

Please quote as: Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A. & Leimeister, J. M. (2024). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. *Information Systems Frontiers*, 26, 729–754. doi: 10.1007/s10796-023-10375-9



Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers

Sofia Schöbel¹ · Anuschka Schmitt² · Dennis Benner³ · Mohammed Saqr⁴ · Andreas Janson² · Jan Marco Leimeister^{2,3}

Accepted: 15 January 2023 / Published online: 20 April 2023
© The Author(s) 2023

Abstract

Conversational agents (CAs) have come a long way from their first appearance in the 1960s to today's generative models. Continuous technological advancements such as statistical computing and large language models allow for an increasingly natural and effortless interaction, as well as domain-agnostic deployment opportunities. Ultimately, this evolution begs multiple questions: How have technical capabilities developed? How is the nature of work changed through humans' interaction with conversational agents? How has research framed dominant perceptions and depictions of such agents? And what is the path forward? To address these questions, we conducted a bibliometric study including over 5000 research articles on CAs. Based on a systematic analysis of keywords, topics, and author networks, we derive “five waves of CA research” that describe the past, present, and potential future of research on CAs. Our results highlight fundamental technical evolutions and theoretical paradigms in CA research. Therefore, we discuss the moderating role of big technologies, and novel technological advancements like OpenAI GPT or BLOOM NLU that mark the next frontier of CA research. We contribute to theory by laying out central research streams in CA research, and offer practical implications by highlighting the design and deployment opportunities of CAs.

Keywords Bibliometric analysis · Chatbot · Conversational agent · Voice assistant · ChatGPT · Large language models · Generative artificial intelligence

1 Introduction

Industry and service providers have found interest in conversational agents (CAs), oftentimes colloquially called chatbots. Research on such agents is a rapidly growing field, with an increase in publications over the last few decades, especially fueled through the development of CAs such as Amazon's Alexa that build upon voice as a modality, so called voice bots (Schmitt et al., 2021; Seaborn et al., 2022; Kendall et al., 2020). Further, recent developments such as the release of the ChatGPT beta from OpenAI pave the way towards a disruptive general AI where conversational partners are general assistants to a

wide variety of tasks (Haque et al., 2022). Historically, the first chatbot was developed in 1966 by Weizenbaum (1966). Since then, numerous studies have discussed and analyzed the relevance and effectiveness of chatbots and other types of CAs. With the advent of artificial intelligence (AI) and the deployment of more complex models such as neural networks (Sedik et al., 2022; Goudos et al., 2019; Kushwaha & Kar, 2021), CAs offer the potential to be sophisticated interactive systems, assisting not only customers to handle their complaints but also acting as learning assistants in tutoring students or employees, for instance (Schmitt et al., 2021; Wambsganss et al., 2021).

The growing relevance of CAs is further supported by the global market of CA commercialization, which is expected to grow by 24.3% until 2025. This equals a total market net worth of 1.25 billion USD. Despite technical advances and CAs' market penetration, current research points towards unexpected challenges and previously unaddressed questions. For example, organizations experience issues integrating chatbots into customer service (Adam et al., 2020; Behera et al., 2021) because of customers' skepticism

Anuschka Schmitt and Dennis Benner have contributed equally to this work.

✉ Sofia Schöbel
sofia.schoebel@uni-osnabrueck.de

Extended author information available on the last page of the article.

toward chatbots who view such bots as unnatural, impersonal, or deceptive. In addition, CAs' adoption can be impeded by communication and interaction problems, as well as ethical concerns (Heyselaar & Bosse, 2019). With many decades of CA research and new questions arising, the results of extant research can serve as a foundation to sketch out the frontiers of existing CA research, while such an overview also enables to identify new avenues for CA research. Therefore, with this study, we aim to provide an overview of the past, present, and future of CAs and aim to answer the following research questions:

RQ1: How has research on CAs developed over time?

RQ2: What are common research streams around CA research?

RQ3: What are future research directions and potential novel frontiers for the design and use of CAs?

To answer our research questions, we present the results of a bibliometric study that considered over 5,000 studies on CAs. In our study, we provide an overview of leading authors and countries and demonstrate the development of keywords in this research domain. Furthermore, we discuss implications for future research by introducing five historical waves of CAs marked by unique technological advances and research focus each.

We contribute to theory and practice by presenting a consolidated overview of the state-of-the-art regarding the usage of CAs. This perspective is additionally supported by implications and a discussion of future research. Practitioners find support in implications about what to care about when designing and deploying CAs. In addition, by presenting different waves of the development of CAs over time, we intend to support practitioners in reacting effectively to future trends in CA design.

2 Related Research

2.1 Conversational Agents as a System Class and their Historical Roots

CAs are used as communication tools between humans and technical devices. A CA involves a technological artifact that uses natural language to engage in a dialogue with the user, usually to fulfill a task or provide assistance (Følstad & Skjuve, 2019; Laban & Araujo, 2019). The first agent was developed in 1966 by Joseph Weizenbaum, who created a computer program that could communicate with humans via a text-based interface (Weizenbaum, 1966). These text-based communication programs were followed by voice-based interaction systems and embodied CAs (McTear et al., 2016). In our bibliometric study, we discuss and consider these various types and instantiation of agents. Many different terms are used concerning communicative technology.

The term “conversational agent” is oftentimes used interchangeably with the terms “intelligent personal assistant” (Hauswald et al., 2016), “smart personal assistant” (Knote et al., 2021), “chatbot” (Brandtzaeg, 2018), or “conversational agent” (Feine et al., 2019). The relevance of CAs has emerged in many disciplines and tasks, such as in customer service that can be available 24/7, in which a bot assists customers in handling their complaints (Qiu & Benbasat, 2009; Behera et al., 2021). In addition, with the rise of AI, especially natural language processing (NLP), new possibilities exist to work with and change our technology. CAs enable us to find a new and convenient way of accessing services and content by improving interactions between information systems and users (Brandtzaeg, 2018; Behera et al., 2021).

Since the first construction of a CA, many advancements have been made, enabling us not only to create Q&A bots but also to develop intelligent bot solutions. Some CAs use natural language to communicate with users (Feine et al., 2019). New forms of such agents can process compound natural language and thereby respond to increasingly complex user requests (Knote et al., 2021), such as Amazon's Alexa. These assistants operate with a voice interface reacting to individual wake words and questions, like “Alexa, what time is it?” or “Alexa, is it going to rain today?” (Fischer et al., 2019). Research and practice are attempting to make both kinds of bots (textual and acoustic) more human by integrating avatars into text-based solutions (Feine et al., 2019; Purington et al., 2017) or by integrating emotional voices (Knote et al., 2021). Integrating human-like characteristics into agents may cause users to exhibit emotional, cognitive, or behavioral reactions resembling human interactions yet can also cloud users' understanding of the system (Krämer et al., 2005).

2.2 Existing Review Approaches to Study Conversational Agents

To summarize the results of extant research efforts in a particular field of study, researchers typically use literature reviews. As a point of departure for our study, we identified and studied nine literature reviews on CAs (see Table 3). We deemed these reviews detrimental in framing and summarizing research narratives in the CA domain, as well as informing us on extant conceptualizations and overviews of this system class. All of these reviews have been published within the past five years, underlining the current omnipresence of CAs. The majority of reviews focus on studies from 1998 onward, thereby neglecting first papers on CAs such as well-known first text-based agent ELIZA (Weizenbaum, 1966). In addition, we quickly

realized that extant reviews mainly take a socio-technical design perspective by exploring specific design themes such as CA team collaboration (Diederich et al., 2022; Poser et al., 2022; Stieglitz et al., 2021; Abedin et al., 2022) or CAs for education (Smutny & Schreiberova, 2020). While these reviews allow for strong theoretical and practical implications regarding the design of such agents for particular contexts, we are missing a holistic overview of the conceptual and empirical foundations of CA research which also considers the underlying technical advances related to CAs.

The number of studies analyzed in literature reviews ranges from 29 articles (Rheu et al., 2021) to 262 articles (Diederich et al., 2022). Whereas some of the literature reviews have a more centralized view on topics involving CAs, for example, the role of trust supporting components (Rheu et al., 2021; Vössing et al., 2022), others discuss the meaning of disembodied chatbots in human-computer interaction (Chaves & Gerosa, 2021) or characteristics of the adoption of AI-based CAs (Rzepka & Berger, 2018; Elshan et al., 2022). More and more questions involve a discussion of how to make CAs more useful by experiencing how to make them more human-like (Chaves & Gerosa, 2021) or about how to make them a better collaborative partner in communication (Diederich et al., 2022; Elshan et al., 2022; Poser et al., 2022; Abedin et al., 2022). Thus, literature reviews often provide a closer look at specific topics surrounding chatbots (e.g., trust) and serve as a valuable contribution to the study of specific areas in CA research.

All the literature reviews suggest areas for future research, such as Smutny and Schreiberova (2020), who suggested analyzing how we can support developers to create and offer tools that allow any teacher to integrate chatbots into their classes. Smutny and Schreiberova (2020) analyzed educational chatbots on Meta (formerly Facebook), this perspective could be better supported by including literature to enrich the perspective on the topic. Research has discussed how we can support teachers in using CAs for their teaching (e.g., Winkler and Söllner (2018)). Other studies suggest better analyzing the behavior in the communication of chatbots by differentiating and considering various variables, such as gender or personality, to detect individual designs (Bickmore et al., 2020; Ahmad et al., 2022). Discussed across various literature reviews is the aspect to take a closer look at the role of AI for CAs, such as the fit between users and systems for different AI-enabled systems (Rzepka & Berger, 2018).

Compared to other methods of literature screening, reviews often only consider a limited, selected number of studies. After 40 years of analyzing the role and meaning of CAs, we should look back to identify which authors and topics are most impactful and to discuss how and

if chatbots could determine future research. Bibliometric analyses support us in obtaining such a perspective by overcoming the limits of literature studies in considering only a limited number of studies. Other than literature reviews, bibliometric studies oftentimes analyze thousands of publications, utilize statistical methods, and, therefore have become a popular method to identify patterns in collected data and work that reveal emerging trends in research as it evolves (Trinidad et al., 2021). A bibliometric approach can thus provide a more exhaustive overview of extant research while simultaneously uncover patterns and details within a research field other research methods are incapable of. To provide an overview of related work, we summarize the results of major CA reviews and the bibliometric study we identified in Appendix B.

So far, only one study has used a bibliometric approach to study CAs (Io & Lee, 2017). In their work, the authors analyzed past research on CAs. They presented an overview of the number of publications for each year and calculated a cluster of keywords, research areas, and keyword co-occurrences. However, the presented insights and suggestions for future research (e.g., analyze other CAs, such as mobile chats, embedded gadgets) only remain on a surface level. With our study, we want to expand the work of Io and Lee (2017) and overcome their limitations in different ways. First, we performed a coding procedure to identify relevant studies using *rayyan.ai* as a tool to determine which study fit our goal. Second, we opened up our study by considering all kinds of CAs, enabling us to detect trends and relationships of a specific group of CAs.

3 Methodology

The search protocol followed the preferred reporting items for systematic reviews and meta-analyses literature search extension (PRISMA-S) method (Rethlefsen et al., 2021). We chose the Scopus database because it covers most articles included in Web of Science and includes a larger selection of technical and social science articles (Norris & Oppenheim, 2007). The Scopus database has rigorously maintained metadata and a curated list of journals and conferences that are subjected to a strict inclusion process (Baas et al., 2020; Singh et al., 2021). Several search iterations were performed, and the authors thoroughly examined the results to choose the most appropriate formula. The final search formula was “chatbot*” or “virtual assistant*” or (“conversation* agent”) or “smart assistant*” or “smart bot*” or “natural language interface”.

Only journal articles, conference proceedings, or book chapters of conference proceedings that were published in English prior to 2021 (to include complete years in the trend analysis) were included. Reviews, editorials, and other

non-original articles were excluded. The data were downloaded on the 31st of May 2021. A total of 6,397 articles were retrieved. Two reviewers (first author and third author) reviewed 681 articles and agreed on 656, with an interrater agreement Cohen's Kappa of 0.963 (adjusted = 0.892). The two authors discussed the disagreements and resolved them based on a common understanding. Then, a single reviewer (third author) completed the remaining articles. When the author was in doubt, the article was classified as "maybe" and resolved based on a review of the full text. A total of 5,107 articles were judged as matching the inclusion criteria, and one article was a duplicate and removed.

