether they describe system behavior or quality attributes (Aurum & Wohlin, 2005).

In contrast to process-oriented development models, Information System Design Theory conceptualizes requirements as part of a broader theoretical framework that supports the creation of generalizable design knowledge. First introduced by Walls et al. (1992)and further elaborated by Gregor and Jones (2007), ISDT introduces the **meta-requirements**, a high-level abstraction to describe a class of system goals. The meta requirements are formulated more broadly and reflect the essential needs, functions, or outcomes that the system should fulfill.

This abstraction has proven useful across domains. For example, Buyssens et al. (2024) identified common system-level goals as meta-requirements supporting adaptive, user-sensitive interventions across various health and wellness contexts. Wessel et al. (2025) framed meta-requirements as recurring practices that any smart service platform should enable. Similarly, in the domain of AI governance, ISDT guided the articulation of high-level governance objectives that reflect institutional needs and regulatory expectations, making the resulting framework usable across organizations (Gupta et al., 2023). These examples show ISDT supports the translation of complex design challenges into structured, theory-informed requirements. Meta-requirements serve as a bridge between real-world needs and theoretical design principles, helping researchers produce design knowledge that is both rigorous and broadly applicable (Walls et al., 2004).

Figure 10

ISDT Components



Source: Walls et al. (2004)

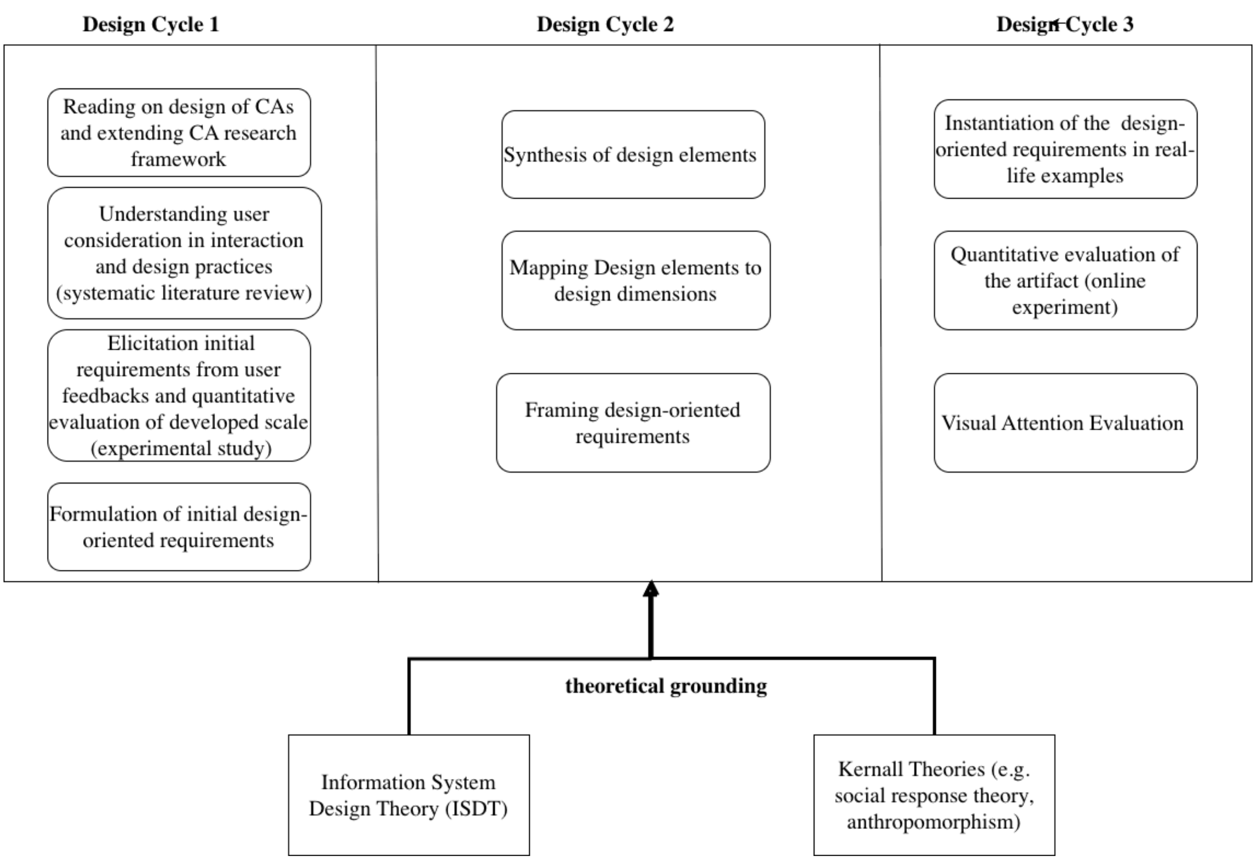
ISDT further distinguishes between the design as a product or process, and defines components of design knowledge accordingly (Walls et al., 2004) (Figure 10). These components are informed by kernel theories, which provide the theoretical grounding for why certain design features are expected to be effective (Jones & Gregor, 2007).

3.3.2. Providing Requirements as Design Knowledge Output

Generating actionable knowledge through designing and evaluating purposeful artifacts lies at the core of the DSR paradigm (Hevner, 2007; Hevner et al., 2024b). DSR emphasizes constructing novel solutions to practical problems and formulating theoretical contributions through artifact evaluation and abstraction (Gregor & Hevner, 2013). Within this paradigm, ISDT provides a conceptual structure for translating complex design challenges into generalizable knowledge by articulating meta-requirements, grounded in kernel theories, and formulating meta-requirements and testable design hypotheses (Walls et al., 1992).

Figure 11

Design Cycles and Related Research Activities



Source: Created by the author.

This dissertation adopts that orientation by developing design-oriented requirements for conversational agents that enhance user experience. These requirements, grounded in empirical and theoretical insights, align with the meta-requirements in ISDT. Rather than developing a single artifact or static solution, the study systematically investigates existing design practices and user experiences to identify which design practices contribute to critical outcomes. The resulting requirements are presented as design knowledge outputs that can inform the design of future CAs across various domains. As a part of artifact development in DSR, the dissertation conducted three design cycles (Figure 11), which have interdependent activities with relevance and rigor cycles. The first cycle

After reading CA literature in the CA research framework, one deeply understands conversational agent design. This extends the research framework, adding a new design dimension and sub-dimensions. Then, all research was structured around an extended research framework. Afterwards, a systematic literature review of leading conferences was conducted to extract initial meta-requirements. Specifically, the study focused on user experience outcomes and design choices, which provided an in-depth understanding of user-centered requirements and identified eight initial meta-requirements from the literature. Meanwhile, the reading and reviewing literature has yielded the primary user evaluation outcomes to measure user experience, which also aids in developing the CA user experience scale for the post-acceptance process. Then, an empirical study was conducted to elicit initial requirements. Specifically, 87 participants interacted with a chatbot and were asked open-ended questions about desired features and improvements. New meta-requirements were added iteratively whenever participants proposed previously unmentioned needs, which resulted in three more meta-requirements. The broader CA design literature was revisited to improve and consolidate the meta-requirement set, ensuring that both empirical user input and established design knowledge were integrated.

In the second design cycle, an actionable design element list was presented to guide designers and analyzed to provide user experience effects. After examining the elements, they are mapped to the associated design dimensions, which are categorized according to design-oriented requirements. Consequently, the third design cycle involved instantiating the meta-requirement in real-world CA examples. Furthermore, users participated in several online experiments evaluating 6 different chatbot versions. This phase led to the validation of 14 meta-requirements based on users' evaluations and qualitative feedback. Lastly, an eye-tracking study was conducted to validate these requirements by analyzing users' visual attention and interaction patterns during task completion with the chatbot. This triangulation of methods ensured that the derived requirements were grounded in user preferences and supported by behavioral evidence.

3.4. Problem Investigation and Meta-Requirement Elicitation

This section presents the initial research activities undertaken to define the problem space, identify gaps in current knowledge, and establish the theoretical and empirical foundations of the study. These activities align with the relevance cycle of the DSR framework by capturing practical needs, user expectations, and contextual design challenges to ensure that the resulting artifacts address a meaningful real-world problem. In parallel, the section informs the design cycle by supporting the early development of artifacts based on identified requirements. It also contributes to the rigor cycle by grounding the research in existing theories and prior empirical findings. The investigation includes insights from a systematic literature review.

3.4.1. Research Problem

The research problem addressed in this thesis concerns the lack of structured and theoretically grounded design requirements for developing conversational agents that improve user experiences in real-world contexts. Although CAs have become increasingly embedded in everyday digital interactions (e.g., from customer service to education), their design often remains ad hoc and driven by functionality rather than user-centered considerations. Prior studies reported persistent issues in CA interaction, including misinterpretations of user inquiries, lack of personalization, repetitive dialogues, and limited contextual understanding (Følstad & Brandtzaeg, 2020; Jo et al., 2023). While some design practices have shown promise (Chandra et al., 2022; Li & Suh, 2021) to improve user experience, a lack of actionable design guidance links such features to measurable user experience outcomes.

From an IS and HCI perspective, this constitutes a relevant design challenge: systematically identifying, validating, and integrating design practices that improve user experience with

CAs. The problem is increased by the growing diversity of agent forms (text-based, voice-based, embodied, etc.) and the rise of large language model-driven systems, which introduce new expectations regarding responsiveness, transparency, and support for critical reflection (Skjuve et al., 2024). Added to that, existing development practices rarely reflect the post-adoption user experience evaluations (e.g, trust and perceived usability), which are crucial for sustained use and long-term engagement (Bhattacherjee, 2001; Borsci et al., 2022a; Hsiao & Chen, 2021; Seeger & Heinzl, 2021). This results in a gap between the growing empirical understanding of what users value in CAs and the design decisions made during agent development.

The following research questions were formulated to reveal CA design requirements and actionable design guidelines to address this gap. The following questions address a specific facet of the design issue. They are linked to relevant research aims and goals, offering a structured pathway for studying various characteristics of CA design and user experience. These are shown in Table 4.

Table 4

Thesis Research Questions, Aims, and Goals

Research Questions (RQ)	Aims and Goals
RQ1. What are the key user considerations when interacting with conversational agents?	**Aim:** To explore which experiential and perceptual factors most influence overall user experience. **Goals:** G1: Conduct a systematic literature review to explore key determinants of user experience. G2: Derive initial user experience requirements based on common expectations and needs.
RQ2: What are the design practices used to improve user experience, and which elements are most prominent in CA design?	**Aim:** To identify actionable design elements from existing CA practices that can be linked to the elicited user experience requirements. **Goals:** G3: Code and classify design elements used in current CA studies. G4: Map design elements to user experience requirements based on theoretical and empirical insights.
RQ3: What meta-requirements in CA design can be identified to enhance user experience?	**Aim:** To formulate meta-requirements that consolidate findings from user considerations (RQ1) and design practices (RQ2). **Goals:** G5: Synthesize insights from RQ1 and RQ2 into higher-level meta-requirements. G6: Articulate how these meta-requirements can inform CA design decisions and guide future implementations.

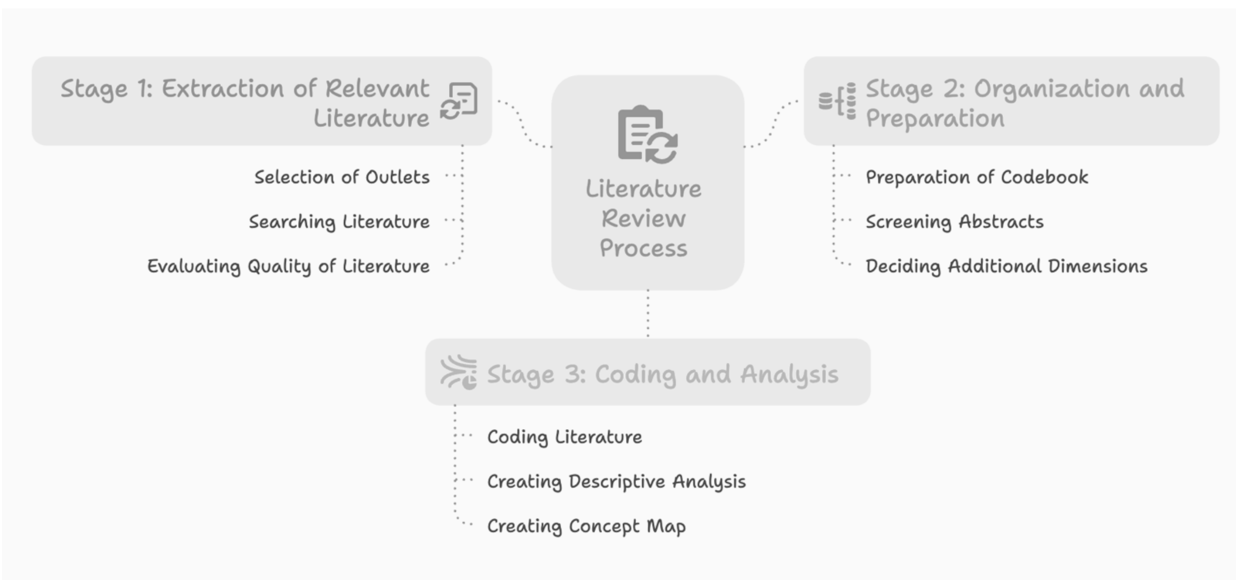
Source: Created by the author.

3.4.2. Systematic Literature Review of Conversational Agent Design and User Evaluations

An SLR was conducted to identify user considerations and design choices in CA systems to derive empirically grounded design requirements for CA development. This review is a formalized process of synthesizing existing knowledge through a transparent, replicable, and structured approach that includes locating, screening, analyzing, and synthesizing primary studies (B. A. Kitchenham et al., 2010; Tranfield et al., 2003). This methodology has become increasingly valued in IS and HCI disciplines, particularly in emerging and interdisciplinary domains like CA design, where research is dispersed across various venues and terminologies (Elshan et al., 2022; Feine et al., 2019; Guan et al., 2025; Khosrawi-Rad et al., 2022; Ling et al., 2021; Mariani et al., 2023; G. R. S. Silva & Canedo, 2024).

Figure 12

Stages of Systematic Literature Review



Source: Adapted from Bandara et al. (2015)

Further, the systematic literature review approach reduces bias by employing well-defined methods for literature selection, coding, and synthesis, while traditional reviews rely on subjective interpretation (B. Kitchenham, 2004). Building on this, SLR is especially appropriate for the current study, since the CA design has diverse conceptual framings and application contexts that influence design choices Moreover, SLRs are increasingly used in exploratory design research (Morashti et al., 2022) to address conceptual fragmentation and map emerging design practices for subsequent artifact development.

In this thesis, the SLR followed a three-stage structured approach adapted from Bandara et al. (2015), who propose a qualitative and systematic technique to extract, organize, and synthesize literature (Figure 12). This approach has been widely applied in IS and HCI studies (Diederich et al., 2022; Schoormann et al., 2023), which is especially suitable for emerging interdisciplinary fields like CA design, where design practices and theoretical constructs are still evolving.

3.4.2.1. Extraction of Relevant Literature

The first stage of the systematic literature review aimed to extract high-quality and relevant empirical studies examining conversational agents' design and their relationship to user experience outcomes. This stage is also organized around three sub-stages: (1) selecting high-quality publication outlets, (2) conducting literature searches using systematic keyword strategies, and (3) evaluating studies using defined inclusion and exclusion criteria.

The careful (1) selection of publication outlets is a foundational step in any rigorous systematic literature review, as it directly influences the quality, reliability, and relevance of the included studies (Tranfield et al., 2003; Webster & Watson, 2002). For this reason, selecting high-quality, peer-reviewed outlets is essential to ensure that the review builds on a robust and validated knowledge base. In HCI and IS, conference proceedings often represent state-of-the-art empirical work, especially in early-stage or exploratory design research (Fisher et al., 2007). Moreover, many highly cited frameworks and foundational models in user experience research have emerged from top-tier conferences rather than traditional journals (Levy & J. Ellis, 2006).

This review ensures methodological rigor and practical relevance by strategically targeting well-established sources (Bandara et al., 2015). To address this, the review included only peer-reviewed journal articles and conference proceedings reported as high-quality sources (Diederich et al., 2022; Levy & J. Ellis, 2006). Thus, the following databases were selected as primary sources: Web of Science (for its comprehensive indexing of multidisciplinary and high-impact journals), AISeL (Association for Information Systems eLibrary) (which archives top IS conferences and journals), and ACM Digital Library (widely used in HCI and computing disciplines, including leading conferences).

Within these databases, the current review focused on empirical studies published in top-ranked IS and HCI journals and conferences, based on the outlet selection criteria proposed by Levy and Ellis (2006) and implemented in prior SLRs in the CA domain. In particular, conferences such as the ACM CHI Conference on Human Factors in Computing Systems, International Conference on Information Systems (ICIS), European Conference on Information Systems (ECIS), and Pacific Asia Conference on Information Systems (PACIS) were prioritized. These venues are well-established platforms for research on user experience, interface design, and digital interaction, and they frequently feature early-stage empirical design work (Fisher et al., 2007; Levy & J. Ellis, 2006).

After identifying relevant publication outlets, the next step involved (2) a literature search that examined the design of CAs and their effects on user experience. A keyword-based query was developed to retrieve studies from selected databases. The following Boolean query was employed to capture a wide range of terminologies in CA-related research across the IS and HCI domains:

(((Conversational OR Virtual OR Interactive OR Intelligent) AND Agent) OR Chatbot OR "Digital Assistant") AND Design

This query syntax was applied to titles, abstracts, and keywords in the selected databases (Web of Science, AISeL, and ACM Digital Library). The inclusion of synonyms of CA (e.g., virtual agent, chatbot, and digital assistant) was informed by prior reviews and taxonomies (Diederich et al., 2022; Rheu et al., 2021). The database search was conducted in January 2024. No restriction was applied to the publication year to ensure historical comprehensiveness. However, the screening phase (described below) retained only empirically grounded and design-relevant studies. The raw search process yielded 1,284 studies across the identified outlets. The results of this initial retrieval process and the number of studies included after full-text screening are presented in Table 5.

To ensure the review's methodological rigor ((3) evaluating quality of literature), the retrieved publications were screened in multiple stages using predetermined inclusion and exclusion criteria, identified in Table 6 (Bandara et al., 2015; Webster & Watson, 2002). After the initial database retrieval, a two-step screening was conducted: (1) abstract and title review and (2) full-text evaluation.

Table 5

Included Outlets and Papers Details

Outlet	**Initial Search Results**	**Included Papers**
ACM CHI Conference on Human Factors in Computing Systems	317	18
European Conference on Information Systems	26	5
Americas Conference on Information Systems	43	1
International Conference on Information Systems	40	9
Pacific Asia Conference on Information Systems	22	0
ACM Transactions on Computer-Human Interaction	175	0
AIS Transactions on Human-Computer Interaction	6	1
ACM Transactions on Information Systems	73	0
ACM Transactions on Human-Robot Interaction	5	0
ACM Transactions on Intelligent Systems and Technology	2	0
ACM Transactions on Interactive Intelligent Systems	66	4
Computers in Human Behavior	154	16
European Journal of Information Systems	6	0
Human-Computer Interaction	8	0
Information & Management	13	1
International Journal of Human-Computer Interaction	54	20
International Journal of Human-Robot Interaction	35	0
International Journal of Human-Computer Studies	125	12
Journal of the Association for Information Systems	6	2
Journal of Management Information Systems	7	4
MIS Quarterly	73	0
After Backward-Forward Search	**14**	**4**
Total	**1284**	**97**

Source: Created by the author.

Table 6

Inclusion and Exclusion Criteria for Study Selection

Criteria Type	Description
Inclusion	Empirical study involving conversational agents
	Examines specific design elements or interaction features
	Measures user outcomes (e.g., trust, perceived usefulness, social presence)
	Published in peer reviewed IS or HCI outlets
	English language
Exclusion	Focuses on backend development or system architecture without user interaction
	Theoretical/conceptual paper without empirical validation
	AI, VR, or robotics studies without explicit CA or user interaction focus
	Papers with no reference to CA design or user-centered evaluation
	Papers focus on user general perceptions of CAs

Source: Created by the author.

After applying the inclusion and exclusion criteria during the abstract screening phase, 264 studies were eligible for full-text review. These studies met all criteria concerning design relevance, empirical methodology, and publication outlet quality. As part of the quality appraisal, citation counts were used as a supporting indicator of scholarly impact, with the highest citation count reaching 969 and the average citation count across the sample being 69. This metric further ensured the relevance and recognition of the selected works within the academic community.

Figure 13

Word Combination Cloud from Full Texts

Source: Created by the author.

To verify the relevance of the selected papers, we conducted a word frequency and word combination analysis on the full texts of the 264 documents using MaxQDA24 after importing them. The most frequently mentioned terms were 'user' (17172 occurrences), 'agent' (14887), 'social' (9347), 'interaction' (8963), 'design' (8730), 'human' (7185), 'chatbot' (6914), 'system' (6476), and 'information' (5671). The most frequent word combinations included 'conversational agent' (2593 instances), 'information system' (2300), 'social presence' (1734), 'virtual agent' (1691), and 'social cue' (1201) (Figure 13). These findings highlight that the paper selection effectively focused on agent interaction, with design being a central theme in IS and HCI research. After completing the full-text evaluation, a total of 97 studies were ultimately included for in-depth coding and synthesis. The complete screening and selection process is illustrated in the PRISMA diagram (see Figure 15).

3.4.2.2. Organization and Preparation

To systematically derive design requirements for CAs, this study employed a staged review process grounded in a tailored research framework (Figure 14). In the initial phase, a baseline conceptual framework was established by adapting and extending the sociotechnical lens proposed by Diederich et al. (2022). This framework was the analytical foundation for identifying relevant design constructs and their relationships to user experience.

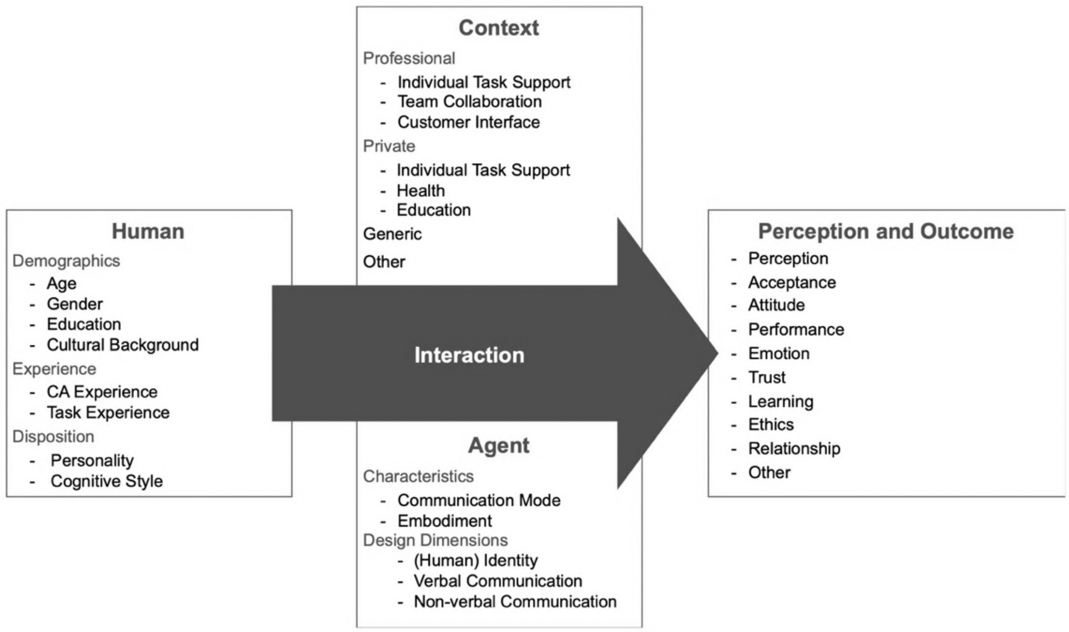
The original framework categorized constructs into Human, Agent, Context, and Outcomes, with interaction as a central mediating process. The framework was initially developed by Zhang & Li (2005) and applies a sociotechnical lens to the analysis of CA research. In the framework, the context dimension considers the environment in which the CA is used. Human dimension refers to user characteristics such as demographic aspects. Agent dimension encompasses how CAs can be designed to fulfill their intended purpose. Lastly, the perception and outcome dimension focuses on the investigated topic regarding the use and impact of technology, showing how users perceive a CA during an interaction and the interaction's impact.

Firstly, we analyzed empirical studies from leading IS and HCI conferences, which provided early insights into the design elements operationalized in real-world CA implementations.

These studies often reported the design choices made (e.g., inclusion of visual identity cues or use of conversational scripts) and their impact on user perceptions. This analysis served

Figure 14

Conversational Agent Research Framework



Source: Diederich et al. (2022)

as the basis for the first version of the codebook, capturing a preliminary list of design elements and their evaluative context. As a result, we expanded the scope of design-related attributes into anthropomorphic design dimensions, drawing from Feine et al. (2019). These include verbal, non-verbal, and identity cues, which capture how the agent mimics human communicative behavior. In addition, we incorporated a broader set of agent characteristics, which describe the core interactional and expressive capabilities of a CA. These encompass the communication mode (text-based, voice-based, or multimodal) (Chattaraman et al., 2019; Rapp et al., 2021), the embodiment of the agent (e.g., virtual vs. physical representations, such as service robots) (Biocca, 1997; Thaler et al., 2020), the expressiveness through which agents convey affect or emotional cues (Al-Natour et al., 2022), and the agent's transparency, referring to its ability to reveal system identity or purpose (Diederich et al., 2020a). In addition to these agent design features, we extended the framework by introducing an agent competency dimension. Agent competency refers to an agent's functional capacity to facilitate interaction. Agent competency reflects the agent's ability to accurately interpret user input, maintain conversational flow, respond to emotional or contextual cues, and recover from failures or misunderstandings (K. M. Lee et al., 2006; Shevat, 2017; Wienrich & Carolus, 2021). It is operationalized through responsiveness, error-handling mechanisms, and the ability to provide clear explanations for CA responses and actions (Chandra et al., 2022).

Building on this, an iterative process of (4) codebook development and (5) abstract screening phases was conducted in parallel. The process continued by deciding conference papers for full-text review based on inclusion and exclusion criteria (extracted in stage 1), followed by reviewing full texts to extract design elements used in the papers and empirically evaluated (Diederich et al., 2022). The resulting codebook included a well-defined set of design elements and a list of user evaluation constructs. This structured codebook served as the analytical foundation for the next review stage (see Appendix A for Codebook). Following that, (5) an abstract screening phase for all papers was performed to assess the applicability of the codebook to a broader sample of studies. This process was carried out by a single researcher using Excel, ensuring consistent application of codes across the dataset. Codes were assigned to studies based on explicitly mentioning design-related elements or evaluation constructs outlined in the codebook. When design dimensions were observed in the abstracts but were not yet covered in the codebook, these were labeled under the category "other," marking them for subsequent review.

Drawing on Bandara et al.'s (2015) and Mayring's (2014) methodological recommendations, 277 studies were retained after abstract screening. However, before transitioning to the full-text coding phase in MaxQDA24, the documents were verified to ensure that the PDF content of these studies was text-readable and thus analyzable. Following this technical check, the final dataset was reduced to 264 studies. While the core of our review focused on design elements and their evaluative context, we extended the codebook to include (6) additional descriptive dimensions, such as

- Research Method (e.g., experimental, design science, focus groups, dimensions),
- Application Context (e.g., education, healthcare, general-purpose),
- Sample Characteristics (e.g., age, gender, experience)

3.4.2.3. Coding and Analysis

A systematic review aims to synthesize what is already known and identify meaningful literature gaps (Levy & J. Ellis, 2006). In line with this goal, the third stage of our review focused on analyzing how CA design elements are operationalized across empirical studies and how these design decisions relate to user evaluation outcomes. To achieve this, we employed a deductive (7) coding approach, guided by the structured codebook developed in the previous stages (Bandara et al., 2015) (see Appendix A).

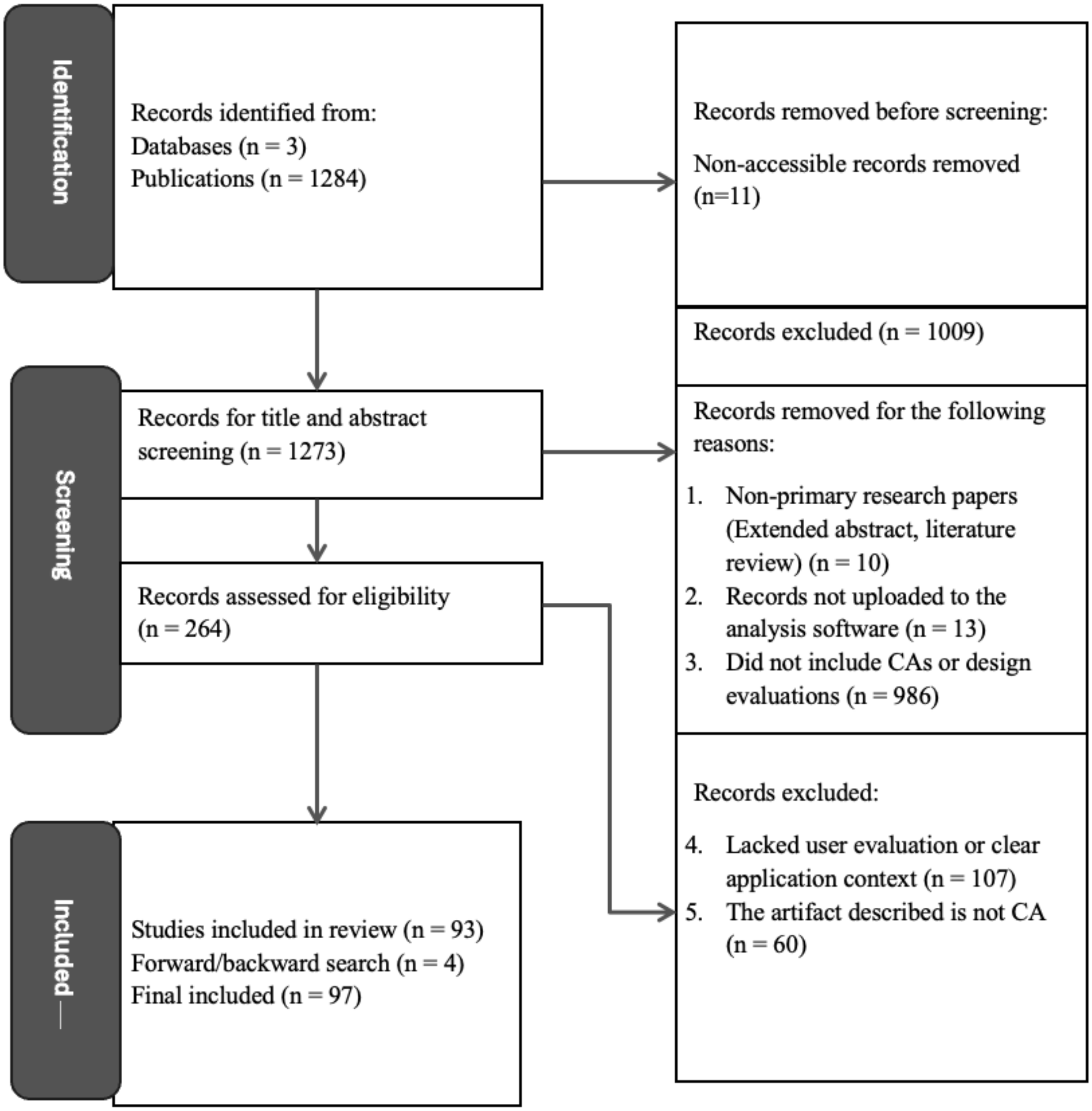
For each category, nodes were created within MaxQDA, and relevant text passages were linked to one or multiple nodes based on content. When studies referred to overarching patterns (e.g., identity cues) but did not specify which design elements were used, we applied the label “unrelated” to denote a lack of implementation detail. In contrast, reviewed studies that specified design features, such as agent gender, voice, or static avatar images, were coded with corresponding subcategories under design elements. Moreover, studies that explored prior user experience without engaging directly with agent design features were also classified as “unrelated.” Finally, studies that did not match the core analytical focus, i.e., did not empirically examine the relationship between design and perception, were marked as “excluded.”

After completing the coding phase, 97 studies were retained as meeting the review’s analytic criteria (Figure 15). Regarding inter-coder reliability, a second coder was trained on the codebook. Following the training, the researcher coded a random sample of 20% of the studies. The agreement rate was calculated for 20% of all studies, resulting in 91.46%. Inconsistencies were discussed collaboratively, and the primary coder revisited these studies to ensure accuracy in coding decisions (Leavy, 2020; Saldana, 2014). The analytical process proceeded in a structured manner.

- In the second step, (8) descriptive analyses were performed across all 97 included studies. This analysis involved: To track evolution trends and provide a summary of study characteristics, including research methods and application domains (e.g., education, healthcare, general-purpose), the frequency distribution of design dimensions and elements was analyzed by publication year.
- Cross-tabulation analysis was conducted.

Figure 15

PRISMA Diagram of SLR



Source: Created by the author.

- Study context (e.g., education, customer service, health),
- Moreover, the focus of design dimensions/elements reveals where specific design strategies are most commonly deployed.

In the third step, (9) concept maps were created to visualize how design characteristics and user evaluation constructs relate. This included:

- o Mapping relationships between design dimensions (e.g., expressiveness, embodiment, communication mode) and specific design elements (e.g., voice tone, avatar, visual feedback),
- o Visualizing the relationships between design dimensions and user evaluation outcomes, categorized by positive, negative, or null effects, as reported in empirical studies,
- o Identifying clusters of design elements under each design dimension consistently associated with positive user outcomes.

3.5. Scale Development

This section explains the theoretical background and practical steps to develop a scale for evaluating user experience after interacting with conversational agents. It introduces the overall user experience model designed to guide CA evaluation. The scale development process is operationalized in two stages. Stage 1, discussed in this section, includes the development process, conceptual grounding, item generation, and initial empirical analysis using exploratory methods. Stage 2, described later in Section 3.5, involves confirmatory testing and validation of the revised scale.

3.5.1. Development Overall User Experience Instrument for CAs

In IS research, capturing user perceptions and behaviors requires a robust instrument, and the development of these instruments is a crucial step. Several well-established approaches to scale development (e.g., (DeVellis, 2016; MacKenzie et al., 2011). Those approaches reflect distinct methodological assumptions. Two widely cited foundational methods are Churchill's (1979) classical psychometric paradigm and DeVellis's (2016) reflective-focused model. In Churchill's (1979) paradigm, construct purification via item reduction and factor analysis was essential for scale development. DeVellis's (2016) model provides a practical guide focused on reflective constructs, scale reliability, and unidimensionality. These approaches assume that the construct under investigation is reflected in its indicators, an assumption aligned with classical test theory and operationalized through psychometric techniques such as exploratory factor analysis, internal consistency testing (e.g., Cronbach's alpha), and construct purification based on item intercorrelations.

These approaches remain dominant among scholars. However, their narrow focus is criticized. For example, MacKenzie et al. (2011) argue that conventional practices may lead researchers to eliminate theoretically meaningful items in pursuit of statistical fit and compromise conceptual integrity. Further, many constructs traditionally modeled reflectively, but they may be better represented using formative indicators, where the indicators define rather than reflect the construct (Diamantopoulos & Winklhofer, 2001). The reason for this misalignment can be measurement error, conceptual ambiguity, or inadequate theorization. DeVellis (2016) explained that when researchers fail to distinguish between "cause indicators" (formative) and "effect indicators" (reflective), it can lead to potential misspecification. Petter et al. (2007) have also revealed that when empirical patterns (e.g., high inter-item correlations) suggest reflective measurement, such evidence can be misleading, especially when the theoretical relationship is formative.

Table 7

Scale Development Process

Phase	Focus Area	Purpose
Conceptual Grounding (3.4.2)	Framing the Theoretical Space	Establish a strong conceptual basis grounded in HCI and IS theories Ensure theoretical alignment between constructs and their intended measurement
Item Development (3.4.3)	Designing Measurement Items	Adapt and synthesize validated items from prior literature Preserve construct validity while contextualizing items for CA interaction
Model Structuring (3.4.4)	Specifying Measurement Logic	Establish the theoretical rationale for measurement but allow empirical findings to inform structural adjustments
Empirical Exploration & Refinement (3.4.5 and 3.5)	Testing Structure with Real Users	Evaluate item behavior and uncover latent structure through real-world engagement; explore dimensional structure
Model Assessment (3.5)	Preparing for Confirmatory Evaluation	Ensure consistency between the theoretical model and empirical findings; validate the structure and test the predictive model in larger sample

Source: Created by the author.

Moreover, poor model fit in reflective models is often interpreted as a failure of item quality. This may signal an inappropriate modeling approach (Diamantopoulos & Winklhofer, 2001).

This concern is amplified by the dominance of structural equation modeling (SEM) software tools (e.g., LISREL, AMOS), which reinforce reflective modeling by default. In line with these arguments, this study acknowledges that scales must align with the theoretical structure of the assessed construct, whether reflective, formative, or potentially a hybrid. Rather than committing prematurely to a specific measurement model, developing the overall user experience instrument for CA in this research was guided by conceptual reasoning and empirical validation. This dissertation is adapted from Mackenzie et al's (2011) approach, while being aware of the considerations provided by scholars, the following process is summarized in Table 7.

3.5.2. Conceptual Grounding

The conceptual foundation of this scale is grounded in the Expectation Confirmation Model (ECM) (Bhattacherjee, 2001), a well-established post-adoption framework widely used to explain continuance intention in information systems. In parallel, construct selection was also informed by the SRL conducted as part of this dissertation. This review identified key user experience constructs frequently evaluated in CA research, providing empirical grounding for selecting and adapting the theoretical model.

ECM has been applied across various domains, including chatbots (Ashfag et al., 2020; Mehrolia et al., 2023), mobile messaging applications (Oghuma et al., 2015), digital assistants (Brill et al., 2019), and mobile social apps (Hsiao et al., 2016). According to the original ECM, users' continuance intention (CI) is primarily influenced by perceived usefulness (PU) and satisfaction (S), which are shaped by confirmation (C) (the extent to which initial expectations are met after system use) (Bhattacherjee, 2001). However, this study adapts and extends the ECM framework to reflect the unique characteristics of conversational agent interactions. In adapting the ECM to evaluate post-interaction experiences with CAs, this study omits the original constructs of satisfaction and confirmation, conceptual considerations and recent developments in CA evaluation literature support this decision. First, perceived usability is measured through the BOT Usability Scale (BUS-11). This scale is explicitly grounded in a definition of satisfaction. Borsci et al. (2022b) developed the BUS scale to operationalize usability in chatbot interactions and frame it within the ISO 9241-11 (ISO, 2018) definition of satisfaction. Specifically, satisfaction is "the extent to which the user's physical, cognitive, and emotional responses that result from using a system, product, or service meet the user's needs and expectations." In this context, usability is not merely a technical attribute but a reflection of the user's satisfaction with interactional quality, as well as confirming expectations. As such, BUS-11 offers a psychometrically validated instrument that captures satisfaction as an outcome of interaction, rather than an abstract attitudinal construct often seen in traditional ECM applications (Aktaş et al., 2025; Borsci & Schmettow, 2024).

Second, recent research in the context of conversational agents has demonstrated that confirmation and satisfaction are not always necessary components in post-adoption models. For instance, several contemporary studies have successfully predicted continuance intention without including these constructs, opting instead for more context-specific predictors such as trust, social presence, and hedonic value (e.g., (Jin & Youn, 2023; Zhang et al., 2023; Zhu et al., 2022)) (Table 3). This empirical trend reflects a broader recognition that alternative constructs may more accurately capture user experience in CA-specific contexts. In line with HCI research, we additionally incorporated constructs that are especially relevant for assessing post-use evaluations of CAs. These include social presence (the perceived interpersonal quality of the interaction) (Haugeland et al., 2022; Lee, 2004), trust (operationalized in two dimensions as goodwill-based and qualification-based trust) (Seeger et al., 2017), and perceived enjoyment (reflecting the hedonic quality of the interaction) (Davis et al., 1992). These additions reflect growing evidence that the CA experience is shaped by utility and relational, affective, and social cues. The adapted model retains perceived usefulness and continuance intention while integrating perceived usability as a central construct that substitutes for satisfaction and expectation confirmation. Including trust, perceived social presence, and enjoyment further contextualizes the model within contemporary understandings of human–agent interaction.

3.5.3. Item Development

For perceived usefulness and continuance intention, items were drawn directly from the original Expectation Confirmation Model (Bhattacherjee, 2001). Trust was measured using items adapted from McKnight et al. (2002) and Seeger et al. (2017), ensuring the inclusion of both qualification-based (competence) and goodwill-based (benevolence) dimensions.

Perceived enjoyment items were adapted from Davis et al. (1992) and Qiu and Benbasat (2008), capturing users' affective and hedonic responses during interaction. For social presence, items were based on Lee (2004)and Haugeland et al. (2022), emphasizing interpersonal engagement and mutual involvement. The complete set of BUS-11 items was retained due to their established psychometric validity. These items comprehensively capture core dimensions of perceived usability, including accessibility, responsiveness, functional interactive conversation, and privacy, while also reflecting the user's satisfaction with the interaction. In total, 36 items were compiled across the seven constructs (Table 8).

Table 8

Items Derived from Literature

Construct Name	Definition and Content	Items	Previous Research
Social Presence	It refers to a psychological state in which users perceive CAs as social beings, allowing them to form meaningful connections with these virtual entities. Referring to: The perception of the existence of other beings. feeling of being fully engaged and immersed in an experience being socially related.	I felt like I was engaged in an active dialogue with the chatbot. My interaction with the chatbot felt like a back-and-forth conversation. I felt as if the chatbot and I were involved in a mutual task. The chatbot was efficient in responding to my activities.	Lee, (2004), Gefen and Straub (2004), Laban and Araujo, (2020), Haugeland et al. (2022)
Chatbot Usability (BUS-11)	Users' assessment on the quality of interaction with CA. Dimensions: Accessibility Functional interactive conversation Responsiveness Privacy	The chatbot function was easily detectable (e.g., the possibility to modify the settings of the chatbot, make the avatar visible or not, etc.). It was easy to find the chatbot. Communicating with the chatbot was clear. The chatbot was able to keep track of context. The chatbot's responses were easy to understand. I found that the chatbot understands what I want and helps me achieve my goal. The chatbot gave me the appropriate amount of information. The chatbot only gave me the information I need. I felt like the chatbot's responses were accurate.	Borsci et al. (2022b), Borsci and Schmettow (2024)

Table Continued

		I believe the chatbot informs me of any possible privacy issues. My waiting time for a response from the chatbot was short	
Trust	It is a multidimensional concept that confirms expectations for future interactions and denotes users' belief in the competence, integrity, and benevolence of CAs. Referring to: fulfilling the expectations of the trustor qualification-based trustworthiness (competence) goodwill-based trustworthiness (benevolence, integrity)	GOODWILL TRUST: I believe that the chatbot would act in my best interest. If I required help, the chatbot would do its best to help me. The chatbot is interested in my well-being, not just its own. The chatbot is truthful in its dealings with me. I would characterize the chatbot as honest. The chatbot is sincere and genuine. The chatbot would keep its commitments. QUALIFICATION TRUST: The chatbot was competent and effective in providing its service. The chatbot performed its role of providing a service very well. Overall, the chatbot was capable and proficient.	(Shneiderman, 2000) Mcknight et al. (2002), Seeger et al. (2017), Seeger and Heinzl (2021)
Perceived Enjoyment	Users' perception of CAs as enjoyable in their own right. Assessment of the hedonic quality of CAs interactions Referring to: intrinsic motivation, pleasure	Interaction with the chatbot is enjoyable. Interaction with the chatbot is fun. Interaction with the chatbot is exciting. Interaction with the chatbot is interesting.	Davis et al., (1992), Koufaris (2002), Qui and Benbasat, (2008)
Perceived Usefulness	Users' perception of the benefits of CAs in their professional or personal lives. Referring to: productivity, effectiveness, overall job performance extrinsic motivation	Using the chatbot improves my performance in retrieving course-related information. Using the chatbot increases my productivity in retrieving course-related information Using the chatbot increases my effectiveness in accessing course materials. I find the chatbot useful accessing course-related information.	Davis, (1989), Venkatesh et al. (2000), Bhattacherjee (2001)

Table Continued			
Continuance Intention	Users' willingness or commitment to continue engaging behavior or action with CAs after its initial acceptance or usage. Referring to: Intention to maintain usage of CAs, willingness to continue using CAs in the future	I intend to continue using the chatbot rather than discontinue its use. If I could, I would like to discontinue my use of the chatbot. *(Reverse coded)* My intentions are to continue using the chatbot rather than seek alternative means.	Bhattacherjee (2001)

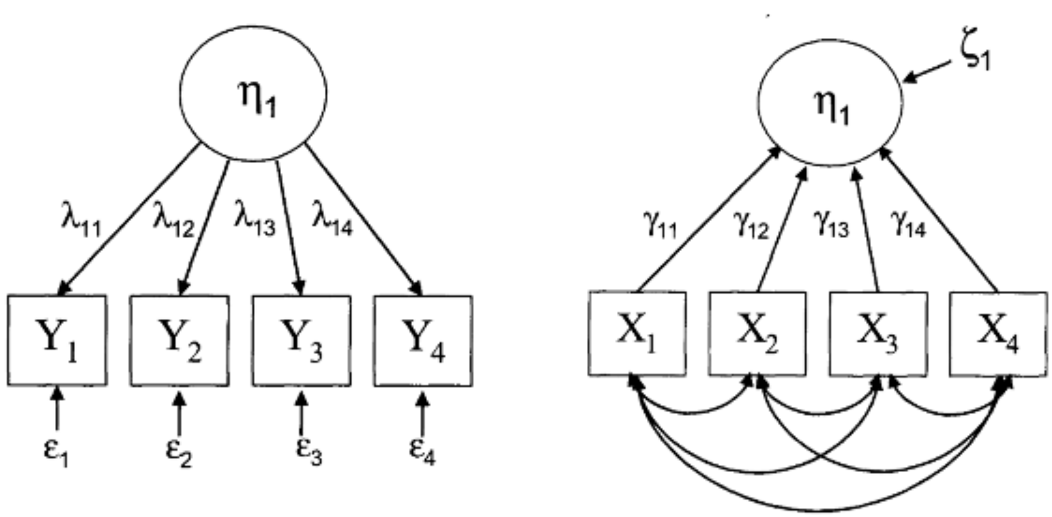
Source: Created by the author.

3.5.4. Model Structuring

This step defines the expected relationships between constructs and their associated indicators based on theoretical rationale. In line with Mackenzie et al. (2011), a theory-driven measurement model specification was initially established, carefully considering the conceptual nature of each construct before applying statistical tests.

Figure 16

Reflective and Formative Measurement Models

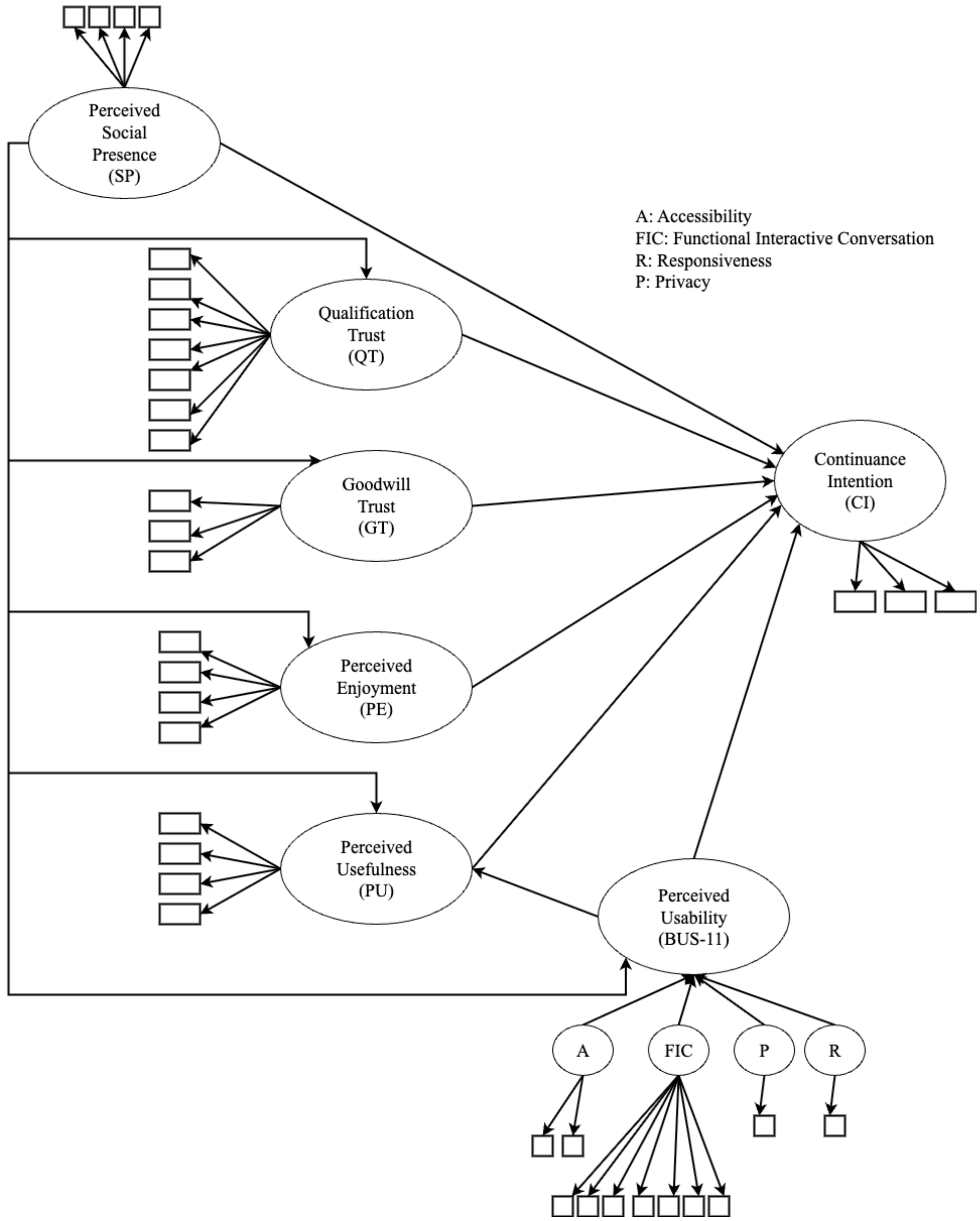


Source: Petter et al. (2007)

The distinction between constructs lies in the direction of causality between the latent variable and its indicators, as well as their theoretical purpose in the model (Diamantopoulos & Winklhofer, 2001; Petter et al., 2007). In reflective models, the latent construct causes the indicators. That is, changes in the underlying construct lead to changes in all observed items. Indicators are manifestations, so each item reflects the same underlying phenomenon.

Figure 17

Proposed Measurement Model



Source: Created by the author.

Indicators can be dropped without significantly altering the meaning of the construct. High internal consistency is expected (e.g., Cronbach's alpha, AVE). For instance, satisfaction, if a person is satisfied, they are likely to agree with various satisfaction-related items (e.g., "I am happy with the service," "The service met my expectations") (See Figure 16). The construct (η_1) influences all indicators (Y_1–Y_4), shown by arrows pointing from the construct to the items. This reflects a reflective measurement model. On the other hand, the indicators form or cause the construct for formative models. Each indicator represents a distinct dimension that contributes to the overall concept. Each item captures a unique part of the concept. Dropping an item may omit critical meaning. Collinearity between items should be examined but is not expected. For example, removing socioeconomic status, which is composed of income, education level, and occupation, would distort the meaning of the construct. The indicators (X_1–X_4) collectively form the construct (η_1), with arrows pointing into the construct (Becker et al., 2012; Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2011; Petter et al., 2007).

