From Human Pragmatic Language Skills to Conversational Agent Design: A Systematic Review of Transfer Strategies

Jiaxiong Hu
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
hujx@ust.hk

Xiwen Yao
Tsinghua University
Beijing, China
yxw20@mails.tsinghua.edu.cn

Zeyu Huang
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
zhuangbi@connect.ust.hk

Danxuan Liang
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, Hong Kong
dliangac@connect.ust.hk

Dongjie Yang
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, Hong Kong
dyangaj@connect.ust.hk

Dingdong Liu
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
dliuak@connect.ust.hk

Junze Li
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
jlijj@connect.ust.hk

Yuanhao Zhang
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, Hong Kong
yzhangiy@connect.ust.hk

Xiaojuan Ma*
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, Hong Kong
mxj@cse.ust.hk

Abstract

While conversational agents' (CAs) semantic and syntactic capabilities have advanced, their pragmatic skills, using language appropriately in context, have emerged as a critical focus in practical applications. Hence, scholars integrate conversational skills derived from human-human interaction into CA designs. However, existing research mainly adopts an empirical approach and focuses on specific CA deployment, making it challenging to identify overarching patterns or develop a comprehensive methodology for transferring human pragmatic skills to CA design. Thus, we conducted a systematic review of 85 studies from primary databases (e.g., ACM, IEEE, etc.), focusing on designing CAs with human-derived conversational skills. We identified skill categories (verbal, paralinguistic, nonverbal), transfer strategies (from dialog data, theories, and via co-design), implementations, and evaluation metrics. We consolidated these insights into a four-stage design process: human skill exploration, definition, transfer, and iterative evaluation. Future research can leverage this to design CAs that achieve conversational goals through contextually appropriate language use.

*Corresponding author

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3792787

CCS Concepts

• **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **HCI theory, concepts and models**.

Keywords

Conversational Agent, Design Process, Conversational Skill, Pragmatic Skill, Systematic Literature Review

ACM Reference Format:
Jiaxiong Hu, Xiwen Yao, Zeyu Huang, Danxuan Liang, Dongjie Yang, Dingdong Liu, Junze Li, Yuanhao Zhang, and Xiaojuan Ma. 2026. From Human Pragmatic Language Skills to Conversational Agent Design: A Systematic Review of Transfer Strategies. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3772318.3792787

1 Introduction

There is an increasing presence of conversational agents (CAs), also known as chatbots, personal assistants, or digital companions, in diverse application domains to provide interactive services through text-based or speech-based dialogues, including, but not limited to, health care [85, 167], education [156], and customer service [14, 103]. Compared to traditional systems such as graphical user interfaces, CAs offer superior user engagement with their interactivity, immediacy, intention of use, and social characteristics conveyed by conversations [51, 75, 156], as well as better accessibility to facilitate individuals with specific accessibility challenges, such as visual impairments or cognitive disorders [1, 192]. The computers are social actors (CASA) paradigm [118] posits that humans tend to interact with computing systems as they do with other humans, applying similar social rules and expectations. It is suggested that the CA's improved conversational skills increase its perceived social presence and affect user behaviors of interaction, which might further lead to increased satisfaction, enjoyment, perceived usefulness, and use of the system [146].

In the scope of this paper, "conversational skills" denote **pragmatic** communication abilities to use language in context, highlighting the parts beyond comprehending and conveying primary meanings of words (semantics) in correct grammatical forms (syntax) [90]. Such conversational skills, empirically identified in human-human interactions, can be divided into verbal behaviors (e.g., topic management, turn-taking, and breakdown repair), paralinguistic behaviors (e.g., fluency, prosody, and vocal quality), and nonverbal behaviors (e.g., facial expressions, gaze, and gestures) [132, 155]. Early development of CAs since the first chatbot ELIZA [173] focused primarily on improving the quality of content generation (semantics) and the accuracy of language forms (syntax). The recent large language model, GPT-4, has surpassed ELIZA and gained the language ability that is comparable with humans in certain setups, such as the game based on the Turing test [70]. Interestingly, one apparent difference between the model and humans is the model's "too perfect grammar" [70]. It is advised that CAs should enable pragmatic conversational skills to meet the user's expectation of their conversational and social intelligence to avoid frustration and dissatisfaction [23]. In the past decade, there has been an increasing recognition of the importance of pragmatic communication skills in CAs across academia and industry, leading to emergent designs of CAs' use of language in context. Take turn-taking as an example. Although conventional voice user interfaces impose a strict request-response structure and only allow users to take the turn to speak after a wake word, such as "Alexa," previous research found that this brings users extra efforts to integrate the CA into daily conversations [131]. This is regarded as a critical difference between human-CA and human-human conversations, which might lead to interruption, inefficiency, and dissatisfaction [131]. Conversely, the absence of such conversational skills can result in poor user evaluations such as reduced user trust, perceived transparency, empathy, and/or suboptimal task performance [99, 178], as observed in baseline CA systems in prior research.

To realize CA conversational skills, researchers employ rule sets, retrieval algorithms, (neural/deep) generative models, to large language models (LLMs) (e.g., [53, 84, 130, 148]). Notably, the development of CAs has become less burdensome with recent advancements in artificial intelligence (AI). Designers and researchers can now implement conversational skills by LLM prompting (e.g., the LLM-based self-disclosure [115]) rather than relying on resource-intensive, technically complex alternatives, allowing focus on high-level designs of conversational skills (e.g., the inner thought framework of CA [100]) instead of technical development. Thus, more complicated conversational skills for specific contexts are explored, such as the interview skills required for hospital admission [99], as well as the administrative skills for conducting tasks for screening cognitive disorders [39, 64].

Human-computer interaction (HCI) scholars stressed the necessity to take inspiration from human-human interaction to design CAs with these special conversational skills. They experimented with different ways to extract skills, from human conversation data [62, 119, 128] or expert feedback [115, 184]. No systematic synthesis of literature, however, has yet explored the whole process by which conversational skills, derived from human interactions, can be translated into CA design. One existing survey develops a taxonomy of CA communicative behaviors and identifies underexplored behaviors that ought to be transferred from human dialogue to CAs [168]. Nevertheless, it did not provide a tactic roadmap for designers to follow. A separate systematic review examines CA design practices, with more emphasis given to the evaluation methods and metrics [186]. While it mentioned some broad categories of design methods (e.g., theory- or framework-based approaches, and participatory design), no specific details were provided about how to implement skill transfer from humans to CAs.

To fill these gaps, we conducted a systematic review of 85 studies to answer the following research questions:

- RQ1: What are the skills that practitioners transfer from human conversations to CA design?
- RQ2: How are conversational skills derived from human dialog to CA design?
- RQ3: How are the skills implemented and evaluated?

Our findings reveal that the skill transfers happened across diverse human-CA interaction scenarios: health, customer service, education, social support, and so on. According to the pragmatic theories [132], we categorized the identified skills that are transferred from humans to CAs into verbal, nonverbal, and paralinguistic skills. Specifically, speech acts, turn-taking, topic, and stylistic variances are identified as sub-categories under verbal skills. The CA's facial expression, gesture, eye gaze, and typing behaviors consist of the nonverbal skills. Last, paralinguistic skills cover aspects of prosody, vocal intensity, and fluency of the CAs. More importantly, we summarized three primary transfer strategies. The reviewed works extract conversational skills from human dialog data, either collected by the researchers or from previous works, through mixed-method analysis. A more flexible strategy is to conduct consultations, interviews, or surveys with human participants to synthesize the conversational skills. Sometimes, the participants are required to craft the skills directly, such as providing the response contents or templates. The most common way is to transfer the conversational skills from the theories of linguistics, communication, psychology, or sociology. The transferred skills are implemented through the traditional model training, such as training a language generation model with the collected dialog data with target conversational skills. To implement the conversational skills in a more controlled way, the reviewed works also build knowledge bases to manage the skills. The more recent trend of utilizing large language models to implement conversational skills is also identified, via approaches from basic prompting engineering to the more advanced chain-of-thought or multi-agent frameworks.