The data were processed, cleaned, and prepared for analysis, so 1) the author names were manually checked and fixed with the aid of Scopus author IDs. Authors' names with different spellings, name changes, and special characters were therefore fixed. 2) Publication venues were cleaned by inspecting and fixing different venues with the same title and spelling variations (e.g., Computers & Education and Computers and Education). 3) Keywords were compiled and cleaned with the aid of Google Openrefine. Openrefine is an open-source software designed to help clean up messy datasets; it offers several NLP algorithms and clustering methods for identifying similar keywords in terms of spelling or phonetics (e.g., "chatbot," "chatbots," "chat bots," "chat-bots," "chat bot"). All such keywords with their identical alternatives were checked carefully by the third author and merged. Another round of keyword cleaning was performed to merge keywords that Openrefine could not recognize, such as "NLP" and "Natural Language Processing," and "AI" and "Artificial Intelligence".

The cleaned dataset was analyzed using the R statistical language Bibliometrix package (Aria & Cuccurullo, 2017). Bibliometrix is an open-source software that offers a wide array of tools for the analysis of bibliometric metadata, such as authors, keywords, citations, countries, and references. Frequencies, plots, and temporal trends were computed and plotted using R statistical language. The top authors were retrieved based on their number of articles within our dataset, and their citations and evolution of citations were computed. Co-authorship networks offer a powerful summarization method for the visualization of collaboration and scientific production that helped build and shape a scientific field. The co-authorship network was constructed based on co-authorship of the same article (i.e., if two authors collaborated on the same article, they were considered connected). To avoid assigning higher weights to manuscripts with a higher number of authors, we chose a fractional counting method to build a weighted co-authorship network (i.e., edge weights are inversely proportional to the number of authors) based on the recom-

mended methods by Perianes-Rodriguez et al. (2016).

Subgroups (i.e., communities) of authors who frequently collaborated were mapped using Louvain modularity (De Meo et al., 2011) to highlight groups who frequently collaborated and their contribution and influence on the field. The resultant network was plotted using Gephi (Bastian et al., 2009) with the Fruchterman Reingold layout algorithm (Fruchterman & Reingold, 1991), and each community was assigned a unique color (Fruchterman & Reingold, 1991). For readability, the network was limited to the top 100 authors, based on a weighted degree threshold of 10. The country collaboration network was constructed in the same way (fractional counting) based on the authors' country affiliation. Hence, two countries were considered connected if two authors affiliated with the two countries collaborated on the same article. The country network was plotted using the Fruchterman Reingold layout algorithm, and each community was assigned a unique color (Fruchterman & Reingold, 1991), and communities of countries who collaborated were identified using the Louvain modularity (De Meo et al., 2011).

A keyword network was constructed using the full counting method, where keywords in the same manuscript were considered connected. The network of keyword co-occurrence was plotted using Gephi and partitioned with Louvain modularity (De Meo et al., 2011) so that keywords that occurred together were connected and colored similarly.

4 Results

4.1 General Results

In the following, we reviewed the ten most cited articles to arrive at an understanding of relevant topics driving CA research (see Table 1).

Based on their citation numbers, we assume that these studies are substantially relevant for establishing and driving a common understanding of CA as a concept and making CAs prominent in the fields of Human-Computer Interaction (HCI) and Information Systems.

Publication dates for those ten manuscripts are dispersed, ranging from 1994 to 2016. This dispersion points toward the importance of the first technology established in this field and the ongoing interest in research on CAs fueled by ongoing advances in related technology.

4.2 Authors and Countries

In a similar vein, we analyzed the most proliferative authors, co-citations and related countries (see Figs. 7, 8 and 9 in Appendix C).

Table 1 Top-cited Manuscripts

Author (year)	Citations	Keywords	Category
Bickmore and Picard (2005)	581	User interfaces-evaluation / methodology, graphical user interfaces, interaction styles, natural language, theory and methods, voice	(Embodied) CA
Androutsopoulos et al. (1995)	384	N/A	Conversational Interface
Graesser et al. (2005)	375	CAs, intelligent tutoring systems, natural language dialogue	CA
Cassell et al. (1994)	340	N/A	CA
Luger and Sellen (2016)	327	CAs, mental models, evaluation	CA
Cassell et al. (1994)	294	Conversational characters, multimodal input, intelligent agents, multimodal output	Embodied CA
Li et al. (2016)	277	N/A	CA
Vertegaal et al. (2001)	251	Attention-based interfaces, multiparty communication, gaze, conversational attention, attentive agents, eye tracking	CA
Bickmore and Cassell (2001)	235	Embodied CA, trust, social interface, natural language, small talk, personality	(Embodied) CA
Graesser et al. (2001)	226	N/A	Intelligent Agent

With 91 articles, Pelachaud is the most productive author. Pelachaud's works have been mainly been published in the 2010s and focus on multimodal dialogue systems, conversation and related emotional and nonverbal competencies, and image classification. With 61 published articles, Pelachaud is followed by Bickmore as the second most productive author who published between 2002 and 2020. Bickmore's works are prominently positioned in two of the most cited articles in CA research (see Table 1). Similar to Pelachaud's, Bickmore's research is scattered over more than 20 years, becoming more relevant around the 2010s. His work encompasses domain-specific CA applications, such as a hospital companion agent (Bickmore et al., 2015) and the deployment of CAs for automated substance use screening (Bickmore et al., 2020), and on how to build user trust through the design of CAs (Bickmore & Cassell, 2001). The third most productive author is Griol, with 47 published articles. Griol's work has mostly been published around the 2010s, focusing on dialogue system design and user analytics (i.e., predicting users' mental states from CA interaction) (Callejas et al., 2011). Compared with the other two authors, however, Griol focuses much more on speech and language and has thus also been published more heavily in respective outlets (Griol et al., 2008). In addition, Griol's manuscripts are prevalent from 2010 onwards.

In a subsequent step, we explored the co-citation references among the important authors (see Fig. 9). Our co-cited reference network, defined by four key clusters, allows us to better understand how the field was built and which theoretical foundations mainly drove CA research throughout time.

One of the key clusters of CA literature dates back to Turing's (Trinidad et al., 2021) well-known paper proposing his test to assess whether a machine can think or, has a conscience. His paper introduced the ongoing discussion around the "humanness" of machines and the technical feasibility and future of machines. Less than 20 years later, Weizenbaum (1966) introduced the first computer program based on natural language communication, called ELIZA, and addressed the question of how CAs should communicate by laying out the differences between online and offline language. These publications represent the first wave of research on CAs, which were interactional interfaces built on rule-based models. Importantly, these papers are fundamental to ongoing advances in CA research and development (Dale, 2016).

A second dominant cluster comprises interactions among researchers (e.g., Hochreiter and Schmidhuber (1997)) is concerned with the underlying technical functions of CAs. Technical methods, such as storing recent inputs for speech processing using gradient models (Heyselaar & Bosse, 2019) or neural network-based response generation (Shang et al., 2015; Kushwaha & Kar, 2021), can be viewed as key contributions to the technical advancements and thus sophistication of CA. This body of literature is composed of the second wave of CA research around NLP and text-based and statistical methods.

The purple-colored research cluster is defined by authors, such as Nass, and is concerned with the personality of CAs. In 1994, Nass et al. claimed through five experiments that computers are social actors (CASA). Following this so-called CASA paradigm, research on anthropomorphism and

perceiving technology as a human makes up a predominant literature stream in contemporary CA research. More scattered is a smaller network around the research of Cassell et al. (1994) and Kopp and Wachsmuth (2004), focusing on the structure of communication and nonverbal behavior. In Cassells et al. (1994) work, elements of human conversation are reviewed, which can be inferred from CAs. Kopp and Wachsmuth (2004) synthesized speech-based utterances with nonverbal modalities to achieve congruent multimodality for CAs.

We further analyzed publications and research focus across countries (see Fig. 9). In general, authors from the same country published together, with an average collaboration score of 2.46. With almost 10,000 total citations, the United States dominate in terms of prevalent work. A cluster of countries follows with a smaller number of total citations, with Germany with 2,595 total citations as the country with the second-most citations. The United Kingdom, France, China, Italy, and Japan exhibit between 1,000 and 2,000 citations. In a similar vein, the mentioned countries are also the most active in terms of collaboration across countries. As shown in Fig. 9, authors from the United States publish their work with researchers from many other dominant countries, including China,

Germany, the UK, Japan, Canada, India, and Australia. We also identified a collaboration network of Western countries, including Germany, France, Italy, Switzerland, the Netherlands, Austria, and Portugal.

4.3 Keyword Analysis

The next step of our analysis included a detailed keyword analysis (see Figs. 1 and 2).

Our top 20 keywords AI, chatbot, neural network, NLP, virtual assistant, and machine learning were only limitedly used until 2015. The use of each of them greatly increased during the last 5–7 years. This is different from the keywords affective computing, embodied CA, emotion, natural language interface, and virtual agent. These keywords have no linear or clear development, but all of them increase over time. Lastly, deep learning has been relevant from the beginning onward, similar to the keyword dialogue system. A comparison of keyword development will demonstrate that some keywords are more dominant than others.

The keyword most prominent from 2003 to around 2016 was “embodied conversational agents”. The development of this keyword has decreased since 2017. Other than this,