Figure 17 shows the hierarchical research model. According to the model, perceived usability was conceptualized as a second-order latent variable, composed of four first-order dimensions: Accessibility, Functional Interactive Conversation, Privacy, and Responsiveness (Borsci & Schmettow, 2024). These first-order constructs were each measured reflectively, and the higher-order construct was modeled using the repeated indicator approach to minimize potential bias in the parameter estimation of the higher-order structure (Becker et al., 2012). While alternative modeling approaches such as the two-stage method are available, the repeated indicator technique was preferred due to its straightforward implementation and ability to preserve measurement model consistency. All other constructs in the model (Perceived Social Presence, Perceived Usefulness, Perceived Enjoyment, Qualification Trust, Goodwill Trust, and Continuance Intention) were conceptualized as first-order reflective constructs, each connected to their respective indicators using reflective logic. No formative constructs were specified in the model, as each construct represents a latent psychological perception assumed to influence, rather than be composed by, its observed measures. This modeling strategy builds upon with the theoretical foundation that prioritizes internal consistency, convergent validity, and construct clarity (Jarvis et al., 2003; MacKenzie et al., 2011; Petter et al., 2007).

3.5.5. Empirical Exploration & Refinement

The fourth step of scale development is conducted in two stages. First, an experimental study was designed to validate and explore the overall user experience model developed in this research. In the second stage, presented in Section 3.6, the adjusted instrument was tested in a larger-scale online experiment to confirm its structure and psychometric properties. Moreover, the first empirical study was conducted to evaluate the theoretical model and item performance in a real interaction.

3.5.5.1. Study Design and Procedure

Initial scale exploration studies are often conducted as pilot studies, involving sample sizes ranging from 25 to 100 participants (Kunselman, 2024). In line with this guideline, preliminary data were collected to assess the proposed user experience model through a within-subjects experimental design. A total of 87 university students participated in the study between March 1 and June 4, 2025. Participation was voluntary and conducted by the ethical protocols approved by the University of Twente (application number: 250437). The study aimed to capture participants' immediate experiential responses after interaction with a chatbot.

Each participant engaged with the same set of tasks, which was constructed around three distinct information-retrieval tasks, but the order of presentation was randomized to mitigate potential order effects (Lazar et al., 2017; Nielsen, 1993). Participants completed all three tasks in each session, which took approximately 20 to 30 minutes. The study was conducted in a controlled experimental setting to ensure procedural consistency and to standardize the participant experience. Each participant was seated at a workstation equipped with two computers. The first computer was used to interact with the chatbot, and the other to complete post-task questionnaires via the Qualtrics platform. This configuration was purposefully designed to prevent copy-pasting chatbot responses into the survey and to encourage participants to articulate answers in their own words. A researcher was present throughout each session to assist with procedural or technical issues, without offering help regarding task content or influencing user responses.

The procedure followed a three-phase structure. In the first phase, participants completed a short pre-task survey capturing demographic information and their prior experience with chatbots. The second phase continues with the chatbot's interaction and completing the assigned tasks. After each interaction, they responded to task-specific follow-up questions (used to calculate task success rate) and rated the task's difficulty. In the final phase, participants completed a post-task survey evaluating their experience with the system using a set of items reflecting the proposed overall user experience model. The sample characteristics of individuals for pilot study are summarized in Table 9.

Table 9

Participants' Demographic Information

Profile of Respondents (N=87)	Distribution	Frequency	%
Gender	Female	38	43.67%
	Male	48	55.17%
	Not Prefer to Say	1	1.14%
Age	18-24	44	50.6%
	25-30	37	42.5%
	31-35	6	6.9%
Educational Background	Bachelor	41	47.2%
	Master	19	21.8%
	PhD	27	31.0%
Chatbot Experience	Yes	84	96.55%
	No	3	3.45%
Frequency	Once in a month	3	3.6%
	Twice or three times in a month	6	7.1%
	Once a week	9	10.7%
	Almost Daily	33	39.3%
	Daily	15	17.9%
	Couple of times a day	7	8.3%
	Continuously throughout the day	11	13.1%

Source: Created by the author.

As part of the requirement exploration, participants were asked an open-ended question at the end of the interaction session to indicate what features they would like to see in the chatbot. This step aimed to complement the scale-based evaluation with qualitative insights into user expectations, informing the subsequent formulation of design requirements.

3.5.5.2. Evaluated Conversational Agent Overview

The study was conducted using a new generation educational chatbot (called BuddyGPT) specifically developed to assist students in retrieving course-related information. The system was designed as a document-grounded chatbot, meaning it generated responses exclusively based on instructor-provided course materials (e.g., syllabi, lecture slides, and assignment documents). This ensured all responses remained context-specific, course-aware, and verifiable (Sedrakyan et al., 2024a).

Importantly, educational theories emphasizing self-regulated learning, cognitive engagement, and critical reflection informed the system's design. BuddyGPT encourages responsible AI use by embedding source transparency and contextual relevance by following recent design principles outlined by Sedrakyan et al. (2024b). The system delivers information to foster active engagement with course content and support the learning process (Sedrakyan et al., 2024a; Sedrakyan, et al., 2024b). Technically, the BuddyGPT was powered by the GPT-4 language model via the OpenAI API. All responses provided by BuddyGPT were filtered to ensure that the chatbot only provided answers that could be explicitly grounded in the uploaded documents.

3.5.5.3.Bayesian Exploratory Factor Analysis (BEFA) for Scale Refinement

Bayesian Exploratory Factor Analysis (BEFA) was applied to investigate the latent structure of the developed measurement scale. BEFA provides a probabilistic alternative to classical factor analysis (FA). Further, BEFA incorporates prior distributions and uncertainty estimates through Markov Chain Monte Carlo (MCMC) sampling. BEFA offers a probabilistic framework that integrates prior beliefs and quantifies uncertainty in parameter estimates. This is achieved through Markov Chain Monte Carlo (MCMC) sampling, which allows the estimation of both the number of factors and their associated loadings with credible intervals (Conti et al., 2014; Hoofs et al., 2017). In traditional exploratory factor analysis (EFA), the number of factors to be specified; however, BEFA allows the number of latent factors and the strength of item–factor relationships to be estimated simultaneously. This joint estimation approach makes BEFA particularly well-suited for situations involving model uncertainty or limited sample sizes (Conti et al., 2014; Muthén & Asparouhov, 2012). All analyses were conducted in R (version 4.4.2) using the “BayesFM” and “psych” packages (CRAN, 2025). Before analysis, the dataset was screened for missing data, outliers, and inconsistencies; no cases were excluded.

Before running the BEFA model, a preliminary inter-item correlation analysis was conducted to assess the suitability of the data for latent factor modeling. The correlation matrix revealed several moderate to strong positive correlations, suggesting that items were sufficiently related to justify exploratory factor analysis (Conti et al., 2014; DeVellis, 2016). At the same time, the absence of extremely high correlations (e.g., $r > 0.90$) indicated low risk of multicollinearity (Tabachnick & Fidell, 2014). This initially supported dimensionality reduction techniques (Hair et al., 2010). In addition to item-level analysis, correlations were examined between higher-order constructs (e.g., social presence, perceived usefulness, trust (goodwill and qualification-based), perceived usability, and continuance intention). These constructs showed consistently positive relationships, with the strongest associations observed between social presence. They perceived enjoyment ($r = .77$) and the relationship between usability and continuance intention ($r = .69$) (Appendix B.1) (Hair et al., 2010; Tabachnick & Fidell, 2014). Then, a parallel analysis was conducted using the "psych" package to determine an initial range for the number of latent dimensions (Revelle, 2017). This simulation-based technique compares observed and random eigenvalues to determine factor retention. Based on the PA (Appendix B.2), four factors are suggested to explain observed item covariance for subsequent Bayesian modeling (Hayton et al., 2004).

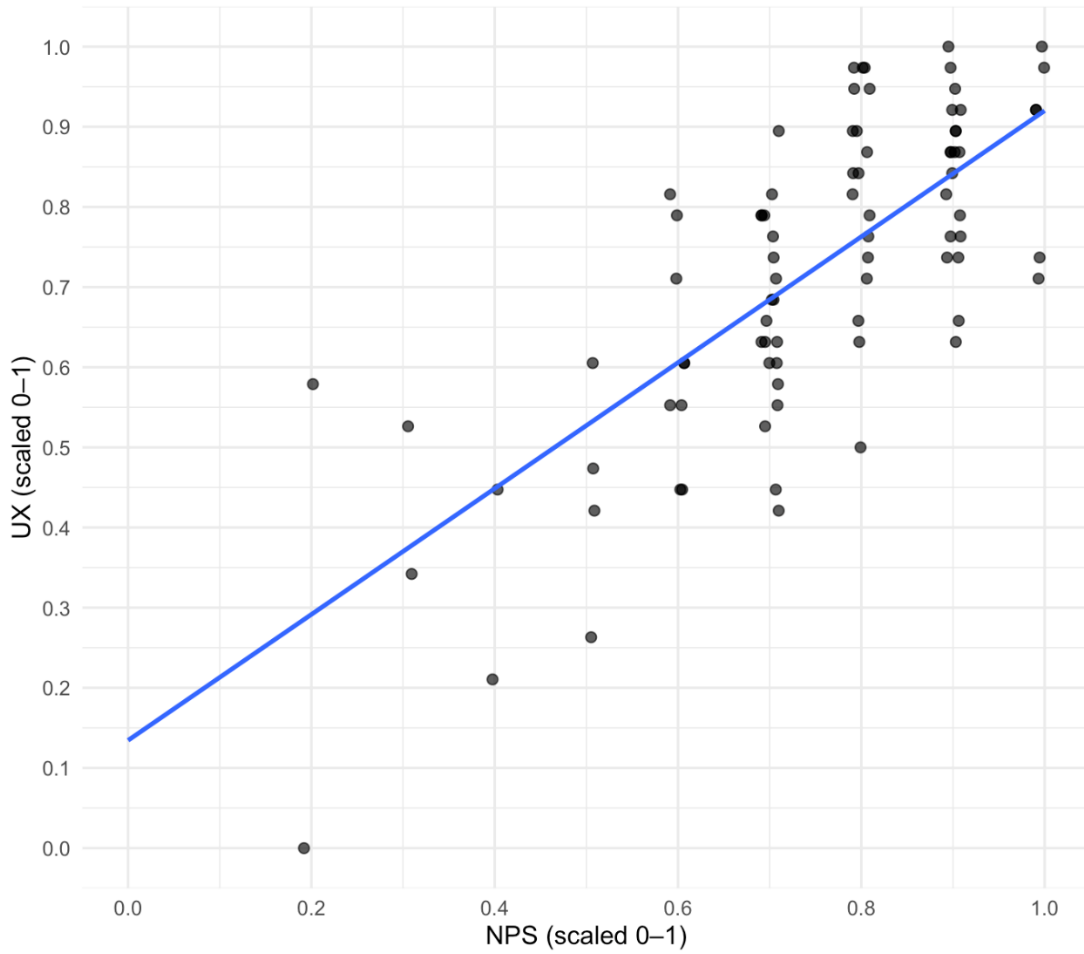
The BEFA was conducted using 50,000 MCMC iterations, including a burn-in period of 5,000. The maximum number of factors was set to 4 in the initial model, and posterior estimates were generated for factor loadings, indicator probabilities, and inter-factor correlations. Posterior indicator probabilities provided insight into which items meaningfully contributed to each latent factor, with values closer to 1 indicating higher confidence in nonzero loadings (Appendix B.3). Items were retained or flagged for removal based on two criteria: posterior probability of loading < 0.60 and high residual (idiosyncratic) variance. The posterior correlation matrix revealed that all four extracted factors were extremely highly correlated ($r \approx 0.98$–0.99). This indicates that the factors were not empirically distinct but instead reflected a common underlying construct (Appendix B.4).

Following this refinement procedure, a revised model was specified using the retained item set (Appendix B5). A second BEFA was run with a reduced factor limit (Kmax = 2), reflecting the revised structure. This two-factor model yielded high posterior loadings and low error variances across all retained items. These results support the empirical coherence of the reduced item set and indicate a strong latent structure. Finally, the correlation between the two retained factors was very high ($r = 0.998$). These findings suggest they may reflect closely related dimensions or a unified underlying construct. This observation will be further examined in the validation study in Section 3.6.

To assess convergent validity, we computed Kendall's tau-b rank correlation between the proposed overall user-experience score and the individual-level Net Promoter Score (NPS). The association was strong and positive, $\tau_\beta = 0.578$, $z = 7.26$, $p = 3.83 \times 10^{-13}$ (two-tailed), supporting convergent validity (see Figure 18). The result is consistent with Pearson's correlation ($r = 0.765$, $p < .001$), indicating that higher UX scores are associated with higher likelihood-to-recommend.

Figure 18

Correlation between Overall User Experience Scale and NPS



Source: Created by the author.

3.6. Empirical Study – Validation Study

This section presents the empirical studies conducted as part of the rigor cycle of the DSR approach. Two complementary empirical studies were designed: an eye-tracking study and an online experiment. Firstly, the chatbot search and selection were performed (section 3.6.1), using design elements list in Table 20. Secondly, the laboratory experiment was conducted as a pilot study for scale validation (section 3.6.2). Thirdly, the eye-tracking experiment was designed to explore users' attention (section 3.6.3). Lastly, online data collection was conducted after interacting with service chatbots to reestimate the overall user experience scale (section 3.6.4), as well as to investigate variations in user perceptions across different chatbot designs.

3.6.1. Service Chatbot Selection and Design Exploration

The first stage of the online study involved selecting the service chatbots and designing scenarios that would allow systematic evaluation of their interaction styles. Chatbots were identified through a structured search of internet sources, industry websites, and prior academic studies of chatbot interaction. Selection criteria emphasized:

1. Public availability of the chatbot to ensure accessibility for participants (allowing interaction without user information, such as forcing the user to enter an email address, or asking for a flight reference code).
2. Diversity in design characteristics allows coverage of different design dimensions (including defined design elements that are explained in the following section 4.2.2, in Table 20).
3. Relevance to service contexts, such as customer support, information retrieval, or task assistance, is consistent with typical user scenarios (Aktaş et al., 2025; Borsci et al., 2022b; Borsci & Schmettow, 2024)

Based on these criteria, six service chatbots were chosen for the study. Each chatbot represented a distinct set of design elements, enabling comparative analysis of user experience outcomes. Table 10 (see also Appendix C.1) presents a comparative overview of the systems against the predefined list of design elements. This mapping clarifies which elements were present, absent, or implemented in distinctive ways, thereby structuring the basis for subsequent evaluation. Further, to ensure the accuracy of the design element mapping, the comparison table was developed through an expert evaluation process. Two domain experts in HCI independently assessed the presence or absence of each design element in the selected chatbots. Each expert was provided with the predefined list of design elements derived from the literature review and asked to evaluate whether these elements were present, absent, or partially implemented in each chatbot.

Following independent coding, the results were compared to identify discrepancies. The intercoder agreement reached 88.81%, indicating a high level of consistency between coders.

Differences were discussed until consensus was reached, ensuring inter-coder reliability and minimizing subjective bias (Krippendorff, 2019; Schreier, 2012). This process provided a validated and transparent foundation for comparative analysis.

Table 10

Selected Chatbot Evaluations

Chatbot	Design Evaluation Result After Agreement (N=35 design elements)
US Citizen (USCIS)	9/35
Lufthansa	8/35
Seattle Ballooning	15/35
Kia	8/35
UTwente	12/35
Wanderlog	13/35

Source: Created by the author.

3.6.2. Pilot Study for Scale Validation

The pilot study was conducted in a laboratory, a within-subjects pilot with 22 participants. Each participant interacted with two chatbots (Seattle Ballooning, high-rated; USCIS, low-rated) and completed the predesigned tasks for both systems; presentation order was counterbalanced via randomization to mitigate sequence effects (Lazar et al., 2017). Participants were 23–39 years old ($M = 27.7$), identified as 12 female, 9 male, 1 other. 82.8% reported using chatbots almost daily or more. Table 11 summarizes task success and chatbot evaluation ratings. Users tended to rate Seattle higher than USCIS. Wilcoxon signed-rank test was performed to compare medians; results were just above the α=.05 threshold, $V = 130.5$, $p = .052$ (Table 12). The paired t-test converged (t (21)=1.99, p=.060). The paired effect size indicated a small positive effect favoring Seattle (Hedges g_paired=0.35, 95% CI [−0.02, 0.72]).

The 16-item overall UX composite showed excellent internal consistency, $\alpha = .95$ (95% CI [.93, .97]) with mean inter-item correlation (r=.58) and G6(smc)=.98. Item diagnostics indicated comparatively weaker corrected item–total correlations for b2 (r_drop=.39) and enj4 (r_drop=.52); removing any single item left α essentially unchanged (α-if-deleted = .95–

.96), suggesting redundancy and supporting potential scale shortening in the next iteration (DeVellis, 2016).

Table 11

Pilot Evaluation Results

Chatbot	Number of Successful Interaction	Number of Unsuccessful Interaction	Success Rate	UX Evaluations	
				Mean	SD
USCIS (Low)	15	7	68.2%	75	18.5
Seattle Ballooning (High)	16	6	72.7%	81.4	15.8

Source: Created by the author.

Table 12

Pilot Study Item Results

Metric	Mean	SD	Cronbach α	Metric	Mean	SD	Cronbach α
b2	85,45	22,56	0.39	enj4	59,55	24,2	0.52
b3	84,09	20,5	0.84	t8	84,09	19,09	0.88
b4	77,73	25,23	0.79	t9	80,91	18,28	0.85
b5	82,73	20,04	0.65	t10	79,55	20,9	0.92
b6	83,18	22,8	0.87	Us1	77,73	23,71	0.71
b9	83,18	21,11	0.80	Us4	81,82	20,15	0.83
sp4	84,09	23,85	0.76	CI_1	72,27	26,66	0.70
enj1	67,73	22,91	0.73	CI_2	66,82	26,57	0.66

Source: Created by the author.

Parallel analysis on the 16 pilot items (N=44) supported a single dominant factor: only the first empirical eigenvalue exceeded the 95th percentile of simulated eigenvalues for both PC and FA criteria (Figure 19).

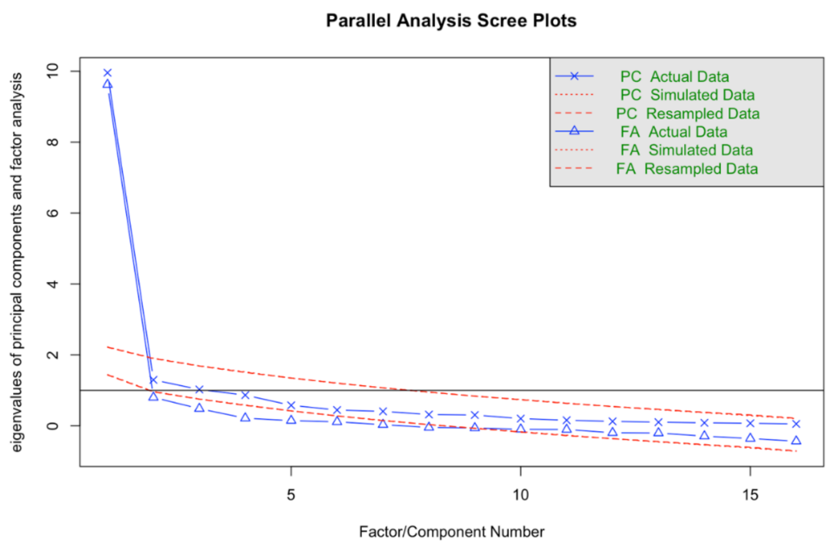
3.6.3. Visual Attention Study: Eye-tracking

User experience with CAs is usually measured with self-reports. Prior studies demonstrated that design choices of CAs (such as low anthropomorphic or high, mechanic or humanlike conversational style) shape user perception (J. Chen et al., 2024; Følstad & Brandtzaeg, 2020; Wang & Benbasat, 2007). Some studies support high anthropomorphism, others support low

(Ciechanowski et al., 2019; Nowak, 2004). Findings on conversational abilities are also mixed. One reason is that perception includes processes outside awareness, so self-reports are not always accurate (Luan et al., 2016). Bias can also come from time pressure, peer pressure, low awareness or motivation, and difficulty expressing thoughts (Denovan et al., 2023; Fulmer & Frijters, 2009; F. Guo & Zhang, 2020).

Figure 19

PA to Determine the Number of Factors



Source: Created by the author.

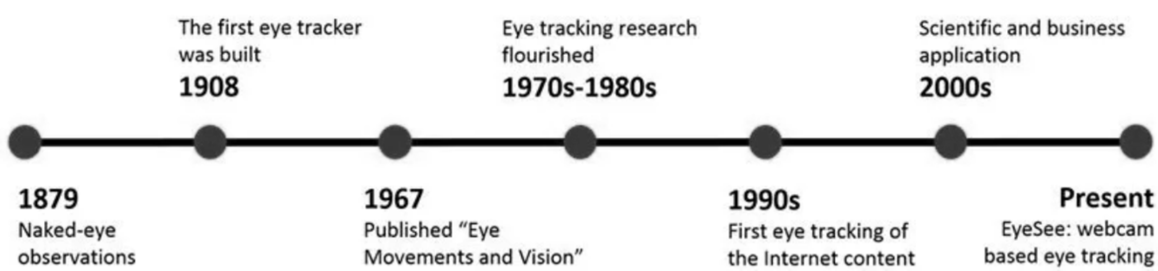
To address these limits, researchers use psychophysiological measures, such as eye-tracking, EEG, and fNIRS, to link self-reports with responses that are intrinsic or unconscious. Specifically, eye-tracking is widely used in the HCI and IS domains. For instance, Albert and Tedesco (2010) compared what participants said they had looked at on website homepages with eye-tracking data. Their reports generally matched, but about 5–10% of the time, participants claimed to have looked (or looked "a lot") at elements without any fixations. This showed that self-reports are reasonably reliable but not perfect. Current literature included eye-tracking studies such as websites, robots, virtual world design and development, human behavior and emotions exploration during the interaction with technology, and usability (Aktaş & Korkmaz, 2025; Sundstedt & Garro, 2022). However, how CAs design affects psychophysiological responses remains unexplored. Therefore, this thesis uses eye-tracking to test how service-chatbot design choices affect visual behavior. The next section describes the eye-tracking method and the study design.

3.6.3.1. Eye-tracking Tools

The first eye-tracking studies were naked-eye observation when French ophthalmologist Louis Émile Javal noticed user reading behavior with quick movements (saccades) and short pauses (fixations) in 1879 (Figure 20). The first eye tracker was built in 1908 to track the reading process, which included contact lenses. The lens was following the movements of the eye. Following, a Russian psychologist, Alfred Lukyanovich Yarbus, showed a correlation between eye movement and people's interest. During the 1970s-1980s, eye-trackers provided more accurate eye movements. In the 1990s, user experience studies were conducted by companies using eye-trackers for internet content. After 2000, eye-tracking technology evolved and was applied in different domains, including usability testing, understanding diagnosis, psychology research, and marketing (Mikac, 2022).

Figure 20

History of Eye-tracking



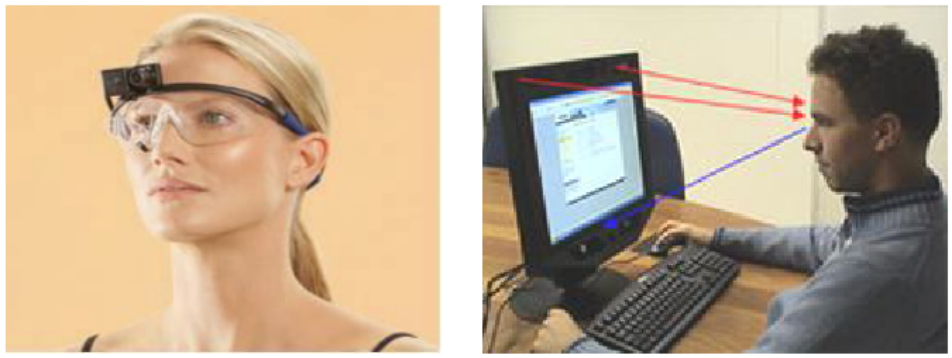
Source: Mikac, (2022)

Eye-tracking tools record eye movements to understand visual behavior and cognitive processes. Eye trackers estimate where someone is looking (gaze) and follow eye movements (e.g, fixations and saccades) (Hessels et al., 2024), using a high-resolution camera to show reflections. They are widely used in fields such as marketing, psychology, usability, and user experience research. Tools range from high-precision eye-tracking hardware such as Tobii Pro, iMotions, and Gazepoint to mobile eye-tracking glasses. Screen-based trackers for desktop and laptop tasks, while wearable trackers are for natural viewing and movements

(Figure 21). Moreover, mobile-device testing solutions are used with phones and tablets. These tools were used in controlled labs while participants were asked to perform realistic tasks on PCs, tablets, or smartphones (Punde et al., 2017).

Figure 21

Eye-tracking Types (Left Wearable, Right Screen-based)



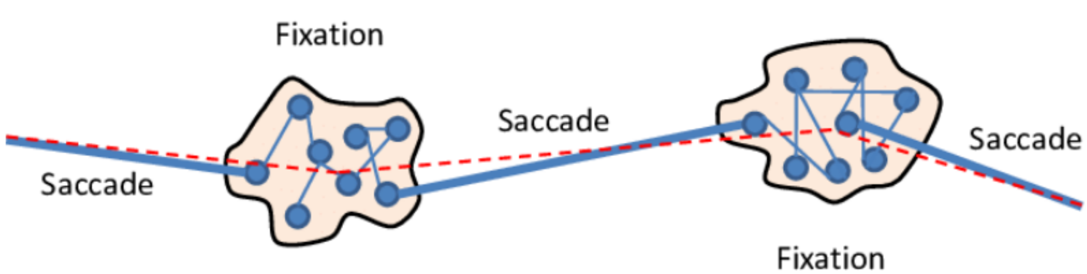
Source: Razeghi (2010)

3.6.3.2. Eye-tracking Metrics

The eye-mind hypothesis says that what people look at (fixate) is what is in their mind (Just & Carpenter, 1976). Following the eye-mind hypothesis, the eye-tracking metrics can recognize the cognitive process when people read documents or view images. Further, the eye-tracking metrics can be used to track visual behaviors to understand cognitive processes (Mele & Federici, 2012; Pei et al., 2022).

Figure 22

Eye-tracking Trajectories



Source: Santhoshikka et al. (2021)

Eye tracking determines where visual gaze is directed and whether it fixates on a stimulus. The two main movement types are fixations and saccades. Fixations are periods in which gaze is held steady to keep the image of a stationary object on the retina; they are marked by miniature eye movements (tremor, drift, microsaccades) and are the primary events used to infer cognitive processing. The important detail about fixations is (1) the duration of fixations ranges from 50 to 600 milliseconds. (2) Their frequency is usually less than 3 Hz. (3) Visual information is absorbed through fixations. (4) Fixations are analyzed by their location and by calculating their frequency and duration in a designated area of interest. (5) The fixation duration indicates the effort needed to process visual information. Saccades are rapid shifts that reposition the fovea to a new location; they can be voluntary or reflexive and last about 10–100 ms, during which visual sensitivity drops markedly (often described as being effectively "blind" in transit). These properties make fixation-based measures central for inferring processing, with saccades indexing changes of focus. During saccadic eye movements, both eyes move together in the same direction (Figure 22) (Duchowski, 2007; Santhoshikka et al., 2021).

Table 13

Eye-tracking Measures

Metric	What it captures	Cognitive Process
Fixation count	Total number of gazes stops during exposure	Fewer → more efficient, guided search; more → broader search / higher complexity.
Fixation duration	Length of a single fixation (ms)	Longer → greater processing demand/effort or level of engagement (ranges from 150 to 300 ms)
Mean fixation duration	Average length of fixations	Needs further investigation
Time to first fixation	How long it takes users to first notice a particular object	Shorter → faster orienting/attentional capture to that region
First fixation duration (on object)	Length of the very first fixation on a target	greater duration means the greater the level of engagement
Number of saccades	Count of rapid shifts between fixations	More → more searching / attentional switching; fewer → steadier focus.
Saccade amplitude	Angular distance of eye movements between fixations	larger = broader exploration; Meaningful visual cues
Saccade duration	Time from saccade start to end (ms)	Less efficient scanning
Saccade direction	Movement orientation (e.g., left/right/up/down)	Indication of search strategy
Saccade length	Spatial distance between fixations (px or °)	Less efficient searching
Dwell time on the AOI	Time inside AOI, including revisits	Greater → the greater level of interest (preferably greater than 500 ms)

Table Continued		
Gaze duration (per AOI / total)	Sum of fixation time within an AOI (or across AOIs)	Difficulty extracting or interpreting information from element
Number of Gaze time per AOI	Frequency of returning to a specific AOI after looking elsewhere	Higher revisit count → uncertainty, re-evaluation, or sustained interest.

Source: Adapted from Razeghi (2010), Tullis & Albert (2013), Punde et al (2017), Sharafi et al (2020), Burch et al. (2021), Santhoshikka et al. (2021), Sundstedt & Garro (2022), and Chen et al. (2024).

Additionally, Area of Interest (AOI) is used to discover visual attention with static (images, still objects) and dynamic (videos, scrolling through a webpage) stimuli. AOIs are defined regions on the display (screen/scene) used to analyze gaze on that specific part of the interface. When marking the AOIs, eye-tracking tools summarize attention with AOI-based metrics (Table 13). AOIs are defined by users/researchers to filter data based on AOI (Pernice & Nielsen, 2009; Rayner, 2009; Sundstedt & Garro, 2022).

Eye tracking is employed to explore visual attention, cognitive load, and emotional responses. Among these, visual attention is the most widely used application for eye trackers. When people think of eye movements, they often picture looking side to side or up and down. In behavioral research, two additional movements provide unconscious insights: pupil movements and eyelid movements. Pupils change with light, emotionally charged stimuli, and changes in cognitive load. Eyelids, especially blinking, also offer information about cognitive processes and implies the nonverbal communication during social interactions (Duchowski, 2007).

3.6.3.3.Eye-tracking Research Procedure

Many researchers using eye trackers are interested in eye movement because their research questions relate to attention. For instance, Wang et al. (2014) showed that objectively measuring how long and where people looked is proof of how they felt while interacting with a complex website. Etzold et al. (2019) indicated that eye movement patterns revealed delays, confusions, and hesitation in an online booking platform. Regarding CAs, Seeger et al. (2017) used eye-tracking metrics to examine cognitive processes related to anthropomorphism across different substitution types. Using fixations and eye movements, they identified anthropomorphic cues that might distract users during task completion. Similarly, Chen et al. (2024) used eye-tracking (fixation count, duration, dwell time) and showed that anthropomorphic chatbots attract more fixations and longer fixation/dwell times, shaping visual behavior and perception.

Eye-tracking studies include different data collection stages to achieve valid outcomes. The main stage is bringing together visual behavior data and relevant data that will be analyzed to make a conclusion. After deciding on a screen or a wearable eye-tracking setup, the real-time monitoring is conducted during the eye-tracking experiment (Razeghi, 2010). Then, calibration is a critical step in eye-tracking research. It configures the eye tracker's optics and threshold parameters so that visual targets are identified accurately. To this end, the participant is shown a series of points that cover the screen's extreme viewing angles (upper-left, upper-right, center, lower-left, lower-right, etc.). While the participant fixates these points, the system uses pupil and corneal-reflection signals to compute the thresholds and the mapping to the point-of-regard. This enables the device to reliably interpolate gaze position between those extreme points and estimate gaze across the entire display (Duchowski, 2007). Following the successful calibration stage, which can be required to be done several times based on experiment settings, the data collection starts. The eye-tracking systems collect raw data, which is a stream of 2d points or points-of-regard (sampling rate usually 50 to 250 Hz) to reveal visual attention (Poole et al., 2005). Data collection process handled by eye-tracking software, capturing data in video format, or storing data in coordinates. Eye-tracking provides massive data, which requires a highly complex process to analyze. Clustering is one of the solutions to analyze massive eye-tracking data. There are some algorithms that process the clustering based on fixations and saccades, which is an essential part of eye-tracking study to interpret the data (Enderle, 2012; Tullis & Albert, 2013).

Following, eye-movement data visualization comes as a core part of eye-tracking research. The most common visualizations are heatmaps and gaze paths. Heatmaps display fixation "hot spots" by aggregating gaze across participants. Gaze paths depict the sequence of fixations and saccades, showing how a stimulus is explored over time; they are useful for identifying what is looked at first and where attention concentrates at the individual level. Quantitative analysis requires AOIs, which can be static or dynamic. For static stimuli, AOIs are defined by outlining regions on a snapshot. For dynamic stimuli, AOIs are more labor-intensive because object position, size, or shape changes over time, often requiring frame-by-frame tracking. Some software eases this workload via gaze mapping and AutoAOIs, reducing manual coding. Without such support, AOI mapping may rely on multiple coders and can introduce consistency issues due to differences in AOI placement (Burch et al., 2021; Hennessey, 2024; Pei et al., 2022; Sundstedt & Garro, 2022).

3.6.3.4. Eye-tracking Advantages and Disadvantages

The eye-tracking technique is widely accepted to reveal information about cognitive processes and identify where and what exactly people have looked at to explore visual attention. This can tell us about personal interest, the design quality of a prototype, and detect individual differences. Moreover, eye-tracking also captures valuable insight to test hypotheses about design. More importantly, eye-tracking provides objective and quantitative data that conventional models can miss (Çöltekin et al., 2009).

The eye-tracking also has some limitations, which lead to disadvantages while doing research. Firstly, it is hard to afford considering the cost and setup. Unfortunately, cost and ease of use are always traded off against spatiotemporal precision. In addition to that, the system should be operated properly by experts. Valid and reliable eye-tracking results are only achieved when it is applied correctly (Razeghi, 2010; Tullis & Albert, 2013) . Second, accuracy hinges on calibration. Typical gaze-direction accuracy is about 0.5–1° and often requires repeated calibration; weak calibration inflates AOIs and limits fine-grained analysis (Sharafi et al., 2020). Accuracy is further affected by participant and optical factors (e.g., eyelashes, glasses, contact lenses) and by head position: modern screen-based systems permit some head movement, yet labs still use chin rests to stabilize the head when higher accuracy is needed (Dollée, 2022). Third, absence of a fixation does not prove an element was not seen because extrafoveal vision can support perception without eye movements; moreover, attention can shift within roughly a degree of the measured point of regard (Razeghi, 2010). Finally, dynamic stimuli pose analysis challenges: mapping gaze to moving objects is labor-intensive, data formats and metrics are not fully standardized, and drift and data volume complicate processing (Pernice & Nielsen, 2009; Rayner, 2009).

3.6.3.5. Eye-tracking Study Design

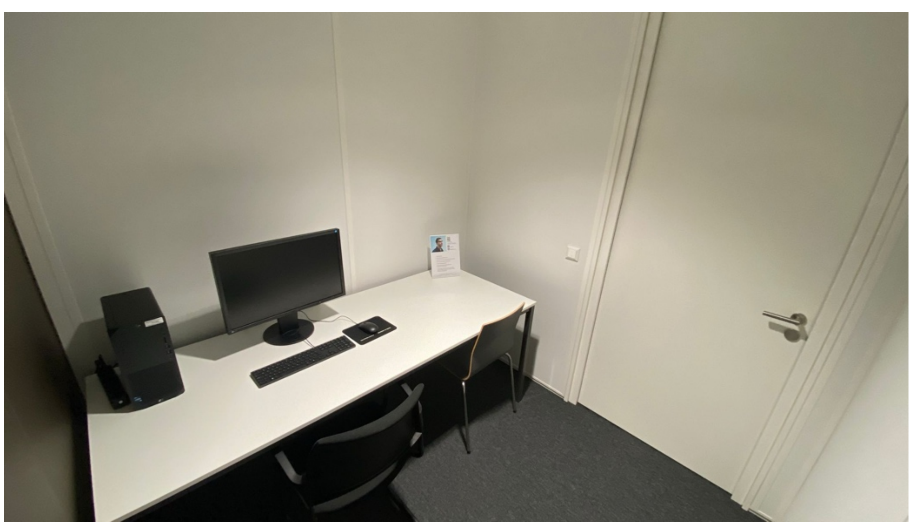
The experiment was designed to investigate users' visual attention while interacting with two chatbots differing in their design quality. The chatbots were selected based on the design evaluation results presented in Section 3.6.1. The high is featuring rich design elements and low is featuring poor design elements. The study was conducted in a human factors laboratory that provided controlled environmental conditions, including sound insulation, adequate illumination, and comfortable temperature (Figure 23). Each participant was informed about the purpose and procedure of the study, and online consent was obtained prior to participation via Qualtrics.

Participants first completed a short demographic questionnaire administered via Qualtrics. The order of chatbot interactions was randomized to minimize potential learning effects. Each participant was presented with two task scenarios:

- USCIS Chatbot: Participants were asked to obtain information about the Green Card application process.
- Seattle Ballooning Chatbot: Participants were asked to find details about flight duration, price, and landing location.

Figure 23

Experimental Environment for Eye-Tracking Study

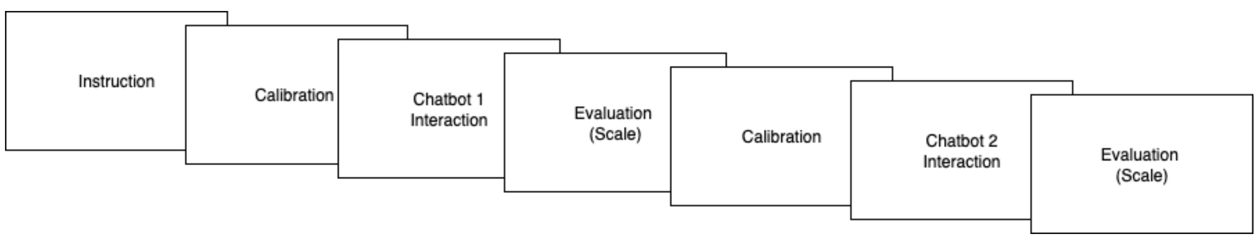


Source: BMS Lab (2025)

Quantitative data on users' gaze behavior were collected using Tobii Pro Fusion, a screen-based eye tracker with a sampling rate of 120 Hz. Before each interaction, a five-point calibration procedure was performed to ensure that the gaze position error remained below 0.5° of visual angle, consistent with Tobii Pro Lab standards. The experimental sequence followed a systematic structure (Figure 24). First, participants interacted with the first randomly assigned chatbot while their eye movements were recorded. Immediately after completing the task, they evaluated the chatbot using the Overall User Experience Scale. The calibration procedure was then repeated, after which participants interacted with the second chatbot and completed the same evaluation process. Eye movement data were collected only during the chatbot interaction phases, excluding questionnaire responses. The session ended once participants had viewed and evaluated both chatbots. On average, the total duration of the experiment per participant was approximately 15 minutes.

Figure 24

Flow of Eye-tracking Experiment

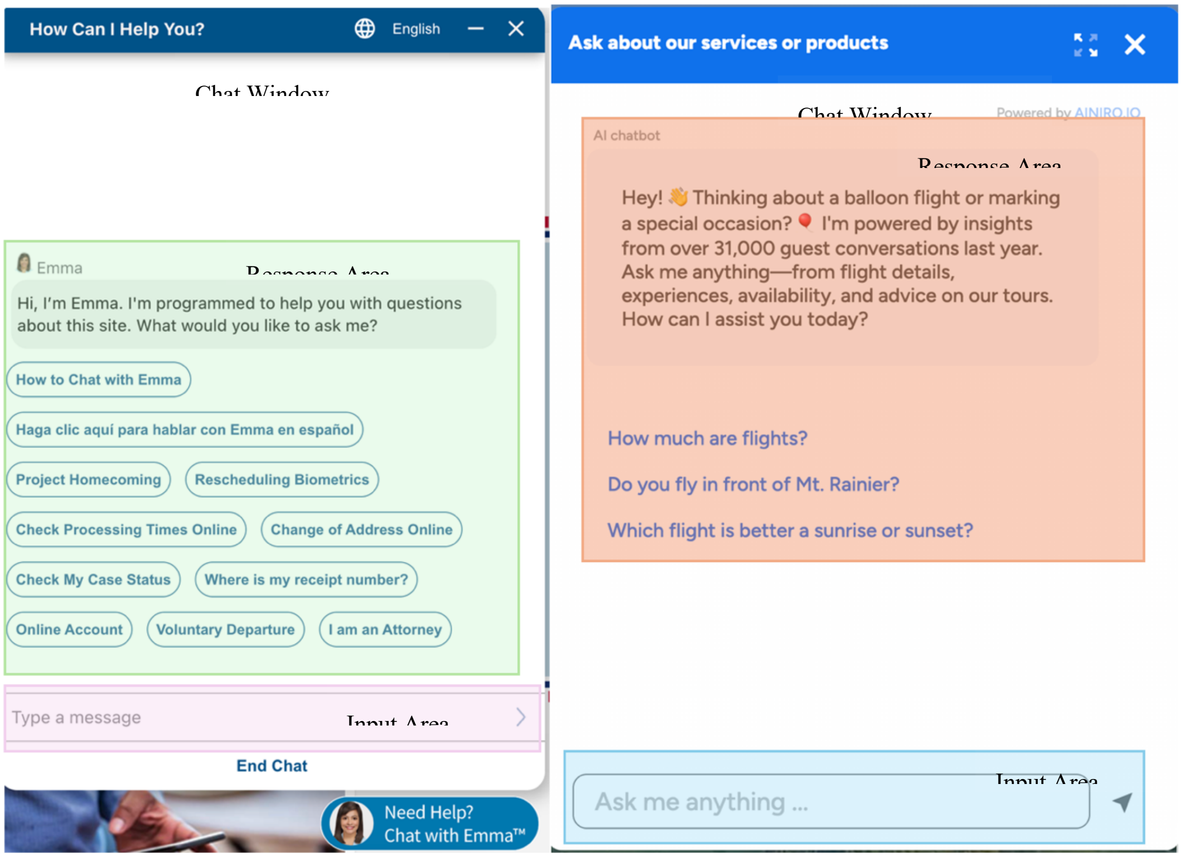


Source: Created by the author.

In this study, users' subjective perceptions and visual attention behaviors were examined while interacting with chatbots in predefined scenarios. The post-interaction evaluations were collected using a five-point Likert scale. For the analysis of eye-tracking data, AOIs were defined for each trial. These AOIs included (1) the webpage area, (2) the chatbot window, (3) the chatbot response area, and (4) the input area in two different stages: chatbot welcoming (introduction) and the following interaction (interaction flow). Figure 25 illustrates the AOI segmentation applied to the chatbot interface. Additionally, the entire webpage was defined as an AOI to assess how effectively the chatbot captured users' attention.

Figure 25

AOIs on Each Chatbot Interface



Source: USCIS (2025); Seattle Ballooning (2025)

3.6.3.6. Eye-tracking Data Analysis

In this dissertation, users' visual behavior during interactions with two randomly assigned chatbots was examined. Eye-tracking recordings were obtained from 22 participants. Data quality control was performed in Tobii Pro Lab to ensure the reliability of gaze metrics. Participants whose recordings had a tracking ratio below 75% or irregular gaze patterns due to excessive head movement were excluded. Outliers based on interaction duration were also removed to maintain data consistency. The final dataset comprised 17 valid participants whose gaze data met the inclusion criteria, resulting in 34 valid recordings. The data were exported from Tobii Pro Lab as Excel files containing AOI-based gaze metrics, which included fixation counts, fixation durations, and time-to-first-fixation values computed.

Fixations were identified and classified using Tobii Pro Lab's I-VT (velocity-threshold identification) (attention filter) fixation filter with default parameters (minimum fixation duration: 60ms, velocity threshold: 30°/s). This algorithm automatically identifies fixations by detecting periods when eye movement velocity falls below the threshold, excluding brief gaze samples that do not constitute meaningful visual processing (Holmqvist, et al. 2011; Salvucci & Goldber, 2020). All fixation metrics (counts, durations, and time to first fixation) were calculated based on these classified fixations.

To examine two chatbot designs, three snapshots per chatbot were used (Raidt, 2009). The first examination explores how accessible each chatbot is, exploring how chatbot bubble attention varies. The website's main page snapshot with a chatbot bubble was utilized for each chatbot. Participants' initial visual engagement with the chatbot interface was analyzed using Time to First Fixation (TTFF) on the chatbot activation element (bubble). TTFF measures the latency between the interface appearance and the first fixation within the target Area of Interest (AOI), reflecting the efficiency of initial attention capture.

To examine potential learning and order effects, Wilcoxon signed-rank tests were used to compare TTFF values between the first and second chatbot exposures within participants, while Mann–Whitney U tests were conducted to assess differences between groups that viewed the Seattle chatbot first versus those who viewed the USCIS chatbot first (Siegel & Castellan Jr., 1988).

In the chatbot introduction (welcoming) interface, five AOIs were defined: button, input field, introduction message, chat window, and webpage, as well as appearance (real person image at the bottom of the chatbot window). To explore visual attention hot zones, the last snapshot includes the button, input field, message area, chat window, and webpage as AOIs. For each AOI, fixation count (number of fixations) and fixation duration (total fixation time in seconds) were extracted in Tobii Pro Lab. Visual attention patterns were compared between the Seattle (high, richer in terms of design elements) and USCIS (low, basic in terms of design element) chatbots using paired Wilcoxon signed-rank tests for each AOI and metric. The relative proportion of attention allocated to each interface element was further analyzed by calculating the percentage distribution of fixation duration across AOIs. Correlation analyses was performed to examine associations between fixation-based metrics and user experience ratings. Spearman's rank correlation was used due to the non-normal distribution of eye-tracking data and the small sample size. Statistical analyses were conducted in R.

3.6.4. Online Validation Study

The online experimental study was conducted to investigate how differences in chatbot design, arising from the implementation or absence of design elements, affect user experience outcomes and support the validation of the proposed scale. The first step involved a systematic chatbot exploration, drawing on internet sources and prior research on chatbot interactions. Following the selection of suitable chatbots (Section 3.6.1), tailored scenarios were developed for each chatbot (Appendix C.2), accompanied by control questions to ensure comparability across conditions (Aktaş et al., 2025; Borsci & Schmettow, 2024). The study adopted an experimental design with predefined scenarios, ensuring participants engaged with chatbots under similar conditions (Lazar et al., 2017). The experiment was administered online using the Qualtrics platform, facilitating the survey distribution and scenario presentation.

3.6.4.1. Participants & Procedure

The present research employed an experimental research protocol using a within-subject design to assess the overall user experience of the service chatbots. A within-subject design was selected to allow each participant to interact with multiple chatbots, enabling the identification of differences in chatbot performance and user responses within the same individual (Lazar et al., 2017; Nielsen, 1993). The study protocol was reviewed and approved by the Humanities & Social Sciences (HSS) Ethics Committee of the University of Twente. All research activities complied with the university's ethical guidelines for human participants' studies.

In this study, a list of publicly accessible English-speaking chatbots from various domains (e.g., travel, education, government) was compiled that did not require membership or login. Six unique chatbots were selected. For each chatbot, a scenario and task were created to guide participants toward a goal-oriented interaction and prompt specific information retrieval (Appendix C.2). The study used a two-part survey. The first part collected demographic information. In the second part, participants selected one of the six chatbots, received a scenario and task tailored to that chatbot, and were instructed to interact with it accordingly. Afterward, they were asked to confirm whether they completed the task, regardless of the outcome. The control question was asked to ensure the participant interacted with the chatbot to complete the task. Following, they evaluated each chatbot using the overall user experience 16-item scale (developed in Section 3.5).

A common rule of thumb for sample size is that researchers should aim for a sample size of N approximately 200 or greater (Jackson, 2001). Some research suggests that sample sizes between N = 200–400 produce better Confirmatory Factor Analysis (CFA) results compared to N = 100, though returns diminish after N = 400 (DeVellis, 2016; Hoelter, 1983; Jackson et al., 2009). However, the quality of CFA outcomes is enhanced by factors beyond raw sample size (Gaborieau & Pronello, 2021; Lambert & Newman, 2023). Participants in this study were recruited through the University of Twente's test subject pool and online advertisements. Participation was voluntary. In total, 155 participants engaged with the study. However, some participants encountered technical issues during their sessions and couldn't finish the tasks or related questions. After removing data from participants who couldn't access the chatbot (e.g., due to a system issue with the Kia chatbot), the final dataset included 363 valid observations from 113 participants for analysis. Demographically, the average age of the group was 26.08, and the gender of participants was 69 women, 41 men, 2 non-binary/third gender, and 1 did not prefer to say. The nationalities represented in this study included Turkish (N=32), Dutch (N=20), German (N=15), Chinese (N=14), English (N=7), Italian (N=5), Spanish (N=5), Polish (N=2), Hebrew (N=3), Greek, Hindi, Indonesian, Hungarian, Slovak, Russian, Romanian, Ukrainian and Other (all N=1).

3.6.4.2. Data Analysis

The analysis of the post-interaction questionnaire responses focused on validating the proposed overall user experience scale. All analysis was conducted using R. Prior to factor analysis, the dataset was examined for missing values, outliers, and distribution.

Descriptive statistics were first calculated to summarize the data distribution and provide an overview of participants' evaluations. The mean and standard deviation were computed for each questionnaire item, and comparative descriptive analyses were conducted across the six chatbot conditions to identify initial differences in user experience ratings. Following this, a regression analysis was performed to examine whether the overall user experience ratings significantly differed depending on the chatbot used. In addition to the subjective evaluations, participants' task performance outcomes were analyzed. Task success rates were calculated based on control question responses for each chatbot, serving as an indicator of interaction effectiveness (Abu Shawar & Atwell, 2007; ISO, 2019). This enabled to compare potential differences between users who successfully completed their assigned tasks, and those who did not.

The analytical process then proceeded to the validation of the proposed measurement model. Bayesian Confirmatory Factor Analysis was performed to assess the overall user experience scale.

3.6.4.3. Bayesian Confirmatory Factor Analysis

Bayesian factorial analysis is more reliable compared to classic approaches due to (i) the stabilized estimation with priors and (ii) gives robust model-fit checks in small samples (Hoofs et al., 2017). The quality of the one-factor measurement model was evaluated through a systematic assessment of convergence diagnostics, model fit, factor loadings, and psychometric properties of Bayesian indices.

According to Table 14, convergence was assessed using $\hat{R}$ (potential scale reduction factor), with values below 1.10 indicating adequate convergence (Gelman et al., 1996). Model fit was evaluated using the Posterior Predictive P-value (PPP), where values between 0.05 and 0.95 indicate acceptable fit (Gelman et al., 2014). Factor loadings were evaluated based on their posterior distributions, with lower 95% credible interval bounds exceeding 0.60 considered substantial (Borsci & Schmettow, 2024). Convergent validity was assessed using Average Variance Extracted (AVE), calculated from posterior mean standardized loadings, with values above 0.50 indicating adequate validity (Fornell & Larcker, 1981). Internal consistency was evaluated using coefficient omega (ω), calculated from posterior mean factor loadings and residual variances, with values above 0.70 indicating acceptable reliability (Nunnally & Bernstein, 1994).

Table 14

Bayesian Confirmatory Factor Analysis Quality Criteria

Criterion	Threshold	Interpretation
	Convergence	
$\hat{R}$ (R-hat)	< 1.10 (ideally < 1.05)	Chain convergence
ESS	> 400 per chain	Effective independence
	Model Fit	
PPP	0.05 - 0.95 (optimal: 0.40-0.60)	Global fit
WAIC/LOO-IC	Lower is better	Predictive accuracy
	Loadings	
Standardized λ	Lower 95% CI > 0.60	Item-factor strength
95% CI	Excludes zero	Statistical significance
	Validity & Reliability	
AVE	> 0.50	Convergent validity
Omega (ω)	> 0.70 (good: > 0.80)	Internal consistency
CR	> 0.70	Composite reliability
	Discriminant Validity	
HTMT	< 0.85 (acceptable: < 0.90)	Factor distinctiveness
Factor r	< 0.90	Factor independence

Source: Created by the author.

CHAPTER 4. RESULTS AND FINDINGS

This chapter presents the results of the research activities within the DSR approach. The study yields three main contributions: a structured understanding of user experience in conversational agent interactions, a validated instrument for measuring this experience, and user experience-oriented design requirements to guide future CA development.