Our work contributes to a systematic review of the literature about transferring human conversational skills to the design of conversational agents. More importantly, we propose the design process for effective and efficient skill transfers by integrating our findings with the Double Diamond design process [31, 32]. The four-stage process includes human skill exploration, skill definition, skill transfer, and iterative evaluation. The process balances creativity and implementation of the CA design with two rounds of divergent and convergent processes. It depicts a comprehensive method for future design that requires the CA to be human-like, socially appropriate, and effective in service delivery

2 Related Works

2.1 From the Theory: Computers Are Social Actors

The anthropomorphic design of conversational agents (CAs) finds its roots in an early human-computer interaction (HCI) framework, the CASA theory (Computers Are Social Actors) [117, 118]. At its core, this theory suggests that we are inclined to engage with computers not as mere instruments, but as social beings, especially when they bear the hallmarks of human interaction, such as the natural language woven into CAs' conversations. This anthropomorphic bent, however, is not merely a quirk of interaction. It carries tangible advantages for user experience and task performance in contexts. In retail settings, for instance, CAs endowed with human-like linguistic cues, such as using more first-person descriptions, foster greater trust, stronger purchase intent, word-of-mouth, and higher satisfaction than their non-anthropomorphic conditions [81]. The reason, in part, lies in the texture of their communication: an anthropomorphic style exerts a significant influence on social presence, that intangible sense of "being with" another, which in turn weaves connections to user trust, empathy, and satisfaction [66]. Today, it is artificial intelligence technologies, most notably large language models, that have propelled these CAs beyond previously imagined limits, pushing the boundaries of what they can achieve [148]. Empowered by such AI-driven text generation tools, CAs now mirror human conversation with a naturalness and complexity once elusive: their responses flow with the cadence of real dialogue, navigating nuance and complexity in ways that blur the line between machine and human interaction [68, 69]. This leap toward humanlikeness aligns with a broader design imperative, as noted in prior reviews: CAs are increasingly crafted with both verbal and non-verbal human behaviors, each a deliberate brushstroke in the portrait of anthropomorphism [168].

2.2 Anthropomorphic Skill Design of Conversational Agents

Turning first to verbal skill design, prior studies highlight how CAs' linguistic behaviors shape user perceptions and interaction quality. These verbal behaviors or skills occupy an explicit, tangible role in the back-and-forth between CAs and users. Consider, for instance, a CA skill design that embeds memory into dialogue: recalling personal details from a conversation, even including intentional lapses to mimic human forgetfulness [137]. This careful calibration of remembering and forgetting, the research found, heightened both user enjoyment and the CA's believability, as if the CA carried the faint, familiar imperfection of a human mind. Beyond memory, researchers have probed the texture of language itself, exploring how human-like linguistic styles shape interaction. Assistant CAs, for example, have been designed with empathetic language, using tones that mirror care or understanding. In contrast, survey CAs adopt more formal cadences, each style tailored to its context [63, 75, 169]. These variations, like the dynamics of human speech, underscore how verbal skills serve as deliberate tools to bridge the gap between machine and person.

Beyond verbal behaviors, researchers have also examined how non-verbal skills contribute to anthropomorphic CA design. In human conversation, non-verbal cues carry as much weight as verbal signals that shape connection, trust, and understanding. It is this truth that guides the design of CAs, which are increasingly endowed with non-verbal skills to mirror the rhythms of human interaction. Consider gaze, for instance: in task-assisting contexts, a CA's deliberate gaze movements have been shown to strengthen user trust, as if the agent's attention, even in digital form, fosters a sense of reliability [159]. Facial expressions, too, serve as tools of empathy: CAs designed to display such cues can cultivate an empathetic profile, bridging the gap between machine and human emotion [112].

2.3 Previous Review Works Related to Conversational Agents

Across existing review studies, three major strands become apparent, although each addresses different aspects of CA design. The first strand focused on human-like design: identifying both verbal and non-verbal human-like behaviors in CAs as mentioned earlier [168], or distilling the characteristics of CAs' conversational and social intelligence [23]. Yet such inquiries remain sparse, overshadowed by the far larger body of reviews that fixate on domain-specific niches, such as elderly support [65, 78], customer service [17, 177, 183], and health care [26, 79, 85, 167]. Others turn a business lens to the subject, dissecting adoption motivations, engagement drivers, and their downstream outcomes [14, 103]. This scholarship maps the commercial pulse of CA integration but stops short of broader design questions. A bigger strand of review works fixates on the technical parts: implementing CAs or outlining technical frameworks for their development [53, 84, 148]. One such survey, for instance, parses the architectural backbone of CA design, categorizing conversation generation into rule-based, retrieval-based, and generation-based methods [130]; another recent work sketches human-centered approaches to algorithm and model development [126]. Within HCI, too, reviews have probed CA research from distinct angles: one centers on usability evaluation [44], while others unpack critical user experience factors—engagement, satisfaction, privacy, and trust, all threaded through the literature [56, 134, 136]. Yet for all this attention, a quiet gap persists: the review works have largely overlooked the design methods that underpin anthropomorphic CA development. The question of how to design CA, through intentional, theory-driven design choices, remains underexplored.

Alongside these review strands, a smaller but important body of work has examined the design methods behind CA development. Much of the existing work on CA design has leaned on empirical inquiry, grounding its insights in prototype systems and user studies to distill actionable design implications [13, 27, 64]. Yet alongside this empirical vein, a smaller body of work has delved into the theoretical methodologies that shape CA design, each probing distinct facets of the process. In healthcare, for instance, one paper revisited longstanding knowledge engineering methodologies, proposing an updated lifecycle that weaves knowledge gathering and usability testing into its core [21]. To gather the domain-specific knowledge that animates a CA's conversation design, researchers often turn to familiar tools: crowdsourcing, interviews, and co-design with experts, each a thread in the fabric of intentional dialogue scripting. A study exploring polyadic human-CA interactions, for example, notes that participatory design often takes center stage in such design scenarios [186]. A rapid review dedicated to co-designing CAs unpacks the techniques, outcomes, and benefits of this collaborative approach, shedding light on how shared creation shapes better CAs [141]. Beyond specific methods, reviews have cast a wider net, cataloging application domains, user-CA relationship dynamics, and evaluation metrics that anchor CA design [37, 152, 186]. Theoretical frameworks, too, have shaped this landscape, such as the Natural Conversation Framework, drawing inspiration from the patterns of human dialogue [111], while other work has leaned on established conversation theories to structure CA interaction [37, 186]. Taken together, these efforts sketch a field increasingly drawn to anthropomorphic design, a trend evident in the focus on human-like skills, conversational fluency, and social acumen. Yet for all this attention to what makes CAs feel human, a critical question lingers, half-formed: the process by which human conversational techniques, including the subtle, learned skills that animate face-to-face exchange, might be translated into the mechanics of human-CA interaction remains unclear. It is this gap that underscores the need for a deeper inquiry into the translation of human dialogue into CA design.