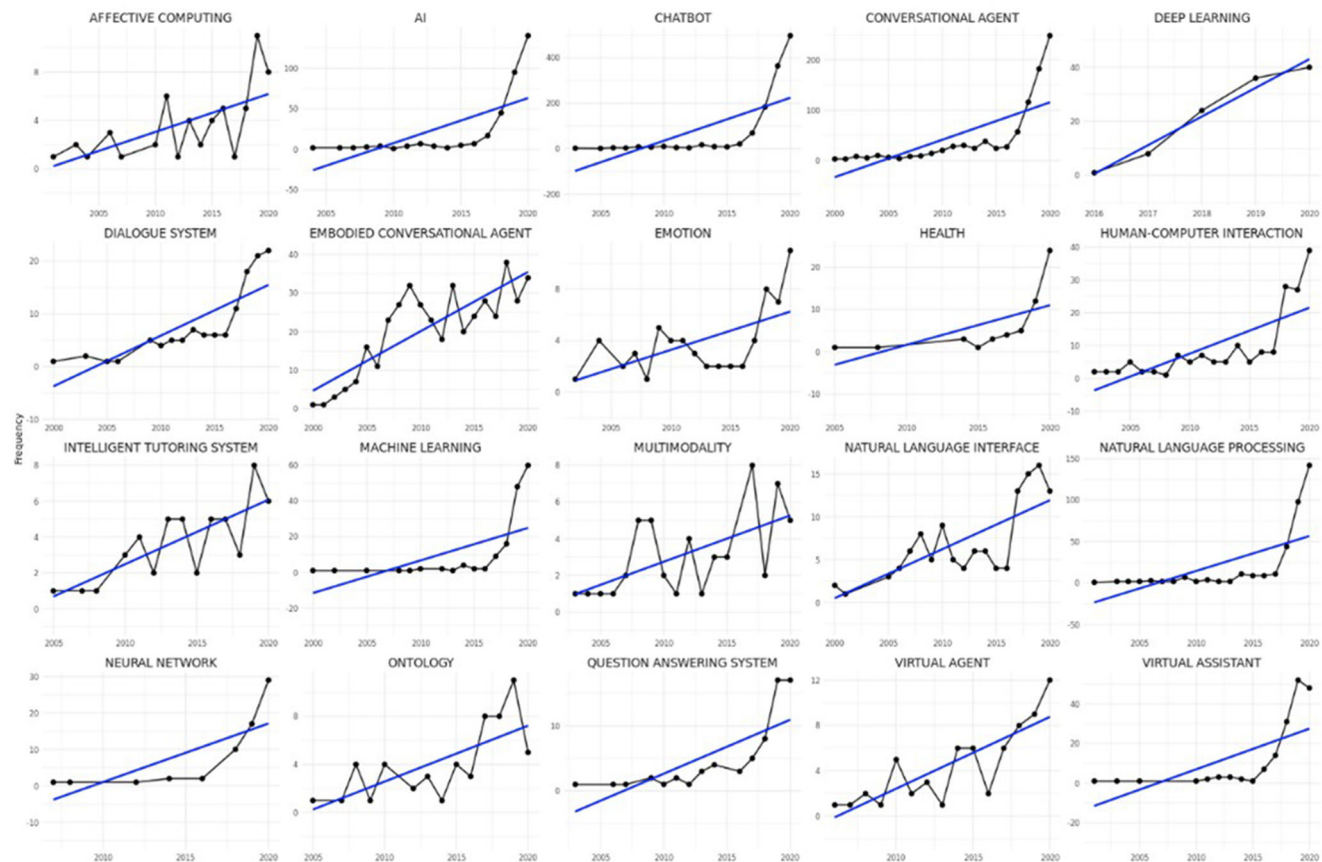


Fig. 1 Distribution of the Top 20 Keywords per Year

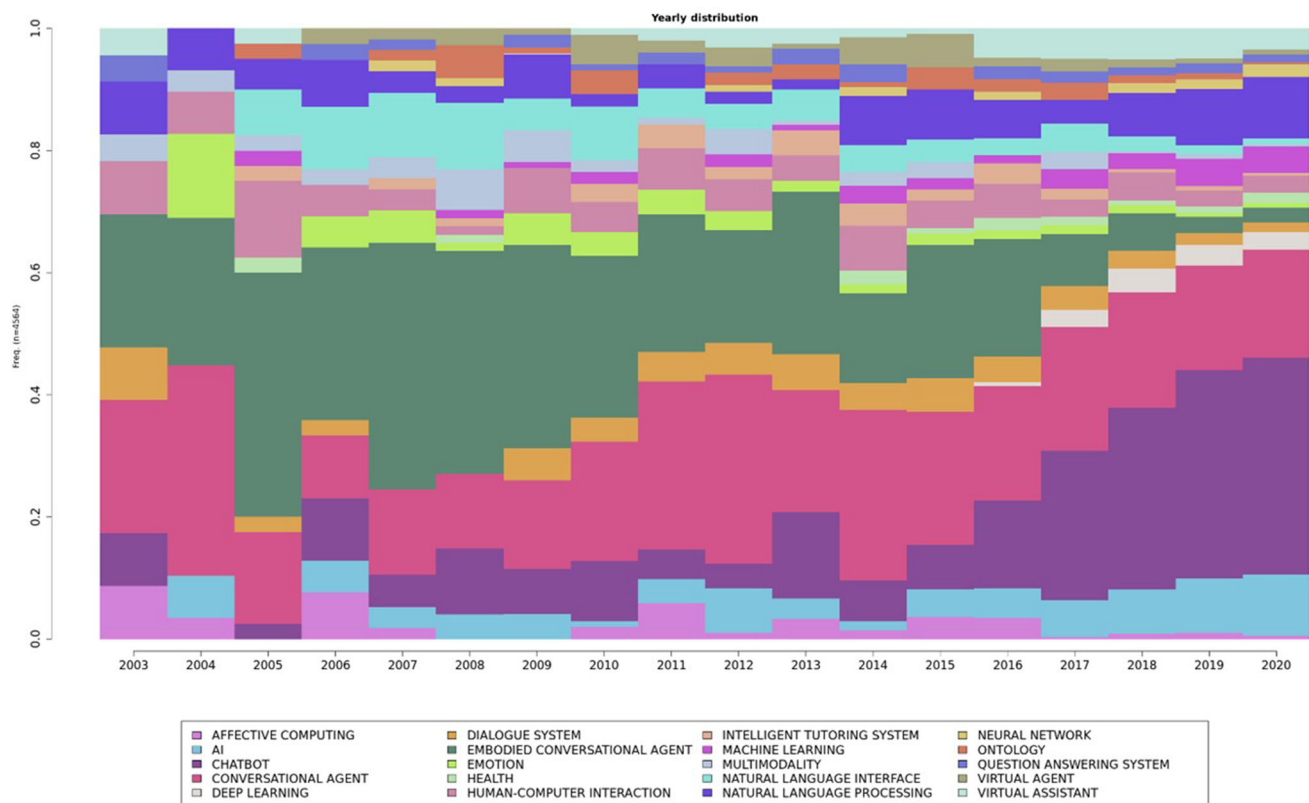


Fig. 2 Yearly Distribution of Keywords

the keyword “conversational agent” has been used more often in studies beginning in 2007. This is similar to the keyword chatbot, which has become more and more relevant beginning in 2015. All other keywords, such as Human-Computer Interaction or NLP, have been relevant over the last couple of years on a continuous level and to a high frequency.

4.4 Keyword Cluster

Turning toward the 50 most prominent keywords in CA literature, the definition of the system class represents the most relevant type of keywords. 11 out of the 50 top keywords refer to a particular system class, such as chatbot, embodied CA, or virtual assistant.

The most prominent are text-based CAs, with the chatbot(s) being most frequent, followed by embodied CAs. These different definitions and connotations of CAs point toward the multitude of types and diversity of CAs, also dependent on the modality of the CA or the publishing outlet of a paper. This diversity also illustrates the fragmentation of CA research and lack of agreement on the terminology. Another prominent keyword cluster is the reference to

the technical model of the CA discussed. Approximately 26% of the top keywords provide insights into the statistical models, pointing toward the importance of technical advances for the proliferation of CAs. AI, machine learning, and natural language (processing) dominate the keywords, which also illustrates how this technical interplay makes the architecture of CAs unique. Approximately 30% of the 50 most prominent keywords illustrate the research focus. Some keywords include related technology relied upon, such as virtual or augmented reality and the Internet of Things. The most relevant research focuses are trust and anthropomorphism, and agent personality and emotions. Interaction-related questions around the multimodality of CAs and the dialogue structure also dominate the keywords in the research focus. Oftentimes, the research domain of a publication is indicated in the keywords. Domains dominating the CA literature include HCI, affective computing, and ontology-related matters. Lastly, keywords indicate the interaction context of a publication (12 out of the 50 top keywords). CA studies seem to be deployed mostly in question answering, information retrieval, and evaluation interactions. E-learning and mental health represent relevant interaction contexts, which have also seen great

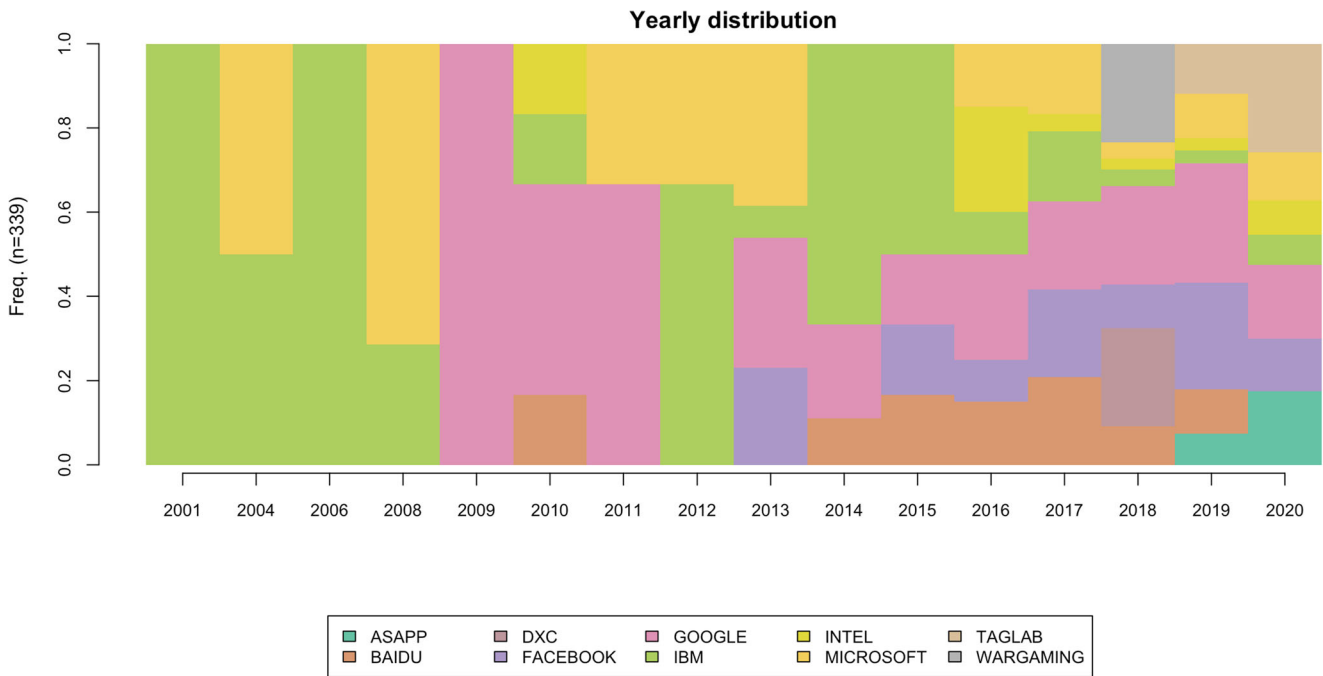


Fig. 3 Relative Yearly Distribution of Corporate Affiliations

prominence in recent publications (Androutsopoulos et al., 1995; Chaves & Gerosa, 2021; Følstad & Skjuve, 2019; Hauswald et al., 2016).

Turning toward keyword clusters (Fig. 10), we see that certain tasks and CAs are commonly associated with specific terminology. The term “conversational agent” is mostly linked to terms concerned with the conversational nature and dialogue of the interaction, such as natural language and dialogue management. Tasks such as question answering and information retrieval are associated with CAs. Lastly, smart personal assistants, such as Amazons Alexa, are linked to the term as well. The term “voice user interface” appears within the CA cluster, yet is not linked to any further words. An equally extensive keyword cluster exists around the term “chatbot”. Chatbots are predominantly studied in interaction contexts around education, e-learning, and deception detection. Technical models, such as NLP, deep learning, and machine learning, appear here. The chatbot cluster is closely linked with a smaller cluster around the term AI, where other technical keywords, such as automation, sentiment analysis, and big data, are found.

The term “embodied conversational agent”, is more concerned with additional modalities, including the avatar or facial expression of the agent. In this context, the personality and emotions of agents are commonly discussed.

Most technical advancements seem to have been empirically investigated in the context of text-based CAs. Voice-based CAs find little appearance in our keyword cluster, which may be the result of a certain choice of

dominant keywords or voice-based CAs not having been researched as extensively as text-based ones. The term “conversational agent” does not necessarily seem to act as an overarching term for the different types of agents. Moreover, certain tasks and contexts are connoted with certain terminology.

4.5 Corporate Affiliations

Given the commercial availability of CAs and the increasing research focus within large technical companies, we deem corporate affiliations a noteworthy aspect to be mentioned when discussing the past, present, and future of CAs (see Fig. 3).

Publications by corporate affiliations were dominated by IBM and Microsoft until 2008, followed by Google from 2009 until 2011. From 2012 onwards, publications by commercially affiliated researchers have become increasingly fragmented (i.e., through a greater number of players and Asian companies, such as Baidu, becoming apparent). Turning toward the temporal development of corporate-related keywords, we identified a steady increase in the frequency of these keywords since 2000 and a steep increase since 2017 (see Fig. 4). More specifically, Google is most frequently listed as a keyword in 2020 ($N = 78$), followed by Met¹ ($N = 49$) and Microsoft ($N = 41$). Whereas Eastern players (i.e., Baidu, ASAPP) have increased in appearance, western players still dominate. Regardless, corporate

¹<https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/>

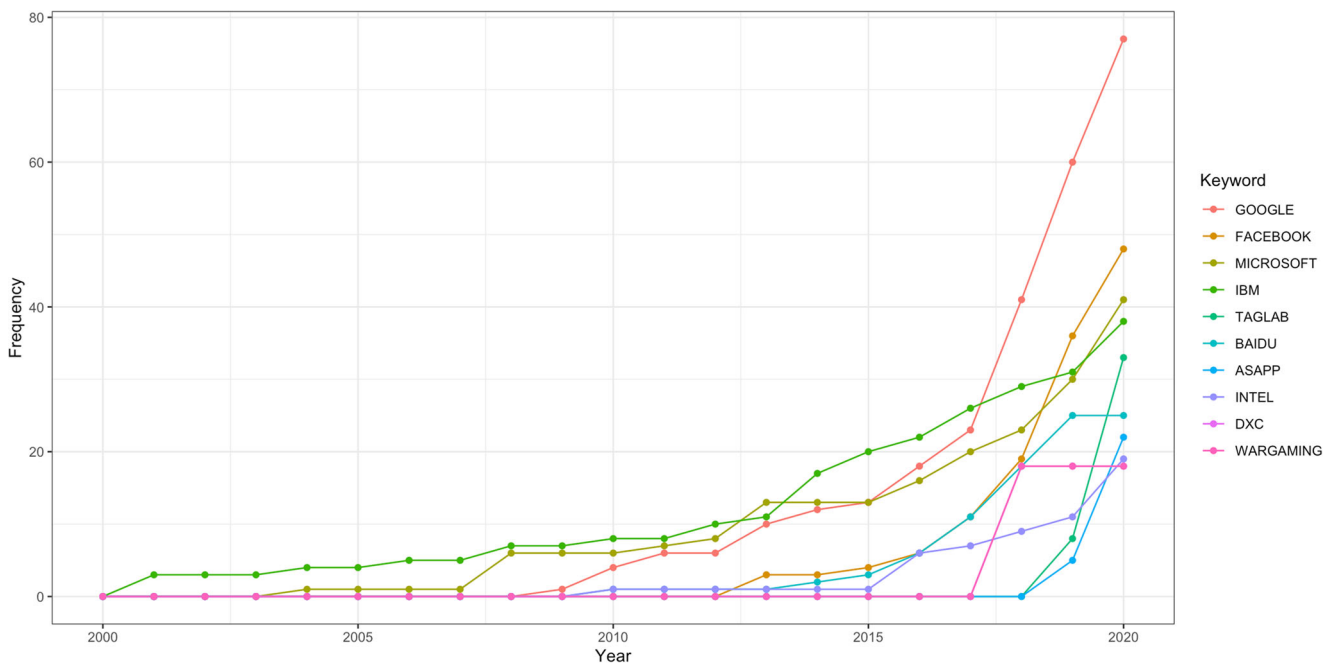


Fig. 4 Development of Corporate-affiliated Keywords over Years

players currently dominate fundamental research on the underlying technologies required to advance CA research. Especially natural language technologies and large language models such as GPT by OpenAI (Brown et al., 2020b), the Galactica model by Meta (Taylor et al., 2022) or the LaMDA model by Google (Thoppilan et al., 2022) are essentially driven and funded by corporate players.