4.1. Relevance Cycle

This section corresponds to the relevance cycle in Design Science Research, which ensures that the problem definition and resulting design goals are grounded in the practical needs of the application domain. It connects the research activities with the real-world environment, ensuring the artefact addresses a relevant and meaningful problem (Hevner et al., 2024b). This thesis's relevance cycle focused on identifying CA design practices influencing user experience. The study examined existing design practices and user expectations through a systematic literature review, establishing a well-founded problem space for the subsequent design activities.

4.1.1. Systematic Literature Review Results

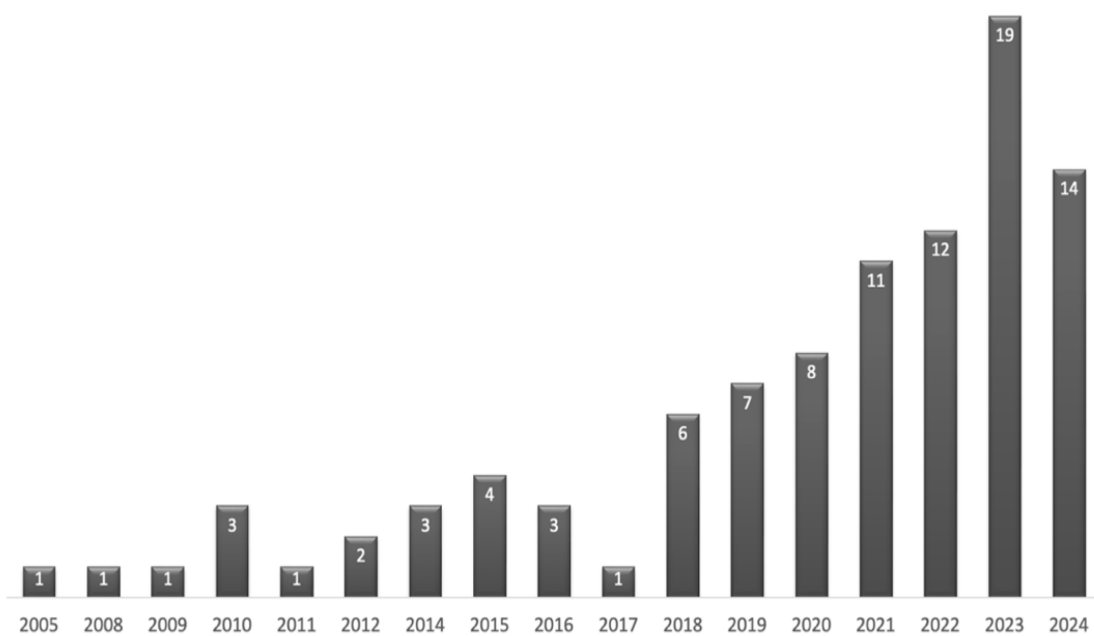
This section reports on findings from the systematic literature review, which served to contextualize the research problem and inform subsequent design activities. The analysis focused on identifying the design elements and dimensions influencing user experience in CA interactions. Based on 97 peer-reviewed studies, the study provide a structured overview of the reviewed literature's characteristics, reported design elements, their associated user evaluation outcomes, and the classification of design elements into broader design dimensions.

4.1.1.1.Design Practices in Existing Conversational Agents

Analysis of the 97 articles collected from WoS, AISeL, ACM Digital Library between 2000 and 2024 reveals a limited number of publications prior to 2018 as illustrated in Figure 26, with articles published from 2005 to 2014 accounting for a total of 9 papers (9.27%), articles published from 2014 to 2018 accounting for a total of 11 papers (11.13%), and articles published from 2018 to 2024 accounting for a total of 77 papers (79.38%). In line with publication trends on 'conversational agent design', articles that have been collected for analysis in this study and published at the beginning of the year 2024 alone have accounted for a staggering 14 papers (14.43%) of the total, indicating the continuous rapid increase in 'conversational agent design' research publications in recent years. The literature analysis presented in Table 5 demonstrates the top journals and conferences with the highest number of publications respectively as follows: International Journal of Human-Computer Interaction (20), ACM CHI Conference on Human Factors in Computing Systems (18), Computers in Human Behavior (16International Journal of Human-Computer Studies (12), International Conference on Information Systems (9), Journal of Management Information Systems (4), and ACM Transactions on Interactive Intelligent Systems (4). This literature review was conducted with journals with highly ranked publications (Diederich et al., 2022; Levy & J. Ellis, 2006).

Figure 26

Number of Studies Over the Years



Source: Created by the author.

The predominant research methodology employed in conversational agent design was a laboratory experiment (60,8% of the papers, N=59). Design science research was detailed in

2.1% of the papers (N=2). Most of the studies used laboratory experiments with a between-subjects design, a within-subjects design, or a within-subjects design for one task and a between-subjects design for the other two tasks of the experiment. These experiments mostly used questionnaires as a measurement tool, and some also collected qualitative feedback at the end of the experiment. Online experiments were also used in 30.9% (N=30) of the studies, and feedback was collected through questionnaires. Table 15 presents a detailed breakdown of the research methodologies featured in our collection of papers.

Several studies included an extensive participant pool; for example, Ben Mimoun et al. (2017) engaged 554 participants, and Hess et al. (2009) involved a similar number of 550 participants. Nonetheless, most research involved smaller sample sizes, around 200 participants or fewer. The study participants were mainly adults (77.3%, N=75), while some studies focused on younger or older adults (14.4%, N=14) (Beer et al., 2015). Fourteen papers recruited university students to evaluate conversational agents. Four papers studied children for educational purposes using virtual characters (Tinwell & Sloan, 2014) or voice-based agents (Xu et al., 2022). Only two papers included female participants, while the others focused on males and females. For instance, Al-Natour et al. (2021) investigated virtual advisors' desired characteristics for female customers in product recommendations.

The main application domains analyzed in the papers included health (13.4%, N=13), virtual environment (13.4%, N=13), service encounter (12.4%, N=12), and e-commerce (11.3%, N=11), which means that these conversational agents assist users with task completion. Chatbots, featured prominently in health (N=9) and service encounter-related studies (N=8), are widely employed to provide assistance and guidance owing to their continuous, round-the-clock availability. Additionally, within the e-commerce domain, N=5 of the studies, conversational agents are utilized to offer product recommendations (Qiu & Benbasat, 2008) and to perform functions akin to a sales assistant (Beldad et al., 2016). Studies on e-commerce mostly featured text-based agents (64%), while studies on virtual environments primarily focused on voice-based agents (62.5%).

Table 15

Distribution of CA Design Characteristics, Contexts, and Evaluation Methods Across Communication Mode

Code System	Voice-based	Text and voice based	Text-based
Agent Design			
Agent Competency			
Explanation Facilities	-	1	11
Handling System Failure	-	-	5
Responsiveness	-	-	5
Agent Characteristics			
Expressiveness	9	-	16
Transparency	1	-	8
Embodiment	9	3	1
Anthropomorphic Design Dimensions	-	-	-
Identity	3	2	27
Non-Verbal Cues	7	3	21
Verbal Cues	6	1	21
Context			
General	1	-	-
Social Task Support	2	-	1
Finance	-	-	1
Virtual Environment	5	2	1
Professional Task Support	4	1	6
Private Task Support	3	-	6
Team Collobration	-	-	2
Multiple	2	-	1
Service Encounter	-	1	11
Customer Service	1	-	6
Ecommerce	1	2	7
Tourism	-	-	1
Education	3	-	3
Health	3	1	9
Gender			
Gender Both	26	8	52
Male	-	-	-
Female	-	-	2
Age			
University Students	6	2	5
Adults	12	4	41
Older Adults	1	1	-
Young Adults	4	1	7

Table Continued			
Children	2	-	1
Research Method			
Field Study	-	1	-
Focus Group	-	-	-
Interview	-	1	-
NeuroIS	2	-	3
Design Science Research	-	-	2
Laboratory Experiment	25	6	23
Online Experiment	2	1	27

Some dimensions may not total 97, as certain studies did not report these aspects. In other cases, totals may exceed 97, as some studies were counted in multiple categories.
Source: Created by the author.

Eleven studies addressed private tasks, such as enhancing the music-listening experience (Cai et al., 2022), while nine contributed to professional tasks, such as assisting with interviews (Kang & Gratch, 2014), and three studies examined social task support that enables social engagement activities (Krämer et al., 2018). Across these domains, the agents predominantly interacted through text. The primary function in each context was to assist the user by automating interactions and offering immediate, context-sensitive help, whether for personal, professional, or social purposes.

As a final, only two studies were used in team collaboration with a text-based agent acting as a virtual team member. These emphasized the agent's team support through its emotional capabilities and social competence (Benke, 2020; A. Silva et al., 2023). Across all these domains, conversational agents aim to improve user experience by providing timely and relevant assistance.

4.1.1.2.Findings on User Evaluation Outcomes

As a part of the systematic literature review, the studies were coded based on the outcomes used to evaluate user experience with conversational agents. Several studies reported multiple constructs simultaneously. Table 16 summarizes the distribution of user evaluation outcomes identified in the review. The distribution shows that constructs related to perception, trust, and acceptance are among the most frequently evaluated dimensions, each appearing in over 20 studies. For instance, Riquel et al. (2021) investigated the positive impact of using social cues (names and avatars) in chatbots designed for donations on the perception of social presence. Additionally, Cai et al. (2022) examined the effects of agent initiative strategies on trust. Hildebrandt et al. (2023) varied a chatbot design (verbal and non-verbal cues, avoidance, and negative language) to measure the effect on intention towards negative WoM.

Constructs such as attitude and performance were also prevalent. Emotion-related outcomes appeared less often, while constructs like learning and relationship were rare, possibly reflecting their higher dependency on domain-specific implementations or more extended interaction periods. These findings indicate that different design strategies can influence user experience across various domains and agent types. In particular, the frequent and positive evaluation of perception-related constructs (e.g., perceived humanness, anthropomorphism, and social presence) highlights the importance of explicitly addressing user perceptions in agent design. To complement the construct-level analysis, a more granular examination of specific user evaluation outcomes was conducted (Appendix D.1). This analysis revealed that the most frequently studied outcomes include perceived social presence, trust, and perceived humanness, which were positively affected by design variations across diverse studies. In contrast, outcomes like self-disclosure and privacy concern exhibited mixed or low responsiveness, suggesting more complex user expectations or design limitations.

Table 16

Frequency of User Evaluation Outcomes

CONSTRUCTS	FREQUENCY	NUMBER OF PAPER	POSITE EFFECT RATE
Perception	73	51	74%
Trust	43	33	72%
Acceptance	28	23	68%
Emotion	25	20	56%
Attitude	27	24	77%
Performance	17	17	71%
Learning	5	5	60%
Relationship	2	2	100%
Ethics	0	0	-
Other	30	25	63%

Source: Created by the author.

4.2. Design Cycle

The design cycle in DSR bridges the identified problem space with the creation and refinement of the artefact. This thesis applied the design cycle to formulate and evolve user experience-oriented design requirements for CAs. Insights were drawn from an analysis of design practices reported in leading conference papers, followed by empirical user feedback gathered through prototype interaction. These inputs were then synthesised with more comprehensive literature review findings to produce a set of structured requirements linked to specific design dimensions and associated elements.

4.2.1. Conversational Agent Design Dimensions

Critically, the study started with labeling extracted literature with design dimensions, which are described as research focused on improving design in these studies. Due to the different interpretation of design dimensions, this process followed a mixed approach. When studies directly stated their design focus (e.g., transparency, verbal cues), these were coded accordingly. In cases where design dimensions were not explicitly mentioned, interpretation was required (Mayring, 2014; Schreier, 2012). This interpretive process was carried out by the first author, who applied theoretical sensitivity (Strauss & Corbin, 1998) to align each study with an appropriate design dimension, informed by the adapted framework from Diederich et al. (2022) (Table 17). This approach was necessary to ensure comparability across studies that investigated similar constructs using different terminologies or operationalizations. While such coding introduces subjectivity, it is a well-established practice in qualitative synthesis and interpretive content analysis (Miles et al., 2014; Schreier, 2012). Moreover, all coding decisions were documented, and representative examples were reviewed to enhance interpretive transparency and conceptual clarity.

The following section presents the results organized under design dimensions, which were adapted and extended from the agent research framework proposed by Diederich et al. (2022). As summarized in Table 15, the reviewed literature reflects an increasing research focus on conversational agents. Interest in anthropomorphic design and agent characteristics has remained consistent, aligning with the findings of Hildebrandt et al. (2022). The concept of agent competency gained attention with the work of Qui and Benbasat (2016), who examined transparency through explanation mechanisms in recommender systems. By incorporating agent competency into the adapted framework, this study builds upon Hildebrandt et al.'s (2023) structure and chronicles the theme's expansion over time, supporting our research insights.

Table 17

Examples for Design Dimensions Coding Validity: Examples and Interpretation Alignment

Design Dimension	Example	Source
Explanation Facilities	"The AI assistant also explained the reasoning for an answer by providing explanations. "	Mehrotra et al. (2024)
Handling System Failure	No need to interpretation all studies stated the purpose as error handling	
Responsiveness	"we designed a CA that may guide users to listen to themselves (i.e., sense the resonance with music) and express their personal feelings and thoughts the music has elicited "	Cai et al., (2023)
Expressiveness	"The three instances of the CA were identical except for the specific cues and phrases related to their communication pattern, following the suggestions of Karpman (1968). The chatbot that applied a helper communication pattern uses "Please" and "Thank you," frequently uses the word "help," and uses positive language (e.g., "good" and "great"). In comparison, the chatbot resembling a victim communication pattern is frequently apologizing (e.g., using "Sorry"), uses passive language, and uses negative language (e.g., "Sadly" and "Unfortunately"). "	Brendel et al. (2020)
Transparency	"MyAdvisor communicates its limits (Heuristic 1.1). For instance, if the student asks for something that MyAdvisor cannot do, the student is told that MyAdvisor cannot help with that request. Further, to help the student, they are given options for what services the advisor can offer. "	Kuhail et al. (2023a)
Embodiment	"Experimental Stimuli: Receptionist Human (top); Doctor Avatar (middle); Nurse Bot (botom)"	Gambino et al. (2019)
Verbal Cues	No need to interpretation all studies stated the verbal communiocation or vebal cues	
Non-Verbal Cues	"For CA with visual social cues, we used a visual representation of a white female academic advisor with facial expression, gaze etc."	Cui et al. (2020)
Identitiy	"The VSA presented to the experimental participants was in the form of photographed middle-aged salespersons (both male and female) wearing eyeglasses and sporting a business attire, based on the finding that older looking individuals may be perceived as more knowledgeable about products than younger salespersons"	Beldad et al. (2016)

Source: Created by the author.

The analysis of design dimensions highlights three core areas that structure how conversational agents are configured: agent competency, agent characteristics, and anthropomorphic design features. As shown in Table 18, the growing number of studies across these dimensions demonstrates the increasing attention paid to design variation and its potential to influence user perceptions and interactions. While earlier studies tended to emphasize anthropomorphic features such as embodiment, identity, and verbal/non-verbal cues, more recent work has extended into agent competency, particularly the role of explanation facilities and handling system errors. This trend reflects an evolving interest in designing socially engaging and functionally competent agents.

Table 18

Studied Design Dimensions in Reviewed Papers

Agent Design	**2005**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**Total**
Explanation Facilities										1					2	2	5	2	**12**
Handling System Failure														1	2		2		**5**
Responsiveness													1			1	3	1	**6**
Expressiveness					1	1		1	1			2	1	2	2	4	5	7	**27**
Transparency											1	1			1	2	3		**8**
Embodiment	1			2	1	1		1	1				3	2		1	2	1	**16**
Identity			1	1						1	1		2	5	6	2	8	5	**32**
Non-Verbal Cues		1	1	1		2		1	1	1		3	2	4	6	3	4	1	**31**
Verbal Cues				1		1		1	1			1	1	5	7	4	6	1	**29**
SUM	**1**	**1**	**2**	**5**	**2**	**5**	**0**	**4**	**4**	**3**	**2**	**7**	**10**	**19**	**26**	**19**	**38**	**18**	**166**
N = Documents	**1**	**1**	**1**	**3**	**1**	**2**	**0**	**3**	**4**	**3**	**1**	**6**	**7**	**8**	**11**	**12**	**19**	**14**	**97**

Source: Created by the author.

Generally, anthropomorphic design dimensions refer to making the agents have qualities and behaviors similar to humans, which are detectable by individuals (Diederich et al., 2022; Feine et al., 2019; Seeger et al., 2018; Seeger & Heinzl, 2021). Anthropomorphic design dimension research, often intertwined with humanlike design and humanness, focuses on integrating humanlike characteristics into conversational agents. In our collection, 52.6% of the papers (N=51) examined topics related to the anthropomorphic design dimensions (95 codes for associated patterns). The research we reviewed strongly emphasized different cues within the same contexts. Specifically, 28 papers examined verbal cues in 7 service encounters and 6 to health. Non-verbal cues were highlighted in 34 papers, with 9 addressing service encounters, 6 health, and 8 task support. For identity cues, 32 papers were analyzed, with service encounters again leading with 9 papers, then health with 7, and task support and e-commerce with 4 each.

Concerning anthropomorphic design dimensions, the types of agents analyzed included 31 virtual agents and 50 chatbots. Studies on chatbots and virtual agents concentrated similarly on identity (20 for chatbots, 7 for virtual agents), non-verbal (14 for chatbots, 13 for virtual agents), and verbal cues (16 for chatbots, 11 for virtual agents). Research on recommender agents, intelligent agents, robots, virtual humans, and digital assistants in an anthropomorphic design context was limited, with only up to four papers examining these cues. Additionally, a single study on robots focused on anthropomorphic design, specifically including verbal cues. Moreover, these studies largely favored text-based communication modes for agent communication.

Diederich et al. (2020) definition in literature for agent characteristics refers to the specific traits of the conversational agent itself, including how it communicates, through voice, text, or both, and its form, which might be a physical, interactive virtual character, a static avatar (image), or no image at all. Our interpretation further extends these characteristics to incorporate transparency and expressiveness. Transparency involves the agent revealing its capabilities and limitations to users (Diederich et al., 2020a; Shevat, 2017). Expressiveness is another essential trait, allowing the agent to disseminate information and display emotions, thus deepening the richness of user interactions (Al-Natour et al., 2011; Pauletto et al., 2013). In exploring the agent characteristics, we reviewed forty-seven papers, including expressiveness (27 papers), embodiment (16 papers), and transparency (8 papers). The communication modes for all these papers are detailed in Table 15. In this context, studies on text-based agents predominantly focused on the expressiveness pattern (16 papers), whereas those on voice-based agents mainly examined embodiment (9 papers).

Agent competency encompasses a conversational agent's capacity for understanding and processing information, relating to users, handling errors, and articulating actions. This study extends upon the agent design framework, adding agent competency due to the literature emphasizing the significance of agent competency (Chandra et al., 2022; Luger & Sellen, 2016; Ternyak, 2023). After 2021, there has been an increase in research focused on enhancing agent competency, evident from the data presented in Table 15. Our analyses identified 21 papers investigating agent competency, focusing 57.14% on explanation facilities, 23.8% handling system failures, and 28.6% on responsiveness patterns. Most of these studies (96%) concentrate on text-based agents, with only one examining voice-based agents (Table 12).

4.2.2. Design Elements Exploration

This section presents a detailed account of individual design elements identified across the reviewed literature. The conference papers were first reviewed during the literature review to explore design choices and how scholars examined them as a part of the codebook development. Then, the elements and user evaluations are analyzed using crosstab analysis in MaxQDA software. This effort indicated how scholars have used various design elements to transform user experience beyond the agent's core functional capabilities (Table 19). For example, Wang and Gratch (2010) investigated user perceptions of virtual humans by integrating elements such as gaze, head nods, and posture mimicking. Guo et al. (2021) have benefited from the features of expressive speech acts and reasoned utterances to mitigate privacy concerns and foster user trust for a customer service agent. Similarly, Brendel et al. (2023) conducted an online experiment in a service encounter context, enhancing the chatbot with positive language and emoticons to shape the emotional and behavioral dimensions of user experience. Then, we conducted an exploratory review to investigate detailed design elements in the entire study corpus. This study conducted an iterative process to derive design elements. As a result of this effort, the final list of design elements is defined in Table 20. As with any other literature study, this dissertation makes no promise that this list is exhaustive. However, the 44 design elements revealed in this study are appropriate as a starting point for methodically studying the concrete features employed in conversational agent design. They provide a formal foundation for future research to analyze, compare, and develop agent designs across various interaction scenarios.

Further, Table 21 shows the distribution of design elements across various agent types while offering insights into which features are commonly implemented in different conversational agent designs. Each column represents a distinct agent type (e.g., Chatbot, Virtual Agent, AI Recommender), while rows correspond to the 44 identified design elements. Each element’s implementation frequency is indicated by color intensity, with darker shades representing more frequent use. The table reveals that chatbots and virtual agents implement various design elements. These systems frequently adopt gender, static avatar, name, greeting, positive language, response delay, and interactivity.

Table 19

Association Between Design Elements and User Evaluation Outcomes

Agent Design Elements	Performance	Learning	Trust	Enjoyment	Acceptance	Perception	Attitude	Other	SUM
Static Avatar	5	1	2	3	6	9	6	2	34
Self-Introduction	2		4	3	3	8	5	2	27
Name	1		4	2	3	9	6	1	26
Typing Indicator	1		3	2	2	9	6	2	25
Response Delay			3	2	3	7	6	2	23
Gender			1	2	4	8	8		23
Emoticons	1		2	1	4	6	5	2	21
Positive Language	1	2	1	1	1	4	2	3	15
Self-referencing			2	1	1	4	5	1	14
Apologizing				2		4	5	2	13
Greeting	1		2	1	3	2	3	1	13
Lexical Diversity	1	1		1		5	3		11
Button	3		3	1	1		1	1	10
Negative Language				1		3	4	1	9
Virtual (interactive)			2	1	2	1		1	7
Persuasive Language	1		1	1	1	1	2		7
Outlining the logical processes			2		4	1			7
Interactivity	2		1			1		2	6
Harsh Response				2		3	1		6
Expressive Speech Acts			4					1	5
Facial Expression			2	1	1				4
Avoidance Language				1		1		1	3
Tone Modulation		1						2	3

Source: Aktaş & Akbıyık, (2025)

In contrast, recommender agents have employed design elements like reasoned utterances, transparency, and explanation facilities. Virtual agents and virtual humans emphasize elements that support embodiment and multimodal presentation, as reflected in the consistent use of eye movement, gestures, facial expressions, and tone modulation. This mapping highlights how agent purpose and form influence design choices. It suggests that designers tailor specific elements to align with the agent's functional goals and domain expectations (e.g., service, recommendation, assistance). To summarize, this analysis emphasize the importance of context-aware design and supports the argument that conversational agent features must be selectively and strategically deployed based on agent type and intended user experience (Rau et al., 2023).

4.2.3. Mapping Design Elements into Design Dimensions

The present study further organized the design elements utilized in the literature for implementing each design dimension. The following sections provide a detailed overview of each design dimension and its respective elements. Rather than a rigid classification, this section presents an interpretive mapping of observed design elements to theoretical design dimensions, enabling a structured yet flexible understanding of how specific design choices align with core constructs in agent design.

4.2.3.1. Anthropomorphic Design Dimensions

In anthropomorphic design efforts, researchers utilized design elements such as static avatars (Gambino et al., 2019; Go & Sundar, 2019), gender representations (Benbasat et al., 2020; T. Cui et al., 2020), names (Liu & Yao, 2023), self-introductions (Chung et al., 2023; Diederich et al., 2019), typing indicators (Appel et al., 2012; Gnewuch et al., 2018), response delay (Gnewuch et al., 2022; Greulich & Morana, 2023), emoticons (Riquel et al., 2021a; Seeger & Heinzl, 2021), and greetings (Greulich & Morana, 2023) to operationalize verbal, non-verbal, and identity cues. This review corroborates existing literature, underscoring these elements as foundational for designers seeking to embed conversational agents with human-like qualities (Diederich et al., 2022; Schuetzler et al., 2021; Seeger et al., 2018).

Regarding verbal cues, the self-introduction element is widely encountered in verbal cues studies. Incorporating positive or negative language (Brendel et al., 2020) and lexical diversity (Chung et al., 2023; Lembcke et al., 2020) has been suggested to enrich verbal cues. Additionally, the dataset also shows limited use of pause fillers (N=1) (Von Der Pütten et al.,

2010) and respect terms (N=1) (A. Zhang & Patrick Rau, 2023) to enhance the perception of verbal cues in conversational agents (Figure 25).

As for non-verbal cues, the primary design elements employed were response delays, typing indicators, and emoticons. In addition to this, eye movements (Appel et al., 2012; Cafaro et al., 2016), gestures (Hess et al., 2009; Kang & Gratch, 2014), displaying listening behavior (Kang & Gratch, 2014), and facial expressions (Harjunen et al., 2018; Hyde et al., 2015) serve as distinct non-verbal design patterns. In the context of identity dimension, key components such as gender, static avatars, and names are crucial in representing identity. These elements help shape the agent's persona, making it more relatable and recognizable to users. They contribute to the user's perception of the agent's character and can influence interactions by providing a more personalized experience (Schuetzler et al., 2019).

Table 20

Design Elements Definitions in Reviewed Studies

Design Element	Example	Sources
Eye movement	pupil dilation, eye contact, gaze	(Fasya et al., 2024; Krämer et al., 2018; Von Der Pütten et al., 2010)
Smile	Smiling behavior	(Fasya et al., 2024; Straßmann et al., 2018)
Facial Expression	Blushing, mouth movement, angry, neutral, and happy	(Guadagno et al., 2011; Harjunen et al., 2018)
Touch	skin contact in VR using with gloves	(Harjunen et al., 2018)
Gesture	nods and hand movements, posture shifts, body movement	(Hess et al., 2009; N. Wang & Gratch, 2010)
Head Movement	a head nod, shake, or wobble	(Krämer et al., 2018)
Button	Button option for user response, pre-defined answer	(Cai et al., 2023; Janson, 2023; Jung & Cho, 2022)
Response Delay	dynamic delays in responses	(Hildebrandt et al., 2023b)
Typing Indicator	[…], blinking dots	(Hildebrandt et al., 2023; Seeger & Heinzl, 2021)
Emoticons	😀, 👌	(Diederich et al., 2021; Seeger & Heinzl, 2021)
Typing Error	"Instead of Hannover, hanover"	(Bührke et al., 2021)
Name	using a human name such as Anna, Gerald, etc.	(Hildebrandt et al., 2023b; Janson, 2023)
Gender	choosing female or male agent	(Diederich et al., 2021; Hildebrandt et al., 2023)
Static Avatar	digital representation of a character (Virtually static), representation with an image.	(Diederich et al., 2022; Seeger et al., 2021; Von Der Pütten et al., 2010)
Virtual (interactive)	digital representation of a character (Virtually interactive)	(Diederich et al., 2022)
Age	Older, younger	(Harrington & Egede, 2023)
Race	ethnicity such as Asian	(Qiu & Benbasat, 2010)
Pause-filler	"erm,", "hm"	(von der Pütten et al., 2010)
Respect Words	"Mr, Miss"	(Stein et al., 2020; A. Zhang & Patrick Rau, 2023)
Displaying listening behavior	"Tell me more about that!", "posture shifts and head nods"	(Appel et al., 2012; Hwang & Won, 2021)

Design Element	Example	Sources
Table Continued		
Voice Tone	humanlike voices: Friendly voice, unfriendly voice	(Stein et al., 2020; A. Zhang & Patrick Rau, 2023)
Tone Modulation	Yay!, Oh! "Always, Really?"	(Ceha & Law, 2022; Hu et al., 2018; Rhim et al., 2022)
Others' Comments	Conversational agent displays others' comments on the topic	(Cai et al., 2023)
Self-referencing	I think that..., I suggest, I guest, "… can I do..", "I address you.." In my experience, I think …	(Cui et al., 2021; Hildebrand et al., 2023b; Riquel et al., 2021a)
Self-Introduction	"I am OTTO. Your personal assistant. I want to make solving captchas more pleasant and easier for you."	(Brendel et al., 2023)
Human Intervention	Transfer to a real person	(Shevat, 2017; Stein et al., 2020)
Greeting	"Good Afternoon", "Hello, How are you?"	(Cai et al., 2023; Jung et al., 2022)
Apologizing	"Sorry to …,", "I apologize for the inconvenience."	(Brendel et al., 2020; Hu et al., 2018)
Harsh Response	"That's not nice", "Please don't use that kind of language," "Well, that's not going to get us anywhere"	(Chin & Yi, 2019)
Asking User Feedback	"Does this song resonate with you?"	(Cai et al., 2023)
Taking responsibility for errors	"Also, you need to know that I am still new and learning. So, mistakes can happen, but I do my best to avoid them"	(Hildebrandt et al., 2023a)
Humorous response	"I am such a klutz. Maybe I need to get sime bot languagae lessons."	(Yang & Kankanhalli, 2023)
Negative Languagae	using the verbs: Sadly, Unfortunately	(Brendel et al., 2020; Riquel et al., 2021a)
Positive Language	Good, Great, Very Nice. Passionate, Empathetic, Polite utterances	(Brendel et al., 2020; Ceha & Law, 2022)
Persuasive language	Self-monitoring, praise, social facilitation. "With this reservation, you already took four e-bike rides this week and travelled more than 6.5 kilometers - without any traffic jam or Co2 emission. Keep it up!" "Two colleagues of yours will be taking the e-bike at the sam time tomorrow"	(Diederich, Lichtenberg, et al., 2019)
Avoidance Language	"I don't know what you mean", "Let me know what you want to chat again." ,"Got it, I'll stop. Goodbye."	(Chin et al., 2020; Chin & Yi, 2019)
Lexical Diversity	using many different unique words, or using abbrevations, assertive statements. Instead of ""The book is not available now" this: "Unfortunately, the book you choose is not available now, but there is another book that is ready to be borrowed"	(Chung et al., 2023)
Interactivity	awareness of the participant's previous responses, and the support of multi-turn conversations in turn. Echoing respondent's answer, remebering user's name "I'm sorry to hear that! If you haven't already please reach out to us here «url»"	(Morana et al., 2020; Rhim et al., 2022; Setlur & Tory, 2022)

Design Element	Example	Sources
Table Continued		
	remember user's name" chatbot: “Please hold while I connect you to a representative.” (After 20 seconds) chatbot: “Sorry, no one’s available right now. Would you like me to send an email? They will respond in 24 hours.”, “Previously, you mentioned . ."	
Direct addressing	Providing verbal feedback: "Do you think …" " I see". "Appreciate your input!"	(Goldberg & Cannon-Bowers, 2015; Guadagno et al., 2011; Guo et al., 2021; Sameh et al., 2022)
Expressive Speech Acts	convey feelings and social attitudes “I understand that you may feel anxious right now”, “I know it sounds frustrating, but it is important to keep your property safe” "But, dont"t worry!!"	(Guo et al., 2021; Sameh et al., 2022)
Outlining Logical Process	describe agent actions (static explanations) Ex: Product recommendation "Nowadays many products exist that can help treat cold sores and reduce the frequency with which they appear. Yet, given that you don’t suffer from this specific problem, there is no need for me to include any cold sore related products."	(Sameh et al., 2022)
Reasoned Utterance	provide the reasoning behind agent's action "Some skin care products are best absorbed when the skin is dry, and some others complement an oily skin. Also, to optimize your skin care and get the best possible results you need to adapt your skin care products and routine to your skin type. Whether your skin is normal, dry or oily will determine the frequency with which you need to cleanse and moisturize, and whether you need to do something specific to protect your skin"	(Sameh et al., 2022)

Source: Created by the author.

Table 21

Frequency of Design Elements Across Agent Type in Reviewed Studies

Design Elements	Chatbot	Virtual Agent	Recommender Agent	Digital Assistant	Virtual Humans	Robot	Intelligent Agent	Design Elements	Chatbot	Virtual Agent	Recommender Agent	Digital Assistant	Virtual Humans	Robot	Intelligent Agent
Static Avatar	17	11	3	0	0	0	0	Smile	0	4	0	0	3	0	0
Self-introduction	20	6	0	1	2	0	0	Negative Language	3	2	0	1	0	1	0
Name	19	7	0	1	1	0	0	Persuasive Language	5	0	1	0	0	0	0
Gender	14	10	2	1	1	0	0	Voice Tone	1	3	1	0	1	0	0
Virtual (interactive)	1	11	1	1	5	2	1	Head Movement	0	2	0	0	3	0	0
Self-referencing	13	5	0	0	1	0	2	Tone Modulation	2	1	0	1	0	0	0
Emoticons	11	4	1	1	0	0	0	Displaying Listening Behavior	1	3	0	0	0	0	0
Greeting	11	2	1	1	1	0	0	Human intervention	2	0	0	0	1	0	0
Interactivity	13	1	0	0	0	0	1	Taking responsibility for errors	2	0	0	1	0	0	0
Button	11	2	1	1	0	0	0	Race	1	0	2	0	0	0	0
Typing Indicator	12	2	0	1	0	0	0	Harsh Response	1	0	0	1	0	1	0
Facial Expression	1	9	1	1	3	0	0	Avoidance Language	2	0	0	1	0	0	0
Response Delay	10	3	0	1	0	0	0	Physical	0	0	1	0	0	1	0
Expressive Speech Acts	5	4	3	2	0	0	0	Free Text Entry	1	0	0	0	0	0	0
Positive Language	6	4	1	1	0	1	0	Proactive Guidance	0	0	1	0	0	0	0
Outlining the logical processes	5	3	4	0	0	0	1	Asking User Feedback	0	0	1	0	0	0	0
Eye Movement	0	6	1	0	4	0	0	Others Comments	0	0	1	0	0	0	0
Apologizing	6	1	0	3	0	0	0	Respect Words	0	1	0	0	0	0	0
Reasoned Utterance	3	2	3	0	0	0	2	Pause-filler	0	1	0	0	0	0	0
Direct Addressing	7	0	0	0	1	0	1	Age	1	0	0	0	0	0	0
Gesture	0	4	2	0	2	1	0	Touch	0	1	0	0	0	0	0
Real Person Image	4	2	0	1	0	0	0	Typing Error	1	0	0	0	0	0	0

Source: Created by the author.

Figure 27

Anthropomorphic Design Dimensions and Associated Design Elements

Multiple codes could be assigned to the same study (e.g., identity and verbal cues coded together). The color gradient is applied per row, meaning that blue indicates lower frequency

and red indicates higher frequency of association within that element. In addition, the size of the squares reflects the strength of the association between the design element and the dimension.
Source: Created by the author.

4.2.3.2. Agent Characteristics

Agent characteristics refer to the communicative and expressive traits embedded into conversational agents to enhance their interpersonal presence and engagement quality. These characteristics include expressiveness, transparency, communication mode, and embodiment, which are frequently operationalized through specific interactional behaviors and interface design choices.

The expressiveness dimension in conversational agents incorporates various elements to enhance interaction. These include expressive speech acts (Al-Natour et al., 2021), which help convey emotions; apologies (Brendel et al., 2020), which address errors or misunderstandings; and a mix of positive and negative language to reflect diverse emotional states (Ceha & Law, 2022; Diederich et al., 2020b) (Figure 28). Furthermore, emoticons, harsh responses, lexical diversity, and tone modulations were employed to make the agent expressive. Additionally, emoticons, harsh responses, lexical diversity, and tone modulations enrich the agents' expressiveness. Transparency was addressed in 8 studies, often through explicit system feedback or the agent's self-disclosure of limitations. For example, in the study by Kuhail et al. (2023a), the agent informed users when it could not fulfill a request and suggested alternative actions. This communicative strategy increased users' understanding of the agent's boundaries and supported more realistic expectations. To incorporate transparency into design, studies frequently implemented the self-introduction and emoticons elements (Brendel et al., 2023; Gnewuch et al., 2022; Pietrantoni et al., 2022; Riquel et al., 2021a) (Figure 28).

In embodiment research, a significant focus was on virtually interactive agents (e.g., doctors or salespeople), which utilized gestures (Krämer et al., 2018; Shinozawa et al., 2005), facial expressions (Stein et al., 2020), smiles (6 occurrences) (Guadagno et al., 2011; N. Krämer et al., 2013), and eye movements (9 occurrences) (T. Cui et al., 2020; Razavi et al., 2022). Papers using static avatars frequently included elements such as gender (Gambino et al., 2019; Kim & Sundar, 2012), name (21) (Brendel et al., 2020; Pietrantoni et al., 2022), self-introduction (17) (Diederich & Benedikt Brendel, 2019; Janson, 2023), typing indicator (16) (Brendel et al., 2023; Diederich et al., 2020b), response delay (16) (Greulich & Morana, 2023; Pietrantoni et al., 2022), and self-referencing (T. Cui et al., 2020; Hildebrandt et al., 2023a). Finally, studies focusing on physical embodiment predominantly concentrated on the gesture element (Shinozawa et al., 2005).

Figure 28

Agent Characteristics Dimensions and Associated Design Elements

Multiple codes could be assigned to the same study (e.g., identity and verbal cues coded together). The color gradient is applied per row, meaning that blue indicates lower frequency and red indicates higher frequency of association within that element. In addition, the size of the squares reflects the strength of the association between the design element and the dimension.

Source: Created by the author.

Figure 29

Agent Communication Modes and Associated Design Elements

Multiple codes could be assigned to the same study (e.g., identity and verbal cues coded together). The color gradient is applied per row, meaning that blue indicates lower frequency and red indicates higher frequency of association within that element. In addition, the size of the squares reflects the strength of the association between the design element and the dimension.
Source: Created by the author.

Regarding communication mode, the review corpus largely included text-based agent research (Figure 29). The analysis revealed that text-based agents were designed with social cues and conversational aspects to enhance interaction quality. Features such as expressive speech acts, lexical diversity, and reasoned utterances were incorporated to strengthen this ability. Voice-based agents, on the other hand, were designed with interactive features, especially visual ones, including facial expressions and eye movement, to complement vocal features and make interaction more human-like. In addition, voice tone was also utilized. The review included a limited number of both-modality agents, and the analysis could not reveal distinctive elements for this group.

4.2.3.3. Agent Competency

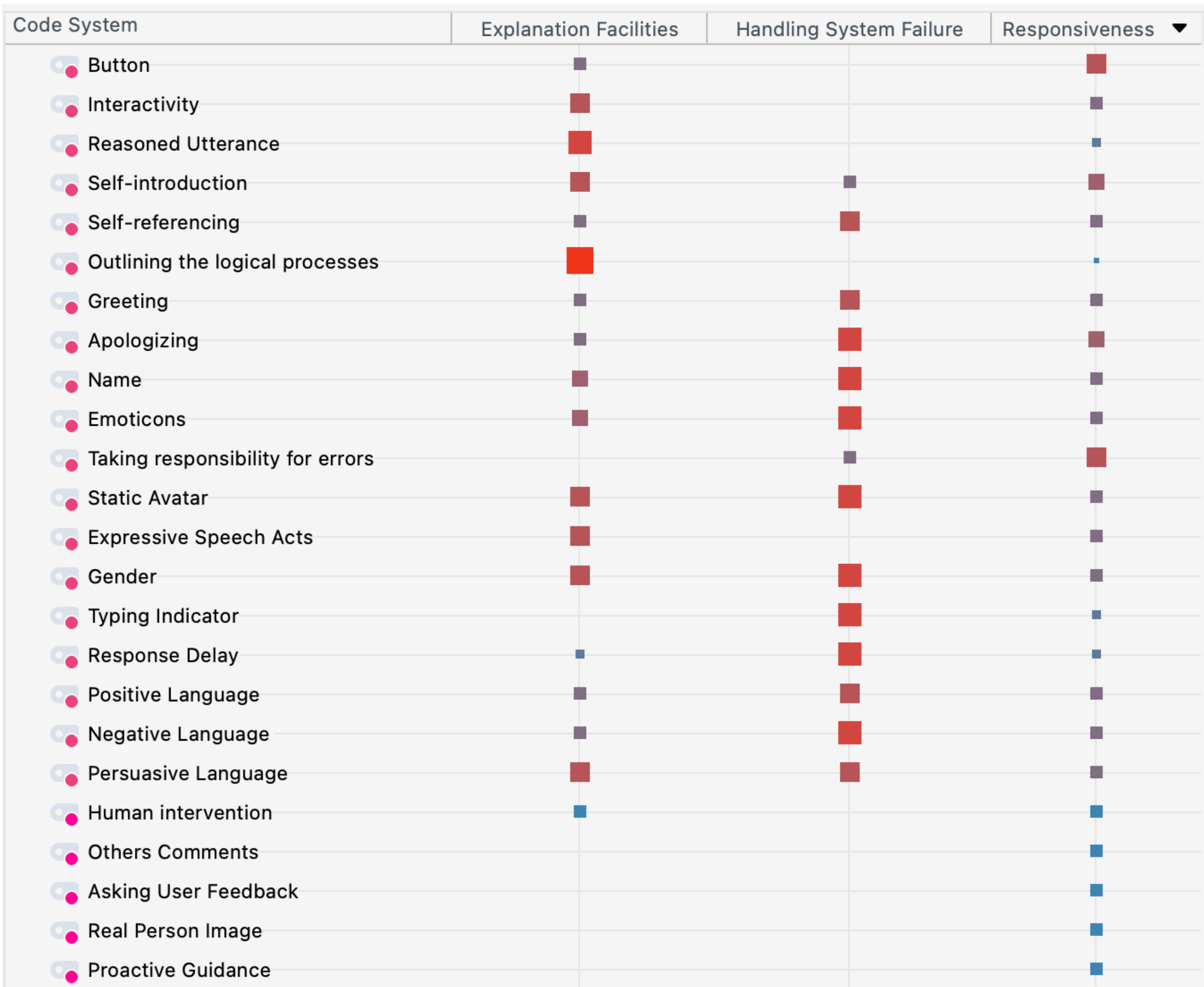
Agent competency refers to the perceived capability of a conversational agent to perform its intended functions effectively, particularly in terms of communication quality and task support. This dimension encompasses explanation facilities, handling of system failures, and responsiveness, all of which reflect how competently the agent manages interactions and supports user goals.

Explanation facilities were implemented in 12 studies. These features allow the agent to clarify its reasoning or justify its responses. For instance, Mehrotra et al. (2024) described a system where the AI assistant elaborated on its recommendations by explicitly stating, "The AI assistant also explained the reasoning for an answer by providing explanations." Such transparency in reasoning is particularly relevant in domains like recommendation, education, or healthcare, where users often need to understand why a particular response or suggestion was made to develop trust and confidence in the system. In recent research on explanation facilities, several key elements have been prominently featured (Figure 30). One such element is outlining logical processes to detail the reasoning behind agent responses or actions explicitly, thereby enhancing transparency and user understanding (Greulich &

Morana, 2023; Nguyen, et al., 2021). Reasoned utterances element (Al-Natour et al., 2021; Mehrotra et al., 2024) involves agents providing detailed explanations that articulate the rationale behind agent actions or recommendations. Furthermore, interactivity and expressive speech acts have been utilized to convey information considering emotions (J. Guo et al., 2021; Kang & Gratch, 2014).

Figure 30

Agent Competency Dimensions and Associated Design Elements



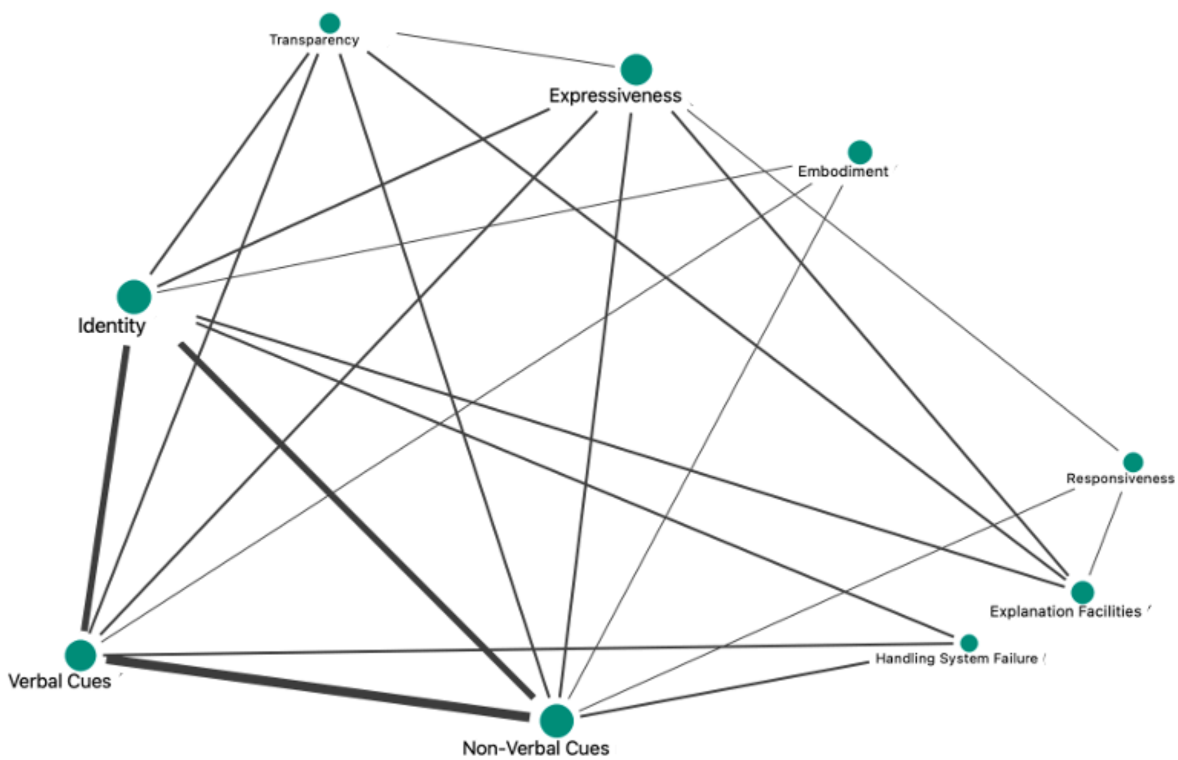
Multiple codes could be assigned to the same study (e.g., identity and verbal cues coded together). The color gradient is applied per row, meaning that blue indicates lower frequency and red indicates higher frequency of association within that element. In addition, the size of the squares reflects the strength of the association between the design element and the dimension.

Source: Created by the author.

The integration of specific linguistic strategies in agent design has been critically examined in a study on how systems handle failures. This includes using apologies to address errors or misunderstandings (Diederich et al., 2020b) and employing negative and positive language to manage user expectations and emotions (Diederich et al., 2021). Additionally, the study reviewed the utilization of avoidance language (Hildebrandt et al., 2023b) to avoid conversations about potential system limitations or errors. Humorous responses (Y. Yang & Kankanhalli, 2023) were also considered for their role in mitigating frustration during system failures (Figure 30). Regarding responsiveness, buttons have been utilized (Cai et al., 2023) to provide a direct method for users to interact with the agent, facilitating more straightforward navigation and decision-making. Interactivity elements (Go & Sundar, 2019; Poser & Bittner, 2023) have supported multi-turn conversations and tracked users' previous responses.

Figure 31

Design Dimensions Network Map



Source: Created by the author.

The responsiveness pattern also incorporates taking responsibility for errors, other users' comments, and asking for user feedback (Figure 30). These elements contribute to the user's perception of the agent as competent. While not all studies labeled these strategies as "competency" dimensions, their functional role in enhancing communication, managing failure, and providing intelligible responses justifies their classification under this dimension.

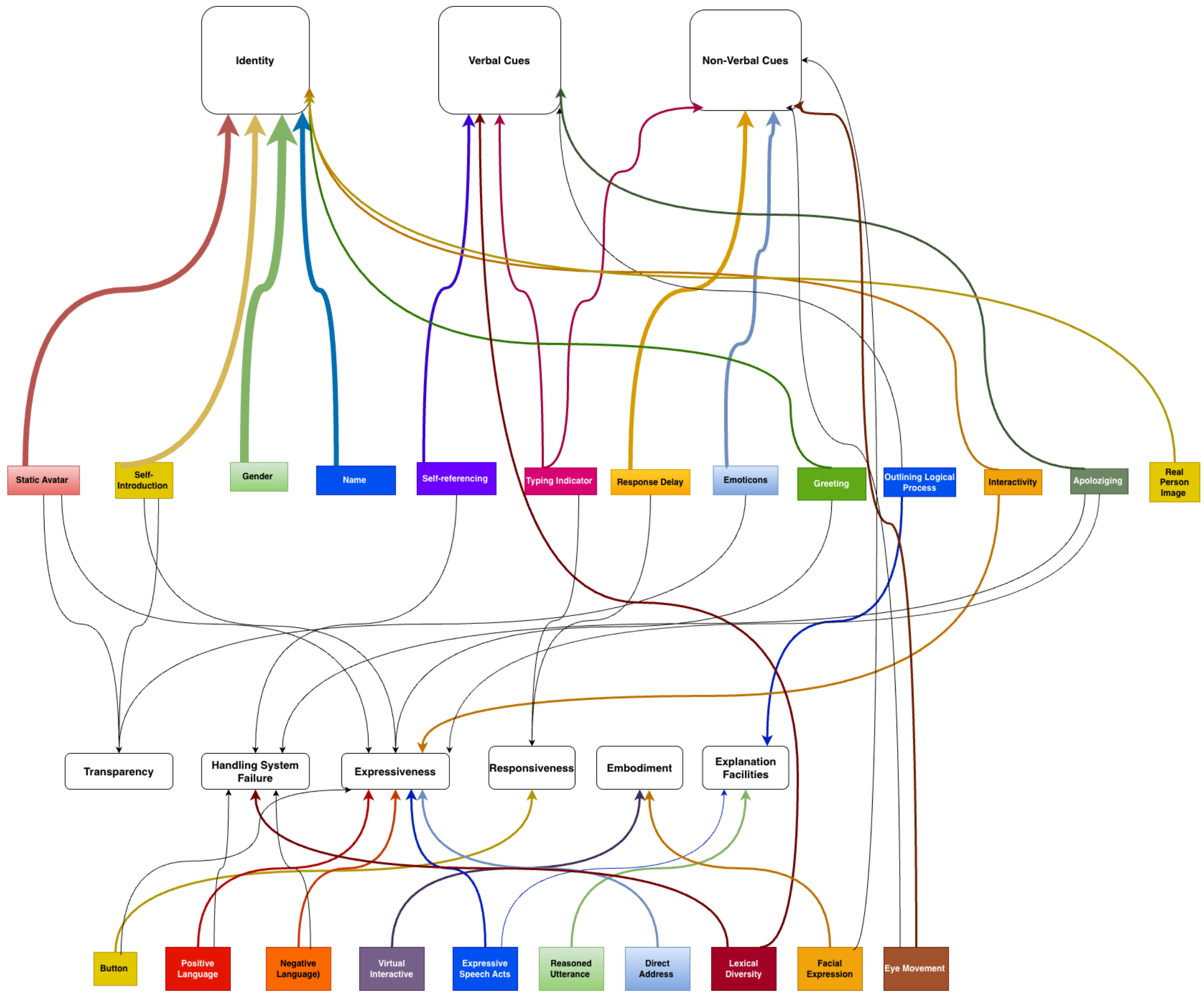
4.2.4. Empirical Instantiation of Design Dimensions

In many cases, reviewed studies combined multiple design dimensions to enhance the design, as illustrated in Figure 31. For instance, to handle conversational system failures, scholars incorporate verbal cues, using apologies and use emoticons, self-referencing, name, or gender (Riquel et al., 2021b). While studies try to enhance verbal communication using lexical diversity, they also incorporate identity cues with name, gender, emoticons, or a static avatar (Diederich et al., 2020b). This integration reflects the holistic design goal of enhancing user experience (Hassenzahl & Tractinsky, 2006).