3 Methodology

Inspired by previous HCI literature review works [151, 163], we follow the PRISMA guidelines [127] to implement a four-stage systematic review process as fig. 1 illustrates: 1) Identification through database searches; 2) Screening with explicit rules; 3) Eligibility assessment through full-text review; and 4) Final inclusion based on relevance criteria.

3.1 Paper Collection

3.1.1 Identification. We conducted systematic searches across six major academic databases: ACM Digital Library (ACM-DL), IEEE Xplore Digital Library (IEEE-XDL), Taylor & Francis Online (T&F), Web of Science (WoS), ScienceDirect (SD), and Scopus. These databases were selected for their strong coverage of human-computer interaction, conversational AI, and computer science, ensuring comprehensive capture of relevant literature across multiple disciplinary perspectives.

The search query is composed of two parts. Inspired by the previous works of conversational agent researches [168, 172, 186], we first searched for papers with conversational agent terminology **in title** using the following query: *"(conversation AND (agent OR AI OR assistant)) OR chatbot OR socialbot OR (digital AND (assistant OR agent)) OR (intelligent AND (assistant OR agent)) OR (virtual AND (agent OR assistant)) OR (personal AND assistant)"*. It allowed us to cover diverse CAs, both dyadic and polyadic, embodied and non-embodied. In order to look for papers that potentially include the conversational skills or strategies designed by authors themselves, we organized a secondary search query through searching **in all fields** with these keywords: *"conversational skill*" OR "conversational strateg*" OR "conversational theor*" OR "conversational framework*" OR "communicational skill*" OR "communication strateg*" OR "communication theor*" OR "communication principle*" OR "communication mechanism*" OR "communication framework*" OR "communication behavior*" OR "conversational dynamic*" OR "conversation dynamic*"*. We focused exclusively on English-language papers published in peer-reviewed conferences and journals, while excluding literature reviews, books, extended abstracts, workshop papers, and other non-peer-reviewed materials. This search was completed on July 30th, 2025, by using these queries throughout the six academic databases, which resulted in 1,024 relevant papers after removing duplicates.

3.1.2 Screening. In the screening stage, our workflow requires rapidly evaluating a large number of paper titles and abstracts and distinguishing CA–related work from adjacent but irrelevant fields. Prior research has shown large language models (LLMs) as effective tools for systematic reviews, particularly for screening and extracting information from academic documents [8, 101, 144]. Thus, we implemented an LLM-assisted screening approach. Prior work suggested that the GPT-4 models are promising in the title and abstract screening task [104, 122]. Given that the OpenAI platform offers accessible APIs that support reproducible batch processing, we deployed our system via the Azure OpenAI GPT-4o Chat Completions API with 0.1 temperature to reduce variability [1]. Specifically, the screening on title and abstract considered two key criteria: 1) The paper focuses on specific conversational behaviors, techniques, or strategies designed for CAs, and 2) the paper suggests a pathway from human conversational skills to CAs, such as informing skill design with theories from the literature. As suggested by prior work [104], we applied a similar layered strategy to compose the LLM prompt to check if the title and abstract mentioned any CA and human conversational skill, and skill evaluation with confidence scores (high/medium/low) respectively. We optimized the LLM prompt iteratively as suggested [101], to verify whether the LLM's output was in the correct format and with minimized hallucinations. Then, we manually assigned papers with *"INCLUDE"*, *"EXCLUDE"*, or *"CHECK FULL TEXT"* labels, referring to the LLM output and the original title and abstract, following the requirement of human verification for AI-assisted literature review [8, 153]. Those with all high scores or all low scores could be quickly labeled as included or excluded after reading the title and a few sentences of the abstract. Titles and abstracts with mixed scores were carefully examined. To ensure the consistent understanding of the screening criteria, we checked the inter-rater reliability with the intraclass correlation coefficient ICC(3,1)=0.55 (fair agreement) [29, 105] among five coauthors based on a subset of 50 papers. Then we independently screened the rest papers. After excluding papers with *"EXCLUDE"* labels, we retained 233 papers out of 1,301 initially identified papers (17.9%). During this phase, many papers from unrelated fields (e.g., telecommunications engineering, military communications, and information systems without conversational focus) were excluded to maintain research scope alignment.

[1] (Accessed by November 29, 2025) https://platform.openai.com/docs/api-reference/chat

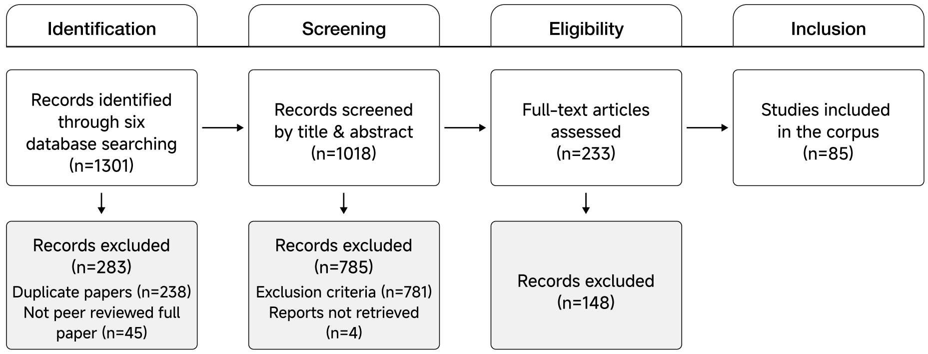

Figure 1: Flowchart of our paper collection process following the PRISMA framework

Process	Criteria	ACM-DL	Scopus	SD	T&F	WoS	IEEE-DL	Total
Identification	Search term	223	394	166	118	154	246	1301
Screening	Title and abstract	193	253	166	117	69	220	1018
Eligibility	Full paper	71	52	41	38	13	18	233
Inclusion	–	34	17	15	12	5	2	85

Table 1: Number of papers processed across four stages in different databases

3.1.3 Eligibility. To ensure the eligibility of the screened papers, we further excluded papers that 1) focus primarily on appearance design, such as avatars, visual interfaces, or color schemes; 2) only focus on technical implementation; 3) derives entirely from previous CA works instead of human conversational skills; 4) involves no specific CA skills; 5) indicates or suggests no clear process from the human skills to CA designs. Similarly, the GPT-4o assessed the full text and identified the transfer process, i.e., WHY and HOW human conversational skills are transferred to CA design, categorized into three levels: *Highly Relevant, Moderately Relevant, and Minimally Relevant*, which served as an initial reference for us to quickly grasp the paper idea. Then, two coauthors manually checked the full-text papers based on the eligibility criteria. To ensure we have a consistent understanding of the criteria, we assessed the inter-rater reliability on a subset of 20 papers ICC(3,1)=0.8 (excellent agreement) [29, 105].