5 Discussion and Contributions

5.1 Five Waves of Conversational Agent Research

In this section we present our five wave view on CAs and CA research. First, we present the first age of CAs that represents the past of CAs. Second, we present the second age of CAs that reflects the present and future of CAs and CA research. We emphasize on the second half and elaborate on current and novel frontiers for CA research. We present our five-wave view of the historical development of CA research, the current status quo, and a future projection based on our research results in Fig. 5.

Regarding RQ1, we studied how research around CAs developed over time. In this sense, we identified key publications established around CAs. As the term “chatbot” is already disclosed, the first research contributing to the understanding and exploration of CAs was concerned with the humanness of machines and how interaction with such machines resembles and differentiates one from human interaction. Remarkable contributions to that time were

those by Turing and Weizenbaum when the first CAs like ELIZA came into existence. By examining past research, we can observe that CA research has undergone certain incremental steps that most likely have been enabled by technological advancements. We call these incremental or evolutionary steps in CA research “the waves of CA research”.

5.1.1 The Past of Conversational Agents

Circling back to the very beginning of CAs, we can observe that the emergence of CAs focused on rather crude artifacts with predetermined and programmed structure. In this regard, the emergence of CAs, which is rooted in the very first, simple, and rule-based CAs like ELIZA, would mark the first wave of CA research — the “zero hour” wave. This first wave of CAs was concerned with achieving a computer program that can resemble a conversation with a human under specific conditions that are predetermined. The first wave had strict technical limitations due to the computational capabilities of that era and therefore couldn’t produce more advanced artifacts. However, with the technological advancements that followed the decades after ELIZA and the current technological developments, more waves have followed and will follow.

The second wave of CA research presented us with more advanced CAs that, for the first time, made use of NLP and statistical methods. Furthermore, first CAs appeared that could somewhat understand and resemble human emotions via scripted dialogue. In general, the second wave

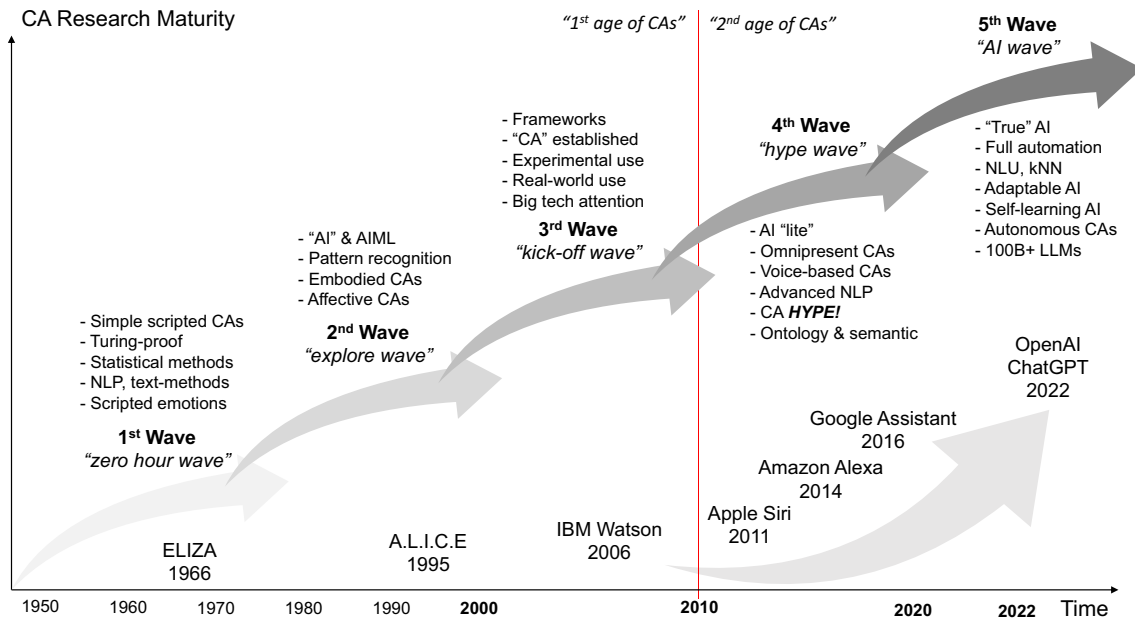


Fig. 5 The Five Waves of Conversational Agent Research

resembles what we describe as the explorational stage of CA research; hence, the “explore” wave. In this regard, the first CAs from Jabberwacky also started to aim at natural, enjoyable, and even hedonistic dialogue that today’s smart CAs have inherited (Lee et al., 2020). Moreover, with the second wave, methods such as pattern recognition and the first simple AI solutions made their wave into CA research. This led to dedicated languages, and somewhat more advanced CAs entering the playing field. The most prominent example from that era is the A.L.I.C.E. (release 1995) and the special language it was programmed with, the so-called artificial intelligence markup language (Wallace, 2009). Based on the former advancements in technology leading toward the first AI applications, the focus of this wave was on developing CAs based on first AI solutions. However, although this wave brought first attempts at AI in CA research, AI itself was still in a very early stage due to technological limitations and the virtual non-existence of voice-based CAs. In addition to these developments, the first embodied CAs with anthropomorphic features emerged and became somewhat popular. Although best efforts to humanize CAs were made (Porra et al., 2020), because of technological limitations, this wave of CAs never managed to overcome the uncanny valley of CAs, thus marking a significant gap in research and practice that in some areas persists until today.

With the third wave of CA research, the entire topic of CAs gained significant momentum enabled by the technological advancements in the 2000s and the mid-2000s, particularly effectively launching the CA developments that we still witness today; hence, we describe

this wave as the “kick-off” wave. This kick-off is reflected in both the development of CAs and related technology (e.g., frameworks, implementations) and the emergence and distribution of keywords. For instance, the keywords “AI,” “chatbot,” and “human-computer interaction” started to emerge in 2000–2005. Regarding “AI,” although the first approaches in CA research existed before 2000 (e.g., AIML), none of these represented “true” AI capabilities that first started to emerge about 2005—at least in the domain of CA research. This development translates to the emergence of further technologies and implementations, such as IBM Watson in 2006. With the release of Watson, IBM became the first big tech company to release a sophisticated product in the domain of CAs. That development started to draw significant attention from other big-tech companies in the following years. This also marked a significant milestone in the domain of CA research, which one may see as the steppingstone into the realm of advanced CAs that are omnipresent today (e.g., Alexa). Another important keyword that emerged in this wave of CA research is “multimodality”. The emergence of this keyword at that time highlights that what should come after, which is what we see to be omnipresent today voice-based and multimodal CAs.

5.1.2 The Present of Conversational Agents

With the revolution of the mobile phone market (i.e., the release of the Apple iPhone in 2007) and general consumer electronics that followed during the 2010s and the technological developments that came with it, the

steppingstone for the fourth wave of CA research was laid out. In the 2010s, many advanced CAs enabled by significant advancements in technology, particularly in AI and NLP, became mainstream and even omnipresent in our everyday lives. Moreover, traditional text-based CA applications have started to shift to voice-based CAs. CAs enjoyed a real hype from academia to the corporate that swept over into the real world. First, AI “lite” CAs emerged that made use of more than simple rule-based, statistical, or light “AI” methods. These implementations of CAs were able to use a wider variety of natural language, including first natural language understanding (e.g., the meaning behind the language) and communication modalities (e.g., first voice-based CAs, voice versus text (Rzepka et al., 2021)), in contrast to their predecessors. These voice-based CAs made them tangible to the broader population, given that they can be operated without the need to type in the text. This in return led to a vast increase in popularity and, subsequently, further development, research, and financial investment into these agents, as not only researchers but also big tech players like Google or Amazon realized the potential. Nowadays, many people cannot imagine a life without CAs like Amazon’s Alexa or Apple’s Siri (Kendall et al., 2020).

Looking back at our keyword analysis, some keywords, in particular, can be observed to have skyrocketed in the 2010s, specifically since 2015. The keywords “chatbot” and “conversational agent” highlight the real hype CA research received from 2015 onwards. Moreover, from a more technical point of view, the keywords “AI,” “NLP,” and “machine learning” showcase what future road CA research may take. Particularly, the keyword “deep learning” that first emerged in CA research after 2015 emphasizes that potential AI-route CA research may take in the following years of the current decade. Overall, machine learning and AI seem to be two of the core topics of current wave of CA research and the current as generation CAs as well as novel CA artifacts (Suta et al., 2020).

This development is reflected in the current technologies being used for developing CAs. For instance, the conversational AI framework Rasa (Bocklisch et al., 2017), that was recognized by Gartner to be one of the key providers and technology offerings (Revang et al., 2022), can make use of the advanced capabilities mentioned above. In this regard, Rasa can be build to include spaCy (Honnibal et al., 2020) and BERT (Devlin et al., 2018) for its’ natural language pipeline. These technologies allow Rasa to overcome the limitations of past frameworks that failed to understand natural language to a large extend and heavily relied on click-based solutions. This development also highlights the importance of NLU technologies for the present and future of CAs and CA research as well as the quickly and vastly growing capabilities of these technologies that enable next

generation CAs and thus further pushing the frontiers of CA research.

Following up on these technological advancements and continuously evolving CAs academics and professionals alike have to wonder where the journey leads and how they address the upcoming challenges or leverages the opportunities. Concluding the fourth wave of CA research, we have observed a stronger focus on natural language understanding and the emergence of novel and advanced voice-based CAs. Based on our research, we project that this development will continue and discuss a potential future route and their frontiers for the fifth wave of CA research next.

5.1.3 The Future of Conversational Agents

As we look at the frontiers of CA research and developments, we especially have to look at the future of CAs. Building upon the four distinct waves of CA research, we outline the future of CAs as the fifth wave that may bring even more remarkable changes for the interaction of CAs and also how we interact with Information Systems in general. Considering the development of keywords over time, we predict that topics around “true” or “general” AI will significantly gain in importance. These topics about true AI include full automation and autonomous CAs, which may be enabled by future advancements toward self-learning and real-time adaptable AI applications. Current examples include for instance the Google conversational technology “LaMDA”² (Language Model for Dialogue Applications Thoppilan et al., (2022) that provides through training of a transformer model on conversations new capabilities for natural language interactions. We specifically point out the frontier to a “general” AI as Google LaMDA gained wide media coverage through a Google developer who claimed that LaMDA is sentient and, in essence, is a self-aware person (Thoppilan et al., 2022).

Moreover, the actual understanding of natural language and the intentions behind it (i.e., what humans mean vs. sarcasm) may become more important and powerful, based on advanced deep learning and NLU methods (Suta et al., 2020). Especially the development of generative AI and according large language models bring in beta versions of the future of CAs. Most remarkable is the release of the beta version of OpenAI’s ChatGPT based on further developments of GPT (Stokel-Walker, 2022; Brown et al., 2020b) end of November 2022, called by experts as the most disruptive AI technology in 2022 and the future (Haque et al., 2022). This technology is enabling a wide range of tasks during a conversational interaction, such as writing essays (Sun et al., 2022), helping with coding

²<https://blog.google/technology/ai/lamda/>

(Castelvecchi, 2022) and other creative tasks (Davenport & Mittal, 2022). A widespread adoption of these developments would potentially enable CAs to become even more human-like, to the point where humans may mistake a CA for a real human as in the above introduced example of Google LaMDA (Thoppilan et al., 2022). We can observe this trend quite clearly when looking at the development of certain keywords, particularly “AI,” “machine learning,” “natural language processing,” “neural network,” and “deep learning.” All these keywords showed either a huge spike (i.e., hype) in the last five years or continuous steep development.