Furthermore, previous research on CA design defines each dimension implementation using design elements independently (Diederich et al., 2020a; Seeger et al., 2017); this dissertation shows that many design elements span multiple dimensions (Figure 32). According to Figure 32, static avatar usage refers to giving identity cues to agents. The reviewed literature indicates that a static avatar is employed to make the agent transparent. In real implementations, designers do not treat verbal, expressive, or competency features as discrete categories. Critically, these overlaps offer opportunity to explain why some features, even when minor, have strong effects on user perception and trust.

Figure 32

Mapping Design Elements and Dimensions



Source: Created by the author.

4.2.5. Design Effects on User Experience

This review has revealed many user evaluation outcomes (Appendix D.1). Sections 4.1.1 and 4.1.4 present design choices associated with user perceptions and outcomes after interaction with CAs. In this section, the review deepens the understanding of how design affects user experience, including reviewed publication findings on CA user interaction, labeling findings as positive, negative, or no-significant effect. This is demonstrated in Table 22 for each design dimension. A total of 441 evaluations were identified across eight outcome categories: perception, trust, attitude, acceptance, emotion, performance, learning, relationship, and other. Each outcome is shown with its frequency across relevant design dimensions, and a percentage indicating the proportion of evaluations reporting a positive effect is provided for each design dimension.

Regarding positive effect rates, explanation facilities yielded the highest proportion of positive outcomes (97.34%), followed by identity (77.64%), verbal cues (75.36%), non-verbal cues (71.28%), and embodiment (70.67%). Among the evaluated user outcomes, positive effect rates for attitude, perception, and trust ranged between 72% and 77%, depending on the associated design dimension, while relationship showed a 100% positive rate; however, given that this result was based on only two studies, relationship outcomes were excluded from interpretation. Emotion exhibited the lowest overall positive effect rate (56%). The review indicates that studies incorporating dimensions such as explanation facilities, identity, and verbal cues were most frequently associated with favorable user experience outcomes. Negative effects were commonly reported in studies addressing handling system failures, accounting for the highest share of negative evaluations. In contrast, explanation facilities and transparency did not yield any reported negative effects. Additionally, expressiveness (44%) showed the highest rate of "no effect" outcomes, followed by transparency (27.28%) and non-verbal cues (23.53%), indicating that not all design interventions produced measurable user experience impacts.

Table 22

Conversational Agent Design Dimension and User Evaluation Outcomes

Interaction Outcomes	Design Dimensions (percentages show the positive effect rates per dimensions)										
	Explanation Facilities	**Handling System Failure**	**Responsiveness**	**Expressiveness**	**Transparency**	**Embodiment**	**Identity**	**Non-Verbal Cues**	**Verbal Cues**	**Total Evaluations**	**Positive Rates (%)**
Positive Rates (%)	**97.34**	**67.93**	**66.67**	**64.9**	**72.72**	**70.67**	**77.64**	**71.28**	**75.36**	**72.68**	
Negative Rates (%)	**0**	**18.76**	**11.11**	**2.27**	**0**	**2.70**	**5.68**	**5.19**	**6.17**	**4.97**	
No-Effect Rates (%)	**3.46**	**12.51**	**22.22**	**32.44**	**27.28**	**29.73**	**16.68**	**23.53**	**18.47**	**22.35**	
Perception	6	8	3	16	10	8	33	40	39	163	**74**
Trust	6		6	15	2	6	14	9	11	69	**72**
Attitude	2	4		8	1	4	13	10	15	57	**77**
Acceptance	2		9	3	3	3	10	3	4	37	**68**
Emotion	1	2		11	1	6	7	7	5	40	**56**
Performance	6		6	3	3	2	3			23	**71**
Learning				4		1				5	**60**
Relationship				2						2	**100**
Other	3	1		12	3	3	8	8	7	45	**63**

Source: Created by the author.

Regarding the anthropomorphic design dimensions, the primary investigated outcome was perceptions. The research examined the impact of agent designs that included or omitted verbal, non-verbal, and identity cues, evaluating the effects through constructs such as aggression (Brendel et al., 2023), social presence (Cafaro et al., 2016), competence (T. Cui et al., 2021), service satisfaction (Diederich et al., 2020b), and purchase intention (Konya-Baumbach et al., 2023). Concerning the agent characteristics, CAs researchers studied different constructs, for instance, the learning construct (studied only agent characteristics), perception, attitude, and acceptance. Ceha & Law (2022) observed that expressive design features, including positive language and tone modulation, enhance learning gains in pedagogical agents. Additionally, Ciardo et al. (2022) evaluated the impact of erring conditions in physically embodied agents. Lastly, agent competency, our corpus revealed the studies utilized constructs such as satisfaction (Al-Natour et al., 2022), familiarity (Diederich et al., 2020b), and self-disclosure (Al-Natour et al., 2021) in the context of attitude. Additionally, the studies on agent competency focused on performance constructs, highlighting their functional capabilities. For example, Poser & Bittner (2023) studied service quality by applying responsive design (outlining logical processes to disclose information). Similarly, Hernandez-Bocanegra & Ziegler (2023)explored different explanations using reasoned utterances and outlining logical processes for recommender agents and measured effectiveness. Finally, the reviewed studies emphasize that user evaluations are influenced not only by the functional capabilities of conversational agents and how well the agents align with users' social, emotional, and task-related expectations. In particular, users form immediate judgments based on observable features such as response style, tone, and fluency in interaction (Y.-C. Lee et al., 2020; Seeger & Heinzl, 2021). Existing research in human–agent interaction further indicates that misalignment between the agent's behavior and user expectations may result in confusion, frustration, or rejection of the agent (Liao et al., 2018). According to the review, the implementation of CA design must align with the desired interaction standards (Feine et al., 2019; Luger & Sellen, 2016).

4.2.6. Development of User Experience-Oriented Design Requirements

Throughout the design cycles (explained in section 3.3.2), a deep understanding of the conversational agent design to improve user experience elicited 14 design requirements. Then, each requirement was framed with a set of design dimensions. The motivation behind this Design Science Research endeavor was to present an actionable tool that enhances user experience with CA beyond its functions. For the empirical studies, the university students, including bachelor's, master's, and PhDs, participated in the initial requirement elicitation process as CA users. Despite being students, their diverse prior experiences with conversational agents indicate that they were not novice users, thus contributing meaningful and context-aware feedback. Furthermore, the studies reviewed during the literature synthesis phase included diverse user populations, broadening the representativeness of the findings.

4.2.6.1. Design Practices in Reviewed Literature

This study's initial set of requirements was informed by a structured review of design practices reported in leading HCI and IS conferences, following a literature-driven requirement elicitation approach (Carrizo et al., 2014; Hickey & Davis, 2004). The process began with identifying recurring design considerations across empirical and conceptual works. This approach follows the recommendations that require derivation from prior research to extract functional and non-functional needs and capture domain-specific patterns and best practices observed in application contexts.

Conference contributions revealed consistent attention to several thematic areas that map closely to the requirement dimensions used in this research. First, agent identity and social cue design were emphasised as central to fostering trust, recognisability, and engagement (K. M. Lee et al., 2006; Nass & Moon, 2000; Qiu & Benbasat, 2008). Second, natural timing, feedback behaviours, and adaptive interaction flows were identified as important for sustaining user attention and creating human-like interaction experiences (T. Cui et al., 2021; Gnewuch et al., 2018; N. Wang & Gratch, 2010). Third, studies revealed the role of transparent communication of capabilities and limitations to support realistic expectations while reducing user frustration (Luger & Sellen, 2016; Seeger & Heinzl, 2021). Lastly, adaptability to context, support for goal completion, and responsiveness to user input were core functional capabilities discussed across multiple application domains (Aktaş & Akbıyık, 2025).

By synthesising these themes into a structured set of initial requirements, this study builds on the notion that literature- and conference-based analysis can serve as a valuable early-stage requirement elicitation method, bridging empirical evidence with the specific design objectives of a new system (Aurum & Wohlin, 2005; Carrizo et al., 2014). The resulting framework provided a grounded starting point for subsequent validation and refinement through user-centred evaluation in the present research.

4.2.6.2. Users' Feedback on CA Design

Analysis of the open-ended responses regarding perceived benefits and desired features led to the formulation of design requirements as a part of the first design cycle. Participants provided open-ended responses about the benefits of using the chatbot and suggestions for new features after interacting with the chatbot described in section 3.4.5. While some responses described specific interaction episodes (e.g., finding a deadline or clarifying a requirement), others reflected general evaluations of the chatbot attributes, and a smaller number offered high-level improvement ideas without referencing a specific use case.

The most frequently mentioned benefits are summarised in Table 23. Nearly half of the participants reported that the dominant benefit was reduced effort required to locate information, followed by improved access to course-specific or university policy knowledge. Other reported benefits included increased accuracy of information, enhanced understanding or learning, faster responses, and improved engagement with course content.

Table 23

Reported Benefits of Chatbot Use from Open-Ended Responses

Benefits	Number of participants
reduce effort in finding source/information	48
course specific/UNI rules knwoledge	19
accurate information	9
help understanding/learning	7
access to related course content	3
fast response	3
reaching right information	2
referencing source	2
clear structure	1
guidance	1
improve critical thinking	1
improve engagement with course content	1
integration with university system	1
organizational support	1

Source: Created by the author.

Following an inductive thematic analysis (Ezzy, 2002), three main categories emerged after analyzing benefits and suggested features: efficiency and accessibility (46%), information presentation and transparency (32%), and expanded functional capabilities (22%). Participants frequently described the chatbot as a time-saving, reliable tool to access information. These responses emphasized reducing effort to navigate multiple sources, quicker access to deadlines or policies, and the ability to retrieve targeted answers in complex environments.

> "It saves user time and easily reaches the right information." (P54)
>
> "It is much faster and easier than to look on your own and it gives you precise details and instructions." (P31)

Many participants valued clear, structured, and source-transparent responses. Suggestions included bullet points, tables, and visual summaries to improve readability, as well as explicit links or references to the original course materials. This was seen as both reducing cognitive load and enhancing trust.

> "That I know where the information is coming from, and it is very specific to my course." (P45)
>
> "Clearly answer each question raised with bold highlighted key information." (P19)
>
> "Maybe using links to reference the website where it pulls the answer." (P17)
>
> "Link of reference." (P42)

Some responses proposed new or extended functionalities, such as visual aids for explanations, calendar integration, execution of code for programming tasks, and academic support tools (e.g., plagiarism checks, APA referencing). Several participants also suggested adaptive interaction modes, enabling users to select concise or detailed responses based on preference.

> "Providing visual diagrams step-by-step for guidance." (P12)
>
> "Maybe an option to reply to specific information rather than the whole answer." (P8)

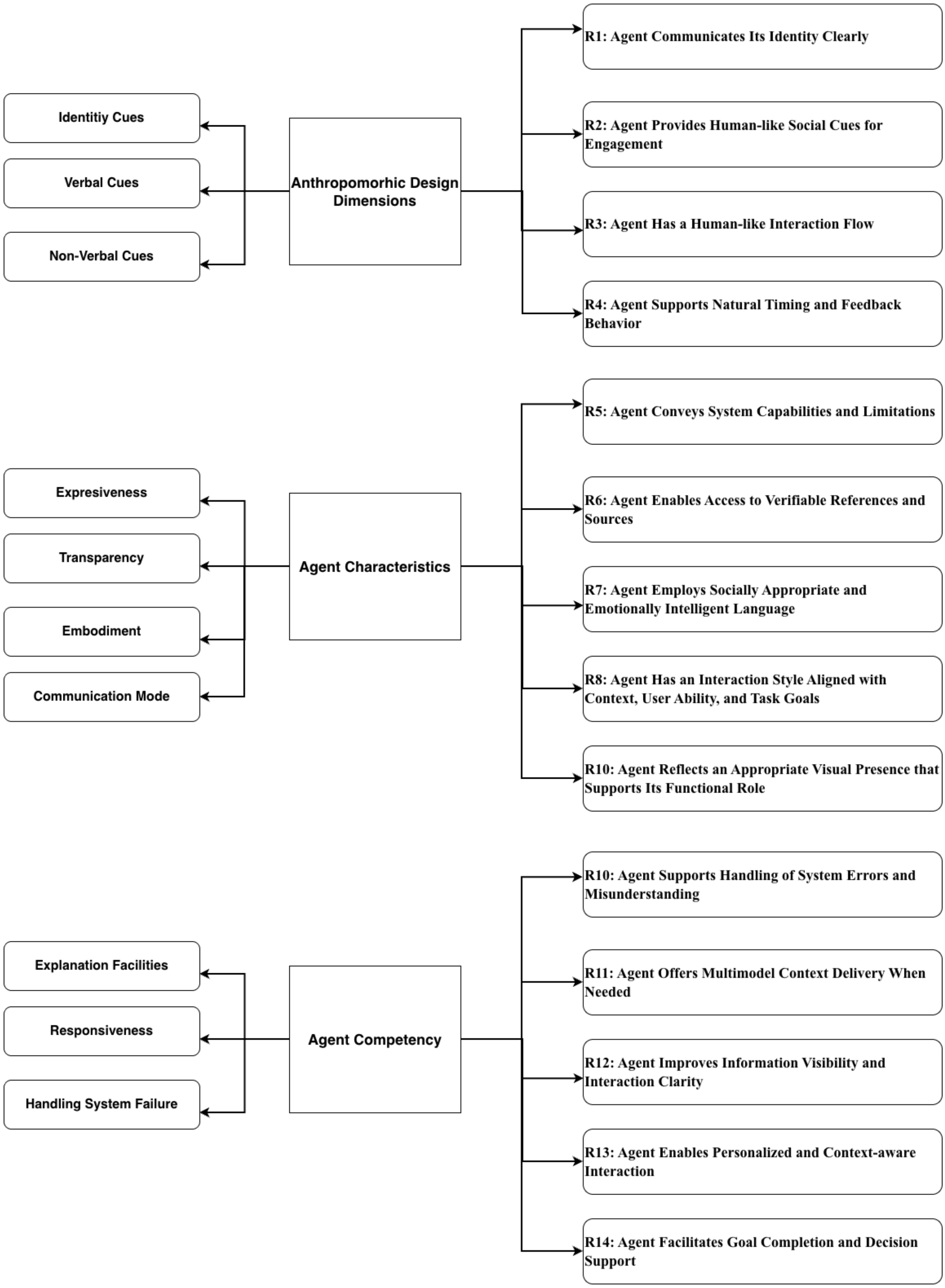
These thematic insights informed the refinement of the requirement set presented later in this section. In particular, user calls for verifiable references, multimodal content delivery, and enhanced information visibility were integrated into the structured requirements framework.

4.2.7. Formulation of Meta-Requirements

This dissertation identified design requirements to enhance the overall user experience with conversational agents. These requirements reflect efforts to support socially and emotionally meaningful interactions through design, rather than merely enabling task completion. In line with the view that design theories should generate prescriptive yet adaptable knowledge (Walls et al., 2004), the requirements formulated here do not assert fixed causal effects. Instead, they represent meta-designs, abstract design representations guided by theoretical insights and empirical observations, which open up pathways for user-centered refinement and contextual application. Further, these requirements are structured to articulate the material properties of the agent (e.g., embodiment, response features) and the user actions they are intended to support (e.g., trust formation, engagement). This dual orientation helps to bridge the gap between abstract design concepts and real-world application, emphasizing the generative role of requirements as heuristic design knowledge. As such, the requirements proposed in this study offer actionable, theoretically informed guidance for CA design while remaining open to reinterpretation across different domains and user contexts. The meta requirements were organized around three design dimensions in Figure 33. This structuring supports the abstraction and generalization of findings within the DSR paradigm.

Figure 33

Meta-Requirements Informed by the Agent Design Dimension



Source: Created by the author

4.2.7.1. Designing an Agent with Humanlike Qualities is Crucial

Anthropomorphic design reflects a long-standing interest in how humanlike design elements shape user perceptions, grounded in media equation theory and related perspectives on social responses to technology (Nass & Moon, 2000; Reevess & Nass, 1996). Humans tend to anthropomorphize digital agents by attributing social roles and human traits even in minimalistic designs. As such, designing agents with features that reflect human identity, social behavior, and communicative cues can significantly improve user experience. The literature review has revealed that 75% of the studies addressing anthropomorphic features reported overall positive effects. More specifically, these design choices were most strongly and positively associated with outcomes such as acceptance (positive evaluation rate: 90.5%), attitude (85%), trust (87.5%), and perception (79%). This suggests that anthropomorphic design supports interaction's relational, affective, and social aspects more robustly than performance outcomes, which showed only a 60% positive rate (with just 10 evaluations).

Requirements 1-4 (Figure 33) refer to the humanlike interaction aspects for CAs. R1 implies identity clarity, which is supported by findings that emphasize the role of agent identity (e.g., name, gender, and visual representation) in increasing perceived similarity and trust. For example, identity cues enhanced social presence and empathy when tailored to users' cultural backgrounds and expectations (Ben Mimoun et al., 2017; Benbasat et al., 2020). R2 focuses on providing human-like social cues (Feine et al., 2019), which is grounded studies that found positive impacts from non-verbal cues such as gesture, gaze, head movement, smile, and listening behavior. These cues improve realism and foster stronger perceptions of social presence (Appel et al., 2012; Gnewuch et al., 2018). R3 addresses human-like interaction flow, which is closely linked to interactional behaviors such as typing indicators, turn-taking, and response delays. These features help agents better simulate conversational rhythm and attentiveness. Supporting this, verbal cues such as greetings, apologies, and lexical variety improved users' enjoyment, trust, and willingness to reuse the system (Chin & Yi, 2019; Chung et al., 2023; Morana et al., 2020). R4 complements the anthropomorphic design perspective by addressing how agents manage conversational rhythm. Specifically, natural timing and feedback behaviors (such as pauses and typing indicators) are essential non-verbal cues that make interactions feel more humanlike. These behaviors help users feel heard and socially engaged (Araujo, 2018). Our review shows that such cues have been widely adopted to enhance perceived social presence and emotional connection (Appel et al., 2012; Gnewuch et al., 2018).

These findings suggest that anthropomorphic design should not be treated as aesthetic augmentation but rather as a strategic approach to improving user engagement, trust, and emotional connection. R1–R4 capture these imperatives by addressing the importance of identity, humanlike behavior, and social communicative rhythm, which contribute to richer and more satisfying interactions with conversational agents.

4.2.7.2. Agent Characteristics as the Cornerstone of Design

The agent characteristics dimension refers to design properties that define how the agent functions within its interaction context, including the style, tone, and manner in which it communicates information (Diederich et al., 2022). According to our SLR results, agent characteristics showed a moderate (68%) overall positive effect rate. Specifically, acceptance (76.7%), perception (71.4%), emotion (60.9%), and trust (63.4%) were positively associated with characteristic-related design elements. No negative effects were reported for trust outcomes. These results highlight the relevance of agent characteristics for improving key user experience outcomes in conversational agent interactions.

Given the high variability of user input in natural language interactions compared to graphical interfaces (Brandtzaeg & Følstad, 2017), conversational agents must be capable of effectively managing and conveying information flexibly yet coherently. Requirements 5–9 (Figure 33) address these communicative functions, ensuring the agent can deliver information, adapt to user input, and maintain interaction quality across varied expressions. R5 focuses on agent transparency, particularly in communicating system capabilities and boundaries. The literature confirms that transparency was one of agent design's safest and most positively rated elements, with 0% negative effects reported. Transparent agents help users develop accurate mental models of the system's functionality, reducing confusion and managing expectations (Reimann et al., 2025). Self-introductions, disclaimers, and upfront communication about limitations were common strategies to implement this requirement. R6 further extends the notion of transparency by advocating for the inclusion of verifiable references and links (Sedrakyan et al., 2024b). This was especially highlighted in participant feedback during the empirical studies, where users requested access to sources or supporting evidence to increase confidence in the agent's answers.

R7 centers on the agent's ability to employ emotionally intelligent and socially appropriate language. The literature on expressiveness shows that agents using empathetic, polite, and affect-sensitive language improve attitudes and emotional outcomes. For instance, agents that adjust tone, avoid harsh language, or express politeness and understanding tend to promote more positive user perceptions (Al-Natour et al., 2021, p. 201; Ceha & Law, 2022). However, over-apologizing or emotional exaggeration can have the opposite effect, underlining the importance of calibrated emotional expression (Bowman et al., 2024; Brendel et al., 2020). R8 introduces the need for interaction styles to be adaptive and aligned with user abilities, context, and task goals, communication modes, whether text or voice, should be selected based on the target environment. For example, voice-based agents are more prevalent in healthcare and virtual environments, while text-based agents dominate service and e-commerce settings. The literature review found that design effectiveness varied across these modes, especially in expressiveness and embodiment. This highlights the importance of aligning interaction modalities with user context and intended agent functionality (W. Terblanche et al., 2023). R9 complements this by calling for the agent's visual embodiment to reflect its functional role. Embodiment research in the reviewed studies often emphasized physical traits such as eye gaze, gestures, and smile, which increase perceived presence and trustworthiness (Felnhofer et al., 2019). The reviewed literature also noted that mismatches between embodiment and purpose can undermine agent credibility. Thus, visual presence should be strategically designed to align with the agent's communicative purpose and social role.

4.2.7.3. Agent Competency Elevates Conversational Agents to New Heights

As conversational agents are increasingly deployed across diverse domains, their role extends beyond basic functionality such as information retrieval or navigation. They are expected to demonstrate greater adaptability, responsiveness, and task-support capabilities (Jiang et al., 2024). Integrating the concept of agent competency into the CA design framework (Diederich et al., 2022), as suggested by Candra et al. (2022), addresses this shift in expectations. In this research, this dimension reflects a move from humanlike traits toward task- and interaction-oriented intelligence (Ternyak, 2023). The result of the

SLR analysis revealed that agent competency studies were concerned with performance and trust, most notably, which yielded a 100% positive evaluation for trust and a 92.7% positive evaluation for performance outcomes. In addition, the positive effects are reported for perception (76.6%), attitude (100%), and acceptance (64.3%) outcomes. Critically, emotion-related evaluations had a much lower positive association (18.75%, 40.6% no effect, 40.6% negative), suggesting that while agent competency effectively supports task performance and trust, it may not be sufficient to foster emotional engagement.

Requirements R10–R15 (Figure 33) define the core capabilities that support agent competency.

The agent's ability to handle system errors and misunderstandings was addressed in R10. According to studies, using polite language, apologies, or humor when a problem arises can reduce tension and help create a positive user experience (Brendel et al., 2020; Diederich et al., 2020a). R11 recommends giving information in multiple formats (e.g., text, links, examples) to clarify interaction and support users with different needs or preferences (Shevat, 2017). Similarly, R12 focuses on visibility and interaction clarity, facilitating information handling within the interface (e.g., expandable sections). R13 encourages personalized and context-aware responses. Responsively designed agents reference prior user input, ask follow-up questions, or use persuasive language to improve satisfaction and trust (Cai et al., 2022; Poser & Bittner, 2023). R14 centres on supporting users in completing tasks and making decisions. This includes scaffolding user tasks, suggesting next steps, or summarizing options. Such features are central to improving task efficiency and usefulness. For instance, providing explanations with logical reasoning contributes significantly to perceptions of transparency, trustworthiness, and competence, while giving recommendations or feedback to users (Hernandez-Bocanegra & Ziegler, 2023). Collectively, requirements 10–14 reflect the evolving expectation for CAs to go beyond basic interactivity by demonstrating task-oriented competence, contextual sensitivity, and responsive communication.

4.3. Rigor Cycle

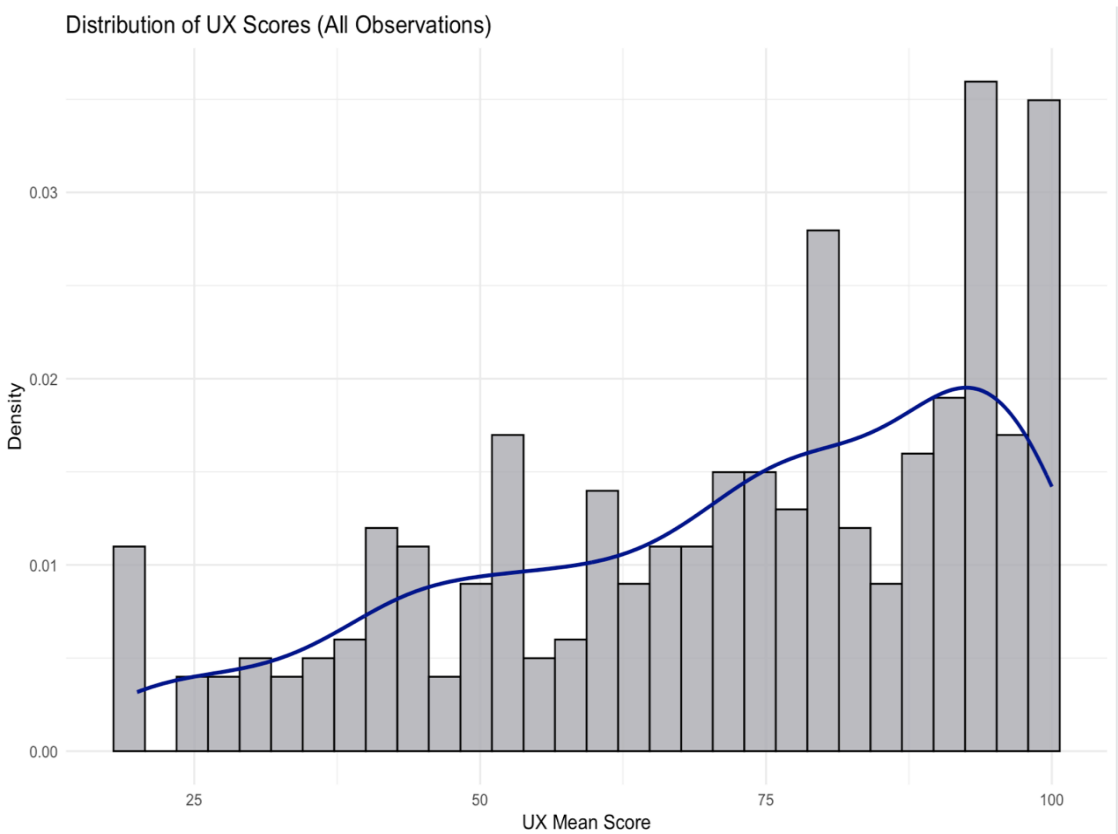
The rigor cycle aimed to establish the validity, reliability, and generalizability of the design artifact of this thesis and its associated measurement model through empirical evaluation. This phase focused on two complementary objectives: (1) validating the psychometric robustness of the user experience measurement scale, and (2) empirically examining the effects of different chatbot design implementations on user experience outcomes. In doing so, the rigor cycle provided systematic evidence to demonstrate that the developed artifact and constructs are both theoretically and empirically supported (Gregor & Hevner, 2013; Hevner, 2007).

4.3.1. Scale Validation

In this phase, data were collected through an online questionnaire consisting of three phases: (1) introduction and demographics, (2) chatbot selection and interaction, and (3) post-interaction evaluation. Participants evaluated between one and six chatbots, with most participants (N=106 of 155) evaluating three chatbots on average. This resulted in 363 independent evaluations. Each evaluation represented a distinct chatbot interaction and was treated as an independent observation for scale validation purposes.

Figure 34

Histogram for User Evaluation Scores



Source: Created by the author

Descriptive statistics and exploratory data analysis were conducted to assess data characteristics. The Kolmogorov-Smirnov test was employed to examine normality (appropriate for sample sizes > 30), supplemented by histogram visualization (Figure 34). Results indicated a significant departure from normality ($p < .001$), with the distribution displaying negative skewness and multiple local peaks. A substantial proportion of scores clustered at the higher end (80–100), while few participants reported low user experience scores (< 40). Moderate evaluations were also present in the midrange (40–60).

Descriptive statistics for all 16 items are presented in Table 24. Mean scores ranged from 62.56 (CI_2) to 81.16 (b5), indicating generally positive evaluations across items. Items b2, b3, b5, and b9 exhibited higher mean scores (above 75), In contrast, the lowest means were observed for CI_1 (M = 63.78) and CI_2 (M = 62.56), reflecting comparatively lower ratings for continuance intention. Standard deviations ranged from 22.57 to 29.18, indicating substantial variability in responses. All items displayed negative skewness values, and Kurtosis values ranged from -1.37 to 1.03, with most items displaying platykurtic distributions (flatter than normal).

Table 24

Descriptive Statistics of Overall User Experience Items (N=363)

Item	Mean	SD	Skew	Kurtosis
b2	78.02	26.10	-0.99	-0.24
b3	75.48	25.87	-0.83	-0.55
b4	72.40	27.20	-0.69	-0.78
b5	81.05	22.79	-1.26	0.81
b6	72.29	27.57	-0.68	-0.86
b9	77.25	25.21	-0.93	-0.25
sp4	73.22	27.68	-0.78	-0.82
enj1	63.80	27.04	-0.19	-1.19
enj4	65.18	26.67	-0.30	-1.08
t8	74.33	27.08	-0.78	-0.7
t9	73.94	26.94	-0.70	-0.8
t10	71.96	26.87	-0.68	-0.8
Us1	68.15	26.96	-0.43	-1.04
Us4	74.49	27.06	-0.88	-0.48
CI_1	63.53	28.72	-0.18	-1.32
CI_2	62.48	29.04	-0.12	-1.37

Source: Created by the author.

Since the instrument was designed to measure a single latent construct, the objective was to confirm whether the observed items load reliably and validly on a common factor. The suitability of the data for factor analysis was verified using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity. The overall KMO value was 0.96, indicating excellent sampling adequacy (Kaiser, 1974). Bartlett's test of sphericity was significant, $\chi^2(120) = 6438.967$, $p < .001$, confirming that the correlations among items were sufficiently large for factor analysis. Additionally, previous scale development study and pilot study resulted in a one-factor structure for a 16-item measurement scale.

The internal consistency of the 16-item scale was assessed using Cronbach's alpha, which yielded an excellent value of $\alpha = 0.97$, substantially exceeding Nunnally's (1978) recommended threshold of 0.70. This indicates strong reliability and suggests that items consistently measure the same underlying construct.

After checking the factor structure of the measurement scale, BCFA was performed using the 'blavaan version 0.5.8' package. Factor loadings were evaluated based on posterior distributions, with a lower 95% credibility bound > 0.60 taken as evidence of sufficient loading strength (Borsci & Schmettow, 2024). The model demonstrated satisfactory convergence ($\hat{R} < 1.05$; ESS > 400). Standardized factor loadings ranged from 0.40 to 0.94 (M = 0.84, SD = 0.14) (Table 25). All 95% credible intervals excluded zero, indicating statistically meaningful item-factor relationships. However, item b2 showed a notably weaker standardized loading ($\lambda = 0.39$, 95% CI [0.31, 0.46]) and high residual variance (85.5%), suggesting it contributes less to the common factor and retains substantial unique variance. Despite this weak loading, b2 was retained in the final scale for two reasons: (1) removing it would minimally affect reliability (ω would decrease from 0.98 to 0.97), and (2) the item provides important content validity by capturing a unique aspect of the user experience (accessibility of CAs) construct that would be lost if removed. The overall scale exhibited excellent psychometric properties: AVE = 0.70, composite reliability = 0.98, and omega = 0.98, all exceeding recommended thresholds. All thresholds met except Posterior Predictive p-value (PPP = .000). As highlighted by Hoofs et al. (2017), the PPP statistic approach zero in large samples, even when the model is well-specified, due to the test's increasing sensitivity to trivial misfit. Consequently, the low PPP in this case is not necessarily indicative of substantive model inadequacy, while the overall evidence still supports the model's practical usefulness and validity.

Table 25

Standardized Factor Loadings from Bayesian CFA

Item	Posterior Mean (λ)	SD	95% Credible Interval	Standardized Loading (Std.all)	Residual Variance (Std.all)	$\hat{R}$
b2	1.000 (marker)	—	—	0.396	0.855	—
b3	2.235	0.353	[1.700, 3.096]	0.833	0.306	1.009
b4	2.374	0.375	[1.805, 3.276]	0.841	0.293	1.008
b5	1.711	0.277	[1.287, 2.372]	0.728	0.470	1.008
b6	2.587	0.403	[1.978, 3.557]	0.901	0.189	1.009
b9	2.256	0.354	[1.720, 3.107]	0.861	0.258	1.008
sp4	2.599	0.406	[1.989, 3.585]	0.901	0.188	1.009
t8	2.621	0.408	[2.009, 3.614]	0.927	0.140	1.009
t9	2.628	0.408	[2.019, 3.621]	0.935	0.126	1.009
t10	2.530	0.395	[1.937, 3.479]	0.904	0.183	1.009
Us1	2.292	0.363	[1.736, 3.171]	0.821	0.327	1.008
Us4	2.535	0.395	[1.941, 3.483]	0.899	0.191	1.009
enj1	2.316	0.367	[1.758, 3.195]	0.826	0.318	1.008
enj4	2.041	0.329	[1.534, 2.834]	0.743	0.449	1.008
CI_1	2.471	0.390	[1.877, 3.400]	0.830	0.311	1.008
CI_2	2.452	0.389	[1.862, 3.379]	0.815	0.335	1.008

Source: Created by the author.

Further, to examine the unidimensional structure of the scale, Explained Common Variance (ECV), Marginal Reliability (ρ^2), Proportional Reduction of Mean Squared Error (PRMSE), and Expected Percentage of True Differences (EPTD) were calculated (Calderón Garrido et al., 2019). According to Table 26, the dimensionality indices collectively indicate that the scale measures a single coherent construct. All values meet or exceed recommended thresholds, confirming that the variance among items is largely explained by one dominant factor, consistent with essential one-dimensionality (Sheng & Wikle, 2007; Strout, 1990).

Table 26

Unidimensionality Indices for the User Experience Scale

Index	Value	Interpretation
ECV	0.68	Proportion of common variance explained by general factor
ρ^2	0.97	Reliability of the general factor
EPTD	97.14%	Percentage of true score differences
PRMSE	0.97	Incremental gain from subfactors ($\approx\rho^2$)

ECV $\geq$ .70–.85, $\rho^2 > 0.80$, EPTD > 0.90, PRMSE $\approx \rho^2$
Source: Created by the author.

4.3.2. Design Evaluations

Following the scale validation, the second stage of the rigor cycle focused on assessing user experience across multiple chatbot designs. This design evaluation examined whether differences in the implementation of design elements led to measurable variations in overall user experience. By combining subjective user evaluations with objective performance metrics (e.g., task success) and eye-tracking metrics, this phase provided a comprehensive understanding of how design choices influence user perceptions and interaction outcomes.

4.3.2.1. Analysis of Selected Service Chatbots

To contextualize the theoretical framework within real implementations, six publicly available chatbots (Table 27) were analyzed based on the agent design dimensions proposed in this dissertation. Each system was evaluated for the presence or absence of design elements (Appendix C.1).

Building on the elements mapping illustrated in Figure 32, this section assesses the service chatbot designs across design dimensions to determine how collectively fulfill design meta-requirements (Table 27). Regarding anthropomorphic design dimensions, this dissertation identified three subdimensions, including identity, verbal, and non-verbal cues. Moreover, section 4.2.3 defined the implementation of these dimensions with corresponding design elements. Importantly, these design dimensions operationalize the meta-design requirements R1-R4, which emphasize (R1) clarity of agent identity, (R2) social cues for engagement, (R3) human-like interaction flow, and (R4) timing and feedback to create natural interaction. Agent characteristics are conceptualized as communication mode, embodiment, transparency, and expressiveness. Implementing related design elements with these dimensions enables the fulfillment of meta-requirements R5-R9. Specifically, (R5) system capabilities and limitations identification, (R6) reliable reference access, (R7) socially appropriate and emotionally intelligent language, (R8) context and task-aligned interaction style, and (R9) appropriate visual presence consistent with its functional role.

Table 27

Service Chatbot Implementation of Design Dimensions

Design Dimension	Represented Chatbot(s)
Anthropomorphic Design	
Identity Cues	US Citizen, UTwente, Lufthansa,
Verbal Cues	Seattle Ballooning, Kia, Lufthansa, UTwente
Non-Verbal Cues	Seattle Ballooning, UTwente, Lufthansa, Wanderlog, Kia
Agent Characteristics	
Transparency	Seattle Ballooning, Wanderlog, UTwente
Expressiveness	Seattle Ballooning, Wanderlog
Communication Mode	All text-based modality
Embodiment	All use single body representation
Agent Competency	
Explanation Facilities	Seattle Ballooning, Wanderlog
Handling System Failure	Seattle Ballooning, Wanderlog
Responsiveness	Seattle Ballooning, US Citizen, Kia, Lufthansa, Wanderlog

The chatbots partially employed elements related to subdimensions
Source: Created by the author.

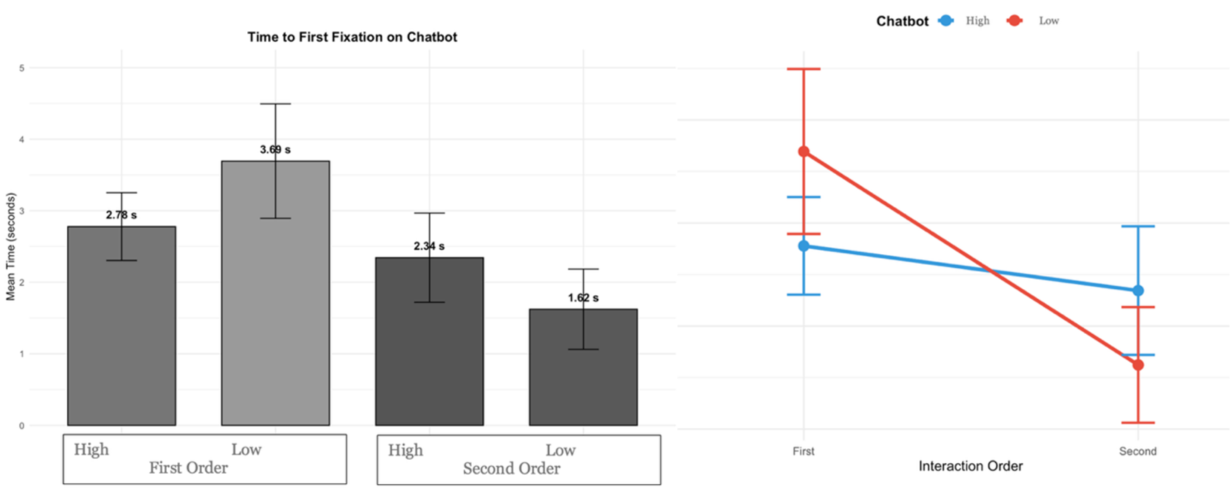
The agent competency dimension reflects the system's functional and cognitive capability to sustain coherent dialogue and manage conversational breakdowns. Section 4.2.3.3 defined its operationalization through elements such as explanation facilities, error handling, and responsiveness. It operationalizes meta-requirements R10–R14, which emphasize that (R10) supporting system failures, (R11) multimodal context delivery, (R12) interaction clarity, (R13) context-awareness, and (R14) facilitating goal completion. These requirements can be met through the combination of design elements that span multiple design dimensions. In practice, elements such as apologizing or lexical diversity simultaneously serve different functions, operating as verbal anthropomorphic cues, enhancing expressiveness, and strengthening competency through handling failures. Similarly, typing indicator contributes both to humanlike interaction flow (R3) and interaction clarity (R12) (e.g. ‘thinking’, ‘responding’). Table 27 illustrates that the selected service chatbots implement a range of observable design elements corresponding to the identified design dimensions. Among them, Seattle Ballooning demonstrates the broadest coverage across all three design dimensions. While not implemented perfectly, its design incorporates elements that collectively reflect these dimensions, indicating a greater level of integration than the other chatbots.

4.3.2.2. Visual Attention Results

The eye-tracking data were processed using a within-subject repeated-measures design. The analysis was conducted in three phases. The first phase included webpage snapshots displaying the chatbot bubble to examine how accessible each chatbot was. The second phase focused on the chatbot’s welcoming page snapshots, which contained different observable design elements. The final phase covered snapshots from the subsequent interaction with the chatbot. The analysis results present users’ visual behavior for each snapshot, including (1) time to first fixation and experimental setting effects, and (2) fixation durations and counts for each Area of Interest, the chatbot window, chatbot response area, button area, input area, appearance representation object (for one chatbot only), and the overall webpage to assess how effectively users engaged with the chatbot.

Figure 35

Time to First Fixation in Chatbot Activation Element by Presentation Order



Source: Created by the author.

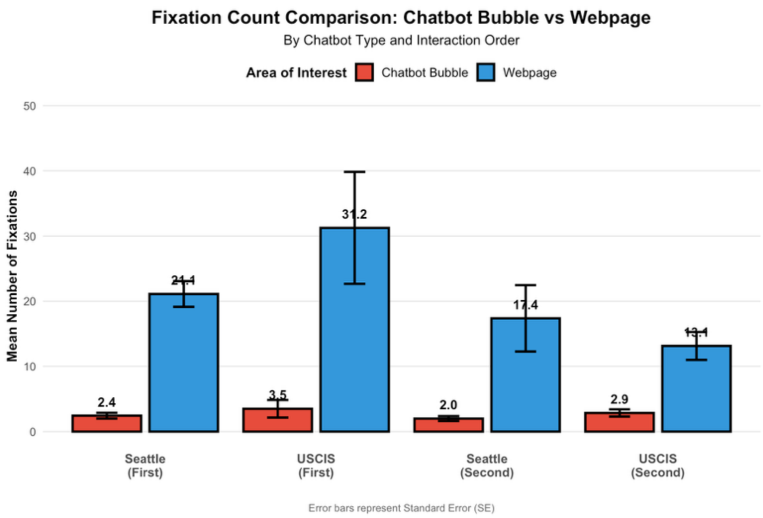
Analysis of time to first fixation (Snapshot 1, Chatbot Bubble) on the chatbot activation element revealed similar initial attention capture between the two chatbots, with Seattle (high) showing a mean TTFF of 2.57s (SD = 1.55) and USCIS (low) a mean TTFF of

2.67s (SD = 2.09). Figure 35 demonstrates time to first fixation by chatbot type and interaction order. The results show that Seattle (high) had shorter TTFF in the first exposure (M = 2.78s) compared to USCIS (Low, M = 3.69s). However, in the second exposure, USCIS achieved faster attention capture (M = 1.40s) compared to Seattle (M = 2.36s), as illustrated by the sharper drop in the red line. Despite these descriptive patterns, statistical analyses revealed no significant differences between chatbot types on TTFF (Wilcoxon: V = 57, p = .89) or order groups (Mann-Whitney U: U = 36, p = .40).

Figure 36 demonstrates fixation counts by AOI, chatbot type. Chatbot bubble fixations (red) remained consistent, showing no difference between chatbot fixation counts (Wilcoxon: V = 20, p = .23). Webpage fixations (blue) also showed no difference between Seattle and USCIS (V = 50, p = .900).

Figure 36

Fixation Count Changes Between AOIs and Chatbot Types

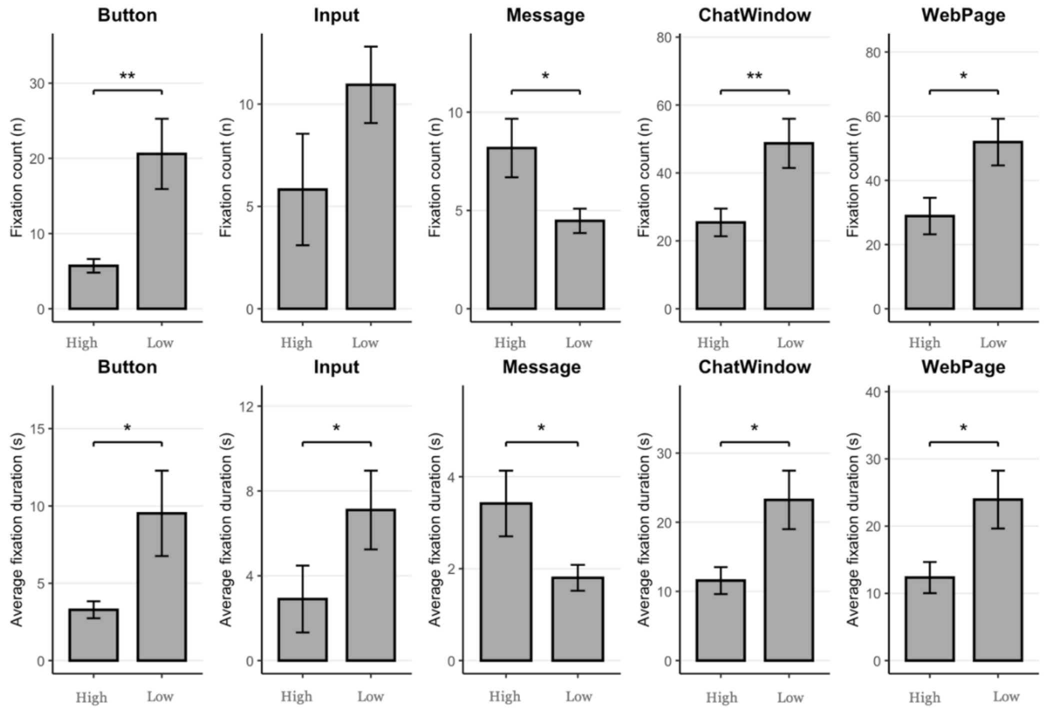


Source: Created by the author.

Figure 37 visualizes fixation counts and durations across five AOIs in the chatbot's introduction (welcoming) message snapshot. Paired Wilcoxon signed-rank tests revealed significant differences between high and low interfaces across multiple AOIs. The Low interface received significantly more fixations and longer viewing times on interactive elements: Button (count: 20.6 vs 5.8, $p < .01$**; duration: 9.6s vs 3.3s, $p < .05$*), Input (count: 10.9 vs 5.8, ns; duration: 7.0s vs 2.9s, $p < .05$*), ChatWindow (count: 48.5 vs 25.4, $p < .01$**; duration: 24.0s vs 12.0s, $p < .05$*), and WebPage (count: 52.7 vs 28.1, $p < .05$*; duration: 24.6s vs 13.6s, $p < .05$*). Conversely, the High interface's Message (chatbot welcoming message) AOI attracted significantly more attention (count: 8.4 vs 4.3, p < .05*; duration: 3.5s vs 1.9s, p < .05*).

Figure 37

Visual Behavior Patterns Between High (Seattle) and Low (USCIS) Interfaces Across Five AOIs in Chatbot Introduction (non-normalized)

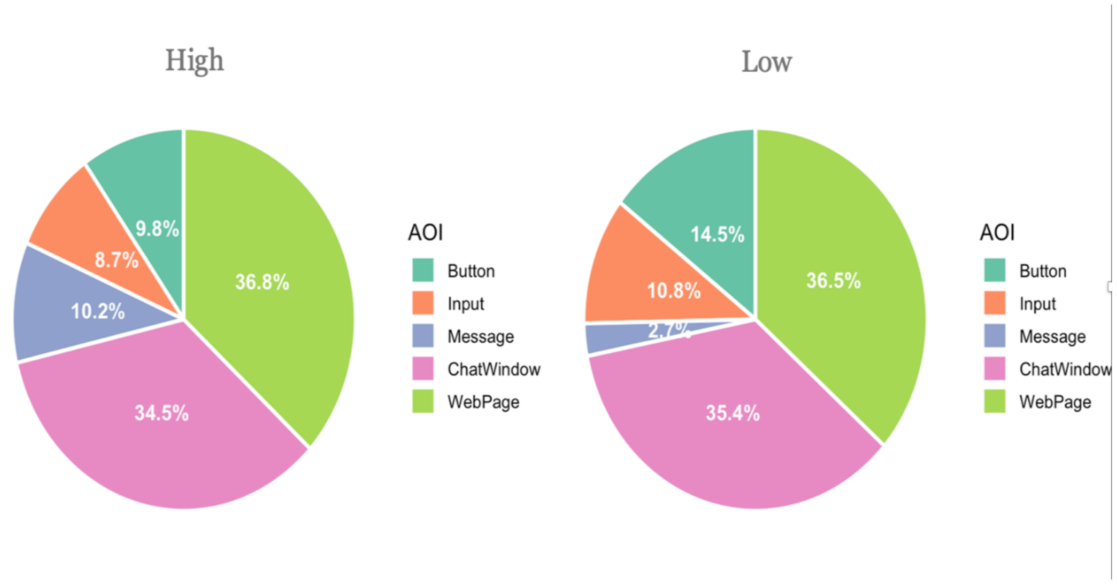


Source: Created by the author.

Figure 38 illustrates the percentage distribution of total fixation duration across AOIs for high and low chatbot interfaces in the chatbot introduction phase. Both chatbots showed similar overall attention allocation patterns, with the most significant proportions devoted to Web Page and Chat Window AOIs. However, notable differences emerged in attention to interactive and informational elements: the low interface attracted proportionally more attention to Button and Input, while the high interface devoted more attention to Message. These patterns suggest that while both interfaces generated similar overall engagement levels, the Low interface directed attention toward interactive elements, whereas the High interface emphasized informational content.

Figure 38

Percentage Distribution of Fixation Duration Across AOIs for High (Seattle) and Low (USCIS) Chatbot Introduction Windows (normalized)

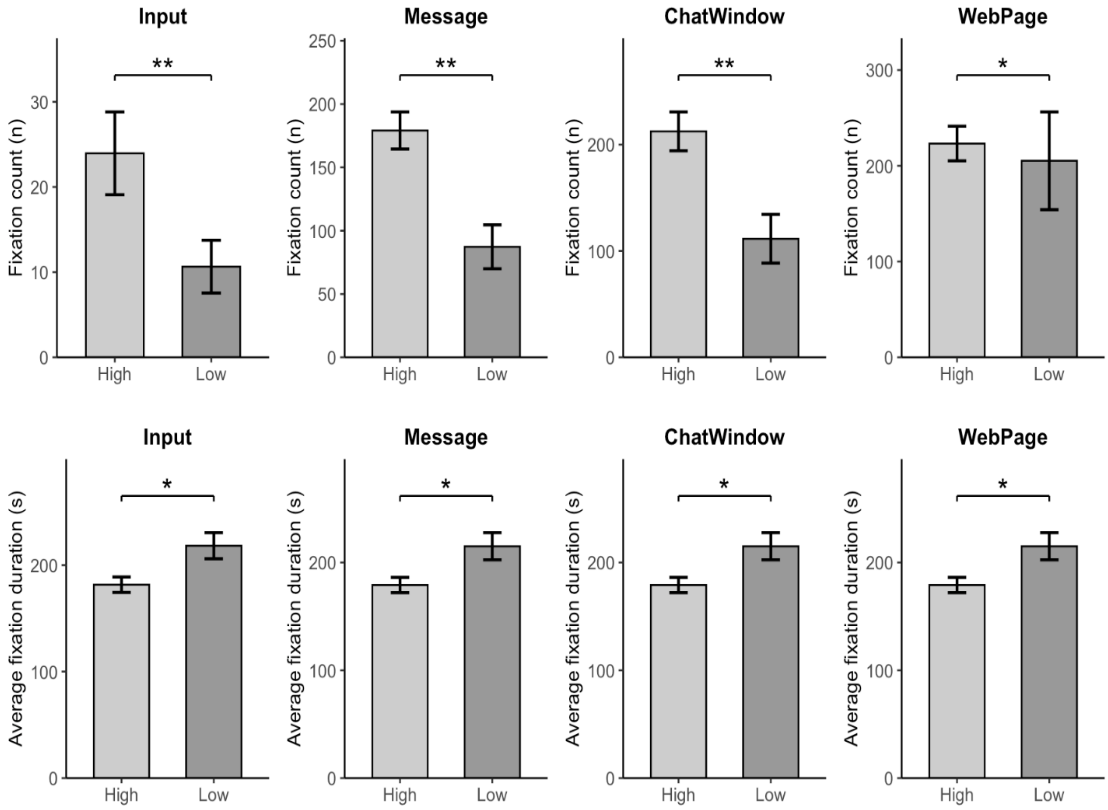


Source: Created by the author.