3.1.4 Inclusion. Following the systematic screening and eligibility assessment process, the final corpus consists of 85 papers. Table 1 presents the detailed breakdown of studies filtered at each stage, while Figure 2 shows the temporal distribution of these papers from 2001 to 2025. Two distinct research surges around 2020 and 2024 overlap temporally with breakthroughs in large language models, namely the release of GPT-3 in 2020 and the ChatGPT-4 series models from 2022 to 2024. Among the included studies, 10 explicitly stated the application of GPT-3 or more advanced LLM models. The majority of the selected studies were published in high-impact conferences and journals, including but not limited to Conference on Human Factors in Computing Systems (CHI), International Journal of Human-Computer Studies (IJHCS), International Conference on Intelligent Virtual Agents (IVA), Special Interest Group on Information Retrieval (SIGIR), Journal of Medical Internet Research (JMIR), Computer Supported Cooperative Work (CSCW), International Journal of Human-Computer Interaction (IJHCI), Journal of Business Research, and Behaviour & Information Technology. This distribution reflects the interdisciplinary nature of conversational agent research, spanning fields from computer science and psychology to business and healthcare informatics. Last, we appraised the included paper quality with the Mixed Methods Appraisal Tool (MMAT) [61] and identified three papers with limited methodological quality for the absence of clear research questions and inadequate evidence-based findings.

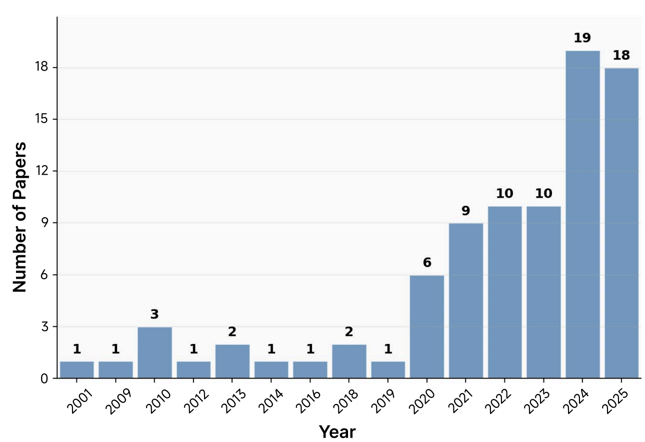

Figure 2: Temporal distribution of the included papers

3.2 Data Analysis

Based on the research questions, our data analysis focuses on data extraction in the following aspects. First, we identify the conversational skills that the paper focuses on. Specifically, we extract information about specific skills, including behaviors, strategies, techniques, functions, features, or manipulations, that are designed or implemented for the CA in the paper. Then we look for the sources from which the authors derive the skills, as well as the approach adopted for the transfer. Specifically, we first engage in close, critical reading of each full paper document: this involves analyzing the study's research questions, methodology, findings, and discussions to gain a deep understanding of how skill transfers are conceptualized, implemented, measured, and contextualized within each work. Second, we extract targeted information about skill transfers and consolidate it in a collaborative table document to organize two key components: summarized key phrases generated through open coding (where we label specific concepts, patterns, or instances of skill transfer in our own concise language) and verbatim original quotes from the papers (to preserve the authors' original intent and provide direct evidence for each coded item). To maintain consistency in coding standards and resolve any ambiguities, the authors conduct at least two meetings per week throughout this phase, using these sessions to align coding protocols, clarify interpretive differences, and ensure all team members apply the same criteria when labeling and extracting data. Finally, we employ a mixed-methods approach, defined here as the integration of deductive and inductive coding, to synthesize the preliminary open codes: this involves using inductive coding to identify emergent patterns and group conceptually similar codes into initial categories based on the raw data, while also applying deductive coding to create the high-level categories: verbal, nonverbal, and paralinguistic skills based on prior work on pragmatics [132]. This dual approach helps merge overlapping codes, resolve inconsistencies, and organize the refined data into coherent, meaningful themes that capture the core insights about skill transfers across the literature.

To facilitate the data extraction, we used the LLM, Doubao [2], to search preliminary key phrases and related parts from the paper documents. We selected this tool for two reasons. First, prior work indicates that its underlying model demonstrates performance comparable to mainstream LLMs (e.g., Deepseek V3, Kimi1.5) with accuracy and reproducibility across multiple languages, including English [49, 190]. Second, and most critically, Doubao, either its website or application version, offers superior source grounding capabilities. Unlike standard text-generation tools, it supports visual traceability by highlighting and linking generated answers directly to their corresponding segments in the original PDF. This feature was instrumental in allowing authors to instantly verify the context and validity of the extracted information. Furthermore, as a freely available tool with proficiency in English academic text processing, it offered a cost-effective solution for our workflow. The prompt includes the target information, such as "check what the conversational agent skills are in the paper", the relevant sections, such as "look at the methodology-related sections" or "the CA skills in related work or background sections should be excluded from the search". The LLM is requested to highlight the related original text in the document first, and then provide a brief summary. Please see the supplementary materials for the full prompts. It is important to emphasize that our extraction workflow remained fundamentally human-led. Doubao served strictly as an assistive tool for locating candidate passages, while the final coding was conducted manually by two researchers. To ensure data integrity, we enforced a strict human-verification protocol consisting of three steps: 1) Checking if the extracted data format follows the requirement in the prompt. 2) Cross-referencing extracted quotes against the full-text PDF to ensure accuracy. 3) Manually reviewing related sections to identify any critical information missed by the LLM.

Regarding the data integrity and copyright issues, we double checked the policy of Doubao [3] and ensured that the data is well protected from other unintended usage. Doubao also provides the local processing function without uploading the original documents to the cloud.

4 Findings

Our findings were organized into four primary themes. First, we summarized the CA application scenarios where the transfers usually happened. Inspired by the taxonomy in pragmatic theories [132], we then categorized all the identified skills transferred from humans to CA designs into three modalities: verbal, nonverbal, and paralinguistic. Second, we identified the primary strategies adopted by researchers to transfer humans' conversational skills to CAs, the approaches of skill implementations, and the metrics for evaluating the transferred skills. Lastly, the major pathways of CA skill transfer were summarized.

4.1 Scenarios

Our analysis reveals that the transfer of human conversational skills to CA design is situated within distinct application scenarios. These scenarios, as summarized in table 2, can be broadly categorized into five primary domains: Health, Customer Service & Commerce, Education, Social Support, and Others.

The **Health** domain represents a significant and diverse area for CA applications. CAs are frequently designed as health information providers, delivering healthcare services as medical chatbots [120, 149], offering advice on specific conditions like acne [67], and providing guidance against public health misinformation [129]. Another major sub-topic within health is behavior change, where CAs act as persuasive counselors or partners. This includes promoting COVID-19 vaccination [121, 166], supporting smoking cessation [60], encouraging physical health [24, 147, 162], counseling for substance use [125], and ensuring medication or treatment adherence [2, 25]. Mental health is another critical area, with CAs designed to facilitate user self-disclosure [86, 88], assisting in counseling sessions [3, 174], offering mental health support as a social partner [47, 96, 107, 108], or improving mental health awareness [115]. Finally, some CAs deal with specific disease-related issues, including autism [182] and neurocognitive disorders [39, 64]. And some are also designed for more general-purpose symptom checking [178] and hospital admission interviews [99].