In addition, topics like embodiment and virtual agents, including avatars and other anthropomorphic approaches, may increase in significance as our ability to model human-like digital artifacts develops. For instance, AI-driven intelligent embodied CAs may become more popular and, with increasing advancements, blur the lines between CAs and robot artifacts (Huang & Rust, 2021). This development is especially fueled through the development of “text-to-output” models such as Open AI’s DALL·E 2 (Ramesh et al., 2022) or other latent diffusion models e.g., Rombach et al. (2021). Interfaces building upon machine learning pipelines that combine sophisticated CA interfaces, such as ChatGPT (Stokel-Walker, 2022), with avatar generators could enable in the future fully customizable CAs with a previously unmatched degree of social presence and anthropomorphism. By these means, adaptability, personalization, and individualization of human-computer interactions via CAs will become more significant (Abedin et al., 2022; Behera et al., 2021). Future developments might include in this context that interactions are not only based on historical data but also real-time data. For instance, the ChatGPT (Stokel-Walker, 2022) CA is trained on a large language model with a cut-off point end of the year 2021. Recent developments around generative AI and LLM illustrate that, just like AI, CAs and their technical sophistication are a moving target. Moving from the niches of computer science to standard technology, users’ expectations regarding CAs’ capabilities will continuously increase. In turn, interacting with CAs will become even more easily accessible, effortless, and integrated in humans’ daily personal and work life.

5.2 Implications for Conversational Agent Research

Although the discussion around the personification and design of CAs frontend remains crucial (Elshan et al. 2023), technology-related topics have become the center of innovative current wave CA research. Advancements in technological models and understanding thereof have made contemporary CAs sophisticated and commercially available to the large public. Public omnipresence, in turn, has introduced tech corporations into the academic research

field, with both of them becoming increasingly intertwined. This is also highlighted in the predominance of our affiliate results and influences research streams with corporate interests.

With RQ2, we aimed to shed light on the current understanding and focus of the predominant research streams studying CAs. Our keyword clusters illustrate the dominance of certain types of CAs, including chatbots and embodied CAs. Interestingly, certain tasks, contexts, and technical models are commonly associated with different terms. These findings either indicate the relevance of certain CAs for particular domains or point toward the fragmentation of CAs across research outlets. While research into voice-based CAs has increased over the past few years, it has not yet been established as research around chatbots.

A better understanding of the distinguishing features of certain modalities and the interplay of certain modalities is required to arrive at a more nuanced understanding of the different types of CAs (Rzepka et al., 2021). In a similar vein, initial research driven by the CASA paradigm and the Turing test found great prevalence in contemporary studies. Our keyword results show that the research focus on CA personalization and anthropomorphization to create a more natural and effortless interaction can, in turn, affect important outcome variables, including usage, satisfaction, and trust (Rheu et al., 2021; Abedin et al., 2022). Trust is also one of the key issue of research on CAs (McTear et al., 2016; Vössing et al., 2022). Considering the vast developments about machine learning, AI, and imprudent exploitation of user data by large corporations, we find that future research should address matters of privacy, trust, and trustworthiness of contemporary CAs.

In this regard, we assume that topics like explainable and ethical AI will gain importance in the future which is supported by recent research (Abedin et al., 2022). Questions of how a CA’s underlying model behaves and affects users become even more important. While a system’s behavior based on machine learning was largely shaped by its input data, large language model introduce a novel facet of opacity to the user. Although we cannot support this with our research results, given that this was not an explicit focus of ours, other research highlights the increasing importance of explainable and ethical AI, particularly in the context of human-computer interactions and CA research. However, what does this mean for the long-term development of CA research? In particular, the underlying functions and interrelatedness within a larger commercial ecosystem are becoming increasingly opaque. What are apparent, understandable, and important to the user are notions of trustworthy system design (transparency, explainability, privacy) and relevance of user trust? We see this already becoming a topic of most recent research e.g., Vössing

et al. (2022) and Abedin et al. (2022). Future research will need to investigate the longitudinal outcomes, habits, and dynamics of users about trust in CAs and the underlying AI, especially when looking at current developments that strongly emphasize data-driven methods for CAs, as our keyword results showcase. Nonetheless, given that CAs, such as Amazon's Alexa, are currently lacking profitable use cases (Mattioli et al., 2022), implications arise that aspects of transparency in conversational interactions should be carefully addressed. For instance, CAs are able to provide recommendations to products in conversational use processes (Schwede et al., 2022), however, ethical questions regarding the transparency and trustworthiness remain when a CA casually during interactions nudges towards product choice (Benner et al., 2022). At the same time, users' expectations regarding the capabilities of CAs' will increase and the need for explainable AI and transparency potentially decrease, especially from a user's perspective. In addition, regardless of whether organizations decide to acquire or build their own CAs, the responsibility lies with them. As a result, we expect to see a shift of research emphasis from trust in the system towards trust in the provider.

Nevertheless, technological changes represent not only challenges like trust but also opportunities to further increase natural and effortless interaction with CAs. However, the underlying technical properties and the conversational nature of CAs are still fairly rarely considered in CA research. As our keyword and cluster analysis revealed, most current CA research focuses on only one specific topic (i.e., dialogue structure, anthropomorphism, and technical models underlying CA). Looking at related studies (see Appendix B), we find that most research is focused on literature only from the year 2000 or newer and does not include more than 100 articles. We find only one comparable study from Io and Lee (2017), who analyzed approximately 4200 articles, however only from the year 1998 onwards. All other studies we find to be comparable have a more traditional literature review approach and seem to focus on some specific topic. For instance CAs in business, collaboration setting or specific characteristics of CAs (e.g., Poser et al. (2022)).

Another important topic we considered when analyzing the domain of CA research was the interplay of academia, research, and the corporate world. We can observe that during the early 2000s, the development of CAs was still driven mainly by academia and independent research. However, this strongly shifted during the 2010s with big corporations (e.g., Google) becoming major affiliates in the domain of CA research. Considering the past developments and current affiliate status in the last year (2020), we predict that affiliations with big-tech corporations will continue to increase. Furthermore, with the increasing interest of Chinese players (e.g., Baidu), we may see a shift toward Chinese affiliations during the next years.

However, based on our observations, American-based corporate affiliates are unlikely to lose their dominant position in the market and CA research. Although our analyses on commercially-funded publications could not consider all potential sources and thus should be interpreted with caution, they provide a first glimpse into emerging sectors. These results call into question which players are and will drive research on CAs. The commercial availability of CAs, as well as the reliance on these commercially available CAs within empirical research (i.e., experimental, laboratory, ethnographic, or diary-based studies), illustrate that companies will have an influential stake in driving and shaping CA research. This is right now also visible through the development that underlying technical papers on large language models such as GPT were previously published (Brown et al., 2020b) but technologies such as ChatGPT now remain proprietary. On the opposite, initiatives such as BigScience (2022) try to counteract this trend to provide open language models such as BLOOM (BigScience, 2022b) as the basis for novel CA developments.

Concerning the academic developments of CA research, regardless of affiliations, most technical advancements have been empirically investigated in the context of a text-based CA. A voice-based CA finds little appearance in our keyword cluster, which may be the result of a certain choice of dominant keywords or voice-based agents not having been researched as extensively as text-based ones. As explained in our five-wave view on CA research, this observation reflects a natural development in CA research, and we expect voice-based CA research to become more popular and potentially become the modality of choice in the future (e.g., Rzepka et al. (2021)). The CA does not necessarily seem to act as an overarching term for the different types of agents. Certain tasks and contexts are connoted with certain terminology.

Furthermore, we want to highlight the development of keywords again, particularly for some keywords which exhibit erratic appearance over time. Considering the steep and steady increase of certain keywords like “embodied conversational agents” and “virtual agent,” CA research seems to show an increasing interest in further anthropomorphizing CAs. This is also reflected in the increasing appearance of keywords like “affective computing” and “emotion,” which both aim to make CAs more human-like by touching on the very human subjects of interpersonal human-human interactions. Similarly, we can observe a real hype of AI and machine learning-related keywords from 2015 until today. Considering concepts like the Gartner hype cycle, we assume that this development represents a novel hype of AI-related technology that we predict will grow even further in the future, given that keywords on these topics show a very strong near-linear increase without deviation. Considering the interest of big

corporate techs in AI and related topics that oftentimes act as affiliates of CA research, we predict this development to be at the core of the fifth wave of CA research toward true AI-driven CAs.

Overall, with our RQs, we intended to clarify the origin, development, status quo, and future of CAs. Thus, we contribute to theory by providing a more nuanced and holistic understanding of the types of CAs and summarizing an overview of waves of the development of CA research and the differentiation and development of keywords into clusters. In doing so, we have highlighted the past, current, and potential future patterns in CA research and the implications that come with the evolutionary development of CA research. From a practical point of view, we highlighted the shift in the design of CAs and a relevant, related technical model and provided an overview of relevant topics to consider when designing and deploying CAs. In particular, we highlighted the noticeable shift toward AI-driven CAs that may incorporate sophisticated self-learning AI techniques for a more autonomous and human-like interaction (e.g., through better natural language understanding, affective computing, and anthropomorphizing).

5.3 Implications for Potential Actionable Directions

The implications for CA research can be translated to actionable directions for practice. To illustrate the implications and describe potential future frontiers of CA research we have compiled a table describing these based on our findings (see Table 4).

First, the technological developments of CA research and also advancements in related fields like AI have raised the bar for technological requirements and are likely to continue this trend. This requires researchers and corporations alike to adapt to the increasing technological demands and expand their resources in order to support novel technologies for novel CA artifacts like AI-driven CAs. We can already witness this development with the advancements that OpenAI and GPT have presented us with. However, looking closer at OpenAI it is evident that this corporate player is not only interested in developing frontier level technology for CA research but also in monetizing it. Such strategic decisions might have larger implications on the CA landscape, including organizations deciding whether to “make or buy” generative AI-based CAs. Considering how past technological advancements have been used by corporate players and monetized, this may raise the concern of the technology going proprietary and paid. This could become a hindrance in further advancing the frontiers of CA research as many researchers or practitioners may be disinclined to comply with these terms.

Considering the raised implications for CA research, CA artifacts incorporating AI should focus on transparency and trust in an ever increasing environment of omnipresent data e.g., Elshan et al. (2022) and Vössing et al. (2022). Data not only allows AI to deliver AI-based core functionality including improved natural language processing and understanding capabilities but also adaptability. For instance, data-driven approaches can be used to individualize and personalize CAs in order to provide an improved experience and performance (Behera et al., 2021). However, personalization and improved experience based on user data must be transparent with its use and explain what data is used when and for what purpose (i.e., explainable AI). This is a trend we can already partially observe in our five waves of CA research presentation. With regard to individualization and personalization of CAs, AI is only one relevant aspect, another relevant aspect is the design of such AI-driven artifacts (Behera et al., 2021).

This, however, has different implications for each type of CA as well as the types of interaction users can have with CAs (e.g., text-based, voice-based, embodied) (Nguyen et al., 2021). The intricacies of each type of CA will have to be detangled for future generations of potentially AI-driven individualized CAs and their designs and user interfaces (Abedin et al., 2022). Particularly with regard to potential challenges with trust or ethical questions and AI, as well as bias in the underlying technologies (e.g., machine learning algorithms). In this regard, characteristics like natural language, tonality (voice only), visual representation (embodied only) and potentially novel emerging design aspects will require deep dive investigation on their own, particularly in context of the raised potential issues and how to design explainable, ethical and unbiased CA artifacts (Abedin et al., 2022; Benner et al., 2022). Last, we expect the questions of how CAs (will) behave and their implications for how humans interact, work, and rely on external advice, to be more important than ever. Thus, we want to encourage fellow researchers to consider our implications for CA research and raised actionable directions.

Overall, there are multiple actionable directions for future research for the current wave of CA research as well as the next generation of CAs and followingly the next wave of CA research.

6 Limitations, Future Research and Conclusion

In this article, we have compiled, organized, and analyzed a vast body of literature on CAs in the IS domain and related fields. In doing so, we have highlighted how research has developed from the very first CA artifacts (e.g., ELIZA)

to current state of the art CAs like Alexa. We have described the past and present CA research (RQ1 and RQ2). Further, we have highlighted potential future research streams according to our keyword analysis, pointing out what topics are currently state of the art and what will be the focus of research in the future (i.e., next developmental generation of CAs), thus answering RQ3. Here, we have pointed out that future research may likely see a strong focus on data-driven and AI-driven approaches, as highlighted by our keyword analysis and five-wave view on CA research.