Figure 39 illustrates fixation counts (top) and average fixation durations (bottom) across four main AOIs during the interaction phase for the high and low-chatbot interfaces. A series of Wilcoxon signed-rank tests revealed significant differences across all AOIs except the header. The high interface chatbot elicited significantly higher fixation counts for all AOIs, indicating more frequent visual scanning behavior within these regions. On the other hand, the low-interface chatbot produced significantly longer fixation durations across the same AOIs. This pattern suggests that although participants fixated less often in the low interface, they tended to dwell longer on specific interface elements. According to Figure 40, both web page contents caught participants' attention, and in terms of being distracted by web content instead of focusing on the chatbot interface. However, distinct attention patterns emerged between conditions. In the high interface chatbot, fixations were more evenly distributed between the Chat Window, Web Page, and Message AOIs. In contrast, the low interface chatbot concentrated attention more heavily on the Web Page, with reduced proportions on the Chat Window and Message.

Figure 39

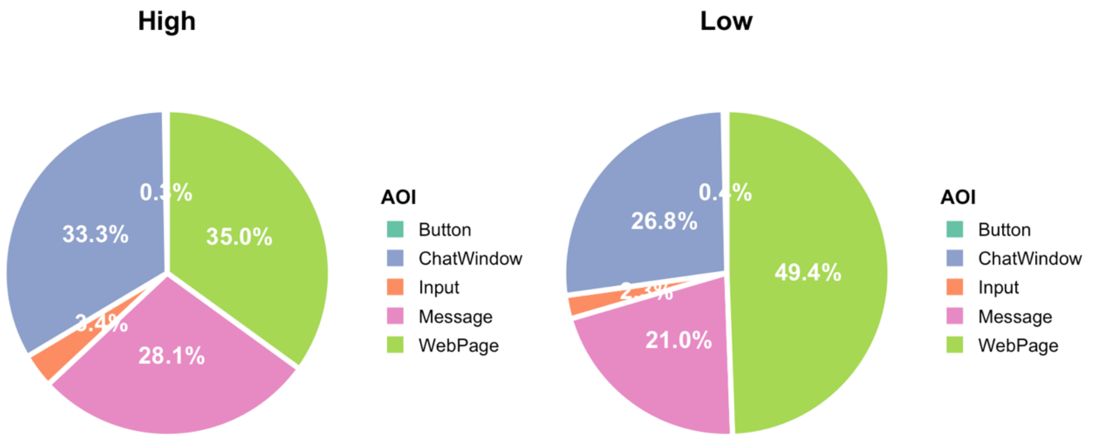
Visual Behavior Patterns Between High (Seattle) and Low (USCIS) Interfaces Across Five AOIs in Chatbot Interaction Flow (non-normalized)



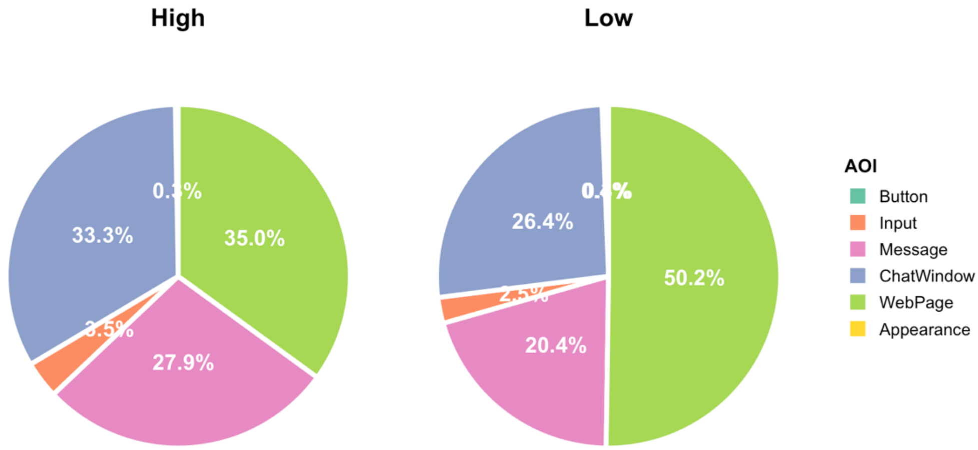
Source: Created by the author.

Figure 40

Visual Behavior Patterns Between High (Seattle) and Low (USCIS) Interfaces Across Five AOIs in Chatbot Interaction Flow



(a)



(b)

(a) Fixation Count (b) Average Fixation Duration
Source: Created by the author.

An overview of the chatbot interaction results was displayed in Table 28. According to Table 28, the chatbot design influenced users' visual attention patterns. The high-interface chatbot prompted more frequent fixations across interactive and message areas, while the low-interface chatbot led to longer fixation durations, indicating more sustained attention on specific regions.

Table 28

An Overview of Measured Differences

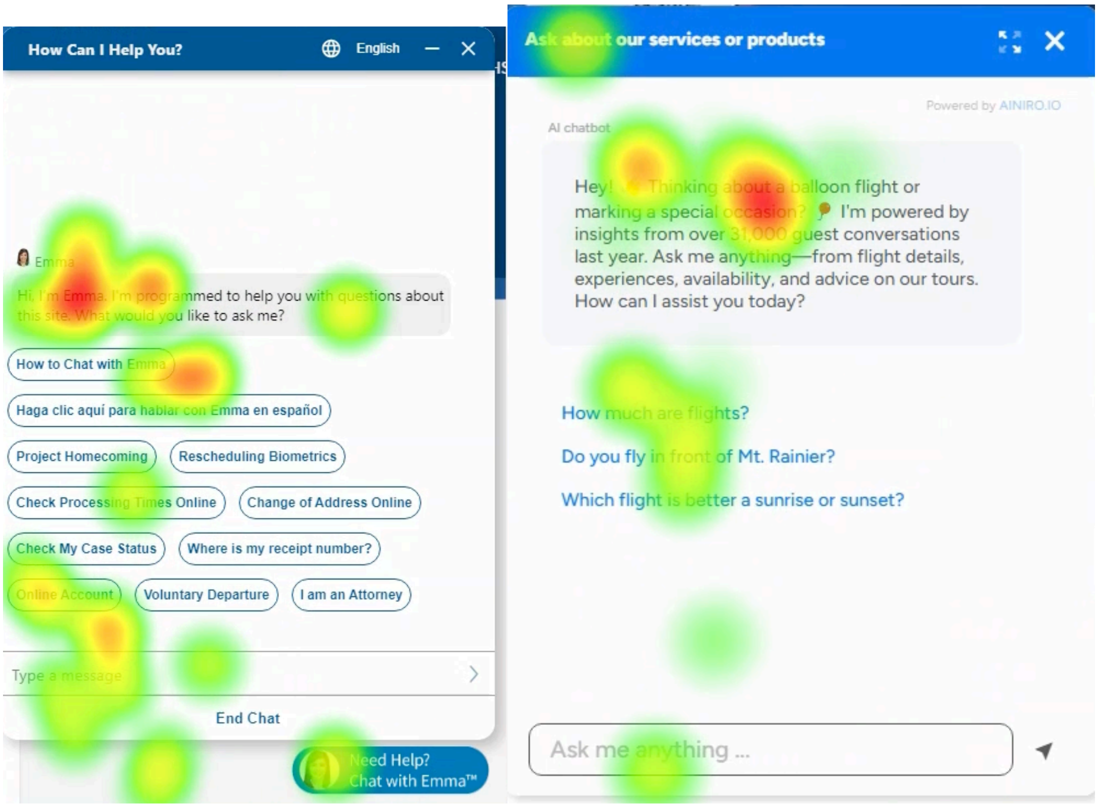
Eye-tracking Metric	AOIs	High Mean	High SD	Low Mean	Low SD	p	Effect Size
Average fixation duration on the AOIs in the Introduction	Webpage	78.95	7.56	78.94	12.71	0.9341	0.001
	Chat Window	78.95	7.56	78.94	12.71	**0.0054**	1.212
	Input Area	10.32	8.94	5.31	6.01	**0.0181**	0.634
	Response Message	61.94	11.67	46.46	16.58	**0.0151**	1.081
	Header	0.69	0.61	1.64	2.42	0.6875	−0.585
	Appearance	NA	NA	0.61	0.39		
Fixation count on the AOIs in Introduction	Webpage	0.818	0.254	0.568	0.336	**0.0083**	-0.467
	Chat Window	0.773	0.244	0.340	0.195	**0.0000**	-0.802
	Input Area	0.083	0.0700	0.029	0.0308	**0.0007**	-0.630
	Response Message	0.657	0.217	0.270	0.144	**0.0000**	-0.826
	Header	0.00	0.0067	0.010	0.0087	0.8054	0.065
	Appearance	NA	NA	0.006	0.0037		
Average fixation duration on the AOIs in the Interaction flow	Webpage	86.80	8.55	68.45	15.98	**0.0084**	1.476
	Chat Window	85.35	9.40	66.04	16.16	**0.0103**	1.497
	Input Area	24.05	20.40	29.92	23.27	0.1953	−0.266
	Response Message	27.78	18.34	11.01	19.69	**0.0052**	0.880
	Button	33.88	23.72	23.05	13.58	0.1531	0.559
	Appearance	NA	NA				
Fixation count on the AOIs in Interaction flow	Webpage	0.099	0.107	0.184	0.0980	**0.0108**	0.516
	Chat Window	0.094	0.0739	0.171	0.0937	**0.0202**	0.478
	Input Area	0.046	0.0640	0.049	0.0377	0.2105	0.319
	Response Message	0.028	0.0181	0.016	0.0086	**0.0438**	-0.422
	Button	0.021	0.0131	0.070	0.0446	**0.0019**	0.648
	Appearance	NA	NA	0.006	0.0051		

TFD and FC data were normalized based on participant interaction duration for each phase

Source: Created by the author.

Figure 41

Chatbot Introduction Phase Eye-tracking Data HeatMap

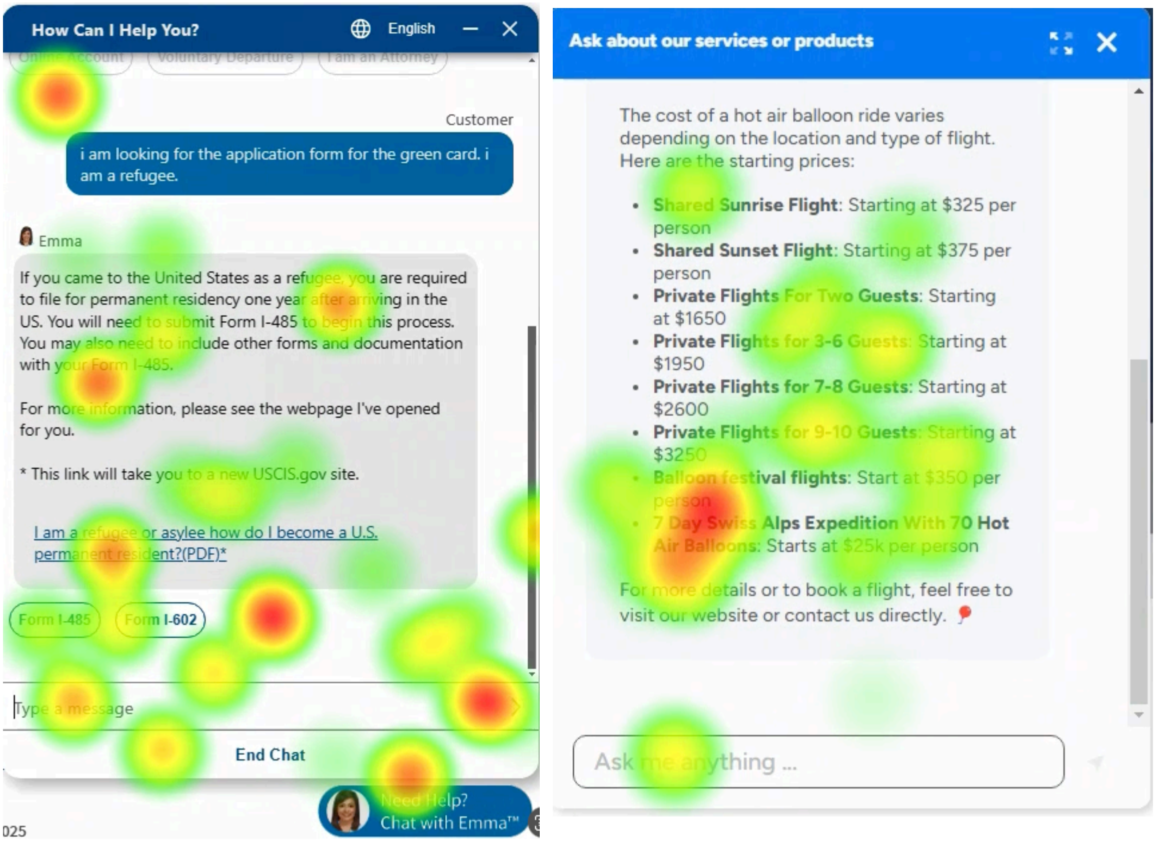


Source: Created by the author.

Figures 41 and 42 demonstrate the fixation density of two chatbots in the introduction phase and subsequent interaction flow, where the users enter inquiries and display chatbot responses (see Appendix E for all snapshots). According to Figure 41, the low chatbot, which is designed with more anthropomorphic dimension-related elements such as a visible agent name and image, elicited stronger fixation density around these humanlike features. In contrast, the high chatbot interface, which employed a more minimalistic design, concentrated users' gaze toward the center of the welcoming message. During the interaction flow (Figure 42), attention in the low chatbot remained distributed across the dialogue area and interactive buttons, reflecting an exploratory scanning pattern, whereas in the high chatbot, fixations were more focused on the system's textual responses.

Figure 42

Chatbot Interaction Flow Eye-tracking Data HeatMap



Source: Created by the author.

Critically, the analysis revealed consistent negative trends between subjective UX evaluations and eye-tracking metrics across both chatbots. Specifically, longer average fixation durations and higher total fixation counts were associated with lower UX scores. For fixation duration, the high chatbot interface showed a moderate negative correlation approaching significance (Spearman's rho = -0.277, p = 0.281), while USCIS demonstrated a weaker but notable trend (Spearman's rho = -0.438, p = 0.090). Regarding fixation count, the low chatbot yielded a statistically significant moderate-to-strong negative correlation (Spearman's rho = -0.576, p = 0.019), indicating that users requiring more visual fixations rated the interface significantly worse. In contrast, the high chatbot showed weak, non-significant correlations for fixation count.

4.3.2.3. User Evaluations Results

Evaluations of participants per chatbot are given in Table 29. The participants' overall user experience evaluations with the chatbot interactions were diverse. According to the findings in Table 28, Seattle Ballooning Chatbot received the highest score, whereas the

University of Twente had a lower rating overall. The regression analysis revealed that the overall user experience evaluations significantly changed due to the chatbot ($F(5, 338) = 7.66$; $p = 7.766e-07$). The Seattle Ballooning chatbot achieved the highest user experience ratings, significantly outperforming the reference chatbot (Kia; $\beta = 14.50$, $p = .001$). In contrast, the University of Twente chatbot received the lowest ratings ($\beta = -7.40$, $p = .071$), followed by USCIS ($\beta = -1.63$, $p = .71$), though these differences were not statistically significant.

Table 29

Descriptive Statistics for Each Chatbot

Chatbot	Number of Observations	Mean	SD	Median
Kia	43	70.0	22.8	72.2
USCIS	54	69.3	22.1	73.8
Seattle Ballooning	56	85.5	15.8	90.6
Wanderlog	41	77.4	21.2	81.2
Lufthansa	77	71.5	23.3	78.8
UTwente	92	63.2	21.5	66.2

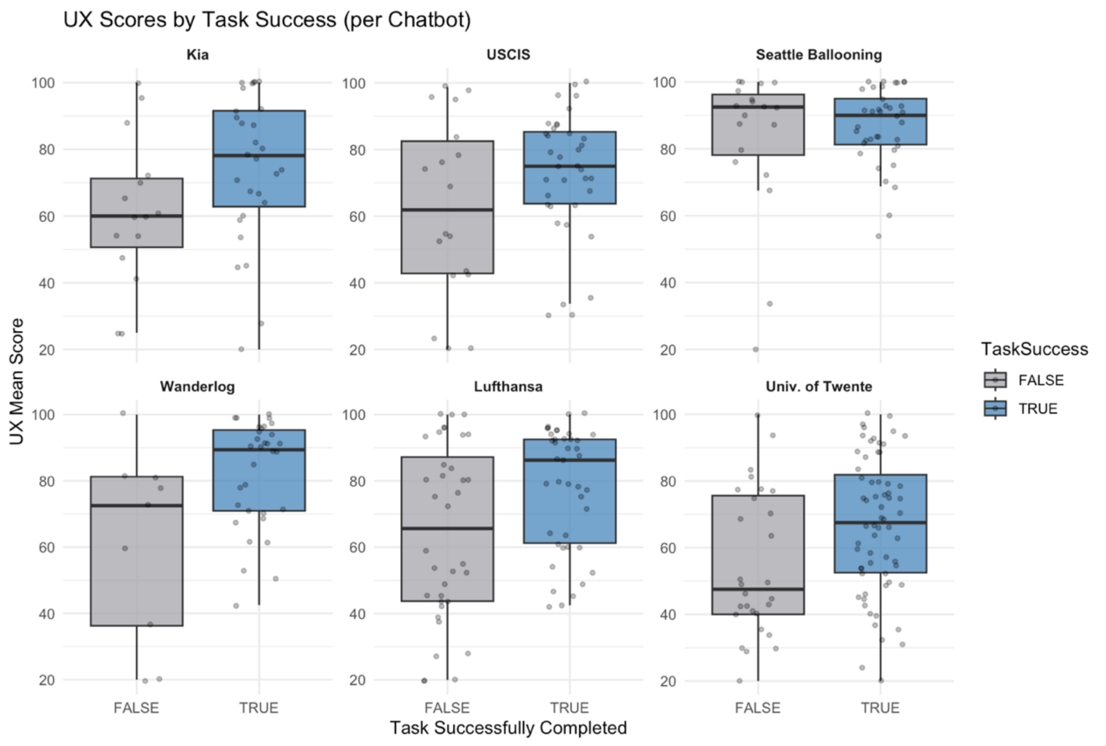
Source: Created by the author.

Regarding the effectiveness of the interaction, a control question was asked after interacting with the chatbot (ISO, 2018). Then, the user evaluation means were calculated for each chatbot, distinguishing between participants who succeeded in the task and those who did not. Figure 43 shows that participants evaluated the interaction more positively when they completed the task successfully across all chatbots. Only for the Seattle Ballooning chatbot were the evaluations similar between participants, regardless of whether they succeeded.

Importantly, these findings are consistent with the qualitative design analysis (Section 4.3.2.1 and Section 3.6.1), showing chatbots that integrate a combination of elements across multiple design dimensions tend to achieve higher user experience. For instance, Seattle Ballooning chatbot demonstrated the most balanced design, combining multiple dimensions (Table 27), which yielded has higher user evaluation (Table 29). On the other hand, the University of Twente chatbot design is also rich in individual observable design elements; however, concentrated within a limited set of design dimensions. The Wanderlog chatbot also achieved relatively positive evaluations, as its design addressed multiple dimensions. Regarding the more positively evaluated chatbots, the inclusion of agent competency–related design requirements appears to make a noticeable difference and enhance perceived interaction quality and overall user experience.

Figure 43

User Chatbot Evaluations Based on Task Success



Source: Created by the author.

4.3.3. Knowledge Contribution

This thesis contributes to the knowledge domain systematically documented through publications and this dissertation. This process is the crucial part of the DSR research, ensuring the robustness and traceability of knowledge generation throughout the research.

The study's contribution to the knowledge base is supported by multiple dissemination activities. The first is the conference paper on CA design elements examination, which provided an early conceptual synthesis of design elements and their relationship with user perceptions across HCI and IS venues (Aktaş & Akbıyık, 2025). This was followed by a systematic literature review paper (under review) that consolidated 97 empirical studies to establish a comprehensive design dimension framework for conversational agents, mapping design elements, dimensions, and user outcomes. Importantly, these studies form the theoretical foundation for the design meta-requirements developed in this thesis.

Furthermore, this research was conducted as part of the TÜBİTAK 2214 International Research Fellowship, enabling collaboration with international experts and exposure to advanced usability and eye-tracking laboratories. These collaborations strengthened the empirical rigor of the findings and supported the development of CA design requirements in real-world settings. Finally, the thesis itself extends the knowledge base by (1) operationalizing a structured framework that links anthropomorphic, characteristic, and competency dimensions to user experience outcomes, (2) developing and validating a measurement scale for assessing CA user experience, (3) evaluating design implementations across diverse chatbots, and (4) outlining 14 meta-requirements to improve effectiveness and user experience of CAs.

DISCUSSION

This section provides an interpretation of the findings and discusses the key results concerning the dissertation's objectives.

Study Overview: Feine et al. (2019) developed a taxonomy for CAs design, and they identified design social cues. Their work was the foundation for designing CAs based on interpersonal communication, which triggers social reactions to CAs. Moreover, Diederich et al. (2022) focused on the CA research dimensions. They defined four dimensions in CA research, including human, context, agent, perception, and outcome. Their agent dimension is also focused on social cues to make agents' communication humanlike, as well as focusing on the communication mode. These works capture the limited aspects of humanlike appearance, as well as verbal and non-verbal cues. A growing body of research has continued to explore the role of humanlike aspects (Appel et al., 2012; Cassell et al., 2001; Chung et al., 2023; Følstad et al., 2018; Go & Sundar, 2019; E. Jin & Eastin, 2024; Li & Suh, 2021; Y. Liu et al., 2024; McKenzie et al., 2003; Seeger & Heinzl, 2021, p. 2021; Urakami & Seaborn, 2023; I. Wang & Ruiz, 2021). Yet, most of these studies remain fragmented, focusing on isolated design aspects rather than providing an integrated design solution.

Critically, the measurement of user experience with CAs remains constrained by conventional scales originally developed for non-conversational systems. Although valuable, such instruments often fail to capture the interactive, social, and affective nature of human–agent communication. A notable effort was made by Borsci et al. (2022b), who developed the BUS-11 scale to evaluate usability specifically in chatbot interactions, which is a precondition for quality of interaction. While BUS-11 focuses on functional and conversational ability, the broader dimensions of user experience, such as trust, enjoyment, social presence, and usefulness, remain insufficiently addressed.

To address these gaps, this dissertation pursues two main objectives. First, it develops a measurement scale specifically designed to assess overall user experience with conversational agents. The proposed scale integrates both pragmatic and experiential dimensions, offering a comprehensive framework for evaluating interaction quality and post-use intentions. Second, the dissertation formulates a set of theory-driven meta-requirements to guide the design of conversational agents, ensuring that design choices are systematically linked to established theoretical foundations.

Research Methodology: The dissertation follows the Design Science Research approach to address the design problem of conversational agents to enhance user experience. DSR provides a structured approach for developing and evaluating artifacts that contribute both theoretical and practical knowledge (Hevner, 2007; Hevner et al., 2024b). DSR cycles relevance, rigor, and design are systematically integrated to ensure that the proposed solutions respond to real-world challenges and build upon existing knowledge.

The research begins with the relevance cycle of DSR, which establish the problem space and the state of the art. Through a comprehensive literature review and systematic analysis of existing CA design studies, the research identifies design practices and their effects on user experience. This stage defines the theoretical and practical motivation for improving CA design by uncovering limitations in current measurement tools and fragmented design knowledge.

Building on this foundation, the design cycle focuses on examining the literature iteratively to identify meta-requirements and design dimensions. These insights were further supported by empirical studies aimed at exploring user experience requirements and developing a measurement scale tailored for CAs. Finally, the rigor cycle ensures the scientific grounding of the research by investigating real-world CA interactions using the developed UX scale and objective data (eye-tracking). This integration of subjective and behavioral evidence validates the theoretical model and demonstrates the practical applicability of the proposed design knowledge for developing more effective and engaging conversational agents. The following discussion interprets these findings in light of existing theories and prior research.

5.1. CA Design Meta-Requirements

This section discusses the 14 meta-requirements derived from the systematic review and empirical study regarding agent design dimensions. Each requirement is interpreted in light of existing theories and evidence, followed by implications for CA design and future research.

5.1.1. Anthropomorphic Design Dimension

5.1.1.1. R1: Agent Communicates Its Identity Clearly

This dissertation contributes to the knowledge base on designing identity in CAs by formulating the requirement that agents must communicate their identity clearly. Existing studies often demonstrate that identity cues such as name, gender, or avatar representation can increase social presence and trust (Beldad et al., 2016; Hess et al., 2009; Mozafari et al., 2021). However, they report inconsistencies across contexts and user groups (Ben Mimoun et al., 2017; Benbasat et al., 2020). The systematic review and user evaluation data provided prescriptive guidance on identity clarity, which is not only a matter of anthropomorphizing agents but of ensuring that these signals are transparent, culturally appropriate, and aligned with user expectations (Diederich et al., 2020a; Janson, 2023; Schuetzler et al., 2019).

A central insight from this research is that who the agent is should align with what it can do. When an agent's identity is consistent with its actual role and capabilities, users are more likely to benefit from identity cues without experiencing disappointment when limitations become apparent. This interpretation resonates with Social Presence Theory and the CASA paradigm, which suggest that people naturally respond socially to identity signals (Hess et al., 2009; Nass & Moon, 2000; Reevess & Nass, 1996). However, the findings indicate that such responses remain more positive and sustainable when the agent's identity is clearly communicated and does not overstate its abilities (Luger & Sellen, 2016).

From a design perspective, identity clarity is achieved through a combination of design elements to present a coherent persona (Figure 27), which is also more beneficial for the replacement of actual human tasks (Seeger et al., 2017). Moreover, identity design should not be seen as static. Instead, the identity cues can be adapted dynamically, adjusting their richness and presentation depending on the user, cultural context, or stage of interaction. For example, minimal cues such as a name and self-introduction may be sufficient to establish trust in initial encounters, while richer identity cues that include real person image, self-referencing, or conversational abilities (Figure 27) may serve as scaffolds in more complex or relational tasks (Diederich et al., 2020a; Følstad & Brandtzæg, 2017; Schlesener et al., 2025). Another example is found in everyday systems such as the ChatGPT mobile app or Apple's Siri, where users can choose whether the agent has a male or female voice and even adjust voice tone. These options show how identity cues can be customized to fit user preferences, reinforcing the idea that identity design should remain adaptive rather than fixed.

5.1.1.2. R2: Agent Provides Human-like Social Cues for Engagement

The presence of human-like social cues strongly shapes engagement with conversational agents. Prior studies have shown that non-verbal behaviors such as gesture, gaze, head movement, smile, and listening behavior enhance realism and strengthen perceptions of social presence (Appel et al., 2012; Gnewuch et al., 2018). Feine et al. (2019) emphasize that such cues are central to creating natural, human-like interactions, and the findings of this dissertation support their value in fostering user engagement.

Specifically, this requirement sheds light on the dynamic role of human-like social cues in sustaining interaction. From the theoretical side, Anthropomorphism theory explains why people attribute social qualities to agents, particularly when social cues are present (Epley et al., 2008; Waytz et al., 2010). These cues are not static features but ongoing signals that shape how users perceive the interaction flow. The CASA paradigm (Nass et al., 1994) and Social Response Theory (Nass & Moon, 2000; Reevess & Nass, 1996) further show that even minimal signals (e.g. gaze, gestures, or politeness markers) can trigger unconscious social responses that sustain engagement throughout a conversation (S. Y. B. Huang & Lee, 2022). Similarly, Social Presence Theory demonstrates that humanlike cues contribute to the perception of interacting with a “real” social entity, thereby reinforcing attentiveness and emotional connection over time (Hess et al., 2009; Qiu & Benbasat, 2008; Schlesener et al., 2025). Drawing on this, these perspectives underline that social cues function dynamically, supporting the continuity of interaction and encouraging users to remain engaged.

Practically, verbal and non-verbal cues are synthesized in this thesis (Figure 27) to provide engaged interaction. A key insight is that even the minimal use of such cues can significantly enhance engagement (Cavedon et al., 2015; T.-C. Lin et al., 2023; Xu et al., 2021). In line with the usage of social cues in design (Gnewuch et al., 2022; Urakami & Seaborn, 2023), users expect involvement in the interaction (Følstad & Brandtzaeg, 2020; Jo et al., 2023; Shevat, 2017). Specifically, they want the agent to appear attentive, responsive, and socially present (Kang & Gratch, 2014; Sundar et al., 2012). Small signals create the impression that the agent is participating actively, which sustains user attention and motivation to continue. Without such involvement, interaction risks becoming transactional and disengaging (Følstad & Brandtzaeg, 2020; Luger & Sellen, 2016; Luria et al., 2019).

5.1.1.3. R3: Agent Has a Human-like Interaction Flow

This requirement emphasizes that the overall conversation with an agent follows a human-like rhythm and structure. Theoretical perspectives provide explanations for why interaction flow matters. Media Naturalness Theory argues that human communication evolved for face-to-face settings, which rely on cues such as synchronicity, immediate feedback, and turn-taking (Kock, 2004). When these features are absent, communication requires greater cognitive effort, increasing ambiguity and reducing engagement (Kock, 2005). In line with this, Chandra et al. (2022) show that conversational agents require cognitive, relational, and emotional competencies to replicate natural flow. According to Social Response Theory, users unconsciously expect human conversational norms in agent interaction (Nass & Moon, 2000). As a result, they anticipate smooth turn-taking, feedback, and repair strategies, and perceive disruptions in flow as violations of these norms (Diederich et al., 2020a; S. Y. B. Huang & Lee, 2022). Further, Expectancy Violations Theory shows that when flow is disrupted, through long silences, abrupt topic shifts, or repetitive loops, users experience negative reactions (Gnewuch et al., 2022; Go & Sundar, 2019).

A useful way to approach the design of conversational flow is through a heuristic lens, taking inspiration from the norms of human-to-human dialogue (Langevin et al., 2021). Human communication provides the baseline of what users unconsciously expect in terms of rhythm, turn-taking, and feedback, and these expectations naturally transfer into human–agent interaction. For example, Morana et al. (2020) designed agents to remember user inputs across turns with social cues, which preserved continuity and contributed to a human-like rhythm of conversation. Such heuristics make the flow of interaction feel natural, beyond isolated greetings or identity cues. This dissertation suggests further investigation into design practices that operationalize this heuristic lens and examines which strategies most effectively support the naturalness of conversational flow.

5.1.1.4. R4: Agent Supports Natural Timing and Feedback Behavior

This requirement highlights the importance of temporal dynamics in conversational interaction. Timing and feedback behaviors which are implemented in design using typing indicator, response delay, or pause-filler, which are frequently implemented in design, shape whether communication feels natural and responsive. When agents respond too quickly or too slowly, the realism of interaction is disrupted, creating either a mechanical impression or user frustration (Hildebrandt et al., 2023a; Seeger & Heinzl, 2021). EVT helps clarify user reactions: when timing or feedback diverges from expected human norms, users experience violations that can reduce trust and satisfaction (Gnewuch et al., 2022).

From a design perspective, this implies that temporal coordination is as crucial as content quality. Simple strategies such as incorporating response delays, typing indicators, or listening behavior cues ("I see," "okay") can convey attentiveness and reduce user uncertainty (Appel et al., 2012; Hwang & Won, 2021; Kang & Gratch, 2014; Von Der Pütten et al., 2010). A key next step is to investigate how different timing strategies affect user perceptions across contexts. For instance, in customer service, rapid responses may be valued, whereas in learning or therapeutic contexts, slight delays and richer feedback may build trust and reflection.

5.1.2. Agent Characteristics Dimension

5.1.2.1. R5: Agent Conveys System Capabilities and Limitations

This dissertation finding revealed that conveying agent capabilities and limitations to make it transparent is a robust design strategy (Luger & Sellen, 2016; Seeger & Heinzl, 2021). When transparency is absent, users overestimate the agent's capabilities. Clear communication of both capabilities and limitations is associated with higher satisfaction and stronger perceptions of reliability (Diederich et al., 2020a). To achieve this, prior research has frequently used design strategies such as self-introduction and self-referencing (Ben Mimoun et al., 2017; Krämer et al., 2018; Pietrantoni et al., 2023). For instance, Kang and Gratch (2014) showed how agents can reduce uncertainty by explicitly stating their limitations. Similarly, Wang and Benbasat (2016), adopting Attribution Theory, demonstrated that users tend to attribute failures to system limits rather than to themselves.

In this sense, clearly stating agent capabilities and limits forms the expectation for all CAs. Thus, this requirement highlights transparency as a stabilizing principle that not only prevents overestimation of the agent's abilities but also strengthens perceptions and trustworthiness across diverse application domains (Pietrantoni et al., 2022; Wang & Benbasat, 2016). With the rise of high-capability agents, broad skills may lead users to assume abilities that exceed what the system can provide. Communicating capabilities and limitations helps manage such expectations and supports more calibrated trust.

5.1.2.2. R6: Agent Enables Access to Verifiable References and Sources

These requirements bring attention to the traceability of information provided by conversational agents. Theoretical perspectives help explain this. For instance, Information Behavior Theory emphasizes that users seek credible and transparent information, and the availability of references supports this process by enabling independent verification (Sin & Munteanu, 2020). Further, Attribution Theory explains how people assign causes for outcomes.Regarding CAs, when no sources are given, users may attribute errors or misinformation directly to the agent, which undermines trust. When references are visible, however, users attribute the content to the cited material, which shifts responsibility to the external source. This way, the agent is perceived as more transparent and trustworthy because it shows where its information comes from (Wang & Benbasat, 2016).

Practical implementations illustrate this principle. The U.S. Citizenship and Immigration Services (USCIS) chatbot, for example, directs users to official web pages answering inquiries. Users are able to see and read policy details that they asked the chatbot. In the educational domain, Costea and Sedrakyan (2025) combined the explanability features of the agent with the source graph to encourage students to evaluate the credibility of the responses, which consistent with learning theories. Similarly, generating CA responses with explanations such as where the responses are coming from improves trust in agents (Al-Natour et al., 2010).

At the same time, LLM-based agents face the challenge of hallucination, where responses may appear fluent and confident, but they are not completely correct (Alkaissi & McFarlane, 2023; Emsley, 2023). Giving clear references is therefore helpful and a safeguard to allow users to verify information and keep trust in the system (Russell et al., 2025; Sedrakyan et al., 2024a). In this sense, verifiability should be treated as a core design practice. Whether a CA is rule-based or LLM-based, embedding references, links, or explanations into responses allows users to verify information, calibrate expectations, and reduce the risks posed by misinformation.

5.1.2.3. R7: Agent Employs Socially Appropriate and Emotionally Intelligent Language

These requirements center on the agent's ability to convey information in a manner that incorporates social and emotional considerations. Prior studies revealed that employing the emotional aspect in agent language fosters positive perceptions (Bowman et al., 2024; Brendel et al., 2020). For instance, when a user encounters a problem or serious situation, a CA should acknowledge the user’s perspective and show it with expressions such as tone modulation, facial expressions, apologizing, or expressive speech acts (Figure 28). According to Media Naturalness Theory, emotionally intelligent agents include emotional and affective signals to resemble natural communication (Chandra et al., 2022; Kock, 2005). In this view, positive or negative language usage and tone modulation can reduce cognitive effort and increase engagement while shaping communication as natural. When agents lack emotional intelligence or socially appropriate behavior, for example, giving humorous responses when someone’s debit card is stolen, interaction risks breaking down instead of sustaining a natural conversation (Dybala et al., 2010). As a well-known perspective, people attribute social norms to agents; if the agent does not behave in a socially appropriate manner, this creates frustration (Brendel et al., 2020; Nass & Moon, 2000).

From a design perspective, training agents in socially appropriate behaviors strengthens their intelligence, whether rule-based or LLM-based (Haas & Moussawi, 2020). This dissertation suggests that agents should adopt adaptive strategies that calibrate their language to the seriousness of the task, as well as to the appropriate level of warmth or friendliness in conversation (Hyde et al., 2015).

5.1.2.4. R8: Agent Has an Interaction Style Aligned with Context, User Ability, and Task Goals

R8 emphasizes that interaction styles, including communication modality and embodiment, need to be aligned with user abilities, context, and task goals. Evidence from the literature indicates that older adults benefit more from voice-based or multimodal agents, as these reduce barriers associated with text-based interaction (Beer et al., 2015, p. 2; Sin & Munteanu, 2019; Zhou et al., 2025). By contrast, studies with younger adults reveal more varied preference that suggests context and task demands can be important than modality alone (Zhang et al., 2009). This alignment is required to ensure that the selected communication modality fits the user’s abilities and situational needs. Further, the literature review demonstrated that interaction style is shaped by the type of task and domain. Chatbots are frequently used for information retrieval tasks, recommender agents are integrated into sales and service roles, while virtual agents are employed in training or guiding roles in education and health contexts (Aktaş & Akbıyık, 2025; Benbasat et al., 2020; J. Kim & Im, 2023; Kim & Sundar, 2012; Lawson et al., 2021).

As noted by Uncanny Valley Theory, interaction with an agent can risk evoking discomfort if the design becomes unrealistic. While the theory is traditionally applied to agents that appear too human-like, it also reminds us of the need to calibrate functionality and design features to the context. For instance, in virtual environments, users may welcome highly human-like virtual agents, whereas in simple information-seeking tasks, such as checking a delivery status with a chatbot, the same level of embodiment may appear unnecessary or even unsettling (Li, 2024; Seymour et al., 2018).

Designing an agent, therefore, requires careful consideration. Agent communication modality defines agent characteristics and the scope and purpose of interaction (Diederich et al., 2022). Once the general scope of an agent's unique expressions is defined, designers should consider incorporating design features that enhance the interaction quality. For example, text-based agents have been extensively examined, particularly in anthropomorphic design and agent competency considerations. Besides, expressiveness and transparency are prominent for text-based agents, while voice-based agents focus more on embodiment and expressiveness (Table 15). These differences point out the importance of choosing the proper communication mode in CA design, as it significantly impacts the agent's effectiveness in interacting with users (W. Terblanche et al., 2023).

5.1.2.5. R9: Agent Reflects an Appropriate Visual Presence that Supports Its Functional Role

The visual presence of an agent, whether represented as a static avatar, a real person image, or a virtual interactive character, follows from embodiment choices. Meta-requirement 9 highlights how such visual design decisions communicate signals about the agent’s role. For instance, in healthcare or banking, users expect visual representations that convey professionalism, authority, and trustworthiness (Sestino & D’Angelo, 2023). By contrast, in social or entertainment contexts, users may more readily accept playful designs such as animal figures or cartoon-like avatars (Fraser et al., 2024; Seymour et al., 2018; Von Der Pütten et al., 2010).

Design decisions about visual presence, therefore, extend beyond aesthetics, which directly influence how the agent is evaluated in terms of role fit and effectiveness (Baylor, 2009). Similarity Attraction Theory supports this view by suggesting that users respond more positively to agents whose visual attributes resemble their own characteristics or expectations, while dissimilarity can lead to repulsion (Qiu & Benbasat, 2010; B. Zhao, 2019). Self-Perception Theory further explains how interacting with certain agent appearances can shape users' own attitudes; for instance, a professional-looking agent may encourage users to perceive the interaction as more serious, while a playful avatar can signal a more casual, social exchange (Baylor, 2009; Bem, 1967). Importantly, if an agent's appearance is too human-like without achieving complete naturalness, it risks evoking discomfort, as highlighted by the Uncanny Valley Theory (Mori et al., 2012). Visual appropriateness thus emerges as a stabilizing principle in CA design, ensuring that agents remain trustworthy, context-sensitive, and consistent with the functional roles they are intended to fulfill.

5.1.3. Agent Competency Dimension

5.1.3.1. R10: Agent Supports Handling of System Errors and Misunderstanding

A critical requirement in CA design is the ability to manage errors or misunderstandings in conversation flow. Diederich et al. (2020) noted that misunderstandings are inevitable in human–agent interaction, making in-conversation repair mechanisms an essential strategy. Følstad and Brandtzæg (2017) emphasized recovery strategies such as explicitly asking users for reformulation when input is unclear. This dissertation suggests that mitigating the effects of errors, even when the agent cannot fully resolve the inquiry, can still foster positive perceptions. Users respond positively when they see that the agent acknowledges the problem and tries to solve it, demonstrating awareness and responsiveness (Riquel et al., 2021b).

Importantly, designers should plan for "sunny-day" scenarios and account for breakdowns as part of natural dialogue. Whether rule-based or LLM-based, agents must be equipped to guide users through errors. For instance, when users express frustration with harsh language, the agent should acknowledge the sentiment and adopt socially appropriate strategies suited to its role and context (Brendel et al., 2020). From a theoretical perspective, Social Response Theory helps explain why error handling matters. Users apply human conversational norms to agents, and when those expectations are unmet, frustration can escalate into verbal abuse or teasing (Brahnam & De Angeli, 2008; Brendel et al., 2020). At the same time, humor and lightness have been shown to function as recovery mechanisms that diffuse tension and restore conversational flow (Ashktorab et al., 2019; Shevat, 2017; Y. Yang & Kankanhalli, 2023).

5.1.3.2. R11: Agent Offers Multimodel Context Delivery When Needed

R11 highlights the importance of providing information across multiple modalities when the task or user needs demand it. While text remains the dominant interaction mode, there are contexts where additional modalities such as voice, visual aids, or interactive elements enhance comprehension and engagement. Research on media richness and communication modes shows that richer channels reduce ambiguity and improve outcomes by aligning the communication format with task complexity (W. Terblanche et al., 2023).

In education and healthcare, for example, visual explanations, diagrams, or spoken feedback improve understanding and trust when users face complex or emotionally charged information (Lawson et al., 2021; Sin & Munteanu, 2019). Multimodal delivery also promotes accessibility by accommodating diverse user abilities: voice interfaces assist those with limited literacy or vision, while visual feedback supports users with hearing difficulties (Beer et al., 2015; Zhou et al., 2025).

From a theoretical standpoint, Uncertainty Reduction Theory suggests that providing multiple modalities gives users more cues to interpret and validate information, which helps reduce ambiguity and fosters confidence in the agent (Kang & Gratch, 2014). Similarly, Social Presence Theory indicates that richer channels (e.g., combining text with voice or visuals) increase perceptions of immediacy and “realness,” which strengthens trust and engagement (Hess et al., 2009; Qiu & Benbasat, 2008).

From a design perspective, multimodal delivery should therefore be seen not as an add-on but as an adaptive strategy that matches the task, user, and environment. Overloading users with unnecessary modes risks raising cognitive demands (Ceha & Law, 2022), whereas selectively integrating voice, visuals, or interactive cues enables more natural, effective, and inclusive interactions.

5.1.3.3. R12: Agent Improves Information Visibility and Interaction Clarity

Information visibility and interaction clarity are central to user experience with conversational agents. Evidence from both the literature and empirical findings shows that structured outputs, such as bullet points, tables, or expandable sections, help reduce cognitive effort and enable users to focus on the most relevant details (Borsci et al., 2022a; Diederich et al., 2020a; Lawson et al., 2021). Participants in the thesis's first experimental study similarly emphasized the importance of clear formatting and highlighted key elements for easier comprehension and faster decision-making.

Theoretical perspectives reinforce these findings. Cognitive Load Theory suggests that minimizing extraneous complexity fosters efficiency and comprehension (Ceha & Law, 2022), while Information Behavior Theory highlights that clarity and transparency support users in verifying and applying information (Sin & Munteanu, 2020). The principles of User-Centered Design further strengthen this requirement, as iterative design and usability testing have been shown to reduce cognitive burden and improve clarity in complex systems by aligning information presentation with user expectations (Mirabdolah et al., 2023). In educational contexts, structured responses also align with self-regulated learning theory, which stresses the role of clarity and feedback in enabling reflection and sustained engagement (Sedrakyan et al., 2024a; Zimmerman, 2000).

Critically, interaction clarity can be enhanced through response structuring, navigational aids, or layered information delivery. Prior CA research also shows that structured and transparent responses increase perceptions of competence and trust (Diederich et al., 2020a; Lawson et al., 2021). Agents can support users in completing tasks effectively by ensuring that information is made transparent and manageable.

5.1.3.4. R13: Agent Enables Personalized and Context-aware Interaction

Personalized and context-aware design support sustaining meaningful interaction with CAs. Responsively designed agents are able to reference earlier conversation turns, ask follow-up questions, or tailor their tone and information delivery to the user’s context (Chung et al., 2023; Janson, 2023). Evidence from prior research shows that users value agents that adapt to their prior inputs, preferences, and situational needs, as these features increase satisfaction and trust (Cai et al., 2022; Poser & Bittner, 2023; Rhim et al., 2022).

Theoretically, Social Identity Theory suggests that when agents personalize cues to reflect group or cultural alignment, users experience stronger belonging and relevance (Liu & Yao, 2023; Pietrantoni et al., 2022). Moreover, users are more likely to engage with and accept technology when their psychological needs are supported, particularly the three needs identified in Self-Determination Theory: autonomy, competence, and relatedness (Cai et al., 2023; Van Lange et al., 2011). Satisfaction of these needs has been found to mediate the effect of system design on engagement and acceptance (Bitrián et al., 2021; Xi & Hamari, 2019).

Design practices further illustrate these benefits. In educational contexts, adaptive scaffolding has been shown to help learners plan, monitor, and reflect, thereby improving self-regulated learning and reducing cognitive effort (Chang et al., 2023; Sedrakyan et al., 2024b). In service domains, agents that adjust tone, style, or level of detail to situational needs foster stronger perceptions of trust and competence (Benbasat et al., 2020; Diederich et al., 2020a). Thus, personalization and context-aware interaction emerge as stabilizing principles in CA design. By tailoring content and delivery to user characteristics, preferences, and situational contexts, agents can sustain engagement, support task success, and build long-term engagement.

5.1.3.5. R14: Agent Facilitates Goal Completion and Decision Support

Supporting users in completing tasks and making decisions is a critical marker of conversational agent competency (Allouch et al., 2021). Well-designed agents act as facilitators, guiding users through structured steps, clarifying available choices, and explaining situations encountered during the interaction (Al-Natour et al., 2022; Chandra et al., 2022; Diederich et al., 2022; Poser & Bittner, 2023). Theoretical perspectives reinforce the value of this approach. Attribution Theory explains that when agents clarify the reasoning behind their responses, users attribute outcomes to transparent system logic rather than to arbitrary errors, thereby strengthening perceptions of competence and reliability (Wang & Benbasat, 2016). In parallel, Uncertainty Reduction Theory suggests that explanatory details help minimize ambiguity, enabling users to make decisions with greater confidence (Kang & Gratch, 2014). Prior work on CA design also underlines that explanation features and logical reasoning foster trust and improve acceptance (Hernandez-Bocanegra & Ziegler, 2023; Luger & Sellen, 2016).

Eventually, this thesis recommends that design strategies for decision support move beyond surface-level answers to provide structured details and contextual clarity. Similar consideration is emphasized by Posser et al. (2022), who revealed that employing explanations about process steps and topic-related terms improves user understanding and supports the decision-making process quality. Thus, CAs should be designed to articulate both the content of their responses and the underlying rationale, ensuring that responses are contextually justified.

5.2. Enhancing User Experience

The main objective of this dissertation is to enhance user experience with conversational agents by considering design choices. To achieve this, the study began by examining how user experience has been understood and evaluated in prior research. However, the literature revealed that existing user experience evaluation models were developed for earlier technologies and do not sufficiently account for the conversational dimensions of CA interactions (Borsci et al., 2022a; Gursoy et al., 2019). Recognizing this gap, the researcher systematically reviewed the literature to explore user experience concerns specific to CAs, as well as the current state of the art in CA design. Critically, most existing models focus on users' perceptions of CAs based on prior experience, rather than evaluating user experience after direct interaction (Ling et al., 2021; Rheu et al., 2021). Moreover, the majority of studies have emphasized adoption-related factors, which address the pre-adoption stage of technology use (Benbasat & Wang, 2005; C. Li et al., 2024; Martins et al., 2014, p. 2024; Vidarshika et al., 2025). In contrast, as users have become more familiar with CAs and gained real interaction experience, there is a growing need for evaluation models addressing the post-acceptance stage. To address this gap, this dissertation developed a new measurement scale for evaluating user experience with CAs in the post-acceptance phase. The proposed scale integrates key constructs identified in the literature, including perceived social presence, perceived usability, trust (qualification- and goodwill-based), perceived enjoyment, and continuance intention behavior. This measurement model represents one of the dissertation's primary artifacts.

In parallel, the dissertation identified 14 design meta-requirements through a systematic literature review and user feedback, grounded in Information Systems Design Theory.

Furthermore, it analyzed and categorized design elements in practice, mapping them to broader design dimension categories. The validated measurement scale was employed to evaluate different chatbots featuring distinct design choices. The following sections discuss the findings from the validation of the measurement scale and examine how user experience varies across chatbots, emphasizing how design choices influence user perceptions and evaluations. In addition, while assessing these design choices, the dissertation integrates objective quantitative data obtained through eye-tracking analysis, providing complementary insights into users' visual attention and interaction behavior.

5.2.1. Unidimensional Overall User Experience Scale

This dissertation followed MacKenzie et al.'s (2011)systematic approach to develop and validate the measurement scale. Three empirical studies were conducted to ensure the scale's theoretical empirical robustness. The first study was a laboratory experiment using a real chatbot (called BuddyGPT). Each participant interacted with the chatbot for approximately 30 minutes and subsequently evaluated their interaction experience. In total, 87 participants took part in this study. The collected data were analyzed using Bayesian Exploratory Factor Analysis, which was selected for its robustness with relatively small samples and its ability to use probability distributions to estimate uncertainty in factor loadings (Hoofs et al., 2017; Schmettow, 2021). The analysis led to the refinement of the initial scale, resulting in a 16-item instrument with two highly correlated factors. The second study involved another laboratory experiment with two service chatbots, yielding a total of 44 evaluations from 22 participants. A second BEFA and Parallel analysis were performed to further assess the structure. This revealed the items converged into a single factor, indicating that user experience could be represented as a unidimensional construct. The third study was an online confirmatory study, in which participants evaluated six different chatbots, producing a total of 363 valid evaluations after data screening and quality checks. Once the dataset met the assumptions for factor analysis, a Bayesian Confirmatory Factor Analysis was conducted with the priors of previous study results. Across three studies employing Bayesian factor analysis approach, the measurement model consistently demonstrated high reliability ($\alpha > 0.90$) and stable factor structure, confirming its validity as a unidimensional instrument for assessing overall user experience with conversational agents.

According to development results, the users evaluate the CAs' interaction as a single, holistic experience instead of a separate perception of usability, trust, or enjoyment. These findings highlight an important fact that social technologies are perceived as a whole, where practical use and emotional connection come together to shape how they feel about the interaction (Diederich et al., 2022; Goernemann & Spiekermann, 2021; X. Yang et al., 2019). The interaction with CAs contains the interaction quality and CA competency into one overall impression (Chandra et al., 2022; Schuetzler et al., 2020). The finding that user experience forms a unidimensional construct supports evidence from prior usability and UX instruments such as the System Usability Scale (SUS; (Brooke, 1996)) and UMUX-Lite (Lewis et al., 2013). They capture a holistic perception of system quality rather than isolated dimensions. Moreover, global measures like the Net Promoter Score (Reichheld, 2011)confirm that users tend to express their judgments through a single, overarching evaluation.

From a methodological standpoint, the successful convergence of three independent studies using a Bayesian approach provides robust evidence of construct validity and model stability. Bayesian estimation allowed the inclusion of prior knowledge and probabilistic uncertainty, addressing sample-size limitations and producing more interpretable posterior credibility bounds (Borsci & Schmettow, 2024). The consistency of the results across exploratory and confirmatory phases supports MacKenzie et al.'s (2011) framework. The validated unidimensional overall user experience scale represents one of the key artifacts produced through the DSR process, contributing to the knowledge domain of conversational agent design and evaluation. Although the model has been validated, future studies are encouraged to test its generalizability across different contexts. Expanding the sample to include voice-based and embodied agents may reveal modality-specific nuances in how users perceive and evaluate their experience.