Then, the **Social Support** domain focuses on the relational and affective functions of CAs. Many agents are designed for casual and

[2](Accessed by November 29, 2025) https://www.doubao.com/chat/

[3](Accessed by September 10, 2025) https://www.doubao.com/legal/privacy

Scenario	Sub-topic	Description	Examples
Health	Health Info Provider	Provide health-related information for users, such as answering questions about healthy habits	[25, 36, 67, 120, 129, 149]
	Behavior Change or Adherence	Persuade users to change unhealthy behaviors or maintain a healthy lifestyle	[2, 24, 25, 36, 60, 121, 125, 147, 160, 162, 166]
	Mental Health	Provide mental health support, such as counseling or consultation	[3, 47, 86, 88, 96, 107, 108, 115, 174]
	Clinical or Disease-related	Facilitate screening, diagnosis, or intervention for certain diseases or symptoms	[39, 64, 99, 178, 182]
Social Support	Casual & Open-Domain Chat	Provide general or non-task-oriented social interactions, such as small talk, storytelling, or acting as a conversational partner.	[46, 82, 100, 109, 123, 188]
	Relational & Emotional Support	Mediate conflicts, enhance specific human-to-human relationships, or provide emotional support	[58, 92, 185]
	Group & Collaborative	Support group conversations or collaborative tasks	[7, 40, 191]
	Elderly Care	Provide companionship and social support for older adults	[50, 97, 140]
Customer Service & Commerce	Operational Support	Address customers' operational needs, such as booking a hotel	[71, 138, 150, 180, 189]
	Sales & Information Provider	Provide counseling, recommendations or suggestions	[55, 72, 76, 87, 91, 113, 116, 119, 165, 181]
	Information Collection	Collect information such as customer survey	[28, 62, 83, 157, 187]
Education	One-on-one Coaching	Act as a one-on-one coach to help users practice and develop specific skills	[95, 124, 142, 143, 184]
	Classroom Teaching	Take on the role of a tutor or lecturer to facilitate learning in traditional academic subjects	[6, 45, 89, 98]
Others	General Scenarios or Generic CA Attributes	Support everyday, general-purpose tasks; Studies investigate fundamental, transferable CA capabilities	[22, 34, 94, 102, 133, 146, 158, 179]
	Workplace & Specialized Domains	Provide assistance in specific professional, industrial, or technical contexts	[48, 52, 145]
	Recommendation	Provide tailored recommendations	[93]
	Games & Entertainment	Facilitate engagement and entertainment in playful, non-utilitarian, and game-oriented interactions	[82]

Table 2: Application scenarios of the CAs from the reviewed studies

open-domain chat, acting as non-task-oriented peers or partners for small talk [82, 109, 188], storytelling [46], and general social conversation [100, 123]. For relational and emotional support, CAs are designed to provide empathetic chat [92], mediate conflicts [58], or enhance communication between long-distance romantic partners [185]. CAs are also being explored in group collaboration and communication. Examples include assistance in group online chats [40], supporting collaborative planning [7], or even social implications with multiple CAs in the group [191]. Another particularly specialized area in this domain is elderly care, where CAs serve as social partners and daily assistants for older adults [50, 97] and act as communication supporters within caregiving relationships [140].

In **Customer Service and Commerce**, CAs are widely deployed to enhance efficiency and customer experience. For basic operational support, they address routine commercial needs such as booking and ordering things [71, 138, 150, 158], managing shipping and delivery [180], and assisting with unemployment benefit applications [189]. Beyond these, CAs function as sales assistants and information providers, offering product details [165] and suggestions [55, 87, 181]. Some of them are more humanized as virtual salespeople for digital devices [116], financial investments [76], and generic online shopping [72, 91, 113, 119]. Furthermore, the last sub-category covers CAs that are designed to be customer information collectors, gathering user data for marketing [28], handling service failure and recovery [83, 157, 187]. The **Education** sector leverages CAs to provide personalized and scalable learning support. One key role is one-on-one coaching, where CAs help users develop specific skills. Apart from one work in which the CA helps develop children's socio-emotional skills [143], examples mostly target language skills such as oral presentation [184], negotiation [124], group discussions [142], and writing [95]. Another role of CA is the tutor or lecturer in classroom teaching. This has been applied to elementary computer science [45], foreign language [98], and generic pedagogical methodologies to increase student engagement [6, 89]. Additionally, learners can converse with multiple CAs with distinct roles, such as a pedagogical host CA for tests and a crowd of student CAs that provide applause and empathy [89]. Another example includes pedagogical CAs that provide instruction on counseling skills and some other role-play CAs that allow users to practice these skills [121].

In addition, several papers explore CAs in more general scenarios, where CAs serve as all-purpose assistants for everyday tasks [94, 102], in smart speakers [133], or for driving cars [34]. Some studies focus on investigating fundamental capabilities like emotion and personality expressivity that are applicable across multiple contexts rather than being tied to a specific scenario [22, 158]. In Workplace and Specialized Domains, agents are designed to function in professional settings like offices [145], manufacturing [52], or to aid in complex tasks such as data science analysis [48].

4.2 Transferred Skills

4.2.1 Verbal Skills. Compared to the other two modalities, verbal skills represent the most extensively studied modality. Their

Category	Sub-category	Description	Examples
Speech Act	Emotion-awareness & Empathy	Express that the CA can feel or understand the user's feelings, or adjust responses accordingly	[3, 36, 67, 71, 86, 89, 92, 96, 99, 129, 147, 166, 174, 189]
	Self-disclosure	The CA discloses information about itself	[88, 93, 115, 124, 140, 147]
	Persuade & Encourage	Use persuasive or encouraging responses to maintain or change user behaviors or opinions	[2, 50, 60, 121, 125, 142, 147, 160]
	Rapport Recovery	Recover rapport between the CA and the user	[76, 157, 187]
	Personalized Response	Differentiate user profiles and adjust responses accordingly	[3, 72, 98, 107, 113, 115, 119, 165, 174]
	Personal Pronoun Use	Tend to use a specific personal pronoun more often, such as "you"	[7, 133, 165]
	Asking Questions	Ask questions to the user	[45, 62, 102, 116, 121, 146, 147]
	Scaffolding	Provide "just enough" support for the user's task maintaining or autonomy	[64, 99, 140]
	Referring to History	Refer to contents that have been mentioned in conversation history	[107, 108, 146]
	Explanation	Explain system or CA outputs	[178]
	Imprecise Statement	Mimic humans' imprecise statements such as describing a date with "few days ago"	[145]
Turn Taking	Response Timing	Decide when to respond to the user	[55, 109, 179, 184]
	Active Listening & Back-channel	Express a state of listening, such as paraphrasing user input, or use Back-channels such as "OK" or "Uh-huh" without taking the turn	[39, 86, 87, 97, 121, 123, 150]
	Proactivity	Proactively initiate a new conversation or join an ongoing conversation in either a dyadic user-CA context or a group conversation	[40, 58, 94, 97, 100, 182, 185]
	Breakdown Repair	Repair the conversation from breakdowns	[34, 52, 64, 180]
Topic	Step-by-step Workflow	Provide step-by-step guidance through the conversation or the task	[48, 64, 83, 95, 99, 140, 143]
	Session Structure	Predefine primary sessions for the conversation	[24, 45, 185]
	Small Talk	Have casual small talk sessions with the user	[86, 121, 147]
Stylistic Variances	Style Adoption	Adopt specific language styles, such as warm, humorous, or formal, in conversations	[25, 28, 91, 138, 149, 162, 178, 181, 185]
	Style Control	Adapt language styles to the CA's personality settings or maintain a specific style across multiple modalities	[7, 120, 165]

Table 3: The taxonomy of CAs' verbal skills, focusing on the lexical or syntactic aspects of language use in text messages or speech

largest cluster is the **speech act**, referring to functional behaviors performed through language use for specific communicative purposes [9, 132]. Among them, emotion-awareness and empathy emerge as a core competency, implemented through various approaches, including emotion-aware responses that adjust based on detected user emotional states [63, 89] and empathetic message generation [63, 71, 166]. Self-disclosure represents another critical skill in many scenarios, such as mental health support, where CAs share information about themselves to facilitate user disclosure [93, 115, 124]. Different from the common context in which the CAs answer questions from the users, recent CAs began to reverse the interaction to ask questions to the users. For example, CAs are equipped with skills to ask follow-up questions that elicit information from users [45, 121, 146]. Then, **turn-taking** refers to the smooth alternation of "speaker" and "listener" roles between communicative partners, including a series of behaviors that regulate when to initiate speech, respond to others, and repair breakdowns [132]. For example, response timing control allows CAs to simulate human-like pauses and deliberation [55], while proactive response capabilities enable them to appropriately initiate conversations rather than merely responding reactively [94, 97, 182]. Then, CAs' active listening and back-channeling behaviors help maintain conversational flow and engagement in scenarios like shopping [87] and disease screening tests [39].