In this context, we have highlighted the developmental waves of CA research from the beginning (i.e., the 1960s) to our current time and what could be the next big wave for CA research (i.e., data and AI). In addition, we have pointed out some limitations in the existing body of research on CAs and the importance of addressing these potential gaps. However, these research gaps should be seen as opportunities, as highlighted by the development of keywords leaning toward AI-driven methods. Moreover, we have briefly highlighted some comparative studies with a similar yet different focus and contribution to provide a more holistic overview of the past, present, and future of CA research.

Although we conducted our research according to established research guidelines and used established statistical methods that we followed as rigorously as possible, our research may not be without limitations. First, although we did our best to include as many relevant articles as possible by choosing the appropriate keywords and outlets and not using limitations like publication year restrictions, we may have missed some niche domain or research stream. Moreover, we may not have included literature that uses more specific keywords exclusively. An example of this would be research that uses "smart personal assistant/agent" as keywords. Second, only two of the authors reviewed the literature in detail and made decisions to include or exclude the literature. Reviewing the articles with more or other reviewers may have led to different results and, thus, a somewhat divergent body of literature.

Overall, our research highlights the past, present, and potential future development of CA research covering implications for research and practice. Both partial contributions come together in our presented five-wave view on CA research, thus highlighting the origins of CAs, their development until today, and potential future developments with the next big wave (i.e., data- and AI-driven CAs). On the one hand, our research contributes to theory by highlighting this development and pointing out our potential future research venues based on our keyword analysis, derived five

waves of CA research and the implications of these findings. On the other hand, our research contributes to practice by highlighting the technical developments based on our keyword analysis and the indicated affiliations with the big-tech corporate world that currently drives CA research in various areas, particularly AI-related topics. Moreover, our five waves of CA research highlight the current status quo and ongoing development future that can form the basis for future CA developments and directions. Here, we find AI and relevant topics related to it (e.g., trust in AI, explainable AI) to be very likely to continue the throughout the next wave.

In general, we hope that our research provides both researchers and practitioners with an exhaustive and holistic overview of CA research with an emphasis on future developments and research opportunities that should and most likely will be addressed within the current decade when considering the fifth wave of CA research.

Appendix

A: Descriptive Results

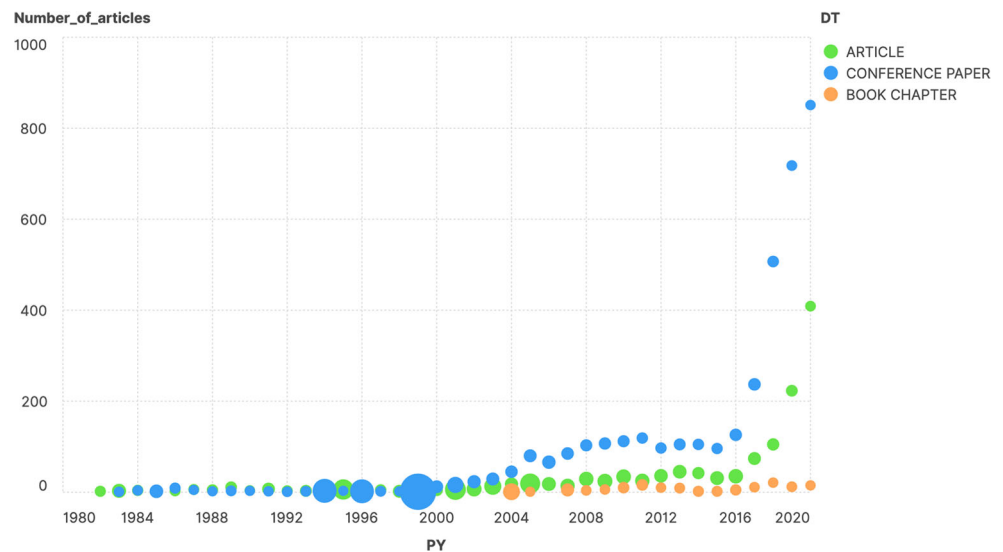
In our study, we present articles published from 1982 to 2021. A summary of the main information is given in Table 2.

Table 2 Summary of General Statistics of Bibliometric Analysis

General Data	
Sources (Journals, Books, etc.)	1,920
Documents	5,106
Average years from publication	5.81
Average citation per document	8.047
Average citation per year per document	1.237
Document Types	
Articles	1,290
Book chapters	130
Conference papers	3,686
Document Contents	
Keywords plus (ID)	14.801
Authors' keywords (DE)	8.327

An overview of the annual scientific production over the years and the average article citations per year is shown in Fig. 6.

Fig. 6 Annual Scientific Production and Citations. (*The size of the circle is proportional to the number of citations*)



Appendix

B: Related Work

Table 3 Summary of Related Literature Reviews and Bibliometric Studies

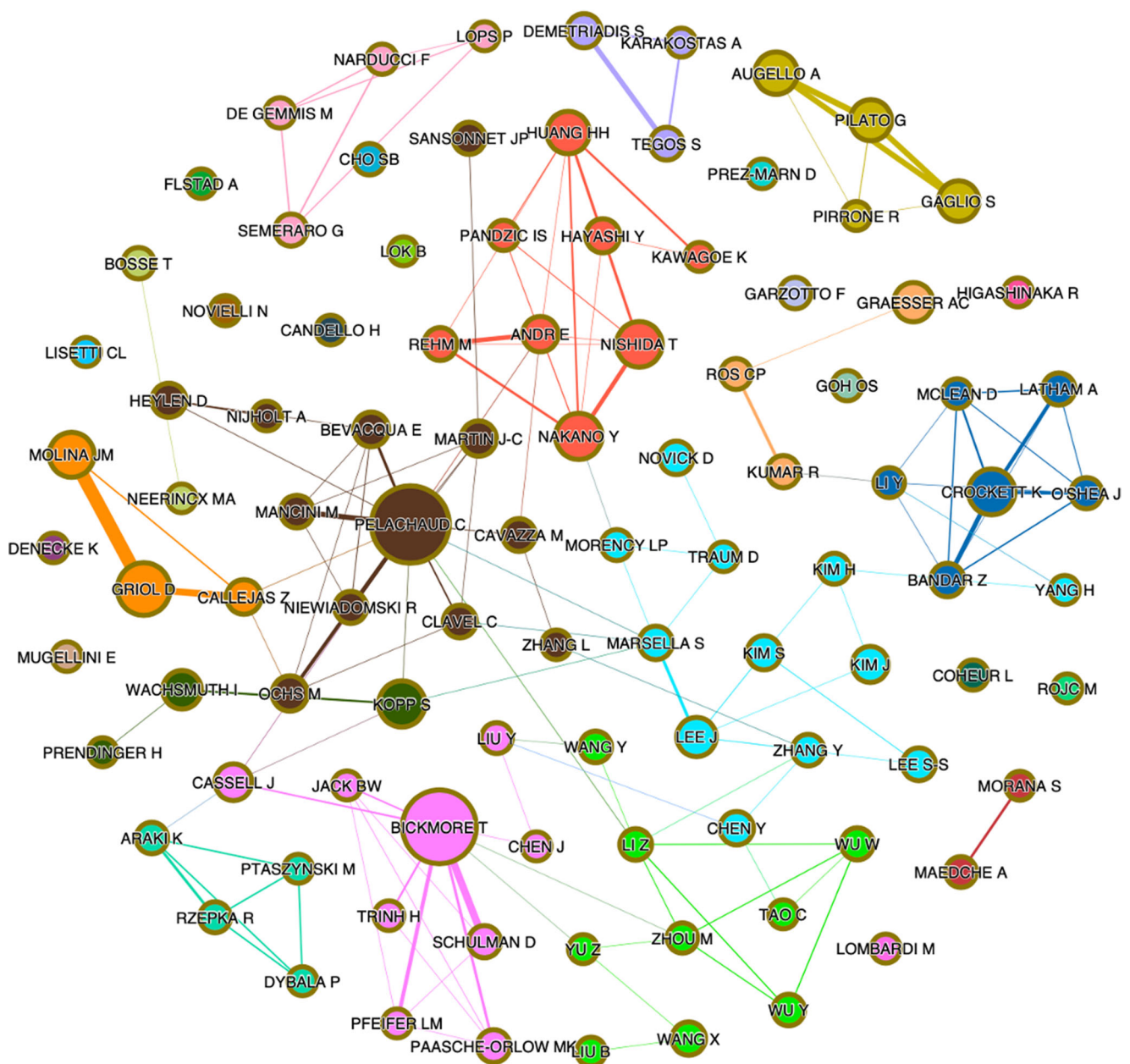
Author (Year)	Method (Sample Size)	Review Period	Focus of Study	Research Implications
Chaves and Gerosa (2021)	Literature review (56 articles)	Up to 2018	Analysis of social characteristics of disembodied chatbots in the HCI context	Analyze if and how moral agency improves communications with bots Analyze social characteristics of embodied chatbots
Diederich et al. (2019)	Literature review (36 articles)	Up to 2018	Discussion of status quo in research involving CAs	CA in team collaboration settings Conduct more design-oriented studies Explore embodied agents Study CA introduction and use in the field
Diederich et al. (2022)	Organizing and assessing review (262 articles)	Up to 2019	Analysis of CA design and research streams	6 research avenues for CA in IS 16 actionable directions for research Call for longitudinal studies
Elshan et al. (2022)	Literature review (107 articles)	No limits	Understanding of User Acceptance of CAs and future research agenda	Conflict of adaptability and standardization Investigation in freedom and autonomy needed Investigating effects of CA components
Rheu et al. (2021)	Literature Review (29 articles)	Up to 2019	Discussion and analysis of trust supporting components	Research should focus on optimization Analyze role of trust and user expectations Optimization and trust in AI by design
Rzepka and Berger (2018)	Literature Review (96 articles)	Up to 2018	Characteristics of adoption of AI-based CAs	Beneficial degree of anthropomorphism Differences in user behavior towards AI Effect of transparency and AI on interaction Negative interactions between user and AI
Smutny and Schreiberova (2020)	Field Analysis (89 CAs)	No limits	Analysis of educational CAs on Meta	Support CA integration in education Content analysis of CAs with users
Van Pinxteren et al. (2020)	Literature review (61 articles)	From 1999 to 2018	Effect of CA behavior on outcomes for service encounters	Analyze user needs and group preferences Analyze categories of communicative behavior Analyze (non-)verbal behavior relations

Table 3 (continued)

Author (Year)	Method (Sample Size)	Review Period	Focus of Study	Research Implications
Io and Lee (2017)	Bibliometric study (61 articles)	From 1998 to 2017	Analyze the past of CA research	New trends to analyze role of CAs Change way of CA development Analyze business roles and use of CAs