5.2.2. Design Analysis Using the Conversational Agent Design Elements List

The development of CA design meta-requirements aims to provide a structured foundation for identifying design choices that influence user experience. These meta-requirements serve as high-level design objectives guiding the systematic exploration of conversational agent design. The development process of the meta-requirements adopted Information Systems Design Theory, which employs kernel theories to identify IS product meta-requirements and explain why specific design features are expected to be effective (Hevner et al., 2024b; Jones & Gregor, 2007; Walls et al., 1992). To achieve this, a systematic literature review was conducted to elicit design requirements for CAs. The inclusion criteria targeted studies that integrated at least one theoretical perspective while exploring user experience with CAs. As a result, all requirements were derived from theoretically tested studies, ensuring conceptual grounding (X. Yang & Aurisicchio, 2021; Zierau et al., 2020). During the SLR, design choices in each study were coded and organized, resulting in a design elements list that ensured alignment with the identified meta-requirements. In addition, empirical studies combine multiple design elements to enhance CA performance. The studies were examined to understand how specific configurations affect user experience. These design elements were then grouped under three design dimension categories: anthropomorphic design dimensions, agent characteristics, and agent competency.

This thesis used the design elements list to evaluate real service chatbot designs. The presence and absence of each design element were independently reviewed by two experts who interacted with the selected chatbots and examined them for evidence of each element. The analysis of real service chatbots revealed that most existing systems tend to incorporate design elements related to anthropomorphic design dimensions, a widely accepted approach in CA design. Feine et al. (2019) describe this tendency as a reflection of interpersonal communication, which relies on social cues similar to human–human interaction. Theories such as Social Response Theory, CASA, Social Presence Theory, and Anthropomorphism Theory further support this observation.

Critically, only one agent (Seattle Chatbot) incorporated design elements primarily focused on the agent competency dimension, highlighting an underrepresented yet critical area of CA design. According to the online evaluations of six service chatbots, the Seattle chatbot provided the most positive user experience. It was followed by the Wanderlog chatbot, which also employed a wider range of design dimensions, indicating that its design better met the identified meta-requirements. In contrast, the Lufthansa and University of Twente chatbots received lower user experience scores and relied mainly on anthropomorphic design elements. Although previous studies in the literature have demonstrated the positive effects of anthropomorphic design on user perception and engagement (Cai et al., 2023; Pietrantoni et al., 2022; Seeger et al., 2017; Zhou et al., 2019), this dissertation takes a more holistic perspective. The findings suggest that while anthropomorphic features can be a cue to natural interaction agents; however, they do not necessarily compensate for limited interaction quality (A.-M. Seeger et al., 2021). Similar to prior observations by Grimes et al (2021) and Følstad and Brandtzæg (2020), the results indicate that when conversational agents appear socially expressive but fail to perform accurately or respond effectively, users' expectations are not met, leading to lower user experience. Further, the results emphasize that design strategies supporting agent competency, such as explanation facilities and responsiveness, produce positive user experiences. With the advancement of AI-based and large language model technologies, these communicative abilities can now be more effectively integrated into CA, improving their interaction quality. Given that CAs function as always-available (24/7) service technologies, enhancing their competency and interaction quality is essential. Therefore, this dissertation suggests that future CA designs should employ competency-driven communication strategies to deliver richer, more reliable, and engaging user experiences.

5.2.3. Comparative Visual Attention Evaluation of Chatbot User Experience

To complement the subjective evaluations and design analysis, an eye-tracking experiment was conducted to examine users' visual attention and interaction behaviors while completing predefined tasks with two service chatbots. The study aimed to understand how design richness, in terms of operationalized as the presence and variety of design elements, affects users' objective behavioral data during conversational agent interaction. Eye-tracking data were collected from participants as they performed identical information-retrieval tasks across both interfaces. Areas of Interest were defined for key interface components: the chatbot window, response area, input message area, interface buttons, and webpages. The interaction was segmented into three distinct snapshots to capture the temporal progression of user engagement. The first snapshot captured the initial webpage with the chatbot bubble visible; the second encompassed the chatbot's welcome message after users clicked the bubble, extending until their first input appeared in the chat window; and the third covered the subsequent interaction flow. This segmentation enabled the identification of changes in visual attention across different interaction phases and the exploration of which interaction aspects captured attention during the conversational exchange (Raidt, 2009).

Based on the design analysis (see Appendix C.1), the two chatbots represented contrasting design approaches. The USCIS chatbot (low design richness) incorporated more identity cues and interactive elements, such as buttons. On the contrary, the Seattle chatbot (high design richness) employed more conversational features and explanatory elements, including greater lexical diversity and reasoned utterances. To understand different chatbot experiences with objective data, the researcher first examined the accessibility of each chatbot using the first snapshot, measuring time to first fixation. The results revealed that both chatbots received similar initial attention. During the experiment, participants interacted with both chatbots in randomized order, and no learning effect was observed. The only notable trend was that the low chatbot included observable elements (e.g., image and name) on the chatbot bubble, which became more noticeable during the second interaction. However, the results showed no significant order effect. Moreover, fixation count showed no significant difference in chatbot bubbles. This step was important to verify because accessibility issues can affect subsequent interactions (Borsci et al., 2022a; Grassini et al., 2025; Saadé & Otrakji, 2007; Stanley et al., 2022; Sunil, 2025). For example, if a user has difficulty locating a chatbot, they may perceive the interaction as more challenging. These factors can shape users' expectations. This dissertation employed chatbots that received similar initial attention. Ensuring comparable initial accessibility allowed later differences in user behavior to be attributed to design features rather than discoverability issues.

In the following, the analysis of the chatbot introduction phase (snapshot 2) is indicate a significant difference between the two chatbots in terms of visual behavior data, specifically fixation counts and fixation durations across the AOIs. The low chatbot produced longer fixation durations across AOIs, whereas the high chatbot showed a higher number of fixations on the AOIs in the chatbot's welcoming snapshot. Similarly, the analysis of chatbot interaction flow (snapshot 3) demonstrated patterns consistent with the introduction phase. These findings suggest that fundamentally different cognitive processes occurred across the two chatbot designs. Longer fixation durations indicate deeper cognitive processing, sustained interest, or increased comprehension effort, while higher fixation counts reflect more active visual scanning and information-seeking behavior.

Furthermore, the percentage distribution of AOIs revealed that attention in the low chatbot was spread across multiple interface areas, while in the high chatbot, users' gaze was mainly focused on the chatbot's message area (J. Chen et al., 2024; Fornalczyk et al., 2021). These findings recommend that differences in interface composition and visual hierarchy influenced how users allocated their visual attention across the chatbot interfaces. Also, the third snapshot further revealed that users were more easily distracted by webpage elements in the low chatbot condition, searching for information across different areas, whereas in the high chatbot condition, users' attention remained focused on the chatbot's responses, and they did not experience difficulty locating the information they were seeking for both chatbots. he task success rates were nearly identical for both chatbots. These findings suggest that design differences influence how users process information rather than whether they succeed. This distinction highlights the importance of user experience design; even when task outcomes are similar, interface design can significantly shape users' overall experience (Giri et al., 2024). The findings of this study did not reveal a significant relationship between overall user experience UX evaluation and fixation duration; however, a significant relationship was observed between fixation count and UX scores in the low chatbot condition. These suggest that longer fixation durations tended to correspond with lower UX ratings, and a higher number of fixations was also associated with lower UX scores, especially in the low chatbot condition. This pattern supports the prior study findings that indicate increased visual effort and frequent gaze shifts may reflect higher cognitive demand (Argunsah et al., 2025; Slomianka et al., 2025).

Consequently, this dissertation advocates a holistic approach to conversational agent design and their evaluations, integrating agent design dimensions that capture meta-design requirements and multiple measurements. The first part of the discussion outlined how CAs should be designed through kernel theory-driven meta requirements, while the second part reflects how these design choices influence user experience in practice. In other words, this dissertation's artifacts connect prescriptive design theory with observed user behavior. Although the empirical findings do not fully confirm all requirements, they partially reflect and support the proposed theoretical principles.

CONCLUSION

This dissertation aims to enhance the understanding and design of Conversational Agents by integrating theoretical, methodological, and empirical perspectives with the Design Science Research approach. Two main artifacts guided this research: (1) formulation of meta-requirements that translate theoretical foundations into prescriptive design knowledge, and (2) development and validation of a measurement scale that heuristically captures user experience with CAs. By addressing these aims, this dissertation contributes to the design and evaluation of CAs, bridging the gap between theory and practical implementation.

Theoretical Implications

Importantly, this dissertation extends existing findings by formulating design knowledge that is both theory-oriented and design-science grounded. Guided by the DSR approach, the theoretical contribution lies in translating fragmented empirical observations into structured, prescriptive knowledge. Previous CA design studies, often informed by theories such as the CASA paradigm and Social Response Theory, have emphasized human-likeness through the use of social cues (e.g., names, avatars, or emotional expression) to simulate human communication (J. Chen et al., 2024; Feine et al., 2019; Munnukka et al., 2022; Seeger et al., 2018; Wagner et al., 2019). However, these studies have largely overlooked that achieving human-like interaction also requires human-level communicative abilities, including natural timing, contextual understanding, and expressive responsiveness, in other words, saying the right thing at the right time, even without knowing the response.

As artificial intelligence and large language models continue to evolve rapidly, the technical barriers to natural language generation have largely been overcome (Kumar, 2024). What remains essential is ensuring that agent interaction feels natural and socially appropriate, while addressing the ethical and communicative concerns brought by AI technologies, such as misinformation and overtrust (Cornell University, 2024; Jenks, 2025; D. Yang et al., 2025). The 14 meta-requirements defined in this dissertation address this gap by combining anthropomorphic design dimensions with HCI theories of natural interaction. In doing so, they move beyond surface-level anthropomorphism, emphasizing that the realism of interaction emerges from communication quality, clarity, feedback, and real flow, rather than from visual or verbal imitation.

Furthermore, the findings highlight that transparency and expressiveness remain crucial determinants of positive user experience. This dissertation's important endeavor was identifying agent competency, which represents an emerging theoretical construct with strong potential for future research. It reflects the agent's ability to communicate meaningfully and adaptively, aligning with the principles of communicative competence in human interaction (Guzman & Lewis, 2020; Wiemann, 1977). Whether the agent performs multiple complex tasks or serves a limited function, communicative competence is essential for sustaining effective, context-aware interaction. By defining generalizable and context-independent meta-requirements, this study contributes to the broader discussion in requirements engineering and human-centered system design. Empirical evidence in system design research consistently shows that only a small portion of requirements directly related to user needs accounts for the majority of user satisfaction (Moriuchi, 2021). However, many technological systems are eventually discontinued because they fail to align with actual user expectations (Altrichter & Benoit, 2025). This dissertation offers a pathway toward more sustainable, acceptable, and human-aligned conversational technologies by emphasizing user-centered requirements and integrating experiential dimensions into CA design.

An additional theoretical contribution of this dissertation lies in its multi-method triangulation, which integrates literature-based design knowledge with empirical evidence from survey and eye-tracking studies. The design elements list, systematically derived from the literature, is a theoretical synthesis that operationalizes abstract design dimensions into observable features. This artifact extends prior taxonomies and provides a measurable link between theoretical constructs. The study contributes to a deeper theoretical understanding of how users perceive, interpret, and visually engage with design features through the combination of survey-based evaluation and eye-tracking analysis. While survey data reveal users' conscious evaluations of experience, eye-tracking offers objective behavioral evidence of attention, cognitive effort, and interaction focus. These complementary methods strengthen the validity of the theoretical claims and illustrate how user experience can be studied as a subjective perception and an observable behavioral process. This triangulated approach bridges design theory with human–

computer interaction research, offering a richer, evidence-based understanding of how design characteristics shape the quality of human–agent communication.

This dissertation is consistent with HCI and IS theories. The results demonstrate how design choices rooted in these theoretical perspectives shape user experience. By combining design theory with user-centered evaluation, the study bridges the gap between artifact design and user experience, illustrating that effective CA design depends on functionality, communication quality, and perceived social connection.

Practical Implications

This dissertation offers several important practical implications for the design and development of conversational agents. The formulated meta-requirements provide a theory-driven foundation for understanding and improving user experience with CAs. These requirements ensure that design choices reflect interaction's functional and experiential aspects. The meta-requirements cover user considerations and communication flow and collectively offer a structured framework that can enhance users' experience with CAs. This approach supports the creation of conversational agents that are usable, efficient, as well as competent. By aligning system behavior with human communication norms, the resulting interactions become more coherent, engaging, and acceptable. Unlike many existing CAs that focus primarily on the surface level of humanlikeness, the proposed framework emphasizes the quality of interaction and communicative competence. Human-like features are employed only to support effective communication, not merely to replace.

The design elements list can be utilized as a design checklist. It reflects the integration of the meta-requirements and the theoretical dimensions of conversational agent design, including anthropomorphic cues, agent characteristics, and agent competency. The design elements list guides and assists CA designers in decision-making, such as selecting appropriate communication modes, balancing human-like and functional features, and ensuring that the agent's behavior reflects the intended interaction goals. Through the design process, CA designers can iteratively improve agents by adapting specific design elements to context, user expectations, and task requirements, thereby enhancing both usability and interaction quality.

The artifact of this dissertation also offers crucial practical implications for organizations. By validating how their conversational agents meet the proposed meta-requirements, organizations can make informed decisions about design improvements and investment priorities. This validation enables them to identify which aspects of their agents enhance user experience and which require redesign or additional support. Specifically, conversational agents deployed in service contexts represent the organization's brand (Bergner et al., 2023). Their interaction quality and the experience they provide directly influence how users perceive the brand's reliability, professionalism, and customer orientation. Therefore, improving CA design based on the meta-requirements can strengthen organizational reputation and foster more positive, trust-based relationships with users.

Conversational agents today range from restricted rule-based systems to competent AI-driven models, some of which are perceived as having almost unlimited abilities. This evolution raises increasing concerns about misinformation and hallucination in agent responses (Alkaissi & McFarlane, 2023; Emsley, 2023). Designing agents in accordance with the proposed meta-requirements ensures that their functionality and communication remain transparent, reliable, and socially appropriate. In doing so, these guidelines help maintain natural and responsible communication, preventing agents from exceeding their intended scope or violating users' expectations.

Limitations and Future Studies

While this dissertation makes several theoretical and methodological contributions to conversational agent design and evaluation, some limitations should be acknowledged to guide future research. First, the meta-requirements proposed in this dissertation were elicited through a systematic literature review complemented by user feedback obtained from empirical studies. This dual approach allowed the requirements to emerge from theoretical and design-oriented considerations, as well as from users' psychological and experiential responses after interaction with CAs. While this process ensured a strong link between theory and real experience, it was limited to the contexts and user groups examined. Moreover, SLR includes a structured coding schema, which also introduces certain limitations, while ensuring transparency and consistency (Bandara et al., 2015). The process required a high level of abstraction in defining and interpreting design dimensions. Although the coding was guided by theoretical alignment, some nuances or contextual variations in individual studies may have been lost during the abstraction process. Future studies could extend this approach by engaging more diverse participants, broadening the scope of design dimensions, and incorporating new or evolving factors that influence user experience.

Second, the empirical studies were conducted with text-based conversational agents. Prior research has reported exploring modality differences in user experience (Wambsganss, 2021). Incorporating only text-based agents may limit the generalizability of the findings to other interaction modalities, such as voice-based or multimodal systems. Future studies should examine different communication modalities to develop a more comprehensive understanding of user experience differences.

Third, the chatbots included in the study represented different contexts and purposes. For instance, some were designed for hedonic interaction, while others focused on pragmatic, task-oriented use. This variation reflects the diversity of real-world chatbot applications; at the same time, this limits the comparability of user experience scores across systems. Differences in task goals, interaction depth, and user expectations can influence how participants evaluate their experiences. Future research could address this limitation by examining context-specific or by grouping chatbots according to their primary function, allowing for more consistent comparisons.

Fourth, the final dataset was collected through an online study where participants could choose which chatbots they wanted to interact with. This flexible design helped attract more participants and made the study easier to complete, but it also meant that the interactions were not randomized. As a result, some chatbots received more attention than others, which may have affected the balance of comparisons (Lazar et al., 2017). Asking participants to test all six chatbots would likely have reduced participation, while limiting the study to only a few chatbots would have narrowed the comparison. Future studies could consider a partially randomized approach to maintain high participation while ensuring a more even distribution across chatbot conditions.

Fifth, future studies could evaluate the proposed user experience model in relation to other established UX and usability frameworks to examine their relative sensitivity in capturing user–agent interaction quality. Such comparative investigations would provide a more nuanced understanding. Moreover, future research could extend the analysis of design features by focusing on the interplay between explanation facilities, expressiveness, error handling, and transparency across different usage contexts and modalities. Examining how these design features jointly influence continuance intention would contribute to a more comprehensive understanding of user experience and strengthen the theoretical robustness of the proposed model and design requirements.

Lastly, the eye-tracking study was limited in its ability to determine which design approach was objectively better. Although the analysis revealed significant differences in visual attention patterns between the two chatbot interfaces under controlled laboratory conditions, eye-tracking data alone cannot explain the underlying reasons for these differences (Bruneau et al., 2002; J. H. Goldberg, 2000). The method shows where and how users allocate attention. Future studies could combine eye-tracking with complementary qualitative or physiological measures to examine how design choices influence user experience.

REFERENCES

Abu Shawar, B., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, *22*(1), 29–49. https://doi.org/10.21248/jlcl.22.2007.88

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Agarwal, R., & Prasad, J. (1999). Are individual differences germane to the acceptance of new information technologies? *Decision Sciences*, *30*(2), 361–391. https://doi.org/10.1111/j.1540-5915.1999.tb01614.x

Aitamurto, T., Zhou, S., Sakshuwong, S., Saldivar, J., Sadeghi, Y., & Tran, A. (2018). Sense of presence, attitude change, perspective-taking and and usability in first-person split-sphere 360° video. *CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Aktaş, N. B., & Akbıyık, A. (2025). Conversational agent design: a comprehensive analysis of research from leading conferences. *International Conference on Intelligent Human Computer Interaction* (pp. 105–121).

Aktaş, N. B., & Korkmaz, C. (2025). Neuro information systems. In *Current and Prospective Approaches, Methods, and Techniques for Management Information Systems Research* (pp. 65–89). Sakarya Yayınclılık. https://www.researchgate.net/publication/381741783_NEURO_INFORMATION_SYSTEMS

Aktaş, N. B., Şişman, B., & Borsci, S. (2025). Unleashing the potential of Turkish chatbots: A study on the validity and reliability of the bot usability scale. *Universal Access in the Information Society*, *23*(4). https://doi.org/10.1007/s10209-025-01211-9

Alagarsamy, S., & Mehrolia, S. (2023). Exploring chatbot trust: Antecedents and behavioural outcomes. *Heliyon*, *9*(5), e16074. https://doi.org/10.1016/j.heliyon.2023.e16074

Albert, W., & Tedesco, D. (2010). Reliability of self-reported awareness measures based on eye tracking. *Journal of Usability Studies, 5(2),* 50-64.

Cornell University (2024, January 17) Should agentic conversational AI change how we think about ethics? Characterising an interactional ethics centred on respect Retrieved August 15 from https://doi.org/10.48550/arXiv.2401.09082

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus 15(2)*. https://doi.org/10.7759/cureus.35179

Allouch, M., Azaria, A., & Azoulay, R. (2021). Conversational agents: goals, technologies, vision and challenges. *Sensors*, *21*(24), 8448. https://doi.org/10.3390/s21248448

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2010). Trustworthy virtual advisors and enjoyable interactions: designing for expressiveness and transparency. *ECIS*, 116.

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2011). The adoption of online shopping assistants: Perceived similarity as an antecedent to evaluative beliefs. *Journal of the Association for Information Systems*, *12*(5), 347–374. https://doi.org/10.17705/1jais.00267

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2021). Designing online virtual advisors to encourage customer self-disclosure: a theoretical model and an empirical test. *Journal of Management Information Systems*, *38*(3), 798–827. https://doi.org/10.1080/07421222.2021.1962595

Al-Natour, S., Benbasat, I., & Cenfetelli, R. T. (2022). Designing caring and informative decision aids to increase trust and enhance the interaction atmosphere. *AIS Transactions on Human-Computer Interaction*, *14*(1), 1–29.

Altrichter, B., & Benoit, S. (2025). Technology discontinuance: A systematic literature review and research agenda. *European Journal of Information Systems*, *0*(0), 1–25. https://doi.org/10.1080/0960085X.2025.2516427

Appel, J., Von Der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, *2012*. https://doi.org/10.1155/2012/324694

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*(2018), 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Argunsah, H., Altıntaş, L., & Şahiner, M. (2025). Eye-tracking insights into cognitive strategies, learning styles, and academic outcomes of Turkish medicine students. *BMC Medical Education*, *25*(1), 276. https://doi.org/10.1186/s12909-025-06855-y

Ashfag, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, *54*(2020), 101473.

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: repair strategy preferences for conversational breakdowns. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, (pp. 1–12). https://doi.org/10.1145/3290605.3300484

Aurum, A., & Wohlin, C. (2005). *Engineering and managing software requirements*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/3-540-28244-0

Avison, D., Fitzgerald, G., & Powell, P. (2006). An opportunity for editors of I.S. journals to relate their experiences and offer advice. The editorial view of David Avison Guy Fitzgerald and Philip Powell, Editors: Of the Information Systems Journal: Second in a series. *European Journal of Information Systems*, *15*(3), 241–243. https://doi.org/10.1057/palgrave.ejis.3000625

Balakrishnan, J., Abed, S. S., & Jones, P. (2022). The role of meta-UTAUT factors, perceived anthropomorphism, perceived intelligence, and social self-efficacy in chatbot-based services? *Technological Forecasting and Social Change*, *180*(December 2021), 121692. https://doi.org/10.1016/j.techfore.2022.121692

Balakrishnan, J., & Dwivedi, Y. K. (2021). Conversational commerce: Entering the next stage of AI-powered digital assistants. In *Annals of Operations Research* (Issue 0123456789). Springer US. https://doi.org/10.1007/s10479-021-04049-5

Bandara, W., Furtmueller, E., Gorbacheva, E., Beekhuyzen, J., Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., Miskon, S., & Beekhuyzen, J. (2015). Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems*, *37*, 154–204.

Barfield, W., & Furness, T. A. (1995). *Virtual environments and advanced interface design*. Oxford University Press. https://doi.org/10.1016/0160-9327(96)88424-9

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3559–3565. https://doi.org/10.1098/rstb.2009.0148

Beaudry, A., & Pinsonneault, A. (2010). The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. *MIS Quarterly*, *34*(4), 689–710. https://doi.org/10.2307/25750701

Becker, J.-M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning*, *45*(5–6), 359–394. https://doi.org/10.1016/j.lrp.2012.10.001

Beer, J. M., Smarr, C. A., Fisk, A. D., & Rogers, W. A. (2015). Younger and older users' recognition of virtual agent facial expressions. *International Journal of Human Computer Studies*, *75*, 1–20. https://doi.org/10.1016/j.ijhcs.2014.11.005

Belanche, D., Casalo, L. V., & Flavian, C. (2019). Artificial intelligence in FinTech: Understanding robo-advisors adoption among customers. *Industrial Management*

& Data Systems, *119*(7), 1411–1430. https://doi.org/10.1108/IMDS-08-2018-0368

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Service robot implementation: A theoretical framework and research agenda. *The Service Industries Journal*, *40*(3–4), 203–225. https://doi.org/10.1080/02642069.2019.1672666

Beldad, A., Hegner, S., & Hoppen, J. (2016). The effect of virtual sales agent (VSA) gender—Product gender congruence on product advice credibility, trust in VSA and online vendor, and purchase intention. *Computers in Human Behavior*, *60*, 62–72. https://doi.org/10.1016/j.chb.2016.02.046

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200. https://doi.org/10.1037/h0024835

Ben Mimoun, M. S., Poncin, I., & Garnier, M. (2017). Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics. *Information & Management Journal*, *54*(2017), 545–559. http://dx.doi.org/10.1016/j.im.2016.11.008

Benbasat, I., Dimoka, A., Pavlou, P. A., & Qiu, L. (2020). The role of demographic similarity in people's decision to interact with online anthropomorphic recommendation agents: Evidence from a functional magnetic resonance imaging (fMRI) study. *International Journal of Human Computer Studies*, *133*, 56–70. https://doi.org/10.1016/j.ijhcs.2019.09.001

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 72–101. https://doi.org/10.17705/1jais.00065

Benke, I. (2020). Towards design principles for trustworthy affective chatbots in towards design principles for trustworthy affective chatbots in virtual teams virtual teams. *ECIS - Research-in-Progress Papers* (pp. 41). https://aisel.aisnet.org/ecis2020_rip/41

Bergner, A. S., Hildebrand, C., & Häubl, G. (2023). Machine talk: how verbal embodiment in conversational ai shapes consumer–brand relationships. *Journal of Consumer Research*, *50*(4), 742–764. https://doi.org/10.1093/jcr/ucad014

Bhattacherjee, A. (2001). An empirical analysis of the antecedents of electronic commerce service continuance. *Decision Support Systems*, *32*(2), 201–214. https://doi.org/10.1016/S0167-9236(01)00111-7

Bickmore, T. W., Pfeifer, L. M., & Jack, B. W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1265–1274). https://doi.org/10.1145/1518701.1518891

Biocca, F. (1997). The Cyborg's Dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2), 1–46. https://doi.org/10.1111/j.1083-%206101.1997.tb00070.x

Biocca, F., Harms, C., & 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), 456–480. https://doi.org/10.1162/105474603322761270

Bitrián, P., Buil, I., & Catalán, S. (2021). Enhancing user engagement: The role of gamification in mobile apps. *Journal of Business Research*, *132*, 170–185. https://doi.org/10.1016/j.jbusres.2021.04.028

Behavioral, Management, Social Science (BMS) Lab. (2025). *Equipment* Retrieved October 4, 2025 from https://www.utwente.nl/en/bmslab/facilities-equipment-and-software/equipment/#equipment

Bødker, S. (2006). When second wave HCI meets third wave challenges. *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, 1–8. https://doi.org/10.1145/1182475.1182476

Boehm, B. (1989). Experiences with the spiral model as a process model generator. *Proceedings of the 5th International Software Process Workshop*, 43–45. https://www.computer.org/csdl/proceedings-article/ispw/1989/00690409/12OmNylKAJ1

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. *International Journal of Human-Computer Interaction*, *31*(8), 484–495. https://doi.org/10.1080/10447318.2015.1064648

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022a). The chatbot usability scale: the design and pilot of a usability scale for interaction with ai-based conversational agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Borsci, S., & Schmettow, M. (2024). Re-examining the chatbot sability scale (BUS-11) to assess user experience with customer relationship management chatbots. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-024-01834-4

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A., & van der Velde, F. (2022b). A confirmatory factorial analysis of the chatbot usability scale: a nultilanguage validation. *Personal and Ubiquitous Computing*, *0123456789*. https://doi.org/10.1007/s00779-022-01690-0

Bowman, R., Cooney, O., Newbold, J. W., Thieme, A., Clark, L., Doherty, G., & Cowan, B. (2024). Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human-Computer Studies*, *184*, 103181. https://doi.org/10.1016/j.ijhcs.2023.103181

Brahnam, S., & De Angeli, A. (2008). Special issue on the abuse and misuse of social agents. *Interacting with Computers*, *20*(3), 287–291. https://doi.org/10.1016/j.intcom.2008.02.001

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *Lecture Notes in Computer Science*, *10673*, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30

Brendel, A. B., Greve, M., Diederich, S., & Riquel, J. (2020). 'You are an idiot!'-how conversational agent communication patterns influence frustration and harassment. *AMCIS 2020*. www.uni-goettingen.de/im

Brendel, A. B., Hildebrandt, F., Dennis, A. R., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, *40*(3), 883–913. https://doi.org/10.1080/07421222.2023.2229127

Brill, T. M., Munoz, L., & Miller, R. J. (2019). Siri, Alexa, and other digital assistants: A study of customer satisfaction with artificial intelligence applications. *Journal of Marketing Management*, *35*(15–16), 1401–1436. https://doi.org/10.1080/0267257X.2019.1687571

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In *Usability Evaluation In Industry* (pp. 207–212). CRC Press. https://doi.org/10.1201/9781498710411-35

Bruneau, D., Sasse, M. A., & McCarthy, J. D. (2002). The eyes never lie: the use of eyetracking data in HCI research. *Proceedings of the CHI 2002: Conference on Human Factors in Computing Systems*. http://hcibib.org/archive/CHI

Bührke, J., Brendel, A. B., Lichtenberg, S., Greve, M., & Mirbabaie, M. (2021). Is making mistakes human? On the perception of typing errors in chatbot communication. *HICSS*, 4456–4465.

Burch, M., Wallner, G., Broeks, N., Piree, L., Boonstra, N., Vlaswinkel, P., Franken, S., & Van Wijk, V. (2021). The power of linked eye movement data visualizations. *ACM Symposium on Eye Tracking Research and Applications*, 1–11. https://doi.org/10.1145/3448017.3457377

Buyssens, H., Leuven, K., & Business, V. (2024). Meta-requirements for the design of a blockchain-enabled multi-sided platform for sustainability and circular economy. *Proceedings of the 57th Annual Hawaii Interna8onal Conference on System Sciences* (pp. 4311-4320). HICSS; Honolulu.

Cafaro, A., Ravenet, B., Ochs, M., Vilhjálmsson, H. H., & Pelachaud, C. (2016). The effects of interpersonal attitude of a group of agents on user's presence and proxemics behavior. *ACM Transactions on Interactive Intelligent Systems*, *6*(2), 1–33. https://doi.org/10.1145/2914796

Cai, W., Jin, Y., & Chen, L. (2022, April 29). Impacts of personal characteristics on user trust in conversational recommender systems. *Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517471

Cai, W., Jin, Y., Zhao, X., & Chen, L. (2023). "Listen to music, listen to yourself": Design of a conversational agent to support self-awareness while listening to music. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp.1–19). https://doi.org/10.1145/3544548.3581427

Calderón Garrido, C., Navarro González, D., Lorenzo Seva, U., & Ferrando Piera, P. (2019). Multidimensional or essentially unidimensional? A multi-faceted factor-analytic approach for assessing the dimensionality of tests and items. *Psicothema*, *4*(31), 450–457. https://doi.org/10.7334/psicothema2019.153

Card, S. K., Moran, T. P., & Newell, A. (2008). *The psychology of human-computer interaction* (Repr). CRC Press Taylor & Francis Group.

Carrizo, D., Dieste, O., & Juristo, N. (2014). Systematizing requirements elicitation technique selection. *Information and Software Technology*, *56*(6), 644–669. https://doi.org/10.1016/j.infsof.2014.01.009

Cassell, J., Nakano, Y. I., Bickmore, T. W., Sidner, C. L., & Rich, C. (2001). Non-verbal cues for discourse structure. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* (pp. 114-123). https://doi.org/10.3115/1073012.1073028

Cavedon, L., Kroos, C., Herath, D., Burnham, D., Bishop, L., Leung, Y., & Stevens, C. J. (2015). "C′Mon dude!": Users adapt their behaviour to a robotic agent with an attention model. *International Journal of Human-Computer Studies*, *80*, 14–23. https://doi.org/10.1016/j.ijhcs.2015.02.012

Ceha, J., & Law, E. (2022, April 29). Expressive auditory gestures in a voice-based pedagogical agent. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3491102.3517599

Cenfetelli, R. (2004). Inhibitors and enablers as dual factor concepts in technology usage. *Journal of the Association for Information Systems*, *5*(11), 472–492. https://doi.org/10.17705/1jais.00059

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To be or not to be …Human? Theorizing the role of human-like competencies in conversational artificial intelligence agents. *Journal of Management Information Systems*, *39*(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441

Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using ai chatbot that supports self-regulated learning in education: goal setting, feedback, and personalization. *Sustainability*, *15*(17), 12921. https://doi.org/10.3390/su151712921

Chattaraman, V., Kwon, W.-S., & Gilbert, J. E. (2012). Virtual agents in retail web sites: Benefits of simulated social interaction for older users. *Computers in Human Behavior*, *28*(6), 2055–2066. https://doi.org/10.1016/j.chb.2012.06.009

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315–330. https://doi.org/10.1016/j.chb.2018.08.048

Chaves, A. P., Egbert, J., Hocking, T., Doerry, E., & Gerosa, M. A. (2022). Chatbots language design: the influence of language variation on user experience with tourist assistant chatbots. *ACM Transactions on Computer-Human Interaction*, *29*(2), 1–38. https://doi.org/10.1145/3487193

Chen, J., Guo, F., Ren, Z., Li, M., & Ham, J. (2024). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International Journal of Human–Computer Interaction*, *40*(14), 3636–3654. https://doi.org/10.1080/10447318.2023.2193514

Chen, M., Wang, X., Wang, J., Zuo, C., Tian, J., & Cui, Y. (2021). Factors affecting college students' continuous intention to use online course platform. *SN Computer Science*, *2*, 114.

Cheng, Y., & Jiang, H. (2020). How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. *Journal of Broadcasting & Electronic Media*, *64*(4), 592–614. https://doi.org/10.1080/08838151.2020.1834296

Chin, H., & Yi, M. Y. (2019). Should an agent be ignoring it?: A study of verbal abuse types and conversational agents' response styles. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). https://doi.org/10.1145/3290607.3312826

Chung, H., Kang, H., & Jun, S. (2023). Verbal anthropomorphism design of social robots: Investigating users' privacy perception. *Computers in Human Behavior*, *142*. https://doi.org/10.1016/j.chb.2022.107640

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*(November 2018), 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73. https://doi.org/10.1177/002224377901600110

Ciardo, F., De Tommaso, D., & Wykowska, A. (2022). Joint action with artificial agents: Human-likeness in behaviour and morphology affects sensorimotor signaling and social inclusion. *Computers in Human Behavior*, *132*. https://doi.org/10.1016/j.chb.2022.107237

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: a case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17. https://doi.org/10.1559/152304009787340197

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics*, *183*(1), 31–57. https://doi.org/10.1016/j.jeconom.2014.06.008

Costea, I., & Sedrakyan, G. (2025). *Designing explainability features for LLM-based educational chatbots to promote reflective learning behavior*. https://research.utwente.nl/en/publications/designing-explainability-features-for-llm-based-educational-chatb/

CRAN. (2025). *Package 'BayesFM.'* CRAN R-Project. https://cran.r-project.org/web/packages/BayesFM/BayesFM.pdf

Cui, B., Wang, H., Ye, K., & Yan, J. (2012). Intelligent agent-assisted adaptive order simulation system in the artificial stock market. *Expert Systems with Applications*, *39*(10), 8890–8898. https://doi.org/10.1016/j.eswa.2012.02.018

Cui, T., Peng, X., & Wang, X. (2020). Understanding the effect of anthropomorphic design: towards more persuasive conversational agents. *ICIS 2020 Proceedings* (pp. 9). https://aisel.aisnet.org/icis2020/user_behaviors/user_behaviors/9

Cui, T., Wang, X., & Qi, J. (2021). Designing anthropomorphic therapeutic conversational agents: an uncanny valley perspective. *ECIS 2021 Research-in-Progress Papers*. https://aisel.aisnet.org/ecis2021_rip/30

Dahanayake, A., & Thalheim, B. (2011). Enriching conceptual modelling practices through design science. *Lecture Notes in Business Information Processing*, 497–510. https://doi.org/10.1007/978-3-642-21759-3_36

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817.

Dascalu, M.-I., Bodea, C.-N., Moldoveanu, A., Mohora, A., Lytras, M., & de Pablos, P. O. (2015). A recommender agent based on learning styles for better virtual collaborative learning experiences. *Computers in Human Behavior*, *45*, 243–253. https://doi.org/10.1016/j.chb.2014.12.027

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

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psycholog*, *22*(14), 1111–1132. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

De Cicco, R., e Silva, S. C., & Alparone, F. R. (2020). Millennials' attitude toward chatbots: An experimental study in a social relationship perspective. *International Journal of Retail and Distribution Management*, *48*(11), 1213–1233. https://doi.org/10.1108/IJRDM-12-2019-0406

Deci, E. L. (1975). *Intrinsic motivation*. In *Plenum Press*.

Dennis, A., Wixom, B. H., & Roth, R. M. (2014). *Systems analysis & design*. John Wiley & Sons.

Denovan, A., Dagnall, N., Drinkwater, K., & Escolà-Gascón, Á. (2023). Evaluating the psychometric properties of the Chronic Time Pressure Inventory using Rasch analysis. *PeerJ*, *11*, e15218. https://doi.org/10.7717/peerj.15218

DeVault, D., Artstein, R., Benn, G., Dey, T., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., & Morency, L.-P. (2014). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 1061–1068).

DeVellis, R. F. (2016). *Scale development theory and applications* (Fourth Edition). In SAGE Publication.

Devitt, S. K. (2018). Trustworthiness of autonomous systems. In J. Kacprzyk (Ed.), *Foundations of Trusted Autonomy* (pp. 161–184).

Diamantopoulos, A., & Winklhofer, H. M. (2001). index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845

Diederich, S., & Benedikt Brendel, A. (2019). Towards a taxonomy of platforms for conversational agent design. *14th International Conference on Wirtschaftsinformatik*.

Diederich, S., Brendel, A. B., & Kolbe, L. (2019). On conversational agents in information systems research: analyzing the past to guide future work. *14. Internationale Tagung Wirtschaftsinformatik. https://aisel.aisnet.org/wi2019/track13/papers/1/*

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020a). Designing anthropomorphic enterprise conversational agents. *Business and Information Systems Engineering*, *62*(3), 193–209. https://doi.org/10.1007/s12599-020-00639-y

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, *23*(1), 96–138. https://doi.org/10.17705/1jais.00724

Diederich, S., Lembcke, T.-B., Brendel, A. B., & Kolbe, L. (2020b). *Not human after all: Exploring the impact of response failure on user perception of anthropomorphic conversational service agents.* In Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference. https://aisel.aisnet.org/ecis2020_rp/110

Diederich, S., Lembcke, T.-B., Brendel, A. B., & Kolbe, L. M. (2021). Understanding the impact that response failure has on how users perceive anthropomorphic conversational service agents: insights from an online experiment. *AIS Transactions on Human-Computer Interaction*, 82–103. https://doi.org/10.17705/1thci.00143

Dilmegani, C. (2025). 50 ChatGPT use cases with real life examples. *AIMultiple*. https://research.aimultiple.com/chatgpt-use-cases/

Dimoka, A., Pavlou, P., Benbasat, I., & Qiu, L. (2010). 14P. Application of neuroimaging methods in is research: An fMRI study of online recommendation agents. *CONF-IRM 2010 Proceedings*. https://aisel.aisnet.org/confirm2010/4

Dollée, I. (2022). *The role of the eyes in the uncanny valley effect: does incongruence between eyes and face influence uncanniness?* University of Twente.

Dooley, J. (2025, June 3). *The future of chatbots powered by AI.* Retrieved September 12. Conveo. https://www.coveo.com/blog/future-of-chatbots/

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. MIT Press.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (Second edition). Springer.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albana, H., Albashraw, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Pagani, M. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*(2023), 102642.

Dybala, P., Ptaszynski, M., Maciejewski, J., Takahashi, M., Rzepka, R., & Araki, K. (2010). Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, *2*(1), 31–48. https://doi.org/10.3233/AIS-2010-0053

Elshan, E., Zierau, N., Engel, C., Janson, A., & Leimeister, J. M. (2022). Understanding the design elements affecting user acceptance of intelligent agents: past, present and future. *Information Systems Frontiers*, *24*(3), 699–730. https://doi.org/10.1007/s10796-021-10230-9

Emsley, R. (2023). ChatGPT: These are not hallucinations – they're fabrications and falsifications. *Schizophrenia*, *9*(1), 52, s41537-023-00379–4. https://doi.org/10.1038/s41537-023-00379-4

Enderle, J. D. (2012). Physiological modeling. In *Introduction to Biomedical Engineering* (pp. 817–936). Elsevier. https://doi.org/10.1016/B978-0-12-374979-6.00013-7

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

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

Eren, B. A. (2021). Determinants of customer satisfaction in chatbot use: Evidence from a banking application in Turkey. *International Journal of Bank Marketing*, *39*(2), 294–311. https://doi.org/10.1108/IJBM-02-2020-0056

Etzold, V., Braun, A., & Wanner, T. (2019). Eye tracking as a method of neuromarketing for attention research—An empirical analysis using the online appointment booking platform from Mercedes-benz. *Intelligent Decision Technologies 2019. Smart Innovation, Systems and Technologies, 143.* Springer, Singapore. https://doi.org/10.1007/978-981-13-8303-8_15

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Felnhofer, A., Kaufmann, M., Atteneder, K., Kafka, J. X., Hlavacs, H., Beutl, L., Hennig-Fast, K., & Kothgassner, O. D. (2019). The mere presence of an attentive and emotionally responsive virtual character influences focus of attention and perceived stress. *International Journal of Human Computer Studies*, *132*, 45–51. https://doi.org/10.1016/j.ijhcs.2019.07.010

Felton, W. M., & Jackson, R. E. (2022). Presence: A review. *International Journal of Human–Computer Interaction*, *38*(1), 1–18. https://doi.org/10.1080/10447318.2021.1921368

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Fisher, J., Shanks, G., & Lamp, J. (2007). A ranking list for information systems journals. *Australasian Journal of Information Systems*, *14*(5), 5–18.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *ACM Interactions*, 38–42.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, *5*(1), 3. https://doi.org/10.1007/s41233-020-00033-2

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11193 LNCS*, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16

Fornalczyk, K., Bortko, K., & Jankowski, J. (2021). Improving user attention to chatbots through a controlled intensity of changes within the interface. *Procedia Computer Science*, *192*, 5112–5121. https://doi.org/10.1016/j.procs.2021.09.289

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Franque, F. B., Oliveira, T., & Tam, C. (2023). Continuance intention of mobile payment: ttf model with trust in an african context. *Information Systems Frontiers*, *25*, 775–793.

Fraser, A. D., Branson, I., Hollett, R. C., Speelman, C. P., & Rogers, S. L. (2024). Do realistic avatars make virtual reality better? Examining human-like avatars for VR social interactions. *Computers in Human Behavior: Artificial Humans*, *2*(2), 100082. https://doi.org/10.1016/j.chbah.2024.100082

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, *21*(3), 219–246. https://doi.org/10.1007/s10648-009-9107-x

Gaborieau, J.-B., & Pronello, C. (2021). Validation of a unidimensional and probabilistic measurement scale for pro-environmental behaviour by travellers. *Transportation*, *48*(2), 555–593. https://doi.org/10.1007/s11116-019-10068-w

Gambino, A., Kim, J., & Sundar, S. S. (2019). Digital doctors and robot receptionists: User attributes that predict acceptance of automation in healthcare facilities. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). https://doi.org/10.1145/3290607.3312916

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega*, *28*(6), 725–737. https://doi.org/10.1016/S0305-0483(00)00021-9

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, *32*(6), 407–424. https://doi.org/10.1016/j.omega.2004.01.006

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*.

Gerlach, J. H., & Kuo, F.-Y. (1991). Understanding human-computer interaction for information systems design. *MIS Quarterly*, *15*(4), 527. https://doi.org/10.2307/249456

Giri, U., Sharma, A., Oza, S., Singh, P., Kumar Angra, P., & Khanna, A. (2024). Understanding customer service chatbot user experience: An experimental study of chatbot interaction design. *2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)* (pp. 118–123). https://doi.org/10.1109/INNOCOMP63224.2024.00028

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). "The chatbot is typing …"-The role of typing indicators in human-chatbot interaction. *Proceedings of the 17th Annual Pre-ICIS Workshop on HCI Research in MIS* (pp. 0–5).

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). The opposing effects of response time in human–chatbot interaction: the moderating role of prior experience. *Business & Information Systems Engineering*, *64*(6), 773–791. https://doi.org/10.1007/s12599-022-00755-x

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *ICIS 2017*. http://ksri.kit.edu

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

Goernemann, E. M., & Spiekermann, S. (2021). A value-based perspective on user experience—how alexa's value dispositions elicit emotional responses. *ECIS Proceedings.*

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior*, *52*, 1–11. https://doi.org/10.1016/j.chb.2015.05.008

Goldberg, J. H. (2000). Eye movement-based interface evaluation: What can and cannot be assessed? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(37), 625–628. https://doi.org/10.1177/154193120004403721

Grassini, E., Buzzi, M., Leporini, B., & Vozna, A. (2025). A systematic review of chatbots in inclusive healthcare: Insights from the last 5 years. *Universal Access in the Information Society*, *24*(1), 195–203. https://doi.org/10.1007/s10209-024-01118-x

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.2753/MIS0742-1222240302

Greulich, R. S., & Morana, S. (2023). Be a miracle-designing conversational agents to influence users' intention regarding organ donation. *ICIS 2023*. https://tu-dresden.de/bu/wirtschaft/winf/isd

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515. https://doi.org/10.1016/j.dss.2021.113515

Guadagno, R. E., Swinth, K. R., & Blascovich, J. (2011). Social evaluations of embodied agents and avatars. *Computers in Human Behavior*, *27*(6), 2380–2385. https://doi.org/10.1016/j.chb.2011.07.017

Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, *30*(4), 4493–4518. https://doi.org/10.1007/s10639-024-12881-y

Guo, F., & Zhang, X. (2020). The impact of brand history on consumers' cognitive process and brand attitude. *Journal of Neuroscience, Psychology, and Economics*, *13*(4), 191–203. https://doi.org/10.1037/npe0000136

Guo, J., Guo, J., Yang, C., Wu, Y., Yang, W., & Sun, L. (2021, May 6). Shing: A conversational agent to alert customers of suspected online-payment fraud with empathetical communication skills. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445129

Guo, Y., Wang, J., Wu, R., Li, Z., & Sun, L. (2022). Designing for trust: A set of design principles to increase trust in chatbot. *CCF Transactions on Pervasive Computing and Interaction*, *4*(4), 474–481. https://doi.org/10.1007/s42486-022-00106-5

Gupta, M., Dennehy, D., Parra, C. M., Mäntymäki, M., & Dwivedi, Y. K. (2023). Fake news believability: The effects of political beliefs and espoused cultural values. *Information & Management*, *60*(2), 103745. https://doi.org/10.1016/j.im.2022.103745

Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, *49*(March), 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Haas, C., & Moussawi, S. (2020). Are Anthropomorphic Intelligent Agents More Intelligent? *AMCIS 2020*. https://aisel.aisnet.org/amcis2020

Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Pearson Prentice Hall.

Hari, H., Iyer, R., & Sampat, B. (2022). Customer brand engagement through chatbots on bank websites– examining the antecedents and consequences. *International Journal of Human-Computer Interaction*, *38(13),* 1212–1227. https://doi.org/10.1080/10447318.2021.1988487

Harjunen, V. J., Spapé, M., Ahmed, I., Jacucci, G., & Ravaja, N. (2018). Persuaded by the machine: The effect of virtual nonverbal cues and individual differences on compliance in economic bargaining. *Computers in Human Behavior*, *87*, 384–394. https://doi.org/10.1016/j.chb.2018.06.012

Harrison, S., Tatar, D., & Sengers, P. (2007). The three paradigms of HCI. *In Alt. Chi. Session at the SIGCHI Conference on human factors in computing systems* (pp. 1–18).

Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology*, *25*(2), 91–97. https://doi.org/10.1080/01449290500330331

Haugeland, I. K. F., Følstad, A., Taylor, C., & Alexander, C. (2022). Understanding the user experience of customer service chatbots: an experimental study of chatbot interaction design. *International Journal of Human Computer Studies*, *161*. https://doi.org/10.1016/j.ijhcs.2022.102788

Hayashi, A., Chen, C., Ryan, T., & Wu, J. (2004). The role of social presence and moderating role of computer self efficacy in predicting the continuance usage of e-learning systems. *Journal of Information Systems*, *15*(2), 139–154.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

HBR. (2023). Data & visuals. *Harward Business Review*. https://hbr.org/data-visuals/2023/03/four-types-of-digital-humans

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

Hennessey, C. (2024, November 25). *Heatmaps vs. fixation maps: Visualizing eye gaze data with gazepoint.* Gazepoint, Retrieved September 20, 2025 from https://www.gazept.com/blog/analysis/heatmaps-vs-fixation-maps-visualizing-eye-gaze-data-with-gazepoint/

Hernandez-Bocanegra, D. C., & Ziegler, J. (2023). Explaining recommendations through conversations: dialog model and the effects of interface type and degree of interactivity. *ACM Transactions on Interactive Intelligent Systems*, *13*(2), 1–47. https://doi.org/10.1145/3579541

Herse, S., Vitale, J., & Williams, M.-A. (2023). Using agent features to influence user trust, decision making and task outcome during human-agent collaboration. *International Journal of Human–Computer Interaction*, *39*(9), 1740–1761. https://doi.org/10.1080/10447318.2022.2150691

Hess, T., Fuller, M., & Campbell, D. (2009). Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. *Journal of the Association for Information Systems*, *10*(12), 889–919. https://doi.org/10.17705/1jais.00216

Hessels, R. S., Nuthmann, A., Nyström, M., Andersson, R., Niehorster, D. C., & Hooge, I. T. C. (2024). The fundamentals of eye tracking part 1: The link between theory and research question. *Behavior Research Methods*, *57*(1), 16. https://doi.org/10.3758/s13428-024-02544-8

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems, 19(2), 4.*

Hevner, A. R., & Chatterjee, S. (2010). *Design research in information systems: Theory and practice*. Springer. https://doi.org/10.1007/978-1-4419-5653-8

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004a). Two paradigms on research essay design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Hevner, A. R., March, S. T., Park, J., Ram, S., & Ram, S. (2004b). Research essay design science in information. *MIS Quarterly*, *28*(1), 75–105.

Hew, J.-J., Leong, L.-Y., Tan, G. W.-H., Lee, V.-H., & Ooi, K.-B. (2018). Mobile social tourism shopping: A dual-stage analysis of a multi- mediation model. *Tourism Management*, *66*(2018), 121–139. https://doi.org/10.1016/j.tourman.2017.10.005

Hickey, A. M., & Davis, A. M. (2004). A unified model of requirements elicitation. *Journal of Management Information Systems*, *20*(4), 65–84. https://doi.org/10.1080/07421222.2004.11045786

Hildebrandt, F., Brendel, A. B., Dennis, A. R., & Sachdeva, A. (2023a). New bots-The influence of a conversational agent’s rookie personality on users’ satisfaction. *ICIS 2023*. https://www.researchgate.net/publication/374388320

Hildebrandt, F., Lichtenberg, S., Brendel, A. B., & Landmann, E. (2023b). Who’s bad?-The influence of perceived humanness on users’ intention to complain about conversational agent errors to others. *ICIS*. https://www.researchgate.net/publication/374388134

Hildebrandt, F., Lichtenberg, S., Brendel, A. B., & Riquel, J. (2023). Conversational agents in service context: Towards a classification of human-like design expectations. *AMCIS 2023 Proceedings (pp. 2.)* https://aisel.aisnet.org/amcis2023/sig_hci/sig_hci/2

Himanshu, R. R. (2021). *Conversational AI market size, share, growth & trends analysis report by component, deployment, type, and technology, and end user: global opportunity analysis and industry forecast, 2021–2030*. Allied Market Research.

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Hoelter, J. W. (1983). The analysis of covariance structures: goodness-of-fit indices. *Sociological Methods & Research*, *11*(3), 325–344. https://doi.org/10.1177/0049124183011003003

Hong, W., Chan, F. K. Y., Thong, J. Y. L., Chasalow, L. C., & Dhillon, G. (2014). A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research*, *25*(1), 111–136. https:// doi.org/10.1287/isre.2013.050

Hoofs, H., Scoot, R. van de, Jansen, N., & Kant, I. (2017). Evaluating model fit in bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement*, *78*(4), 537–568. https://doi.org/10.1177/0013164417709

Hoppe de Sousa, W., Zamudio Igami, M. P., & de Souza Bido, D. (2009). R&D management and the stokes diagram: An exploratory study. *Journal of Technology Management & Innovation*, *4*(4), 95–109. https://doi.org/10.4067/S0718-27242009000400008

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Hsiao, C.-H., Chang, J.-J., & Tang, K. (2016). Exploring the influential factors in continuance usage of mobile social Apps: Satisfaction, habit, and customer value perspectives. *Telematics and Informatics*, *33*(2006), 342–355. https://doi.org/10.1016/j.tele.2015.08.014

Hsiao, K. L., & Chen, C. C. (2021). What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction. *Library Hi Tech*, *40*(4), 929–946. https://doi.org/10.1108/LHT-08-2021-0274

Huang, S. Y. B., & Lee, C. J. (2022). Predicting continuance intention to fintech chatbot. *Computers in Human Behavior*, *129*. https://doi.org/10.1016/j.chb.2021.107027

Huang, S.-H., Huang, C.-Y., Lin, Y.-F., & Huang, T.-H. K. (2023). What types of questions require conversation to answer? A case study of askreddit questions. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544549.3585600

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning.