To a higher level of dialog organization and management, **topic** is the category representing a core verbal component of pragmatic competence, focusing on how CAs choose, present, sustain, and shift conversational topics or themes in line with contextual norms and interaction goals [19, 132]. For example, the CA may manage the topic via structured approaches, including dialogue frameworks that provide step-by-step guidance [95] or sophisticated dialogue session organization systems using decision trees and categorized responses [45]. Meanwhile, **stylistic variances** refer to the ability to modulate their communication styles across different contexts and listeners [132]. Included studies enabled CAs with specific language styles, such as humorous via joke telling by CAs [162, 185]. Sometimes, different styles are compared to investigate their effects under a certain context. In customer service scenarios, for example, warm versus competent [28, 138] and formal versus informal [91] communicative styles of the CAs may affect user experience, such as trust. Moreover, the CA's conversational style is suggested to be adaptive, such as CA personality-based style expression to manifest consistent personality traits [120], and dynamic style adaptation that adjusts communication patterns based on user styles [165].

4.2.2 Paralinguistic and Nonverbal Skills. Compared to verbal skills, paralinguistic and nonverbal skills have received relatively less research attention, yet they play a crucial role in creating natural human-CA interaction experiences. In human pragmatic language skills, they are used to shape the understanding of the message and

Category	Sub-category	Description	Examples
Paralinguistic	Prosody	Vary pitch, loudness, and duration to convey meaning and affect	[7, 22, 87, 165]
	Fluency	Control the smoothness and pacing of speech, including pauses and fillers	[47, 179]
Nonverbal	Facial Expression	Convey emotions through facial expressions	[7, 22, 82, 158, 165, 174, 179, 181]
	Eye Gaze	Direct the eyes to signal attention, regulate turn-taking, or express social and emotional cues	[6, 46, 116, 124, 191]
	Gesture	Use movements of hands, arms, or body that support, complement, or replace verbal communication	[22, 124, 184, 191]
	Typing Behavior	Simulate human typing delays and input styles	[188]

Table 4: The taxonomy of CAs' paralinguistic skills and nonverbal skills, focusing on vocal (but not lexical or literal), physical, or behavioral cues that shape and regulate interaction or complement verbal messages

complete the verbal message via vocal but non-linguistic aspects such as prosody and fluency, or physical and behavioral aspects such as body posture and facial expression [43, 132].

Paralinguistic skills primarily involve acoustic characteristics and temporal features of speech. Prosodic control emerges as a core capability, including the CA's speech length, pitch, and loudness manipulation [7, 165]. For the text-based CAs, the prosodic cues can be simulated with punctuation marks such as "!!!" [87]. Besides, fluency management involves addressing chatbot delays through filled pauses or time lags [47, 179]. On the other hand, CAs equipped with **nonverbal** skills communicate with various visual and behavioral cues. They usually have an embodiment design, such as virtual agents with a human body in VR [46, 47, 184, 191] or a screen [22, 82, 97, 124, 125, 158]. So, arm gestures can facilitate the conversation [22, 191]. It could be only a virtual head if involving no bodily gestures [7, 165]. Thus, facial expression represents the most extensively studied skills, including emotion-driven designs such as smiling with different frequencies [82] and personality-based facial expression updates [158]. Eye gaze skills of CAs can facilitate communicative accuracy and perceived naturalness in the context of information-seeking [6]. Most of the CAs are non-embodied, for example, text-based chatbots [28, 95, 149] or voice-based agents [50, 102, 140], but they can still use emoticons to simulate facial expressions [181].

4.2.3 Skills Synergies. Reviewed studies sometimes implement skills in synergies. For example, social rapport building relies on a synergy of small talk, self-disclosure, and active listening to establish trust, which serves as a foundation for effective persuasion [86, 121, 147]. CAs for clinical or elder care use often combine scaffolding with empathy skills. This combination aims to reduce user frustration and maintain motivation during complex tasks [64, 99, 140]. The synergies are also cross modalities. Verbal skills like empathy and style adoption are frequently reinforced by nonverbal skills such as facial expressions and prosody to create more authentic emotional agents [7, 22, 165, 174].

4.3 Transfer Strategies

We identified three primary strategies for researchers to transfer conversational skills from humans to CAs. First, they analyze the human-human conversation dialog data to extract the skills or the parameters for skill implementation. Or, they apply co-design methods to invite individuals with certain conversational skills to the design process. Last, a more common way is to use the literature as inspiration or motivation for the skill transfer.

4.3.1 Transfer From Human Dialog Data. To extract human conversational behaviors, researchers rely on open datasets from prior work that possibly contain certain conversational skills, such as the NoXi database [20], a corpus of screen-mediated face-to-face interactions, or the Counseling Conversation Analysis dataset [4]. If there is no available dataset, researchers may consider augmenting existing data with large language models (LLMs). For instance, researchers transformed the non-humorous human-crafted dialogues into humorous dialogues with LLMs [162]. Alternatively, researchers create dialog datasets either by collecting from a real-world scenario or building in a lab context. For instance, to extract the elevator pitch skills, researchers attend public talks to observe and videotape any potential elevator pitches with consent from participants [184]. Or, they also recruit human speakers to the lab to demonstrate conversational skills. For example, to observe how older adults apply support interruptions and back-channels, researchers invite them to pair up and discuss provided topics in a university laboratory [97]. To equip the CA with skills to understand human behaviors in conversations, researchers use prototypes, such as a manually controlled CA (Wizard-of-Oz), to stimulate human participants' behaviors [116].

Regarding the approach to extract conversational skills from data, researchers apply qualitative analysis and mixed-methods combining qualitative and quantitative analysis. One example is that to extract professional skills to ask follow-up questions, researchers qualitatively analyzed the dialog data to identify different question types [62]. When the transferred skill requires quantified parameters from data, mixed-method analysis is the option. To decide the back-channel timing, researchers analyzed the time interval distributions of different types of back-channels in human dialogues [39]. Sometimes, specific knowledge about the skills is not required, then, researchers annotate the data for later model training, especially when the skills can be implemented via end-to-end language models, such as the model to generate empathetic responses [96].