Appendix

C: Network Graphs

**Fig. 7** Author Network

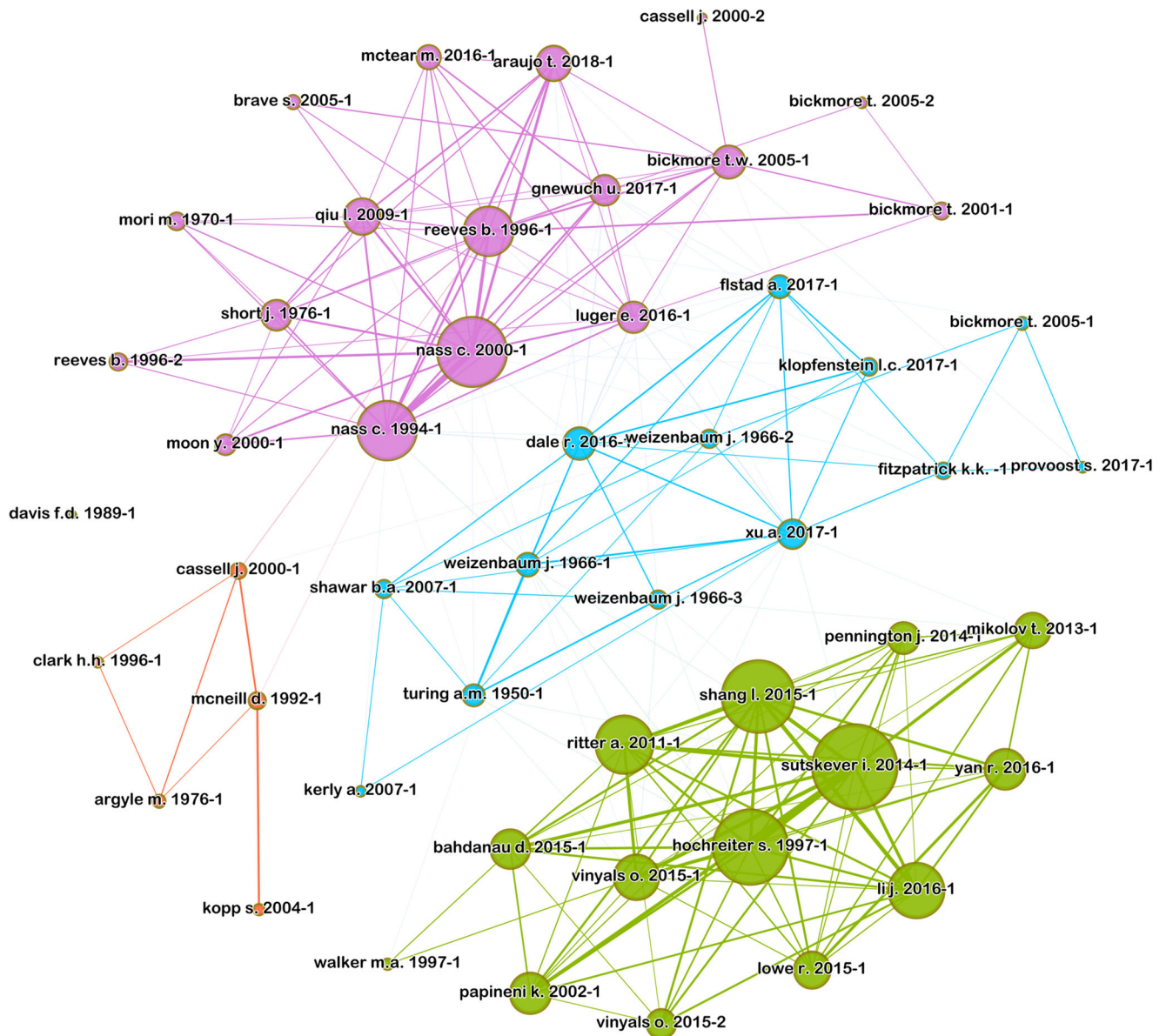


Fig. 8 Co-citation References

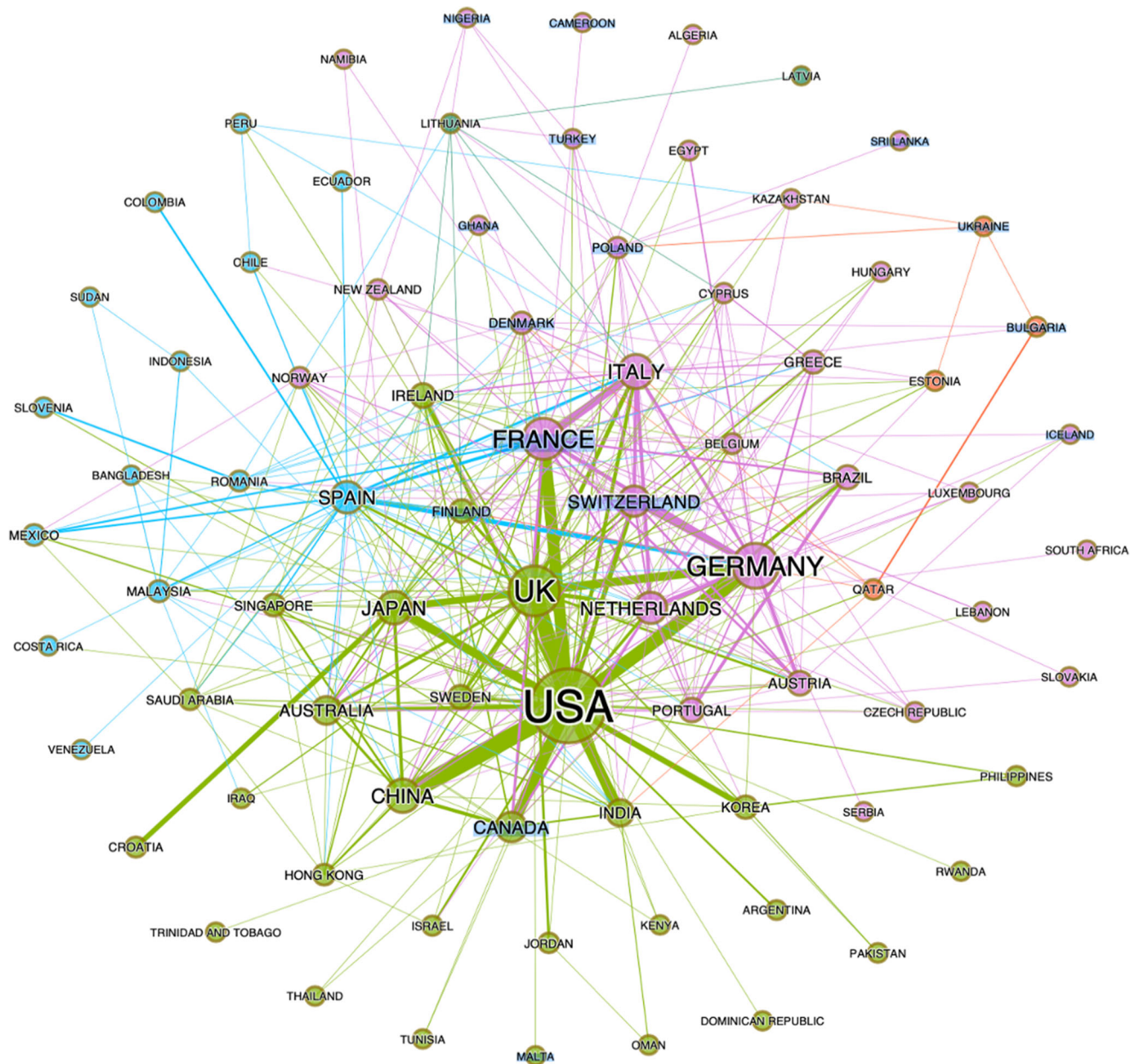


Fig. 9 Country Network



Fig. 10 Keyword Cluster

Appendix

D: Future Frontiers for CA Research

Table 4 Proposed Frontiers for the Future of CA Research

Frontier	Current Status and Directions for Future	
Theory	Status quo:	Many past and current studies rely on old theories not native to CA research (e.g., social presence (Short et al., 1976), computers as social actors (Nass et al., 1994))
	Directions:	We call for native CA research theories that are based on and adapted to the boundaries of human-agent interaction. Researchers may draw on the existing theories but should advance CA research theory towards a stand-alone theoretical base.
Technical	Status quo:	Advanced natural language models and machine learning algorithms are being used but no “true” AI
	Directions:	Research should focus on establishing a “true” conversational AI. Fundamental technical advances have been made, especially with large language models (e.g., BLOOM, GPT) but need to be developed towards true human natural language and especially human thinking including social features e.g., social cues, humor, sarcasm (Allouch et al., 2021).
Design	Status quo:	Most current CAs do not include personalized features, trust or explainable AI features or feature believable humanization
	Directions:	Researchers and developers need to find a way to personalize CAs but at the same time respect ethical boundaries and establish trust in the CA and the underlying mechanisms. In this context persuasive design features may provide a solution, however their boundaries must be respected (Benner et al., 2022)
General	Status quo:	Most CA software, programs, libraries, models, etc. are free and open source. This fosters CA research and enables independent researchers to contribute.
	Directions:	While monetarization is important for corporate players, restriction of access may have serious adverse effects and hinder development and contributions by independent researchers and developers. Therefore, corporate players must find a balance between monetarization and open access
Managerial	Status quo:	Many companies use CAs but most being simple clickbots or less advanced CAs that may even hurt the interaction.
	Directions:	Managers should transform their customer interaction and invest on CAs with actual conversational capability to improve user experience and their market position (Abedin et al., 2022).
Society	Status quo:	CAs already have a significant impact on society which is expected to extend.
	Directions:	Emergent fields of application for CAs as well as transforming existing fields and pushing the boundaries may become even more significant with increased societal impact (Wahde & Virgolin, 2022)

Acknowledgements The authors acknowledge partial funding of the following funding bodies: “Stiftung Innovation in der Hochschullehre” within the project “Universität Kassel digital: Universitäre Lehre neu gestalten”, Swiss National Science Foundation (project ID: 100013_192718), Academy of Finland (Suomen Akatemia) Research Council for Natural Sciences and Engineering for the project Towards precision education: Idiographic learning analytics (TOPEILA), Decision Number 350560, and the fifth author acknowledges personal funding from the Basic Research Fund (GFF) of the University of St. Gallen.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Authors two and three have contributed equally. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

References

- Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-ai interactions. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-022-10313-1>.
- Adam, M., Wessel, M., & Benlian, A. (2020). Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, pp. 1–19.
- Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022). Designing personality-adaptive conversational agents for mental health care. *Information Systems Frontiers*, pp. 1–21. <https://doi.org/10.1007/s10796-022-10254-9>.
- Allouch, M., Azaria, A., & Azoulay, R. (2021). Conversational agents: goals, technologies, vision and challenges. *Sensors (Basel Switzerland)*, vol 21(24). <https://doi.org/10.3390/s21248448>.
- Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). Natural language interfaces to databases—an introduction. *Natural Language Engineering*, 1(1), 29–81.
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: an r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4), 959–975.
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks.
- Behera, R. K., Bala, P. K., & Ray, A. (2021). Cognitive chatbot for personalised contextual customer service: behind the scene and beyond the hype. *Information Systems Frontiers*:1–21. <https://doi.org/10.1007/s10796-021-10168-y>.
- Benner, D., Schöbel, S., Janson, A., & Leimeister, J. M. (2022). Propositions for ethical persuasive design in information systems. *AIS Transactions on Human-Computer Interaction*, vol 14(4).
- Bickmore, T., Asadi, R., Ehyaei, A., Fell, H., Henault, L., Intille, S., Quintiliani, L., Shamekhi, A., Trinh, H., & Waite, K. (2015). Context-awareness in a persistent hospital companion agent. *International Conference on Intelligent Virtual Agents*, pp. 332–342.
- Bickmore, T., & Cassell, J. (2001). Relational agents: a model and implementation of building user trust Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 396–403.
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293–327.
- Bickmore, T., Rubin, A., & Simon, S. (2020). Substance use screening using virtual agents: towards automated screening, brief intervention, and referral to treatment (sbirt). *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pp. 1–7.
- BigScience (2022). BigScience research workshop. <https://bigscience.huggingface.co/>. Accessed 22 Dec 2022.
- BigScience (2022b). BLOOM: a 176B-parameter open-access multi-lingual language model.
- Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). Rasa: open source language understanding and dialogue management. NIPS 2017 Conversational AI workshop.
- Brandtzaeg, P. B. (2018). Chatbots: changing user needs and motivations. *Interactions*, 25(5), 38–43.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodi, D. (2020b). Language models are few-shot learners. arXiv:2005.14165.
- Callejas, Z., Griol, D., & López-Cózar, R. (2011). Predicting user mental states in spoken dialogue systems. *EURASIP Journal on Advances in Signal Processing*, 2011(1), 1–21.
- Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., & Stone, M. (1994). Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, pp. 413–420.
- Castelvecchi, D. (2022). Are chatgpt and alphacode going to replace programmers? *Nature*. <https://doi.org/10.1038/d41586-022-04383-z>.
- Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? a survey on social characteristics in human-chatbot interaction design. *International Journal of Human-Computer Interaction*, 37(8), 729–758.
- Dale, R. (2016). The return of the chatbots. *Natural language engineering*, 22(5), 811–817.
- Davenport, T., & Mittal, N. (2022). How generative ai is changing creative work. Harvard Business Review.
- De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized louvain method for community detection in large networks. *International conference on intelligent systems design and applications*, pp. 88–93.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

- Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019). On conversational agents in information systems research: analyzing the past to guide future work. *International Conference on Wirtschaftsinformatik*, pp. 1550–1564.
- 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 System (JAIS)*.
- 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*. <https://doi.org/10.1007/s10796-021-10230-9>.
- Elshan, E., Ebel, P., Söllner, M., & Leimeister, J. M. (2023). Leveraging Low Code Development of Smart Personal Assistants: An Integrated Design Approach with the SPADE Method. In: *Journal of Management Information Systems*. <https://doi.org/10.1080/07421222.2023.2172776>.
- 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.
- Fischer, J. E., Reeves, S., Porcheron, M., & Sikveland, R. O. (2019). Progressivity for voice interface design. *Proceedings of the 1st international conference on conversational user interfaces*, pp. 1–8.
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164.
- Følstad, A., & Skjuve, M. (2019). Chatbots for customer service: user experience and motivation. *Proceedings of the 1st international conference on conversational user interfaces*, pp. 1–9.
- Goudos, S. K., Tsoulos, G. V., Athanasiadou, G. E., Batistatos, M. C., Zarbouti, D. A., & Psannis, K. E. (2019). Artificial neural network optimal modeling and optimization of uav measurements for mobile communications using the l-shade algorithm. *IEEE Transactions on Antennas and Propagation*, 67, 4022–4031.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). Autotutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618.
- Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22(4), 39.
- Griol, D., Hurtado, L. F., Segarra, E., & Sanchis, E. (2008). A statistical approach to spoken dialog systems design and evaluation. *Speech Communication*, 50(8–9), 666–682.
- Hauswald, J., Laurenzano, M. A., Zhang, Y., Yang, H., Kang, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R. G., & Mudge, T. (2016). Designing future warehouse-scale computers for sirius, an end-to-end voice and vision personal assistant. *ACM Transactions on Computer Systems (TOCS)*, 34(1), 1–32.
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). I think this is the most disruptive technology: exploring sentiments of ChatGPT early adopters using twitter data. [arXiv:2212.05856](https://arxiv.org/abs/2212.05856).
- Heyselaar, E., & Bosse, T. (2019). Using theory of mind to assess users' sense of agency in social chatbots. *International Workshop on Chatbot Research and Design*, pp. 158–169.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). Spacy: industrial-strength natural language processing in python. <https://doi.org/10.5281/zenodo.1212303>.
- Huang, M.-H., & Rust, R. T. (2021). Engaged to a robot? the role of ai in service. *Journal of Service Research*, 24(1), 30–41.
- Io, H. N., & Lee, C. B. (2017). Chatbots and conversational agents: a bibliometric analysis. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 215–219.
- Kendall, L., Chaudhuri, B., & Bhalla, A. (2020). Understanding technology as situated practice: Everyday use of voice user interfaces among diverse groups of users in urban india. *Information Systems Frontiers*, 22(3), 585–605. <https://doi.org/10.1007/s10796-020-10015-6>.
- Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2021). Value co-creation in smart services: a functional affordances perspective on smart personal assistants. *Journal of the Association for Information Systems*, 22(2), 5.
- Kopp, S., & Wachsmuth, I. (2004). Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, 15(1), 39–52.
- Krämer, N. C., Iurgel, I., & Bente, G. (2005). Emotion and motivation in embodied conversational agents. *Proceedings of the symposium agents that want and like, artificial intelligence and the simulation of behavior*, pp. 55–61.
- Kushwaha, A. K., & Kar, A. K. (2021). Markbot – a language model-driven chatbot for interactive marketing in post-modern world. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10184-y>.
- Laban, G., & Araujo, T. (2019). Working together with conversational agents: the relationship of perceived cooperation with service performance evaluations. *International Workshop on Chatbot Research and Design*, pp. 215–228.
- Lee, K., Lee, K. Y., & Sheehan, L. (2020). Hey alexa! a magic spell of social glue?: sharing a smart voice assistant speaker and its impact on users' perception of group harmony. *Information Systems Frontiers*, 22(3), 563–583. <https://doi.org/10.1007/s10796-019-09975-1>.
- Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016). Deep reinforcement learning for dialogue generation. [arXiv:1606.01541](https://arxiv.org/abs/1606.01541).
- Luger, E., & Sellen, A. (2016). Like having a really bad pa the gulf between user expectation and experience of conversational agents. *CHI Conference on Human Factors in Computing Systems*, pp. 5286–5297.
- Mattioli, D., Herrera, S., & Toonkel, J. (2022). Amazon, in broad cost-cutting review, weighs changes at alexa and other unprofitable units. *The Wall Street Journal* (11.10.2022). Accessed 22 Dec 2022.
- McTear, M., Callejas, Z., & Griol, D. (2016). The dawn of the conversational interface. In *The conversational interface*. Springer (pp. 11–24).
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors in Computing Systems*, pp. 72–78.
- Nguyen, T. H., Waizenegger, L., & Techatassanasontorn, A. A. (2021). “don't neglect the user!” – identifying types of human-chatbot interactions and their associated characteristics. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10212-x>.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the web of science for coverage of the social sciences' literature. *Journal of informetrics*, 1(2), 161–169.
- Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: a comparison between full and fractional counting. *Journal of informetrics*, 10(4), 1178–1195.
- Porra, J., Lacity, M., & Parks, M. S. (2020). Can computer based human-likeness endanger humanness? – a philosophical and ethical perspective on digital assistants expressing feelings they can't have. *Information Systems Frontiers*, 22(3), 533–547. <https://doi.org/10.1007/s10796-019-09969-z>.
- 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*. <https://doi.org/10.1007/s10796-022-10265-6>.
- Purinton, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). Alexa is my new bff social roles, user satisfaction, and personification of the amazon echo. *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, pp. 2853–2859.
- Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems. *Journal of Management Information Systems*, 25(4), 145–182.
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP Latents. [2204.06125](https://arxiv.org/abs/2204.06125).
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). Prisma-s: an extension to the prisma statement for reporting literature searches in systematic reviews. *Systematic reviews*, 10(1), 1–19.
- Revang, M., Mullen, A., & Elliot, B. (2022). Magic quadrant for enterprise conversational ai platforms. Gartner Reports. Accessed 09 Jul 2022.
- 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.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-resolution image synthesis with latent diffusion models. [2112.10752](https://arxiv.org/abs/2112.10752).
- Rzepka, C., & Berger, B. (2018). User interaction with ai-enabled systems: a systematic review of is research. *International Conference on Information Systems*.
- Rzepka, C., Berger, B., & Hess, T. (2021). Voice assistant vs. chatbot – examining the fit between conversational agents’ interaction modalities and information search tasks. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10226-5>.
- Schmitt, A., Wambsganss, T., Janson, A., & Söllner, M. (2021). Towards a trust reliance paradox ? exploring the gap between perceived trust in and reliance on algorithmic advice. In *Forty-second international conference on information systems, Austin, Texas* (pp. 1–17).
- Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a contemporary frontier of interaction design. *European Conference on Information Systems*.
- Schwede, M., Zierau, N., Janson, A., Hammerschmidt, M., & Leimeister, J. M. (2022). I will follow you! - how recommendation modality impacts processing fluency and purchase intention. *International Conference on Information Systems*.
- Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2022). Voice in human–agent interaction. *ACM Computing Surveys*, 54(4), 1–43. <https://doi.org/10.1145/3386867>.
- Sedik, A., Maleh, Y., Banby, G. M. E., Khalaf, A. A. M., El-Samie, F. E. A., Gupta, B. B., Psannis, K., & El-Latif, A. A. A. (2022). Ai-enabled digital forgery analysis and crucial interactions monitoring in smart communities, technological forecasting and social change. *IEEE Transactions on Antennas and Propagation*, vol. 177.
- Shang, L., Lu, Z., & Li, H. (2015). Neural responding machine for short-text conversation. [arXiv:1503.02364](https://arxiv.org/abs/1503.02364).
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of web of science, scopus and dimensions: a comparative analysis. *Scientometrics*, 126(6), 5113–5142.
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: a review of educational chatbots for the facebook messenger. *Computers & Education*, 151, 103862.
- Stieglitz, S., Mirbabaie, M., Möllmann, N. R. J., & Ryzski, J. (2021). Collaborating with virtual assistants in organizations: analyzing social loafing tendencies and responsibility attribution. *Information Systems Frontiers*, pp. 1–26. <https://doi.org/10.1007/s10796-021-10201-0>.
- Stokel-Walker, C. (2022). Ai bot chatgpt writes smart essays - should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>.
- Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., & Weisz, J. D. (2022). Investigating explainability of generative ai for code through scenario-based design. In *27th International conference on intelligent user interfaces. ACM digital library. Association for computing machinery* (pp. 212–228). <https://doi.org/10.1145/3490099.3511119>.
- Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*, pp. 502–510. <https://doi.org/10.18178/ijmerr.9.4.502-510>.
- Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., Poulton, A., Kerkez, V., & Stojnic, R. (2022). Galactica: a large language model for science. [arXiv:2211.09085](https://arxiv.org/abs/2211.09085).
- Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Yu, D., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., ... Chi, E. (2022). Le Quoc: LaMDA: language models for dialog applications. [arXiv:2201.08239](https://arxiv.org/abs/2201.08239).
- Trinidad, M., Ruiz, M., & Calderón, A. (2021). A bibliometric analysis of gamification research. *IEEE Access*, 9, 46505–46544.
- Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*.
- Vertegaal, R., Slagter, R., Van Der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *SIGCHI Conference on Human Factors in Computing Systems*, pp. 301–308.
- Vössing, M., Köhl, N., Lind, M., & Satzger, G. (2022). Designing transparency for effective human-ai collaboration. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-022-10284-3>.
- Wahde, M., & Virgolin, M. (2022). Conversational agents: theory and applications. [arXiv:2202.03164](https://arxiv.org/abs/2202.03164). <https://doi.org/10.1142/9789811246050>.
- Wallace, R. S. (2009). The anatomy of alice. In *Parsing the turing test*. Springer (pp. 181–210).
- Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). Arguetutor: an adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems. Association for computing machinery*. <https://doi.org/10.1145/3411764.3445781>.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: a state-of-the-art analysis. *Annual Academy of Management Meeting*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Sofia Schöbel is an assistant professor in information systems at the University of Osnabrück. She has written her Ph.D. about gamification in digital learning from the University of Kassel. Her research focuses on persuasive system design, designing smart personal assistants, digital transformation of services, and the design of interactive processes in digital learning. Her research has been published in different outlets such as the European Journal of Information Systems, Communications of the AIS and in leading information systems conferences such as ICIS or ECIS.

Anuschka Schmitt is a research associate and Ph.D. candidate at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland and currently a visiting researcher at Harvard University's School of Engineering and Applied Sciences. Her research focuses on the psychological aspects of AI ethics and user trust in human-computer interaction. Her research has been published at the International Conference on Information Systems and the Conference on Artificial Intelligence, Ethics, and Society.

Dennis Benner is a doctoral researcher at the chair of information systems at the University of Kassel, Germany. His research focuses on the design of conversational agents, persuasive systems, digital services, and digital education. His research has been presented at leading information systems conferences such as the International Conference on Information Systems.

Mohammed Saqr received a Ph.D. degree in learning analytics from Stockholm University. He works as a Senior Researcher at the University of Eastern Finland, where he is working on artificial intelligence, big data, network science, and scientometrics. His research interests include network analysis, temporal networks, machine learning, process, and sequence mining and temporal processes in general. He is also an active member of several scientific organizations and acts as an academic editor in leading academic publications.

Andreas Janson is a Postdoctoral Researcher at the Institute of Information Management (IWI-HSG) at the University of St. Gallen, Switzerland. He obtained his Ph.D. from the University of Kassel, Germany. His research focuses on service design, smart personal assistants, decision-making in digital environments, and digital learning. His research has been published in leading information systems and management journals such as the Journal of the Association for Information Systems, European Journal of Information Systems, Journal of Information Technology, and Academy of Management Learning and Education.

Jan Marco Leimeister is Full Professor and Director of the Research Center for Information System Design (ITeG) at the University of Kassel, Germany. He is furthermore Full Professor and Director at the Institute of Information Management, University of St. Gallen, Switzerland. His research covers Digital Business, Digital Transformation, Service Engineering and Service Management, Crowdsourcing, Digital Work, Collaboration Engineering and IT Innovation Management. He runs research groups and projects that are funded by the European Union, German Ministries, DFG, various foundations or industry partners. Prof. Leimeister teaches in several Executive Programs. He is a co-founder of several companies and serves as board member, consultant, coach and speaker for numerous companies.

Affiliations

Sofia Schöbel¹  · Anuschka Schmitt²  · Dennis Benner³  · Mohammed Saqr⁴  · Andreas Janson²  ·
Jan Marco Leimeister^{2,3} 

Anuschka Schmitt
anuschka.schmitt@unisg.ch

Dennis Benner
benner@uni-kassel.de

Mohammed Saqr
mohammed.saqr@uef.fi

Andreas Janson
andreas.janson@unisg.ch

Jan Marco Leimeister
leimeister@uni-kassel.de

- ¹ Information Systems, University of Osnabrück,
Osnabrück, Germany
- ² Institute of Information Management (IWI-HSG), University of
St. Gallen, St. Gallen, Switzerland
- ³ Information Systems, Research Center for IS Design (ITeG),
University of Kassel, Kassel, Germany
- ⁴ School of Computing, University of Eastern Finland,
Joensuu, Finland