Journal of Computer Assisted Learning, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Hwang, A. H. C., & Won, A. S. (2021, May 6). Ideabot: Investigating social facilitation in human-machine team creativity. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445270

Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

Hyde, J., Carter, E. J., Kiesler, S., & Hodgins, J. K. (2015). Using an interactive avatar's facial expressiveness to increase persuasiveness and socialness. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, (pp. 1719–1728). https://doi.org/10.1145/2702123.2702465

International Organization for Standardization (ISO). (2018). *ISO 9241-11: Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. https://www.iso.org/standard/63500.html

International Organization for Standardization (ISO). (2019). *ISO 9241-210:2019 Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*. International Organization for Standardization (ISO). https://www.iso.org/standard/77520.html

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A monte carlo investigation. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(2), 205–223. https://doi.org/10.1207/S15328007SEM0802_3

Jackson, D. L., Gillaspy Jr., J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23. https://doi.org/10.1037/a0014694

Janson, A. (2023). How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior*, *149*. https://doi.org/10.1016/j.chb.2023.107954

Japutra, A., Molinillo, S., Utami, A. F., & Ekaputra, I. A. (2022). Exploring the effect of relative advantage and challenge on customer engagement behavior with mobile commerce applications. *Telematics and Informatics*, *72*(2022), 101841. https://doi.org/10.1016/j.tele.2022.101841

Jarvenpaa, S. L., Knoll, K., & Leidner, D. E. (1997). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Information Systems*, *14*(4), 29–64. https://doi.org/10.1080/07421222.1998.11518185

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218. https://doi.org/10.1086/376806

Jenks, C. J. (2025). Communicating the cultural other: Trust and bias in generative AI and large language models. *Applied Linguistics Review*, *16*(2), 787–795. https://doi.org/10.1515/applirev-2024-0196

Jiang, Z., Huang, X., Wang, Z., Liu, Y., Huang, L., & Luo, X. (2024). Embodied conversational agents for chronic diseases: Scoping review. *Journal of Medical Internet Research*, *26*(1), e47134. https://doi.org/10.2196/47134

Jin, E., & Eastin, M. (2024). Gender bias in virtual doctor interactions: gender matching effects of chatbots and users on communication satisfactions and future intentions to use the chatbot. *International Journal of Human–Computer Interaction*, *40*(23), 8246–8258. https://doi.org/10.1080/10447318.2023.2279402

Jin, S. V., & Youn, S. (2023). Social presence and imagery processing as predictors of chatbot continuance intention in human-AI-interaction. *International Journal of Human-Computer Interaction*, *39*(9), 1874–1886. https://doi.org/10.1080/10447318.2022.2129277

Jin, Y., Yang, K., Yan, L., Echeverria, V., Zhao, L., Alfredo, R., Milesi, M., Fan, J. X., Li, X., Gasevic, D., & Martinez-Maldonado, R. (2025). Chatting with a learning analytics dashboard: the role of generative AI literacy on learner interaction with conventional and scaffolding chatbots. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, (pp. 579–590). https://doi.org/10.1145/3706468.3706545

Jo, E., Epstein, D. A., Jung, H., & Kim, Y. H. (2023, April 19). Understanding the benefits and challenges of deploying conversational ai leveraging large language models for public health intervention. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3544548.3581503

Jones, D., & Gregor, S. (2007). The anotomy of a design theory. *Journal of the Association for Information Systems*, *8*(5). https://aisel.aisnet.org/jais/vol8/iss5/1/

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kallel, A., Ben Dahmane Mouelhi, N., Chaouali, W., & Danks, N. P. (2023). Hey chatbot, why do you treat me like other people? The role of uniqueness neglect in human-chatbot interactions. *Journal of Strategic Marketing*, *00*(00), 1–17. https://doi.org/10.1080/0965254X.2023.2175020

Kang, S.-H., & Gratch, J. (2014). Exploring users' social responses to computer counseling interviewers' behavior. *Computers in Human Behavior*, *34*(2014), 120–130.

Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: A cross-sectional comparison of pre- adoption and post-adoption beliefs. *MIS Quarterly*, *23*(2), 183–213.

Kasilingam, D. L. (2020). Understanding the attitude and intention to use smartphone chatbots for shopping. *Technology in Society*, *62*(2020), 101280. https://doi.org/10.1016/j.techsoc.2020.101280

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0482-1

Khosrawi-Rad, B., Rinn, H., Schlimbach, R., Gebbing, P., & Yang, X. (2022). Conversational agents in education—a systematic literature review. *ECIS 2022 Research Papers*, (pp. 18). https://aisel.aisnet.org/ecis2022_rp/18

Kim, H., Suh, K.-S., & Lee, U.-K. (2013). Effects of collaborative online shopping on shopping experience through social and relational perspectives. *Information & Management*, *50*(2013), 169–180. http://dx.doi.org/10.1016/j.im.2013.02.003

Kim, J., & Im, I. (2023). Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior*, *139*. https://doi.org/10.1016/j.chb.2022.107512

Kim, K., Schubert, R., & Welch. (2016). Exploring the impact of environmental effect. *Intelligent virtual agents 16th International Conference*. https://doi.org/10.1007/978-3-319-47665-0

Kim, Y., Reza, M., McGrenere, J., & Yoon, D. (2021). Designers characterize naturalness in voice user interfaces: their goals, practices, and challenges. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445579

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Kitchenham, B. (2004). *Procedures for performing systematic reviews* (Technical No. 0400011T.1; Joint Technical Report) https://www.researchgate.net/publication/228756057_Procedures_for_Performing_Systematic_Reviews.

Kitchenham, B. A., Brereton, P., Turner, M., Niazi, M. K., Linkman, S., Pretorius, R., & Budgen, D. (2010). Refining the systematic literature review process—Two participant-observer case studies. *Empirical Software Engineering*, *15*(6), 618–653. https://doi.org/10.1007/s10664-010-9134-8

Kocaballi, A. B., Quiroz, J. C., Laranjo, L., Rezazadegan, D., Kocielnik, R., Clark, L., Liao, Q. V., Park, S. Y., Moore, R. J., & Miner, A. (2020). Conversational agents for health and wellbeing. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, (pp. 1–8). https://doi.org/10.1145/3334480.3375154

Kock, N. (2004). The psychobiological model: Towards a new theory of computer-mediated communication based on darwinian evolution. *Organization Science*, *15*(3), 327–348. https://doi.org/10.1287/orsc.1040.0071

Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE Transactions on Professional Communication*, *48*(2), 117–130. https://doi.org/10.1109/TPC.2005.849649

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly: Management Information Systems*, *30*(4), 941–960. https://doi.org/10.2307/25148760

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, *139*. https://doi.org/10.1016/j.chb.2022.107513

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, *13*(2), 205–223. https://doi.org/10.1287/isre.13.2.205.83

Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J. (2018). Social snacking with a virtual agent – On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human Computer Studies*, *109*, 112–121. https://doi.org/10.1016/j.ijhcs.2017.09.001

Krämer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you—The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human Computer Studies*, *71*(3), 335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (Fourth edition). SAGE.

Kuhail, M. A., Al Katheeri, H., Negreiros, J., Seffah, A., & Alfandi, O. (2023a). Engaging students with a chatbot-based academic advising system. *International Journal of Human–Computer Interaction*, *39*(10), 2115–2141. https://doi.org/10.1080/10447318.2022.2074645

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023b). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kuhail, M. A., Alturki, N., Thomas, J., Alkhalifa, A. K., & Alshardan, A. (2024). Human-human vs human-AI therapy: An empirical study. *International Journal of Human–Computer Interaction*, 1–12. https://doi.org/10.1080/10447318.2024.2385001

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, *23*(5), 473–483. https://doi.org/10.1016/j.intcom.2011.06.005

Kumar, P. (2024). Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artificial Intelligence Review*, *57*(10), 260. https://doi.org/10.1007/s10462-024-10888-y

Kunselman, A. R. (2024). A brief overview of pilot studies and their sample size justification. *Fertility and Sterility*, *121*(6), 899–901. https://doi.org/10.1016/j.fertnstert.2024.01.040

Laban, G., & Araujo, T. (2020). Working together with conversational agents: The relationship of perceived cooperation with service performance evaluations. *Chatbot Research and Design Third International Workshop, Conversations 2019*, *11970 LNCS*, 215–228. https://doi.org/10.1007/978-3-030-39540-7_6

Lahav, O., Talis, V., Cinamon, R. G., & Rizzo, A. (2020). Virtual interactive consulting agent to support freshman students in transition to higher education. *Journal of Computing in Higher Education*, *32*(2), 330–364. https://doi.org/10.1007/s12528-019-09237-8

Lambert, L. S., & Newman, D. A. (2023). Construct development and validation in three practical steps: recommendations for reviewers, editors, and authors. *Organizational Research Methods*, *26*(4), 574–607. https://doi.org/10.1177/10944281221115374

Langevin, R., Lordon, R. J., Avrahami, T., Cowan, B. R., Hirsch, T., & Hsieh, G. (2021). Heuristic evaluation of conversational agents. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445312

Laumer, S., Maier, C., & Gubler, F. T. (2019). Chatbot acceptance in healthcare: explaining user adoption of conversational agents for disease diagnosis. *In Proceedings of the 27th European Conference on Information Systems (ECIS)*.

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior*, *114*, 106554. https://doi.org/10.1016/j.chb.2020.106554

Lawson-Guidigbe, C., Louveton, N., Amokrane-Ferka, K., Le Blanc, B., & André, J.-M. (2023). Embodying a virtual agent in a self-driving car: A survey-based study on user perceptions of trust, likeability, and anthropomorphism. *International Journal of Mobile Human Computer Interaction*, *15*(1), 1–18. https://doi.org/10.4018/ijmhci.330542

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Elsevier.

Lazarus, R. S., Kanner, A. D., & Folkman, S. (1980). Emotions: A cognitive–phenomenological analysis. In R. Plucthik & H. Kellerman (Eds.), *Theories of Emotion* (1st ed., pp. 189–217). Elsevier Inc. https://doi.org/10.1016/b978-0-12-558701-3.50014-4

Leavy, P. (2020). *The oxford handbook of qualitative research* (Second). Oxford University Press.

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50.

Lee, K. M., Jung, Y., Kim, J., & 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. *Int. J. Human-Computer Studies*, *65*(2006), 962–973.

Lee, S., Oh, J., & Moon, W. K. (2022). Adopting voice assistants in online shopping: Examining the role of social presence, performance risk, and machine heuristic. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2022.2089813

Lee, S. Y., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human Computer Studies*, *103*(January), 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I hear you, i feel you": Encouraging deep self-disclosure through a chatbot. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, (pp. 1–12). https://doi.org/10.1145/3313831.3376175

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information and Management*, *40*(3), 191–204. https://doi.org/10.1016/S0378-7206(01)00143-4

Lembcke, T.-B., Diederich, S., & Benedikt Brendel, A. (2020). Supporting design thinking through creative and inclusive education facilitation: The case of anthropomorphic conversational agents for persona building. *Proceedings of European Conference on Information Systems (ECIS).*

Levy, Y., & J. Ellis, T. (2006). a systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *9*, 181–212. https://doi.org/10.28945/479

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE - *When there's no time for the SUS*. CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2099–2102). https://doi.org/10.1145/2470654.2481287

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human-Computer Interaction 34(12),* 1148-1156.

Li, C., Zhu, W., Xing, W., & Guo, R. (2024). *Analyzing student attention and acceptance of conversational AI for math learning: insights from a randomized controlled* *trial.* Proceedings of the 14th Learning Analytics and Knowledge Conference, (pp. 836–842). https://doi.org/10.1145/3636555.3636895

Li, L., Lee, K. Y., Emokpae, E., & Yang, S. B. (2021). What makes you continuously use chatbot services? Evidence from chinese online travel agencies. *Electronic Markets*, *31*(3), 575–599. https://doi.org/10.1007/s12525-020-00454-z

Li, M., & Suh, A. (2021). *Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology*. Proceedings of the Annual Hawaii International Conference on System Sciences (pp. 4053–4062). https://doi.org/10.24251/hicss.2021.493

Li, S. (2024). Social media users' affective, attitudinal, and behavioral responses to virtual human emotions. *Telematics and Informatics*.

Li, X., Xie, S., Ye, Z., Ma, S., & Yu, G. (2022). Investigating patients' continuance intention toward conversational agents in outpatient departments: Cross-sectional field survey. *Journal of Medical Internet Research*, *24*(11), 1–11. https://doi.org/10.2196/40681

Liao, Q. V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M. P., Wang, D., Muller, M., Shami, N. S., & Geyer, W. (2018). *All work and no play?* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, (pp. 1–13). https://doi.org/10.1145/3173574.3173577

Lin, C.-C., Huang, A. Y. Q., & Yang, S. J. H. (2023). A review of ai-driven conversational chatbots implementation methodologies and challenges (1999–2022). *Sustainability*, *15*(5), 4012. https://doi.org/10.3390/su15054012

Lin, T.-C., Unni Krishnan, A., & Li, Z. (2023). Perception-motion coupling in active telepresence: human behavior and teleoperation interface design. *J. Hum.-Robot Interact.*, *12*(3). https://doi.org/10.1145/3571599

Ling, E. C., Tussyadiah, I., Tuomi, A., Stienmetz, J., & Ioannou, A. (2021). Factors influencing users' adoption and use of conversational agents: A systematic review. *Psychology & Marketing*, *38*(7), 1031–1051. https://doi.org/10.1002/mar.21491

Liu, C., & Hung, K. (2021). A multilevel study on preferences for self-service technology versus human staff: Insights from hotels in China. *International Journal of Hospitality Management*, *94*(April 2020), 102870. https://doi.org/10.1016/j.ijhm.2021.102870

Liu, K., & Tao, D. (2022). The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127*, 107026. https://doi.org/10.1016/j.chb.2021.107026

Liu, W., & Yao, M. (2023). Gender identity and influence in human-machine communication:A mixed-methods exploration. *Computers in Human Behavior*, *144*. https://doi.org/10.1016/j.chb.2023.107750

Liu, Y., & Martens, J.-B. (2024). Conversation-based hybrid UI for the repertory grid technique: A lab experiment into automation of qualitative surveys. *International Journal of Human-Computer Studies*, *184*, 103227. https://doi.org/10.1016/j.ijhcs.2024.103227

Liu, Y., Yan, W., Hu, B., Lin, Z., & Song, Y. (2024). Chatbots or humans? effects of agent identity and information sensitivity on users' privacy management and behavioral intentions: A comparative experimental study between China and the United States. *International Journal of Human–Computer Interaction*, *40*(19), 5632–5647. https://doi.org/10.1080/10447318.2023.2238974

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*(2015), 140–151. http://dx.doi.org/10.1016/j.tourman.2014.09.020

Lombard, M. (1995). Direct responses to people on screen: Television and personal space. *Communication Research*, *22*(3), 288–324. https://doi.org/10.1177/009365095022003002

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. In *Journal of Computer-Mediated Communication 3(2).* https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior*, *56*(2016), 225–237.

Luan, J., Yao, Z., Zhao, F., & Liu, H. (2016). Search product and experience product online reviews: An eye-tracking study on consumers' review search behavior. *Computers in Human Behavior*, *65*, 420–430. https://doi.org/10.1016/j.chb.2016.08.037

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

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

Luria, M., Reig, S., Tan, X. Z., Steinfeld, A., Forlizzi, J., & Zimmerman, J. (2019). *Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents.* Proceedings of the 2019 ACM Designing Interactive Systems Conference (pp. 633–644). https://doi.org/10.1145/3322276.3322340

Maar, D., Besson, E., & Kefi, H. (2023). Fostering positive customer attitudes and usage intentions for scheduling services via chatbots. *Journal of Servie Management*, *34*(2), 208–230.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in mis and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293–334. https://doi.org/10.2307/23044045

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Magyar, G., Balsa, J., Cláudio, A. P., Carmo, M. B., Neves, P., Alves, P., Félix, I. B., Pimenta, N., & Guerreiro, M. P. (2019). *Anthropomorphic virtual assistant to support self-care of type 2 diabetes in older people: A perspective on the role of artificial intelligence.* 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 1(Visigrapp), (pp. 323–331). https://doi.org/10.5220/0007572403230331

Mantei, M. M., & Teorey, T. J. (1989). Incorporating behavioral techniques into the systems development life cycle. *MIS Quarterly*, *13*(3), 257–274. https://doi.org/10.2307/249000

Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*. https://doi.org/10.1016/j.jbusres.2023.113838

Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). “Alexa, let’s talk about my productivity”: The impact of digital assistants on work productivity. *Journal of Business Research Journal*, *142*(2022), 571–584.

Market.us. (2025, November). *Enterprise conversational AI platform market* (Report ID: 165366). https://market.us/report/enterprise-conversational-ai-platform-market/

Martins, C., Oliveira, T., & Popovic, A. (2014). Understanding the internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. *International Journal of Information Management*, *34*(2014), 1–13.

Mathies, C., Chiew, T. M., & Kleinaltenkamp, M. (2016). The antecedents and consequences of humour for service: A review and directions for research. *Journal of Service Theory and Practice*, *26*(2), 137–162. https://doi.org/10.1108/jstp-09-2014-0187

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, *146*, 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Mayer, R. C., Davis, J. H., & Schoorman, D. F. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Mayring, P. (2014). Qualitative content analysis. Theoretical foundation, basic procedures and software solution. *SAGE Open 4(1).* https://doi.org/10.1177/2158244014522633

McKenzie, F. (Rick), Scerbo, M., Catanzaro, J., & Phillips, M. (2003). Nonverbal indicators of malicious intent: Affective components for interrogative virtual reality training. *International Journal of Human-Computer Studies*, *59*(1), 237–244. https://doi.org/10.1016/S1071-5819(03)00049-1

Mcknight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information systems research*, *13*(3), 334-359.

Mehrolia, S., Alagarsamy, S., Moorthy, V., & S., J. (2023). Will users continue using banking chatbots? The moderating role of perceived risk. *FIIB Business Review*, 1–19. https://doi.org/10.1177/23197145231169900

Mehrotra, S., Jorge, C. C., Jonker, C. M., & Tielman, M. L. (2024). Integrity-based explanations for fostering appropriate trust in ai agents. *ACM Transactions on Interactive Intelligent Systems*, *14*(1), 1–36. https://doi.org/10.1145/3610578

Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive Processing*, *13*(S1), 261–265. https://doi.org/10.1007/s10339-012-0499-z

Mikac, M. (2022). *The history behind eye tracking*. Eyelogic Retrieved September 15, 2025 from https://www.eyelogicsolutions.com/history-behind-eye-tracking/

Miles, M. B., Hubermen, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook*. SAGE Publications.

Mirabdolah, A., Alaeifard, M., & Marandi, A. (2023). User-centered design in HCl: enhancing usability and interaction in complex systems. *International Journal of Advanced Human Computer Interaction*, *1*(1).

Mishra, A., Shukla, A., & Sharma, S. K. (2021). Psychological determinants of users' adoption and word-of-mouth recommendations of smart voice assistants. *International Journal of Information Management*, *August*, 102413. https://doi.org/10.1016/j.ijinfomgt.2021.102413

Morana, S., Gnewuch, U., & Jung, D. (2020). The effect of anthropomorphism on investment decision-making with robo-advisor chatbots. *ECIS 2020*. https://www.researchgate.net/publication/341277570

Morashti, J., An, Y., & Jang, H. (2022). A systematic literature review of sustainable packaging in supply chain management. *Sustainability*, *14*(9), 4921. https://doi.org/10.3390/su14094921

Mori, M. (2012). The uncanny valley: The original essay by Masahiro Mori. *IEEE Robotics & Automation Magazine*, *12*(Figure 1), 1–6.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morina, N., Brinkman, W.-P., Hartanto, D., & Emmelkamp, P. M. G. (2014). Sense of presence and anxiety during virtual social interactions between a human and virtual humans. *PeerJ*, *2*, e337. https://doi.org/10.7717/peerj.337

Moriuchi, E. (2021). An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. *Psychology and Marketing*, *38*(1), 21–42. https://doi.org/10.1002/mar.21407

Mostafa, R. B., & Kasamani, T. (2022). Antecedents and consequences of chatbot initial trust. *European Journal of Marketing*, *56*(6), 1748–1771. https://doi.org/10.1108/EJM-02-2020-0084

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). That's so embarrassing! when not to design for social presence in human–chatbot interactions. *In Proceedings of the International Conference on Information Systems.*

Muniady, V., Ali, A. Z. M., & Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Perak, Malaysia, (2020). The effect of valence and arousal on virtual agent's designs in quiz based multimedia learning environment. *International Journal of Instruction*, *13*(4), 903–920. https://doi.org/10.29333/iji.2020.13455a

Munnukka, J., Talvitie-Lamberg, K., & Maity, D. (2022). Anthropomorphism and social presence in human–virtual service assistant interactions: The role of dialog length and attitude. *Computers in Human Behavior*, *135*(2022), 107343. https://doi.org/10.1016/j.chb.2022.107343

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

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

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can computer personalities be human personalities? *International Journal of Human Computer Studies*, *43*, 228–229. https://doi.org/10.1145/223355.223538

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors in Computing Systems*, 72–78.

Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education*, *192*, 104661. https://doi.org/10.1016/j.compedu.2022.104661

Nguyen, T.-T., Chiu, Y. T. H., & Le, H. D. (2021a). Determinants of continuance intention towards banks' chatbot services in vietnam: A necessity for sustainable development. *Sustainability (Switzerland)*, *13*(14), 1–24. https://doi.org/10.3390/su13147625

Nguyen, T.-T., Sim, K., To, A., Kuen, Y., O'donnell, R. R., Lim, S. T., Wang, W., & Nguyen, H. D. (2021b). Designing ai-based conversational agent for diabetes care designing ai-based conversational agent for diabetes care in a multilingual context. *PACIS 2021 Proceedings. (pp. 96)*. https://aisel.aisnet.org/pacis2021/96

Nicolescu, L., & Tudorache, M. T. (2022). Human-computer interaction in customer service: The experience with AI chatbots—A systematic literature review. *Electronics (Switzerland)*, *11*(10). https://doi.org/10.3390/electronics11101579

Nielsen, J. (1993). *Usability engineering*. Academic Press.

Norman, D. (1986). Cognitive engineering. In *User Centered System Design*. CRC Press.

Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, *9*(2), JCMC925. https://doi.org/10.1111/j.1083-6101.2004.tb00284.x

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., & Vogel, D. R. (1991). Information technology for negotiating groups: Generating options for mutual gain. *Management Science*, *37*(10), 1325–1346. https://doi.org/10.1287/mnsc.37.10.1325

Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical Diagnosis of Mental Disorders: A Handbook* (pp. 97–146). Springer US. https://doi.org/10.1007/978-1-4684-2490-4_4

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.

Oghuma, A. P., Chang, Y., Libaque-Saenz, C. F., Park, M.-C., & Rho, J. J. (2015). Benefit-confirmation model for post-adoption behavior of mobile instant messaging applications: A comparative analysis of KakaoTalk and Joyn in Korea. *Telecommunications Policy*, *39*(2015), 658–677.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17(4)*(November), 460–469. https://doi.org/10.1177/002224378001700405

Papas, P. (2018). *Digital customer care in the age of AI* (pp. 1–16). IBM Corporation. https://www.ibm.com/downloads/cas/GQDGPZJE#:~:text=AI

Park, N., Jang, K., Cho, S., & Choi, J. (2021). Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived humanlikeness. *Computers in Human Behavior*, *121*, 106795. https://doi.org/10.1016/j.chb.2021.106795

Pauletto, S., Balentine, B., Pidcock, C., Jones, K., Bottaci, L., Aretoulaki, M., Wells, J., Mundy, D. P., & Balentine, J. (2013). Exploring expressivity and emotion with artificial voice and speech technologies. *Logopedics Phoniatrics Vocology*, *38*(3), 115–125. https://doi.org/10.3109/14015439.2013.810303

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pei, H., Huang, X., & Ding, M. (2022). Image visualization: Dynamic and static images generate users' visual cognitive experience using eye-tracking technology. *Displays*, *73*, 102175. https://doi.org/10.1016/j.displa.2022.102175

Pelau, C., Dabija, D. C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, *122*(April), 106855. https://doi.org/10.1016/j.chb.2021.106855

Pernice, K., & Nielsen, J. (2009). *How to Conduct Eyetracking Studies* (No. 945397498). Nielsen Norman Group Retrieved September 25, 2025 from https://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/

Petter, Straub, & Rai. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623. https://doi.org/10.2307/25148814

Pietrantoni, N., Brendel, A. B., Greulich, R. S., & Hildebrandt, F. (2022). Follow me if you want to live—understanding the influence of human-like design on users' perception and intention to comply with covid-19 education chatbots. *Proceedings of the Forty-Third International Conference on Information Systems (ICIS 2022), Copenhagen, Denmark.*

Pietrantoni, N., Greulich, R. S., & Morana, S. (2023). Be a miracle—Designing conversational agents to influence users' intention regarding organ donation. *ICIS 2023 Proceedings. (pp. 5).* https://aisel.aisnet.org/icis2023/ishealthcare/ishealthcare/5

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *32*(10), 3199–6119.

Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. In S. Fincher, P. Markopoulos, D. Moore, & R. Ruddle (Eds.), *People and Computers XVIII — Design for Life* (pp. 363–378). Springer. https://doi.org/10.1007/1-84628-062-1_23

Poser, M., & Bittner, E. (2023). May the guide be with you: CA-facilitated information elicitation to prevent service failure. *ECIS 2023*. https://www.researchgate.net/publication/370745703

Poser, M., Küstermann, G. C., Tavanapour, N., & Bittner, E. A. C. (2022). Design and evaluation of a conversational agent for facilitating idea generation in organizational innovation processes. *Information Systems Frontiers*, *24*(3), 771–796. https://doi.org/10.1007/s10796-022-10265-6

Prakash, A. V., Joshi, A., Nim, S., & Das, S. (2023). Determinants and consequences of trust in AI-based customer service chatbots. *Service Industries Journal*, *43*(9–10), 642–675. https://doi.org/10.1080/02642069.2023.2166493

Punde, P. A., Jadhav, M. E., & Manza, R. R. (2017). *A study of eye tracking technology and its applications*. 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), (pp. 86–90). https://doi.org/10.1109/ICISIM.2017.8122153

Qiu, L., & Benbasat, I. (2008). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human Computer Studies*, *68*(10), 669–688. https://doi.org/10.1016/j.ijhcs.2010.05.005

Radziwill, N., & Benton, M. (2017). Evaluating quality of chatbots and intelligent conversational agents. *CEUR Workshop Proceedings*, *1982*, (pp. 40–49).

Rajamaran, V. (2023). From ELIZA to ChatGPT history of human-computer conversation. *Resonance*, *28*, 889–905.

Rapp, A., Boldi, A., Curti, L., Perrucci, A., & Simeoni, R. (2024). How do people ascribe humanness to chatbots? An analysis of real-world human-agent interactions and a theoretical model of humanness. *International Journal of Human–Computer Interaction*, *40*(19), 6027–6050. https://doi.org/10.1080/10447318.2023.2247596

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human Computer Studies*, *151*. https://doi.org/10.1016/j.ijhcs.2021.102630

Rau, A., Rau, S., Zöller, D., Fink, A., Tran, H., Wilpert, C., Nattenmüller, J., Neubauer, J., Bamberg, F., Reisert, M., & Russe, M. F. (2023). A context-based chatbot surpasses radiologists and generic chatgpt in following the ACR appropriateness guidelines. *Radiology*, *308*(1), e230970. https://doi.org/10.1148/radiol.230970

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8). https://doi.org/10.1080/17470210902816461

Razavi, S. Z., Schubert, L. K., Van Orden, K., Ali, M. R., Kane, B., & Hoque, E. (2022). Discourse behavior of older adults interacting with a dialogue agent competent in multiple topics. *ACM Transactions on Interactive Intelligent Systems*, *12*(2), 1–21. https://doi.org/10.1145/3484510

Razeghi, R. (2010). *Usability of eye tracking as a user research technique in geo-information processing and dissemination*. University of Twente.

Reevess, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge*, *10*, 19–36.

Reichheld, F. (2011). *The ultimate question 2.0 (revised and expanded edition): How net promoter companies thrive in a customer-driven world*. Harvard Business Review Press.

Reimann, M. M., Kunneman, F. A., Oertel, C., & Hindriks, K. V. (2025). *Transparent conversational agents.* 7th Conference on Conversational User Interfaces, CUI 2025 (pp. 1–12). https://doi.org/10.1145/3719160.3736629

Revelle, W. (2017). *psych: Procedures for personality and psychological research*. Northwestern University. https://CRAN.R-project.org/package=psych

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

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, *126*. https://doi.org/10.1016/j.chb.2021.107034

Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45091-8

Rietz, T., Benke, I., & Maedche, A. (2019). the impact of anthropomorphic and functional chatbot design features in enterprise collaboration systems on user acceptance. *14th International Conference on Wirtschaftsinformatik*.

Riquel, J., Brendel, A. B., Hildebrandt, F., Greve, M., & Dennis, A. R. (2021a). "F*** you!" – An investigation of humanness, frustration, and aggression in conversational agent communication. *Proceedings of the 42nd International Conference on Information Systems*.

Riquel, J., Brendel, A. B., Lichtenberg, S., & Diederich, S. (2021b). Do you feel a connection? How the human-like design of conversational agents influence donation behavior. *International Conference on Wirtschaftsinformatik*.

Robb, A., White, C., Cordar, A., Wendling, A., Lampotang, S., & Lok, B. (2014). A qualitative evaluation of behavior during conflict with an authoritative virtual human. *Intelligent Virtual Agents*. https://doi.org/10.1007/978-3-319-09767-1

Rogers, E. M. (1995). Diffusion of innovations: modifications of a model for telecommunications. *Die Diffusion von Innovationen in Der Telekommunikation*, *17*, 25–38.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human - computer interaction*. John Wiley & Sons.

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1), 364–389. https://doi.org/10.37074/jalt.2023.6.1.23

Rudowsky, I. (2004). Intelligent agents. *Communications of the Association for Information Systems*, *14*. https://doi.org/10.17705/1cais.01414

Russell, J.-E., Smith, A. M., George, S., Pratt, J., Fodale, B., Monk, C., & Brummett, A. (2025). *Unlocking insights: Investigating student AI tutor interactions in a large introductory STEM course*. Proceedings of the 15th International Learning Analytics and Knowledge Conference (pp. 451–461). https://doi.org/10.1145/3706468.3706524

Saadé, R. G., & Otrakji, C. A. (2007). First impressions last a lifetime: Effect of interface type on disorientation and cognitive load. *Computers in Human Behavior*, *23*(1), 525–535. https://doi.org/10.1016/j.chb.2004.10.035

Saldana, J. (2014). *Thinking qualitatively: Methods of mind*. SAGE Publications.

Santhoshikka, R., C R, L., & C, H. (2021). Eye tracking and its applications. *IARJSET*, *8*(8). https://doi.org/10.17148/IARJSET.2021.8824

What is a digital assistant? (2024, October 29). SAP Retrieved June 07, 2025 from https://www.sap.com/resources/what-is-a-digital-assistant#:~:text=A%20digital%20assistant%2C%20as%20the,content%2C%20and%20so%20much%20more

Sarikaya, R. (2017). The technology behind personal digital assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, 67–81. https://doi.org/10.1109/MSP.2016.2617341

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. https://doi.org/10.1093/scan/nsr025

Schlesener, E. A., Shivakumar, V., Breeze, D., Rennison, B., Soehmelioglu, B., & Babu, S. V. (2025). *'Age isn't just a number': Effects of virtual human age and gender on persuasion, social presence and influence in interpersonal social encounters in* VR. 2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 82–92). https://doi.org/10.1109/vr59515.2025.00033

Schmettow, M. (2021). *New statistics for design researchers A Bayesian workflow in Tidy R*. Springer. http://www.springer.com/series/6033

Schoen, J. P. (1976). *Silents to sound: A history of the movies*. Four Winds Press.

Schoormann, T., Strobel, G., Möller, F., Petrik, D., & Zschech, P. (2023). Artificial intelligence for sustainability. A systematic review of information systems literature. *Communications of the Association for Information Systems*, *52*(1). https://doi.org/10.17705/1CAIS.05209

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage. https://www.torrossa.com/en/resources/an/4913035

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. *MIS Quarterly Executive*, *20*(1), 1–15. https://doi.org/10.17705/2msqe.00039

Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, *97*, 250–259. https://doi.org/10.1016/j.chb.2019.03.033

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875–900. https://doi.org/10.1080/07421222.2020.1790204

Sedrakyan, G., Borsci, S., Machado, M., Rogetzer, P., & Mes, M. (2024a, October 12-14). *Design implications for integrating AI chatbot technology with learning management systems: A study-based analysis on perceived benefits and challenges in higher education*. Proceedings of the International Conference on Artificial Intelligence and Teacher Education, Beijing, China https://doi.org/10.1145/3702386.3702405

Sedrakyan, G., Borsci, S., Van Den Berg, S. M., Van Hillegersberg, J., & Veldkamp, B. P. (2024b). Design implications for next generation chatbots with education 5.0. In J.-C. Hong (Ed.), *New Technology in Education and Training* (pp. 1–12). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3883-0_1

Seeger, A. M., & Heinzl, A. (2017). Human versus machine: Contingency factors of anthropomorphism as a trust-inducing design strategy for conversational agents. Davis, F., Riedl, R., vom Brocke, J., Léger, PM., Randolph, A. (eds), *Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation,25.* Springer, Cham. https://doi.org/10.1007/978-3-319-67431-5_15

Seeger, A. M., & Heinzl, A. (2021). *Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service*? ECIS 2021 Research Papers (pp. 6–14). https://aisel.aisnet.org/ecis2021_rp

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2017). *When do we need a human? anthropomorphic design and trustworthiness of conversational agents*. SIGHCI

2017 Proceeding. http://aisel.aisnet.org/sighci2017http://aisel.aisnet.org/sighci2017/15

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2018). Designing anthropomorphic conversational agents: Development and empirical evaluation of a design framework. *International Conference on Information Systems 2018*, *Oracle 2016*, (pp. 1–17).

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, *22*(4). https://doi.org/10.17705/1jais.00685

Sepideh Ebrahimi, M. G., & Benbasat, I. (2022). The impact of trust and recommendation quality on adopting interactive and non-interactive recommendation agents: A meta-analysis. *Journal of Management Information Systems*, *39*(3), 733–764. https://doi.org/10.1080/07421222.2022.2096549

Sestino, A., & D'Angelo, A. (2023). My doctor is an avatar! The effect of anthropomorphism and emotional receptivity on individuals' intention to use digital-based healthcare services. *Technological Forecasting and Social Change*, *191*, 122505. https://doi.org/10.1016/j.techfore.2023.122505

Seymour, M., Riemer, K., & Kay, J. (2017). *Interactive realistic digital avatars—Revisiting the uncanny valley*. Proceedings of the Annual Hawaii International Conference on System Sciences (pp. 547–556). https://doi.org/10.24251/hicss.2017.067

Seymour, M., Riemer, K., & Kay, J. (2018). Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence. *Journal of the Association for Information Systems*, *19*(10), 953–981. https://doi.org/10.17705/1jais.00515

Sharafi, Z., Sharif, B., Guéhéneuc, Y.-G., Begel, A., Bednarik, R., & Crosby, M. (2020). A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering*, *25*(5), 3128–3174. https://doi.org/10.1007/s10664-020-09829-4

Shen, J. (2012). Social comparison, social presence, and enjoyment in the acceptance of social shopping websites. *Journal of Electronic Commerce Research*, *13*(3), 198–213.

Sheng, Y., & Wikle, C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, *67*(6), 899–919. https://doi.org/10.1177/0013164406296977

Shevat, A. (2017). *Designing bots: Creating conversational experiences*. O'Reilly.

Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human Computer Studies*, *62*(2), 267–279. https://doi.org/10.1016/j.ijhcs.2004.11.003

Shneiderman, B. (2000). Designing trust into online experiences. *Communications of the ACM*, *43*(12), 57–59. https://doi.org/10.1145/355112.355124

Siegel, S., & Castellan Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences, 2nd ed* (pp. xxiii, 399). Mcgraw-Hill Book Company.

Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N., & Gombolay, M. (2023). Explainable artificial intelligence: Evaluating the objective and subjective impacts of xai on human-agent interaction. *International Journal of Human–Computer Interaction*, *39*(7), 1390–1404. https://doi.org/10.1080/10447318.2022.2101698

Silva, E. S., & Bonetti, F. (2021). Digital humans in fashion: Will consumers interact? *Journal of Retailing and Consumer Services*, *60*(October 2020), 102430. https://doi.org/10.1016/j.jretconser.2020.102430

Silva, G. R. S., & Canedo, E. D. (2024). Towards user-centric guidelines for chatbot conversational design. *International Journal of Human–Computer Interaction*, *40*(2), 98–120. https://doi.org/10.1080/10447318.2022.2118244

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press.

Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology* (pp. 587–607). The Guilford Press. https://doi.org/10.4324/9781315663531-11

Sin, J., & Munteanu, C. (2019, May 2). A preliminary investigation of the role of anthropomorphism in designing telehealth bots for older adults. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312941

Sin, J., & Munteanu, C. (2020). An empirically grounded sociotechnical perspective on designing virtual agents for older adults. *Human-Computer Interaction*, *35*(5–6), 481–510. https://doi.org/10.1080/07370024.2020.1731690

Sirkin, D., Venolia, G., Tang, J., Robertson, G., Kim, T., Inkpen, K., Sedlins, M., Lee, B., & Sinclair, M. (2011). Motion and attention in a kinetic videoconferencing proxy. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6946 LNCS* (PART 1), 162–180. https://doi.org/10.1007/978-3-642-23774-4_16

Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday*. https://doi.org/10.5210/fm.v29i1.13541

Slomianka, V., May, T., & Dau, T. (2025). Adaptions in eye-movement behavior during face-to-face communication in noise. *Frontiers in Psychology*, *16*, 1584937. https://doi.org/10.3389/fpsyg.2025.1584937

Song, Y., & Han, J. (2009). *Is enjoyment important? An empirical research on the impact of perceive enjoyment on adoption of new technology*. 2009 International

Conference on Information Management, Innovation Management and Industrial Engineering (pp. 511–514). https://doi.org/10.1109/ICIII.2009.582

Spadoni, E., Carulli, M., Mengoni, M., Luciani, M., & Bordegeni, M. (Eds.). (2023). Empowering virtual humans' emotional expression in the metaverse. In *Universal Access in Human-Computer Interaction*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35897-5

Stanley, J., ten Brink, R., Valiton, A., Bostic, T., & Scollan, R. (2022). Chatbot accessibility guidance: A review and way forward. (In X.-S. Yang, S. Sherratt, N. Dey, & A. Joshi), *Proceedings of Sixth International Congress on Information and Communication Technology* (pp. 919–942). Springer. https://doi.org/10.1007/978-981-16-1781-2_80

Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital age: innovative alternatives to student samples. *MIS Quarterly*, *38*(2), 355–378.

Stein, J.-P., Appel, M., Jost, A., & Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *International Journal of Human-Computer Studies*, *142*, 102463. https://doi.org/10.1016/j.ijhcs.2020.102463

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory, 2nd ed* (pp. xiii, 312). Sage Publications, Inc.

Strout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*(2), 293–325. https://doi.org/10.1007/BF02295289

Stryker, C. (2025). *Types of AI agents*. IBM Retrieved October 20, 2025 from https://www.ibm.com/think/topics/ai-agent-types#:~:text=For%20example%2C%20in%20reinforcement%20learning,predict%20and%20optimize%20content%20recommendations

Sugisaki, K., & Bleiker, A. (2020). *Usability guidelines and evaluation criteria for conversational user interfaces: A heuristic and linguistic approach*. ACM International Conference Proceeding Series (pp. 309–319). https://doi.org/10.1145/3404983.3405505

Sundar, S. S., Oh, J., Bellur, S., Jia, H., & Kim, H.-S. (2012). *Interactivity as self-expression: A field experiment with customization and blogging.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 395–404). https://doi.org/10.1145/2207676.2207731

Sundstedt, V., & Garro, V. (2022). A systematic review of visualization techniques and analysis tools for eye-tracking in 3d environments. *Frontiers in Neuroergonomics*, *3*. https://doi.org/10.3389/fnrgo.2022.910019

Sunil, G. (2025). *The psychology behind good UI: How cognitive load affects design*. Aufait UX Retrieved June 20, 2025 from https://www.aufaitux.com/blog/cognitive-load-theory-ui-design/

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (Sixth edition, Pearson new international edition). Pearson.

Talwar, S., Dhir, A., Khalil, A., Mohan, G., & Islam, A. K. M. N. (2020). Point of adoption and beyond. Initial trust and mobile-payment continuation intention. *Journal of Retailing and Consumer Services*, *5*(2020), 102086. https://doi.org/10.1016/j.jretconser.2020.102086

Tan, S. M., & Liew, T. W. (2020). Designing embodied virtual agents as product specialists in a multi-product category e-commerce: The roles of source credibility and social presence. *International Journal of Human-Computer Interaction*, *36*(12), 1136–1149. https://doi.org/10.1080/10447318.2020.1722399

Taylor, M. P., Girard, S., Jacobs, K., Buvat, J., Subrahmanyam, K., Puttur, R., Shah, H., & B., A. (2020). Smart Talk: How organizations and consumers are embracing voice and chat assistants. In *Capgemini* (pp. 1–44).

Terblanche, N., & Kidd, M. (2022). Adoption factors and moderating effects of age and gender that influence the intention to use a non-directive reflective coaching chatbot. *SAGE Open*, *12*(2). https://doi.org/doi.org/10.1177/2158244022109613

Terblanche, W., Lubbe, I., Papageorgiou, E., & van der Merwe, N. (2023). Acceptance of e-learning applications by accounting students in an online learning environment at residential universities. *South African Journal of Accounting Research*, *37*(1), 35–61. https://doi.org/10.1080/10291954.2022.2101328

Ternyak, D. (2023, January 29). *53 Chatbot statistics for 2022: Usage, demographics, trends*. ServiceBell. https://www.servicebell.com/post/chatbot-statistics

Thaler, M., Schlogl, S., & Groth, A. (2020). *Agent vs. avatar: comparing embodied Conversational agents concerning characteristics of the uncanny valley*. Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (pp. 1–6). https://doi.org/10.1109/ICHMS49158.2020.9209539

Tinwell, A., & Sloan, R. J. S. (2014). Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior*, *36*(2014), 286–296.

Toader, R., Mara, M., Toader, C., & Toader, D. (2020). The effect of social presence and chatbot errors on trust. *Sustainability*, *12*(1), 256. https://doi.org/10.3390/su12010256

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2021). Do we trust in ai? role of anthropomorphism and intelligence. *Journal of Computer Information Systems*, *61*(5), 481–491. https://doi.org/10.1080/08874417.2020.1788473

Tullis, T., & Albert, B. (2013). *Behavioral and physiological metrics. in measuring the user experience.* Elsevier. https://doi.org/10.1016/B978-0-12-415781-1.00007-8

Tussyadiah, I., & Miller, G. (2019). Nudged by a robot: Responses to agency and feedback. *Annals of Tourism Research*, *78*, 102752. https://doi.org/10.1016/j.annals.2019.102752

Urakami, J., & Seaborn, K. (2023). Nonverbal cues in human–robot interaction: A communication studies perspective. *ACM Transactions on Human-Robot Interaction*, *12*(2), 1–21. https://doi.org/10.1145/3570169

Vaishnavi, V. K., & Kuechler, W. (2007). Design science research methods and patterns: Innovating information and communication technology. In *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*. CRC Press Taylor & Francis Group. https://doi.org/10.1201/9781420059335

van Bussel, M. J. P., Odekerken–Schröder, G. J., Ou, C., Swart, R. R., & Jacobs, M. J. G. (2022). Analyzing the determinants to accept a virtual assistant and use cases among cancer patients: A mixed methods study. *BMC Health Services Research*, *22*(890), 1–23. https://doi.org/10.1186/s12913-022-08189-7

van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, *28*(4), 695–704.

Van Lange, P. A. M., Higgins, E. T., & Kruglanski, A. W. (2011). *Handbook of theories of social psychology*. Sage.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342–365. https://doi.org/10.1287/%20isre.11.4.342.11872

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

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

Venkatesh, V., Thong, J. Y. L., Chan, F. K. Y., Hu, P. J., & Brown, S. A. (2011). Extending the two-stage information systems continuance model: Incorparating UTAUT predictors and the role of context. *Information Systems Journal*, *21*, 527–555. https://doi.org/10.1111/j.1365-2575.2011.00373.x

Vidarshika, W., Dayapathirana, n., & ranasinghe, a. (2025). *Understanding AI chatbot adoption in education: The role of perceived usefulness, ease of use, and* *anthropomorphic tendencies*. 2025 International Research Conference on Smart Computing and Systems Engineering (SCSE) (pp. 1–6). https://doi.org/10.1109/SCSE65633.2025.11031020

Von Der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

Wagner, K., Nimmermann, F., & Schramm-Klein, H. (2019). *Is it human? The role of anthropomorphism as a driver for the successful acceptance of digital voice assistants*. Proceedings of the Annual Hawaii International Conference on System Sciences (pp. 1386–1395). https://doi.org/10.24251/hicss.2019.169

Walls, J. G., Widermeyer, G. R., & Sawy, O. A. E. (2004). Assessing information system design theory in perspective: How useful was our 1992 initial rendition? *Journal of Infromation Technology Theory and Application*, *6*(2), 43–58.

Wambsganss, T. (2021). Designing adaptive argumentation learning systems based on artificial intelligence. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411763.3443422

Wang, I., & Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction*, *37*(17), 1648–1673. https://doi.org/10.1080/10447318.2021.1898851

Wang, N., & Gratch, J. (2010). Don't just stare at me! *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1241–1250. https://doi.org/10.1145/1753326.1753513

Wang, Q., Jing, S., Camacho, I., Joyner, D., & Goel, A. (2020). *Jill Watson SA: Design and evaluation of a virtual agent to build communities among online learners*. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–8). https://doi.org/10.1145/3334480.3382878

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10. https://doi.org/10.1016/j.dss.2014.02.007

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems ISSN:*, *23*(4), 217–246. https://doi.org/10.2753/MIS0742-1222230410

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems*, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, *86*, 48–60. https://doi.org/10.1016/j.dss.2016.03.007

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

Weber, R. (2006). Still desperately seeking the it artifact. In *Information systems: The state of the field*. J. Wiley & Sons.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 13–23.

Wei, W., Torres, E., & Hua, N. (2016). Improving consumer commitment through the integration of self-service technologies: A transcendent consumer experience perspective. *International Journal of Hospitality Management*, *59*(2016), 105–115.

Wessel, L., Sundermeier, J., Rothe, H., Hanke, S., Baiyere, A., Rappert, F., & Gersch, M. (2025). Designing as trading-off: A practice-based view on smart service systems. *European Journal of Information Systems*, *34*(2), 181–206. https://doi.org/10.1080/0960085X.2024.2308541

Wiemann, J. M. (1977). Explication and test of a model of communicative competence. *Human Communication Research*, *3*(3), 195–213. https://doi.org/10.1111/j.1468-2958.1977.tb00518.x

Wienrich, C., & Carolus, A. (2021). Development of an instrument to measure conceptualizations and competencies about conversational agents on the example of smart speakers. *Frontiers in Computer Science*, *3*. https://doi.org/10.3389/fcomp.2021.685277

Wieringa, R. J. (2014). Design science methodology: For information systems and software engineering. In *Design Science Methodology: For Information Systems and Software Engineering*. Springer. https://doi.org/10.1007/978-3-662-43839-8

Xi, N., & Hamari, J. (2019). Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management*, *46*, 210–221. https://doi.org/10.1016/j.ijinfomgt.2018.12.002

Xiao & Benbasat. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, *31*(1), 137. https://doi.org/10.2307/25148784

Xu, Y., Vigil, V., Bustamante, A. S., & Warschauer, M. (2022). "Elinor's talking to me!": Integrating conversational ai into children's narrative science programming. *CHI Conference on Human Factors in Computing Systems*, (pp. 1–16). https://doi.org/10.1145/3491102.3502050

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*, 104059. https://doi.org/10.1016/j.compedu.2020.104059

Yang, D., Hovy, D., Jurgens, D., & Plank, B. (2025). Socially aware language technologies: perspectives and practices. *Computational Linguistics*, *51*(2), 689–703. https://doi.org/10.1162/coli_a_00556

Yang, X., & Aurisicchio, M. (2021, May 6). *Designing conversational agents: A self-determination theory approach.* Proceedings Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411764.3445445

Yang, X., Aurisicchio, M., & Baxter, W. (2019). Understanding affective experiences with conversational agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*(pp.1–12). https://doi.org/10.1145/3290605.3300772

Yang, Y., & Kankanhalli, A. (2023). Humor touch: should conversational agents express humor when they fail? *PACIS 2023*. https://aisel.aisnet.org/pacis2023

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zehnder, E., Dinet, J., & Charpillet, F. (2021). Social virtual agents and loneliness: impact of virtual agent anthropomorphism on users' feedbacks. *advances in usability, user experience, wearable and assistive technology*. *Proceedings of the AHFE 2021*. https://doi.org/10.1007/978-3-030-80091-8

Zhang, A., & Patrick Rau, P.-L. (2023). Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents. *Computers in Human Behavior*, *138*, 107415. https://doi.org/10.1016/j.chb.2022.107415

Zhang, B., Rau, P.-L. P., & Salvendy, G. (2009). Design and evaluation of smart home user interface: Effects of age, tasks and intelligence level. *Behaviour & Information Technology*, *28*(3), 239–249. https://doi.org/10.1080/01449290701573978

Zhang, B., Zhu, Y., Deng, J., Zheng, W., Liu, Y., Wang, C., & Zeng, R. (2023). "I am here to assist your tourism": Predicting continuance intention to use AI-based chatbots for tourism. Does gender really matter? *International Journal of Human-Computer Interaction*, *39*(9), 1887–1903. https://doi.org/10.1080/10447318.2022.2124345

Zhang, P., & Li, N. (2005). The intellectual development of human-computer interaction research: A critical assessment of the MIS literature (1990-2002). *Journal of the Association for Information Systems*, *6*(11), 227–292.

Zhao, B. (2019). *Consumer behavior analysis and repeat buyer prediction for e-commerce*. Georg-August-Universität.

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019). Transparency in advice-giving systems: A framework and a research model for transparency provision. *IUI Workshops'19, March 20, 2019, Los Angeles, USA*

Zhou, J., Salvendy, G., Boot, W. R., Charness, N., Czaja, S., Gao, Q., Holzinger, A., Ntoa, S., Rau, P.-L. P., Rogers, W. A., Stephanidis, C., Wahl, H.-W., & Ziefle, M. (2025). Grand challenges of smart technology for older adults. *International Journal of Human–Computer Interaction*, (pp. 1–43). https://doi.org/10.1080/10447318.2025.2457003

Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems*, *9*(2–3), 1–36. https://doi.org/10.1145/3232077

Zhou, T. (2011). Understanding mobile internet continuance usage from the perspectives of UTAUT and flow. *Information Development*, *27*(1), 207–218. https://doi.org/10.1177/0266666911414596

Zhou, Z., Li, Z., Zhang, Y., & Sun, L. (2022). Transparent-AI blueprint: Developing a conceptual tool to support the design of transparent AI agents. *International Journal of Human–Computer Interaction*, *38*(18–20), 1846–1873. https://doi.org/10.1080/10447318.2022.2093773

Zhu, Y., Wang, R., & Pu, C. (2022). "I am chatbot, your virtual mental health adviser." What drives citizens' satisfaction and continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China. *DIGITAL HEALTH*, *8*, 205520762210900. https://doi.org/10.1177/20552076221090031

Zierau, N., Bruhin, O., Hausch, M., & Söllner, M. (2020). Towards developing trust-supporting design features for AI-based chatbots in customer service. *Proceedings in the Forty-First International Conference on Information Systems, India 2020 (ICIS)*.