4.3.2 Transfer via Co-design. Researchers often involve human participants with certain conversational skills in the design process to share or summarize their skills in interviews or surveys. For example, retrospective think-aloud protocols are frequently used so the participants can clarify their skill knowledge while watching the replay of the recorded conversation [97, 100]. A more

Primary Strategy	Approach	Description	Examples
Transfer from Human Dialog Data	Qualitative Analysis	Apply qualitative methods such as thematic analysis to extract strategies and guidelines as skills	[62, 182, 184]
	Quantitative or Mixed-method Analysis	Add quantitative analysis to check behavior distribution or extract parameters for CA skill implementation	[39, 58, 64, 97, 116, 119, 123]
	Model Training	Data is for later model training to generate conversational behaviors or skills	[3, 39, 62, 92, 96, 125, 174, 182]
Transfer via Co-Design	Summarize Strategies or Guidelines	Invite human participants to share their conversational experience to inspire CA skill design	[2, 97, 99, 100, 145, 184, 185]
	Craft Dialog Content	Invite human participants to directly craft dialog content, such as predefined responses or templates	[45, 60, 92, 115]
	Collect Feedback	Invite human participants to provide feedback in the skill design phase	[45, 50, 60, 86, 95, 119, 142, 178, 185]
Transfer from Theories	Inform Skill Design by Theories	Use theories from the literature about human conversational behaviors or skills to inform CA skill design; the referred theories encompass linguistics, sociology, psychology, and cognitive science	[6, 7, 22, 24, 25, 28, 34, 36, 40, 46–48, 52, 55, 67, 71, 72, 76, 82, 83, 87–89, 91, 93, 94, 98, 102, 107–109, 113, 120, 121, 124, 129, 133, 138, 140, 143, 146, 147, 149, 150, 157, 158, 160, 162, 165, 166, 179–181, 187–189, 191]
	Motivate Skill Transfer	Use theories, such as Computers As Social Actors (CASA), as the original motivation for the skill transfer	[25, 40, 67, 71, 82, 87, 93, 94, 102, 107, 108, 120, 129, 146, 147, 150, 166, 189]

Table 5: Primary strategies identified for conversational skill transfers from human-human to human-CA conversations

straightforward approach is to invite participants to directly craft the conversational skills by providing response scripts or templates for the CA. This is for the skills that require more professional expertise to deliver information precisely, such as the phrasing and wording by psychologists in mental health scenarios [92, 115]. Additionally, feedback on the CA skill design from participants is valuable, such as the knowledge confirmed by the experts as a part of the customer service skills [119].

4.3.3 Transferred From Theories. A more prevalent transfer strategy is to derive conversational skills from theories in the literature. The theories about how people converse or communicate with others can directly inform the conversational skill design. Reviewed studies often look into the theories in linguistics, such as the specification of temporal expressions [33, 154] and the communication adaptation theory in pragmatics [161], to design the CA's skills of imperfect temporal statement [145] and pronoun adaptation [133]. A larger number of reviewed studies are also inspired by theories related to psychology, cognitive, and social science. For instance, CAs [3, 120, 174] are equipped with therapy or intervention skills for mental health derived from psychological therapy theories, such as the Cognitive-Behavioral Therapy (CBT) [15]. Similarly, the method of motivational interviewing is referred to when researchers design the CAs with skills of behavior change motivation [160]. Moreover, the design of CA's self-disclosure skills [93, 115, 140] follows the Social Penetration Theory (SPT) [5]. Empathy-related theories [54, 57, 115] are frequently mentioned to inform the empathic response skills design for the CAs [67, 96].

Other than deriving skills from theories in the literature, the reviewed works also find theories as their theoretical grounding for the skill transfer. For instance, they usually mention "Computers Are Social Actors" (CASA) [117, 118], positing that humans tend to exhibit social norms, as they interact with other humans, to computers with limited human characteristics. Many reviewed works [87, 150, 166] narrate it as that based on the CASA theory, if the CA exhibits more humanlikeness, it may trigger more social behaviors of the user, leading to positive effects for the human-CA interaction.

4.4 Skill Implementation

We identified four primary approaches that researchers employ to implement conversational skills for CAs. First, traditional model training equips models with skills via data-driven methods. Second, knowledge base development utilizes structured repositories, such as special databases, for CA information retrieval when executing the skill. Third, large language models (LLMs) handle skill management via their powerful language capabilities. Lastly, the major approach is rule-based, using predefined guidelines for consistent skill execution.

4.4.1 Traditional Model Training. Traditional model training used to be a popular method before the widespread adoption of LLMs, such as in previous works before 2019 [116, 123]. If the training data is available, it is still an effective approach for developing conversational skills in CAs, suitable for scenarios where precise intent recognition and targeted response generation are critical. For example, models are trained to discern user emotions [92] or engagement levels [116] to determine appropriate skill timing and selection. Another purpose is skill generation, including both verbal [3, 96] and nonverbal [174] skill outputs. Additionally, model training enables parameter prediction, which improves interactive behaviors by forecasting values such as optimal back-channel timing to enhance the naturalness [39]. Multiple training approaches are also identified. For example, researchers have fine-tuned RoBERTa to recognize emotions and CDial-GPT to generate empathetic replies, effectively integrating empathy as a skill into the system [96]. Direct training, on the other hand, creates purpose-built models from scratch, as exemplified by SVM classifiers that detect user intents and trigger corresponding back-channel responses [39].

4.4.2 Knowledge Base Development. Knowledge bases serve two primary objectives for skill implementation: enabling efficient information retrieval and supporting contextual decision-making. For

Category	Sub-category	Approach	Description	Examples
Traditional Model Development	Model Objective	Intent Classification	Detect specific user intents or behaviors	[92, 116, 182]
		Entity Extraction	Extract critical entities or keywords from user inputs	[92]
		Skill Generation	Generate CA’s responses or behaviors	[3, 96, 123, 174]
		Skill Retrieval	Retrieve appropriate skills from knowledge bases	[39, 62, 125, 182]
		Parameter Prediction	Predict skill parameters, such as response timing	[6, 39]
	Training Approach	Fine-tuning	Fine-tune traditional models with a amount of data	[3, 62, 96]
		Training	Train model weights with existing data	[3, 39, 125, 174, 182]
		Parameter Transfer	Extract critical features or parameters from data	[39, 123]
Knowledge Base Development	Knowledge Structure	General Database	Organize databases by indexing or pairing for CA to retrieve information to execute skills	[39, 62, 72, 109, 115, 116, 119]
		Structured Knowledge	Organize knowledge as graphs or trees	[48, 64, 99, 182]
LLM-based Implementation	LLM Objective	User Understanding	Identify user needs or update user profiles	[99, 109]
		Skill Control	Manage conversational skills, such as deciding which skill to use or adjusting skill parameters	[64, 97, 99, 100, 109]
		Skill Generation	Directly generate contents to implement skills, such as text responses	[25, 99, 109, 115, 160]
	Special Architecture	Chain-of-thought	Break down skills into several steps which align with the chain-of-thought concept	[64, 100]
		Multi-agent	Implement skills using multiple agents with different roles	[97, 109]
Rule-based Only	Rule Execution	Wizard-of-Oz	Simulate CA skills through human researchers instead of automated systems	[2, 45, 58, 82, 87, 94, 107, 108, 116, 184]
		Predefined Workflows and Responses	Rule-based skill management based on predefined flows or scripted responses	[7, 22, 24, 28, 34, 36, 40, 46, 47, 50, 52, 55, 60, 67, 71, 76, 83, 86, 88, 91, 93, 95, 98, 102, 113, 120, 121, 124, 129, 133, 138, 140, 142, 143, 145–147, 149, 150, 157, 158, 162, 165, 166, 178–181, 185, 187–189, 191]

Table 6: Identified approaches employed by the reviewed studies for CA skill implementation

retrieval, as demonstrated in question-answering systems [119] and back-channel selection [39], domain knowledge databases utilize indexing and pairing mechanisms to provide on-demand access to specific responses. For complex reasoning or decision making for skill implementation, domain knowledge with special structures, such as the state machine used in multi-step dialogue management [182], and the scaffolding strategy knowledge graph in CA-administered tasks [64], organizes information through semantic relationships, facilitating context-aware skill execution and dialogue flow management.