Zimmerman, B. J. (2000). *Attaining self-regulation.* In *Handbook of Self-Regulation* (pp. 13–39). Elsevier. https://doi.org/10.1016/b978-012109890-2/50031-7

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

APPENDICES

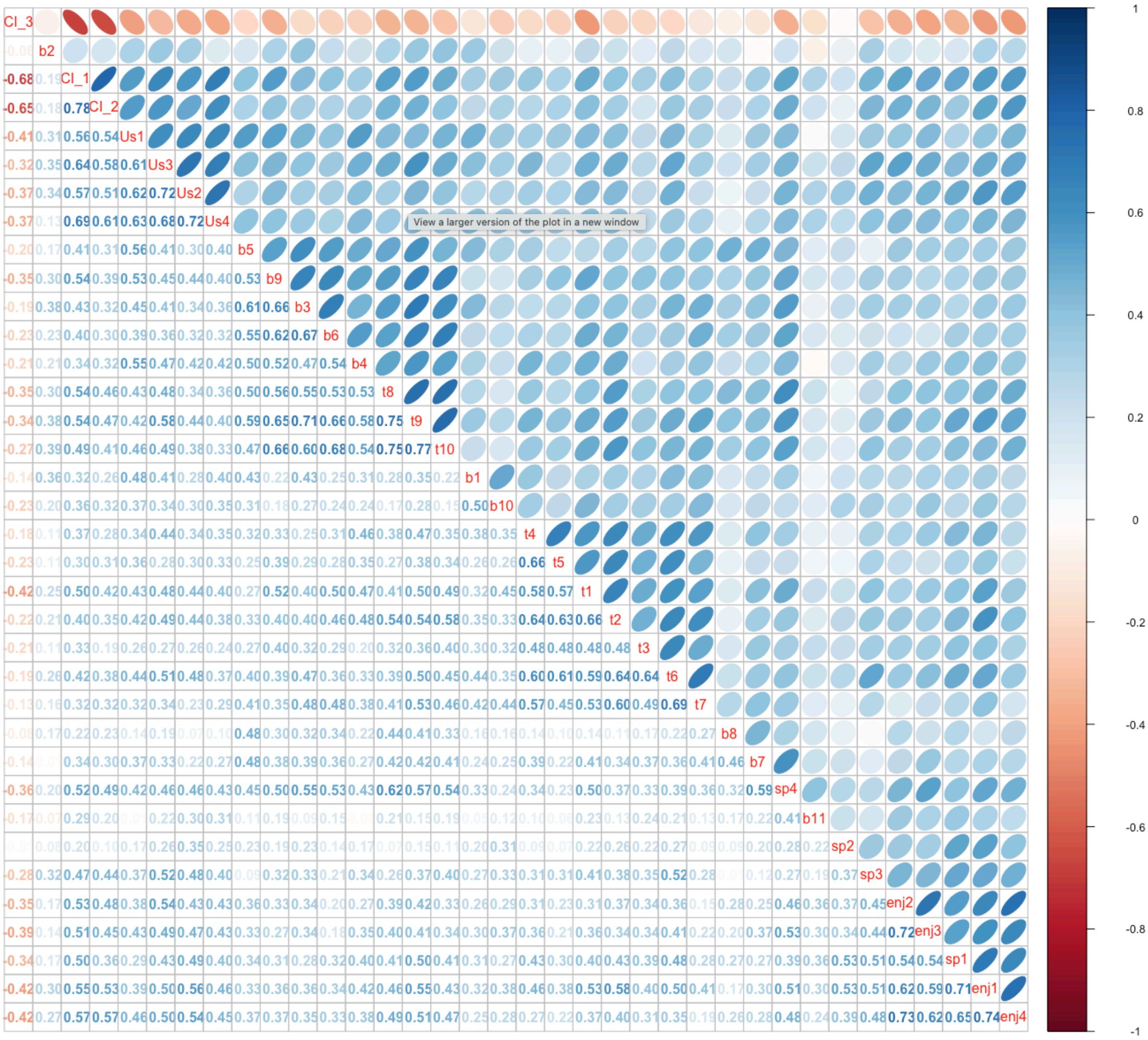
Appendix 1. *Conversational Agent Design Codebook*

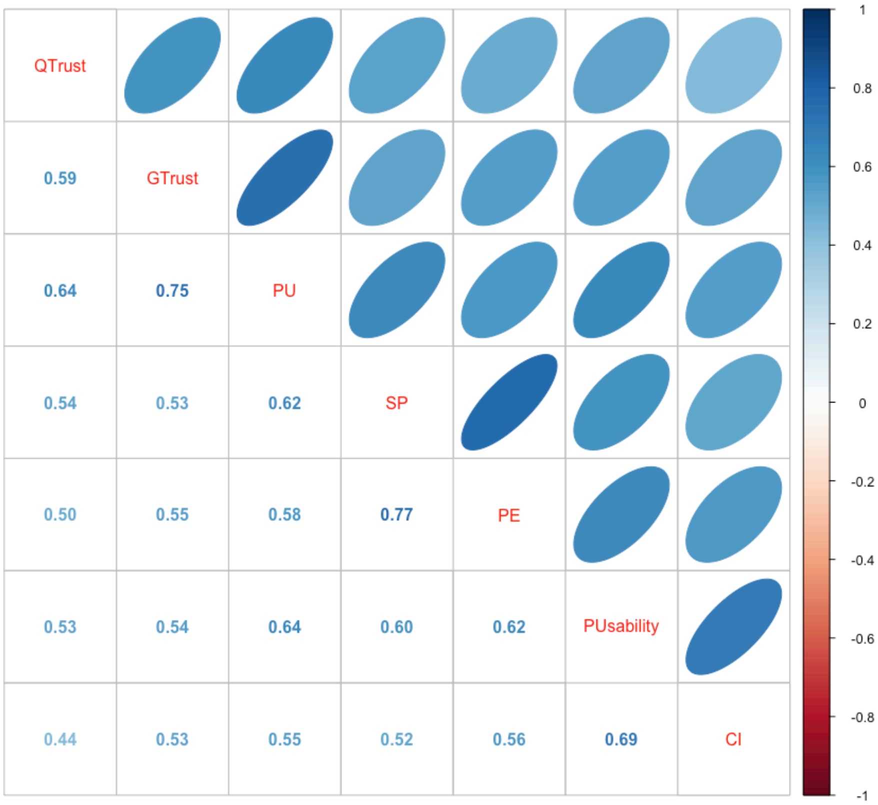
Design Dimension	Desing SubDimension: Definition/Identification	Motivation
Agent Competency	**Responsiveness:** refers to helping users along in a chat when they need it, and learning from past conversations, and improving how the agent talks and responds (productivity). Responsive design makes conversations smoother, quicker, and more helpful (Schuetzler et al., 2018).	To enable fluid and adaptive interactions by providing real-time, context-aware support, especially during task execution
	E.g.: Guidance in conversation with button options, providing relevance and wide variety of responses	
	Explanation Facilities: provides users with clear, understandable, and contextual explanations about the agent's responses, actions, or decisions.	To provide justifications for agent behavior, including clarifying actions, decisions, and recommendations, thereby supporting user comprehension as well as implied openness
	E.g.: Personalized daily health recommendation chatbot: "Hello John, would you please set your exercise goals by choosing among 1) reducing body fat, 2) improving balance, or 3) building muscle?" After accessing the participant schedule data: Good morning, John. Based on your schedule from 10 am to 3 pm today, you will be tired. I recommend you have considered chicken soup for lunch, a high-protein meal." (Park & Youn-kyung, 2019)	
	Handling system failures (Error Recovery Strategies): Happens when the response is unknown. Agent can apologize, changed the topic of conversation, or route the human representative.	To reduce frustration and maintain conversational flow when agents cannot fulfill requests.
	E.g.: "User: I heard that Chewbacca meets Donatella in the next movie. Instead of not saying "Sorry, I have no information regarding that," Changed the topic of conversation. Starwars-bot: That is interesting. Did you know you will be able to see Leia again in this movie?" (Shevat, 2017)	
Agent Characteristics	**Transparency (Self-Identity):** Agent is introducing itself and transparently revealing its capabilities and limitations.	To set accurate user expectations about the agent's capabilities and limitations.
	E.g.: "Hello, I'm BankerBot, your virtual banking assistant. I can provide account balances, recent transaction details, and help with common service inquiries. For anything I can't do, I'll connect you with a human colleague" (Diederich et al., 2020)	
	Expressiveness: Agent conveys information and emotions in a manner.	To enrich the agent's communicative behavior with affective and stylistic nuances that resemble human ability to communicate
	E.g.: "Sorry to hear that you are experiencing this condition. There is nothing to worry about as this very common", Changing voice-tone during the utterance	
	Embodiment: ***Re-embodiment***: single agent to transition between different forms or bodies, enabling it to accompany a user seamlessly across various tasks or services.	To support continuity across multi-platform or multi-agent services
	E.g.: Amazon Alexa, Apple Siri	

	Co-embodiment: an agent collaborates or joins forces with another agent already present within a particular system or device	To create layered and collaborative user experience
	E.g.: Different agents converse each other. (Driving agent and home agent in a single agent)	
	Physical appearance: Physically presence agent	To enhance engagement in co-located environments
	E.g.: Service robots	
	One-for-one: Single body representation	To provide presence representation
	E.g.: Chatbot with an avatar	
Anthropomorphic Design Dimension	**Verbal Cues:** Provides communication using words.	To simulate natural human dialogue and build rapport
	E.g.: Hi, how are you? I am your customer service agent from … I am sorry to hear that.	
	Non-Verbal Cues: Provides communication without use of words. They encompass signals and behaviors that convey meaning and emotion.	To complement verbal communication with affective and behavioral signals
	E.g.: Blinking dots, or “is typing”, Emoticons (😊, 🎅), Response Delay, Eye-Gaze, Look Away, Facial Expression	
	Identity (Personality) Cues: reveal key aspects of an individual’s identity.	To enable personalization and social categorization of agents
	E.g.: Demographic information (gender, age, name, race), humanlike visual representation (image)	

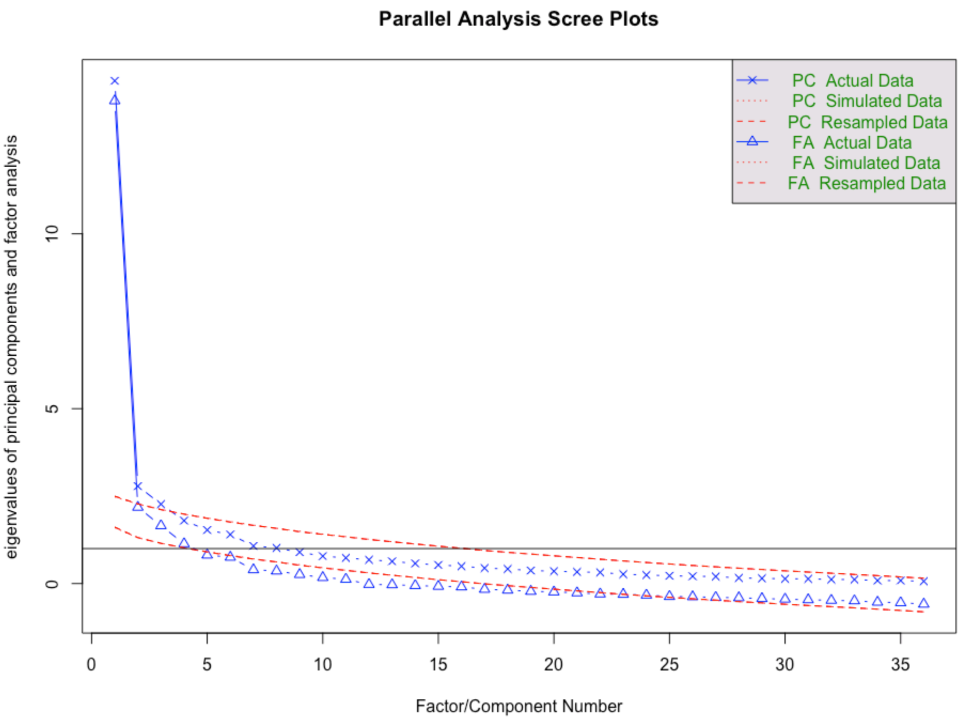
Appendix 2. *Quantitative Analysis*

Appendix 2.1. *Inter-Item Correlation and Factor Correlations*

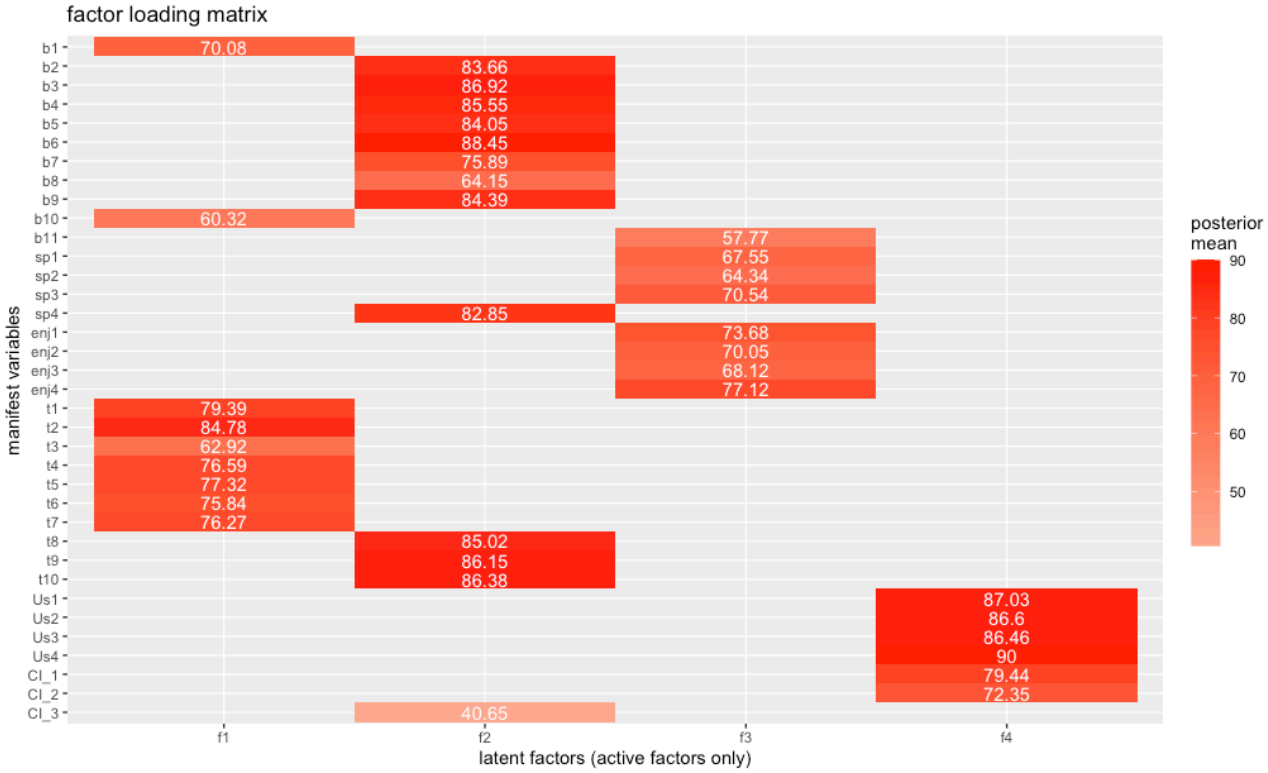



QTrust
0.59
GTrust
0.64
0.75
PU
0.54
0.53
0.62
SP
0.50
0.55
0.58
0.77
PE
0.53
0.54
0.64
0.60
0.62
PUsability
0.44
0.53
0.55
0.52
0.56
0.69
CI
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Appendix 2.2. *Parallel Analysis to Determine Number of Factor (N=87)*



Appendix 2.3. *Posterior Indicator Probabilities of Being a Nonzero Matrix (Factor=4)*

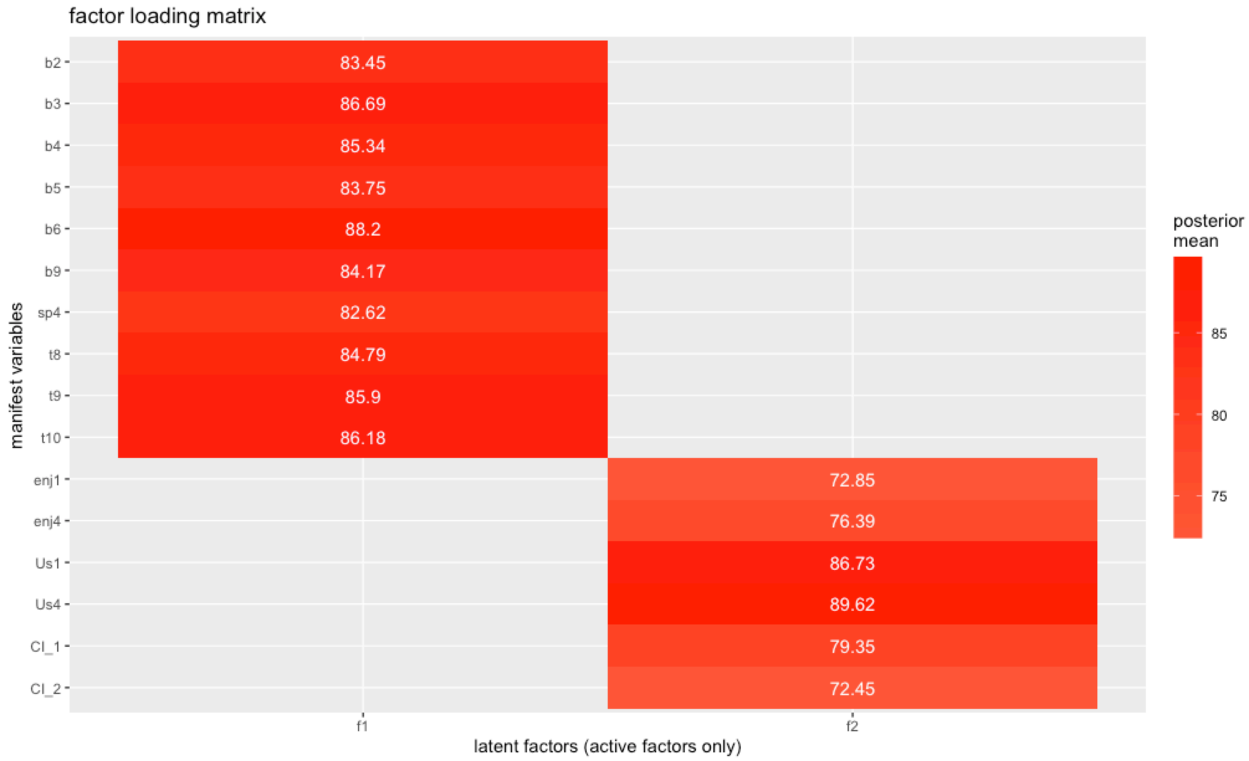


Appendix 2.4. *Retained Item Evaluation*

Construct	Item	Prob.	Mean Loading	Error Variance (σ²)	Decision	Rationale
BUS-11	b1	1.000	69.9	353.6 (HIGH)	DROPPED	Statistically strong prob, but high error variance; could be added for content coverage.
	b2	0.901	83.5	322.2	KEPT	Strong loading & prob.
	b3	0.911	86.7	93.7 (LOW)	KEPT	Very strong, low error.
	b4	0.910	85.4	160.8	KEPT	Strong loading.
	b5	0.911	83.8	219.7	KEPT	Solid loading.
	b6	0.911	88.3	110.3	KEPT	Best PU item.
	b7	0.886	75.7	372.7 (VERY HIGH)	DROPPED	High error variance.
	b8	0.910	63.9	472.7 (VERY HIGH)	DROPPED	Weak loading + high variance.
	b9	0.911	84.2	97.7 (LOW)	KEPT	Very good item.
	b10	0.570	60.0	464.1 (VERY HIGH)	DROPPED	Low prob + high error.
	b11	0.618	57.6	505.3 (VERY HIGH)	DROPPED	Weak item.
Social Presence (SP)	sp1	0.592	67.4	211.7	DROPPED	Below cutoff.
	sp2	0.592	64.2	386.2	DROPPED	Weak + high error.
	sp3	0.363	70.3	248.4	DROPPED	Very weak prob.
	sp4	0.910	82.7	199.7	KEPT	Only strong SP item.
perceived Enjoyment (PE)	enj1	0.592	73.5	99.1	KEPT	Moderate but decent; could retain for coverage.
	enj2	0.592	69.9	128.7	DROPPED	Weak compared to enj1/4.
	enj3	0.592	68.0	224.4 (HIGH)	DROPPED	Too noisy.
	enj4	0.592	76.9	107.3	KEPT	Strongest enjoyment item.
Trust	t1	0.457	79.2	150.7	DROPPED	Prob below .60 cutoff.
	t2	0.457	84.6	92.7	DROPPED	Strong loading but prob < .60.
	t3	0.457	62.7	292.0	DROPPED	Too weak.
	t4	0.457	76.4	134.5	DROPPED	Prob < .60.
	t5	0.457	77.2	159.6	DROPPED	Prob < .60.
	t6	0.457	75.7	157.0	DROPPED	Prob < .60.
	t7	0.457	76.1	187.5	DROPPED	Prob < .60.
	t8	0.911	84.8	117.1	KEPT	Very strong.
	t9	0.911	86.0	82.7 (LOW)	KEPT	Best trust item.
	t10	0.911	86.2	71.3 (VERY LOW)	KEPT	Excellent trust item.
Perceived usefulness (pu)	Us1	0.460	86.9	119.8	KEPT	Added for coverage (good loading).
	Us2	0.460	86.4	91.1 (LOW)	DROPPED	Statistically fine; prob below cutoff.
	Us3	0.460	86.3	108.1	DROPPED	Similar to Us2.

	Us4	0.460	89.8	86.0 (LOW)	KEPT	Strongest PU item.
Continuance Intention (CI)	CI_1	0.459	79.3	218.4	KEPT	Best of CI set despite low prob.
	CI_2	0.455	72.2	333.5	KEPT	Weaker than CI_1 but usable if ≥2 items needed.
	CI_3	0.799	40.6 (LOW)	576.5 (VERY HIGH)	DROPPED	Very weak loading.

Appendix 2.5. *Posterior Indicator Probabilities of Being a Nonzero Matrix after Scale Refinement (Factor=2)*



Appendix 3. *Design Evaluation*

Appendix 3.1. *Presence of Conversational-Agent Design Elements Across Six Benchmark Chatbots (Yes/No Coding)*

Design Element	US Citizen	Lufthansa	Seattle Ballooning	Kia	University of Twente	Wanderlog
Button	yes	yes	yes	yes	no	yes
Response Delay	yes	yes	yes	yes	yes	yes
Typing Indicator	no	yes	yes	yes	yes	yes
Emoticons	no	no	yes	no	yes	no
Typing Error	no	no	yes	no	no	no
Name	yes	no	no	no	yes	no
Gender	yes	no	no	no	no	no
Static Avatar	yes	no	no	no	no	no
Virtual (interactive)	no	no	no	no	no	no
Age	no	no	no	no	no	no
Race	no	no	no	no	no	no
Pause-filler	no	no	no	no	no	no
Respect Words	no	no	no	no	no	no
Displaying listening behavior	no	no	no	no	no	no
Tone Modulation	no	no	yes	no	no	no
Others' Comments	no	no	yes	no	no	no
Self-referencing	no	no	yes	no	yes	no
Self-Introduction	yes	yes	yes	yes	yes	yes
Human Intervention	no	yes	no	no	no	no
Greeting	yes	yes	yes	yes	yes	no
Apologizing	yes	yes	no	yes	yes	yes
Harsh Response	no	no	no	no	no	no
Asking User Feedback	no	no	no	no	no	no
Taking responsibility for errors	yes	no	no	no	no	no
Humorous response	no	no	no	no	no	no
Negative Languagae	no	no	no	yes	yes	no
Positive Language	no	no	yes	no	no	no
Persuasive language	no	no		no	yes	no
Avoidance Language	no	no	no	yes	yes	no
Lexical Diversity	no	no	yes	no	no	yes

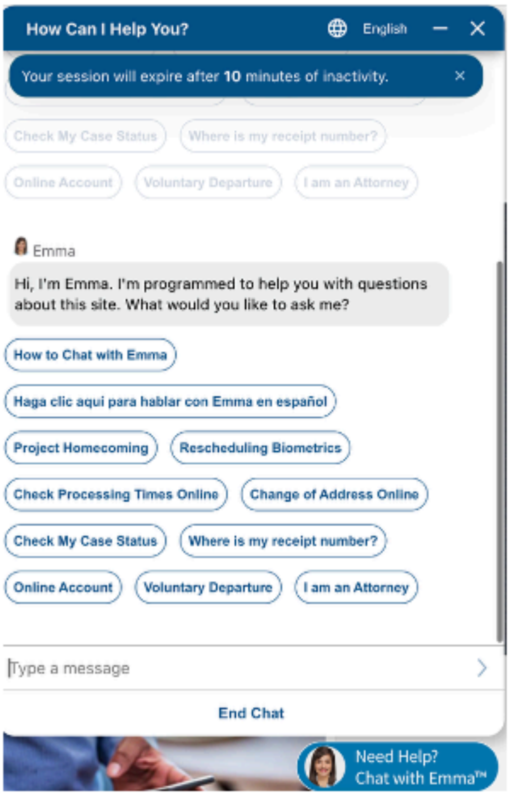
Interactivity	no	no	yes	no	yes	no
Direct addressing	no	no	no	no	no	no
Expressive Speech Acts	no	no	yes	no	no	no
Outlining logical process					no	no
Reasoned Utterance	no	no	yes	no	no	no

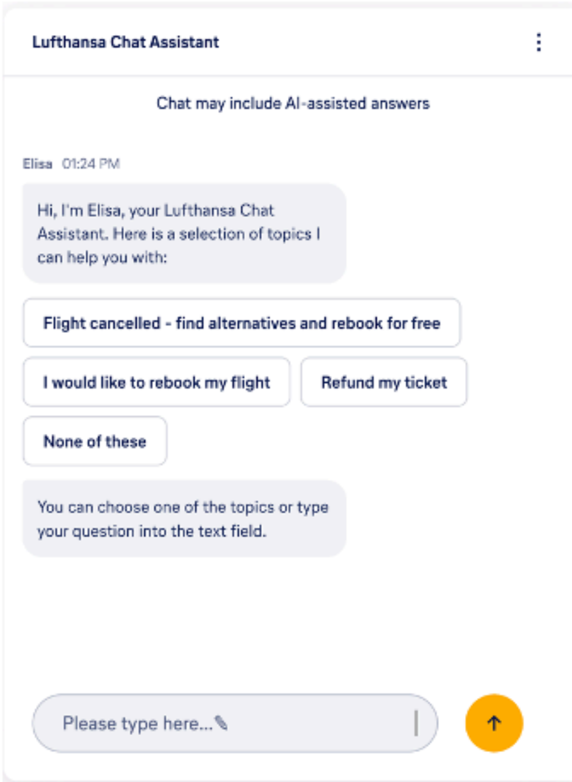
Appendix 3.2. *Selected Chatbots and Scenario's for Online Study*

Chatbots	Scenarios
Kia	You are considering buying a new car and want to explore Kia's options. Use the Kia chatbot to find out the starting price and main features of the Kia Sportage model in the UK.
US Citizen Immigration Service	Imagine you are a refugee in the United States and you want to apply for a Green Card (permanent residence). You decide to use the USCIS website chatbot to find the correct application form for refugees. Please try to use the chatbot to find where you can access this application form.
Seattle Ballooning	Imagine you are planning a special trip in Seattle and are interested in taking a hot air balloon ride. You decide to use the Seattle Ballooning website's chatbot to learn more. Please try to use the chatbot to find the following information: 1. Find the price of a hot air balloon ride in Seattle. 2. Find out how long the balloon ride usually lasts. 3. Find out where does the ride end
Wanderlog	Imagine you are planning a trip to Amsterdam. You want to visit some of the city's most famous museums, and you decide to use the Wanderlog chatbot to plan your visit. Please try to use the chatbot to find out: 1. Which museums should be visited in Amsterdam? 2. What are the entrance fees? 3. Are there student discounts?
Lufthansa Airline	You're planning to book a first-class ticket on Lufthansa from Frankfurt to New York for an important business trip. You want to understand the baggage allowances for first-class passengers before finalizing your booking. Ask the Lufthansa chatbot about luggage allowances for first-class flights from Frankfurt to New York.
University of Twente	Imagine you are a Turkish citizen, and you are a recent bachelor's graduate exploring Master's programs at the University of Twente in **September 2026.** You want detailed information before applying, including admission requirements, deadlines, and tuition fees. Find out details for the Psychology Master Programme using the University of Twente chatbot.

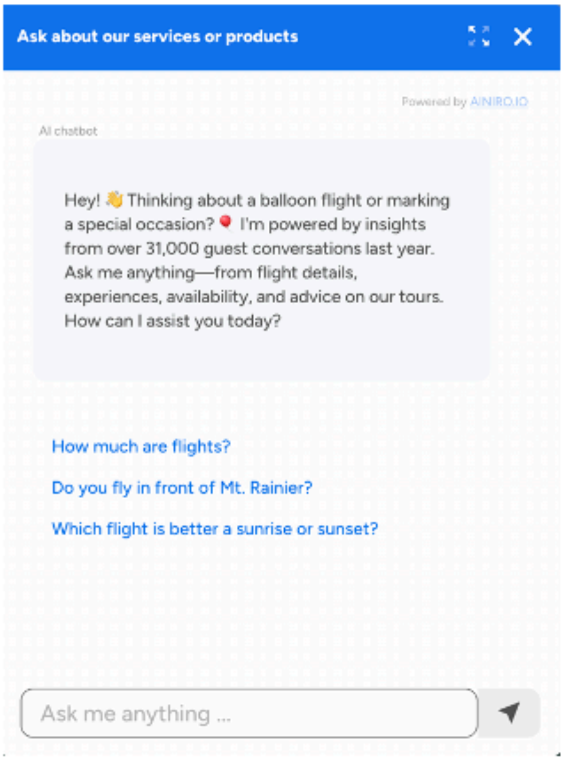
Appendix 3.3. *Selected Chatbot Interfaces*

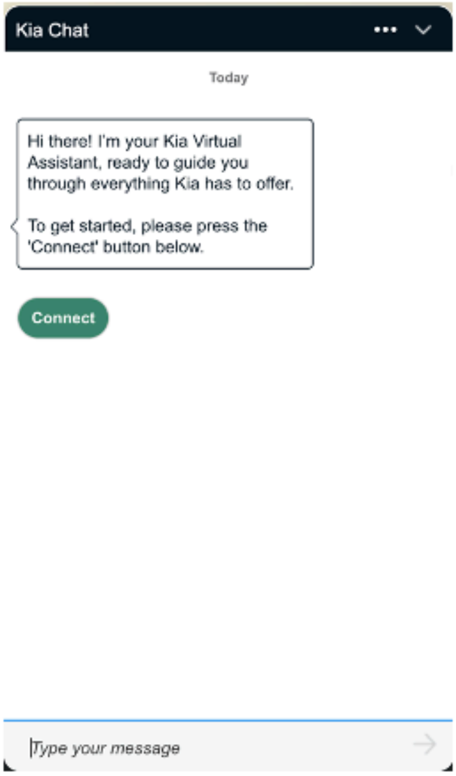
Appendix 3.3.1. *US Citizen and Lufthansa Chatbots*





Appendix 3.3.2. *Seattle Ballooning and Kia Chatbots*





Appendix 3.3.3. *University of Twente and Wanderlog Chatbots*

Appendix 4. *Systematic Literature Review*

Appendix 4.1. *Summary of User-Evaluation Outcomes and Effect Directions Across Studies*

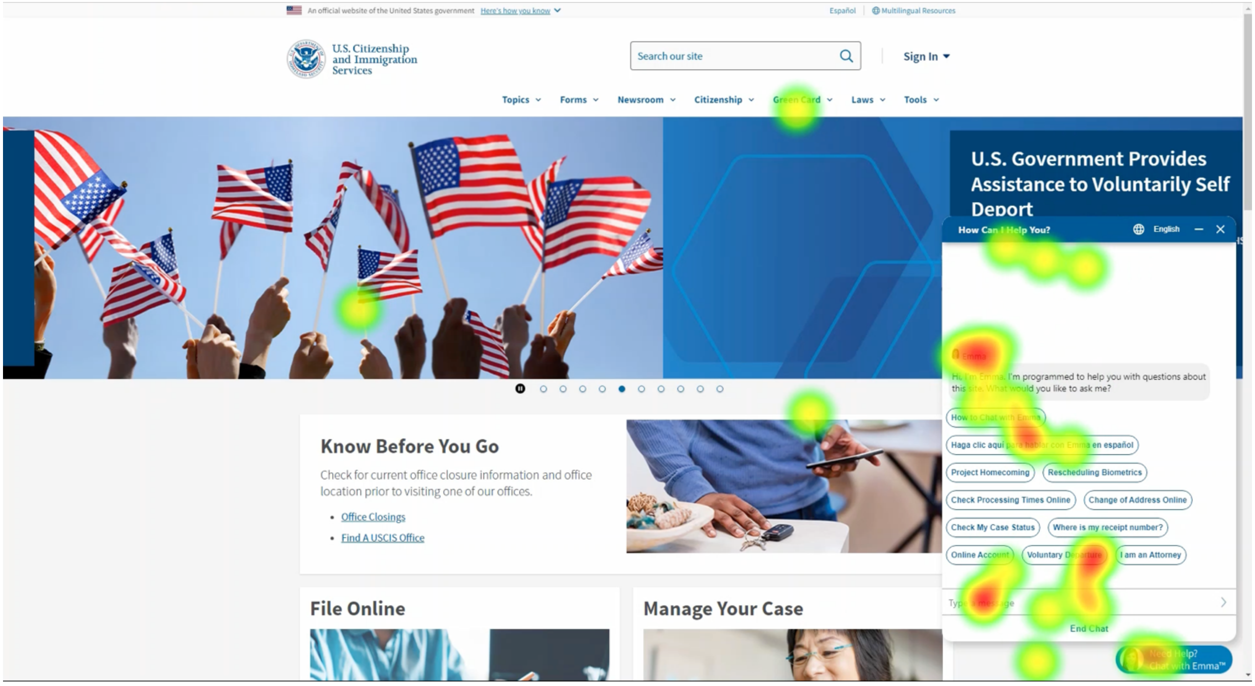
Outcome	User Evaluation Outcome	Negative Effect	No effect	Positive Effect	Total	Positive Effect Rate
Perception	**Perceived Social Presence**	0	9	22	31	70,97%
Trust	**Trust**	0	2	18	20	90,00%
Perception	**Perceived Humanness**	1	3	13	17	76,47%
Acceptance	**Perceived Anthropomorphism**	0	3	7	10	70,00%
Acceptance	**Self Disclosure**	1	5	3	9	33,33%
Attitude	**Satisfaction**	0	0	9	9	100,00%
Acceptance	**Perceived Usefulness**	1	2	3	6	50,00%
Emotion	**Enjoyment**	1	0	5	6	83,33%
Performance	**Perceived Usability**	0	2	3	5	60,00%
Acceptance	**Intention to Use**	0	0	5	5	100,00%
Perception	**Service Satisfaction**	0	0	5	5	100,00%
Acceptance	**Privacy Concern**	2	2	0	4	0,00%
Trust	**Credibility**	0	2	2	4	50,00%
Perception	**Continued use**	0	1	3	4	75,00%
Other	**Compliance**	1	0	2	3	66,67%
Other	**Effectiveness**	0	1	2	3	66,67%
Performance	**Empathy**	0	0	3	3	100,00%
Other	**Frustration**	0	3	0	3	0,00%
Other	**Intention to Donate**	1	0	2	3	66,67%
Perception	**Negative WoM**	0	2	1	3	33,33%
Emotion	**Perceieve Ease of Use**	0	0	3	3	100,00%
Other	**Perceived Transparency**	0	0	3	3	100,00%
Perception	**Purchase Intention**	0	0	3	3	100,00%
Perception	**Rapport**	0	1	2	3	66,67%
Perception	**Uncanniness**	1	0	2	3	66,67%
Perception	**Warmth**	0	1	2	3	66,67%
Trust	**Competence**	0	1	1	2	50,00%
Performance	**Efficiency**	0	1	1	2	50,00%
Emotion	**Emotion Recognition**	0	0	2	2	100,00%
Perception	**Emotions (Guilt, Shame, Anger, etc.)**	0	2	0	2	0,00%
Other	**Helpfulness**	0	1	1	2	50,00%
Learning	**Learning Outcome**	0	1	1	2	50,00%
Trust	**Likeability**	0	0	2	2	100,00%
Other	**Perceived Intelligence**	0	1	1	2	50,00%
Perception	**Perceived Persuasiveness**	0	0	2	2	100,00%
Emotion	**Affective Trust**	0	0	2	1	100,00%

Emotion	**Aggression**	1	0	0	1	0,00%
Emotion	**Anger**	0	0	1	1	100,00%
Perception	**Appropriateness**	0	1	0	1	0,00%
Perception	**Behavioral Intention**	0	1	0	1	0,00%
Trust	**Benevolence Belief**	0	0	1	1	100,00%
Perception	**Boundary linkage**	0	1	0	1	0,00%
Trust	**Cognitive Trust**	0	0	1	1	100,00%
Perception	**Comfort**	0	1	0	1	0,00%
Perception	**Communication**	0	0	1	1	100,00%
Attitude	**Communication Satisfaction**	0	0	1	1	100,00%
Emotion	**Companionship**	0	0	1	1	100,00%
Trust	**Competence Belief**	0	0	1	1	100,00%
Trust	**Dissatisfaction**	0	0	1	1	100,00%
Other	**Eeriness**	0	0	1	1	100,00%
Acceptance	**Effort Expectancy**	0	0	1	1	100,00%
Emotion	**Engagement**	0	0	1	1	100,00%
Learning	**Extrinsic Motivation**	0	0	1	1	100,00%
Perception	**Familiarity**	0	1	0	1	0,00%
Other	**Feedback Acceptance**	0	1	0	1	0,00%
Performance	**Goodwill Trust**	0	0	1	1	100,00%
Other	**Hedonic Motivation**	0	1	0	1	0,00%
Other	**Hedonic Quality**	0	0	1	1	100,00%
Perception	**Information quality**	0	0	1	1	100,00%
Emotion	**Integrity**	0	0	1	1	100,00%
Trust	**Integrity Belief**	0	0	1	1	100,00%
Acceptance	**Intention to Accept**	0	1	0	1	0,00%
Perception	**Intention to Comply**	0	0	1	1	100,00%
Trust	**Intention to reuse**	0	1	0	1	0,00%
Emotion	**Intimacy**	0	0	1	1	100,00%
Other	**Intrinsic Motivation**	0	0	1	1	100,00%
Acceptance	**Perceived Agent Expertise**	0	0	1	1	100,00%
Other	**Perceived Explanation Quality**	0	0	1	1	100,00%
Perception	**Perceived Friendliness**	0	0	1	1	100,00%
Acceptance	**Perceived Informativeness**	0	0	1	1	100,00%
Other	**Perceived Interaction Atmosphere**	0	0	1	1	100,00%
Perception	**Perceived Realism**	0	0	1	1	100,00%
Other	**Perceived Recommendation Quality**	0	0	1	1	100,00%
Acceptance	**Perceived Relatability**	0	0	1	1	100,00%
Perception	**Perceived Reliability**	0	0	1	1	100,00%
Other	**Perceived Responsiveness**	0	0	1	1	100,00%
Other	**Perceived Similarity**	0	0	1	1	100,00%

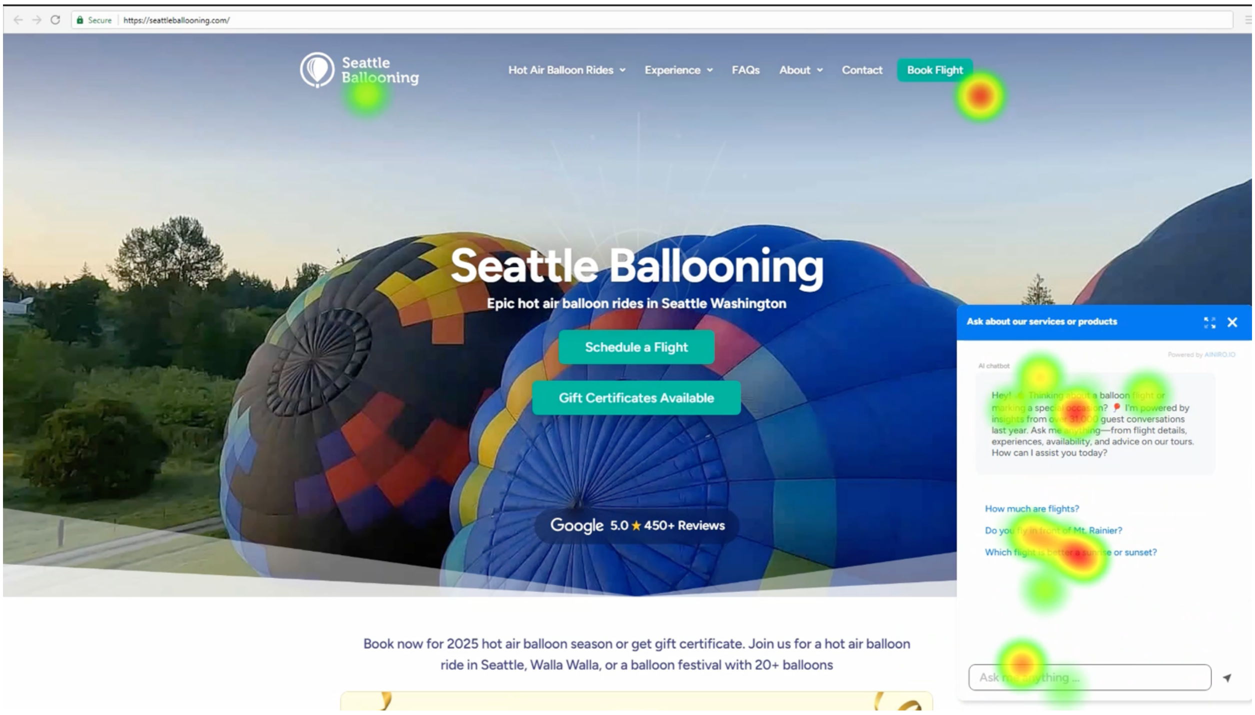
Acceptance	**Perceived Strangeness**	0	0	1	1	100,00%
Perception	**Performance Expectancy**	1	0	0	1	0,00%
Perception	**Physical Presence**	0	1	0	1	0,00%
Perception	**Pragmatic Quality**	0	0	1	1	100,00%
Perception	**Presence**	0	0	1	1	100,00%
Acceptance	**Propensity to trust**	0	1	0	1	0,00%
Trust	**Qualification Trust**	0	0	1	1	100,00%
Perception	**Quality of Interaction**	0	1	0	1	0,00%
Perception	**Recommendation adherence**	0	0	1	1	100,00%
Perception	**Relationship**	0	0	1	1	100,00%
Other	**Risk Awareness**	0	1	0	1	0,00%
Perception	**Self efficacy**	0	1	0	1	0,00%
Perception	**Service quality**	0	0	1	1	100,00%
Perception	**Social Anxiety**	0	1	0	1	0,00%
Perception	**Trusting Disposition**	0	1	0	1	0,00%
Perception	**Understanding Rapport**	0	1	0	1	0,00%
Perception	**Willingness to Donate**	0	0	1	1	100,00%
Perception	**Willingness to engage in social activities**	0	1	0	1	0,00%

Appendix 5. *Eye-Tracking Experiment Figures*

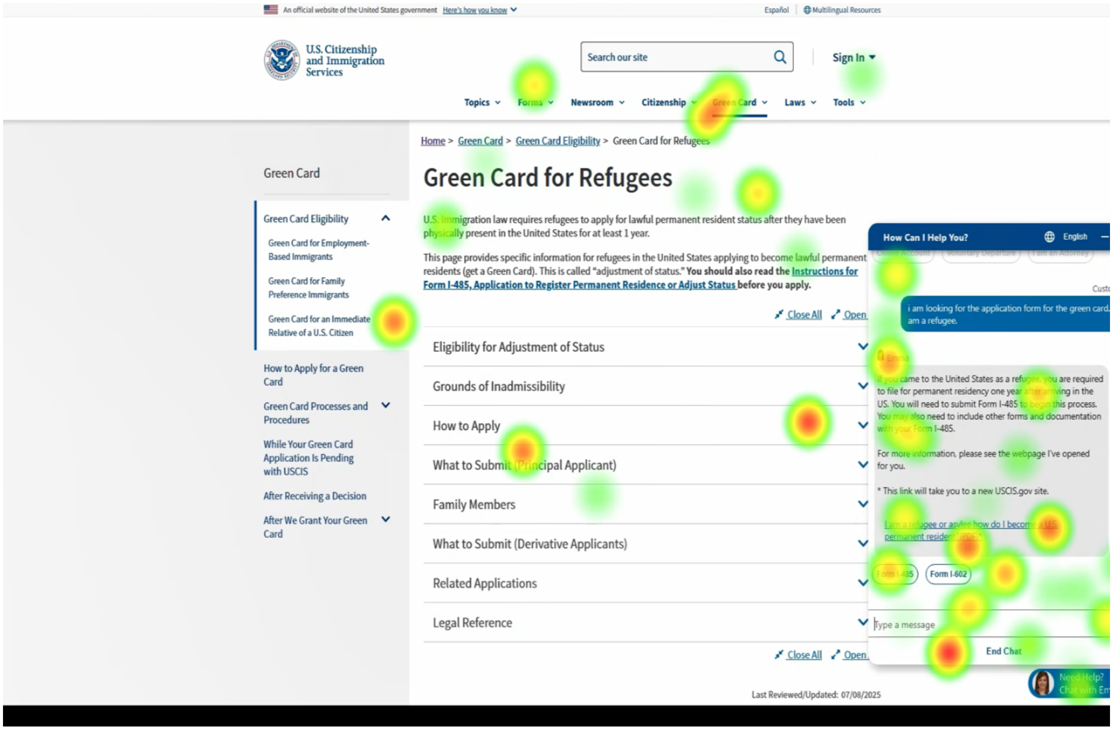
Appendix 5.1. *Low Chatbot HeatMap for Introduction Phase*



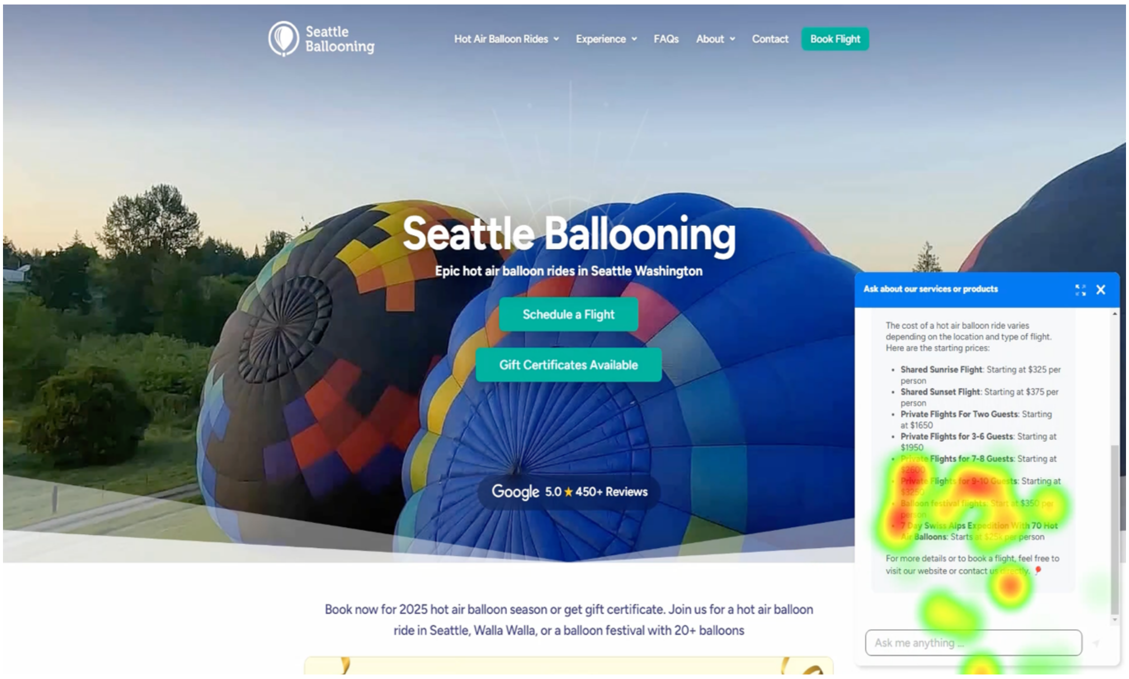
Appendix 5.2. *High Chatbot HeatMap for Introduction Phase*



Appendix 5.3. *Low Chatbot HeatMap for Interaction Phase*



Appendix 5.4. *High Chatbot HeatMap for Interaction Phase*



CURRICULUM VITAE

Name Surname: Neşe BAZ AKTAŞ

EDUCATION STATUS

Degree of Education	Place of Study	Year of Study
Doctorate	Sakarya University / Graduate School of Business / Management Information Systems	2020 - Currently
Master	Dokuz Eylül University / Graduate School of Social Sciences/ Management Information Systems	2015-2019
Bachelor	Ege University / Faculty of Engineering / Electrical and Electronics Engineering	2009-2014

JOB EXPERIENCE

Year	Location	Position
2025-Currently	University of Twente	Researcher
12.01.2026 14:07:00	Bogaziçi University	Research Assistant
2021-2023	Bilgi University	Research Asisstant
2017-2018	Adres Gezgini Yazılım Tasarım Bilişim	Data Analyst

PUBLISHED WORKS

Aktaş, N. B., Şişman, B., & Borsci, S. (2025). Unleashing the potential of Turkish chatbots: a study on the validity and reliability of the bot usability scale. *Universal Access in the Information Society,* 2467–2476. https://doi.org/10.1007/s10209-025-01211-9.

Aktaş, N. B., & Akbıyık, A. (2025). *Conversational Agent Design: A Comprehensive Analysis of Research from Leading Conferences.* International Conference on Intelligent Human Computer Interaction.105-121. Enschede, Netherlands.

Aktaş, B. & Korkmaz, C. (2025). Neuro Information Systems. In A. Akbıyık (Eds.), *Current and prospective approaches, methods, and techniques for management information systems research* (pp. 17–31). Sakarya Yayıncılık.

Aktaş, B., Baz Aktaş, N., & Akbıyık, A. (2022). İşgücü piyasalarında yönetim bilişim sistemleri programlarının farkındalığı: ABD ve Türkiye'deki iş ilanları üzerinden bir değerlendirme. *Journal of Research in Business*, *7(IMISC2021 Special Issue)*, 60-79. https://doi.org/10.54452/jrb.1024997

Baz Aktaş, N., Aktaş, B., & Akbıyık, A. (2021). Koronavirüs'ün (covid-19) Türkiye'de e-ticaret müşteri memnuniyetine etkisi: Trendyol örneği. *Journal of Information Systems and Management Research*, 3(1), 39-50.

Baz, N., Aktaş, B., Akcura, N., Sokullu, R., Uyar, E. (2015). *BikeEge - bicycle sharing system for ege university students.* International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2015, Sofia, Bulgaria.