4.4.3 *LLM-based Implementation.*

LLMs have become a popular approach for implementing conversational skills in CAs since the rise of generative technology, particularly in user understanding, skill control, and generation. They provide benefits over earlier methods, such as better adaptability, improved context awareness, and the ability to create more nuanced responses. For instance, researchers have applied LLMs to interpret user preferences and behaviors [109], generate CA’s self-disclosure to encourage user self-disclosure in counseling dialogues [115], and decide when the CA should take the turn [97]. Two prevalent architectures identified are chain-of-thought and multi-agent. The chain-of-thought method decomposes skills into sequential steps, as demonstrated in frameworks like “inner thoughts,” which uses components such as trigger, retrieval, thought formation, evaluation, and participation to manage the CA’s turn-taking skill in group chat context [100]. In contrast, the multi-agent approach employs several specialized LLM-based agents behind the CA, each dedicated to a specific function, for example, separating interruption handling and back-channel mechanisms into distinct modules [97]. Please note that the user should only interact with one CA, though its skill is implemented via the multi-agent approach, differentiating from the special scenario we mentioned in 4.1 where the user converses with multiple CAs.

4.4.4 *Rules-based Only.*

This category includes the most studies that only adopted rule-based approaches for CA skill implementation. One classic execution method for early-stage CA prototypes is the Wizard-of-Oz [41, 73]. In Wizard-of-Oz studies, researchers simulate CA skills through human operators following a structured feedback strategy or protocol, rather than using an automated system. For example, researchers developed a graphic interface for the human operator to control the CA’s embodied feedback in a reviewed paper [184]. In contrast, the other method, also the most common one, employs an automated system operating through fixed event triggers and corresponding responses, which could be described as a series of “if-else” logics. It could be implemented in platforms like Google DialogFlow [86], where developers predefine intents, entities, and response logic within a structured dialogue framework to enable deterministic and rule-based conversational interactions.

4.5 Skill Evaluation

Category	Sub-category	Examples
General Metrics & Usability	Manipulation Check	[25, 76, 107, 121, 129, 149, 162, 180, 189]
	Linguistic Quality	[3, 24, 62, 92, 99, 100, 160, 166, 188]
	Usefulness, Perceived Competence or Intelligence	[6, 7, 58, 87, 94, 100, 107, 108, 116, 119, 140, 142, 145, 149, 160, 165, 178, 180, 187, 188]
	Ease of Uses	[94, 125, 140, 165, 188]
	User Satisfaction and Likeability	[2, 6, 7, 24, 50, 52, 60, 67, 71, 89, 94, 96, 113, 121, 125, 129, 150, 157, 165, 178, 180, 181, 184, 187–189]
	Engagement	[3, 6, 47, 60, 89, 93, 97, 98, 100, 115, 146, 162, 174, 180, 188]
	Naturalness	[3, 6, 109, 116, 158, 165, 174, 188]
	Preference	[47, 48, 52, 92, 100, 102, 125, 133, 145, 165, 188]
	Intention or Willingness to Use	[47, 55, 67, 87, 91, 109, 120, 129, 149, 166]
	General User Experience or Perception	[34, 39, 40, 45, 47, 48, 82, 83, 88, 93, 95, 97, 100, 102, 115, 116, 124, 133, 142, 147, 160, 182, 184, 185, 188]
	Human-CA Conversation Analysis	[2, 24, 45, 46, 50, 62, 64, 82, 88, 93, 95, 98, 99, 107, 109, 115, 116, 121, 125, 142, 143, 174, 188, 191]
Social & Emotional Intelligence	Anthropomorphism & Social Perception	[2, 7, 46, 47, 50, 55, 58, 67, 71, 86, 100, 109, 113, 116, 143, 145, 146, 150, 165, 174, 178, 188]
	Manner, politeness and Friendliness	[119, 123, 191]
	Relationship, alliance or rapport with User	[6, 58, 60, 83, 86, 87, 91, 98, 120, 125, 138, 174]
	Trust or Reliability	[2, 6, 25, 28, 76, 88, 94, 124, 125, 129, 142, 178, 184, 189]
	Perceived Empathy or Warmth	[60, 71, 92, 99, 108, 149, 178, 181, 184, 187]
Skill Quality and Effects	Specific Quality of Skill Content	[6, 39, 40, 64, 71, 99, 100, 108, 174]
	Effects on Behaviour or Mindset Change	[2, 25, 28, 45, 60, 86, 91, 96, 98, 129, 138, 147, 150, 162, 166, 191]
	Effects on User Emotion Improvement	[7, 24, 50, 83, 108, 143, 174, 179]
	Other Task Gains	[6, 40, 48, 62, 87, 93, 121, 124, 142, 181, 184]
	Potential Downsides	[7, 86, 165]

Table 7: Taxonomy of evaluation metrics for CAs and their skills

Last, we summarised the metrics used to assess conversational agents (CAs) and their skills in table 7. The most common category is **general metrics and usability**, assessing the general user experience of the CA. Reviewed studies often check if their manipulations or the implemented skills were effective. For example, a one-way ANOVA was employed to check the CA style difference between two conditions, making the comparison experiment valid [25]. Then, linguistic quality is a traditional metric to evaluate how well the CA can understand what the user says and express itself clearly. Both objective and subjective measurements can be used, such as perplexity and distinct n-grams indicating fluency and informativeness [92], and the questionnaire to assess the perceived factual correctness and relevance of the CA's interview skills [99]. The participants are usually required to compare multiple conditions and choose their preferred CAs or skills, such as ranking two conditions (CAs with or without feedback modalities) by user preference [47]. If there is no specific evaluation aspect to focus on, researchers also use comprehensive scales, questionnaires, or interviews with open-ended questions to collect user feedback. For example, a 34-item questionnaire measured user perception of the smiling CA [82]. In addition, some studies analysed human–CA conversation logs to verify the proper execution of CA skills or examine interaction patterns. For instance, the user's group joining behaviors were analyzed to evaluate the CA's persuasion skill [191].

Beyond general metrics, many studies assess the CA's **social and emotional intelligence**. Perceived empathy is a typical metric for the user to assess if the CA feels their emotions and responds appropriately, such as the artificial empathy questionnaire, including items like if the CA tried to put themselves in my shoes [71]. The last category is **skill quality and effects**. Some skills have special aspects to evaluate, such as skill usage appropriateness [39, 64] and CA response informativeness [40, 99]. The CA skills' effects on tasks and users are also evaluated, such as their effects on behavior or mindset change [60, 166], work efficiency [48], and emotional improvement [50, 143]. Moreover, the potential negative effects, such as taskload measured by NASA Task Load Index (NASATLX) [59].

4.6 Skill Transfer Pathways

To provide practical implications for future CA design, we created the Sankey diagram based on our data distribution in Figure fig. 3. We also provide an interactive website for better presentation[4]. The visualization reveals that the pathway from theories to rule-based implementation currently predominates. This prevalence likely reflects the historical context of the pre-LLM era, as well as the need to use explicit design rules in early design stages to build controlled prototypes for concept validation. While this traditional pathway serves as a mature reference, the currently underutilized flows from data to LLM-based implementation suggest potential for leveraging generative AI to automate complex skill implementation, indicating a promising direction for future exploration.

[4]https://sankey-ca-skill-transfer.vercel.app/

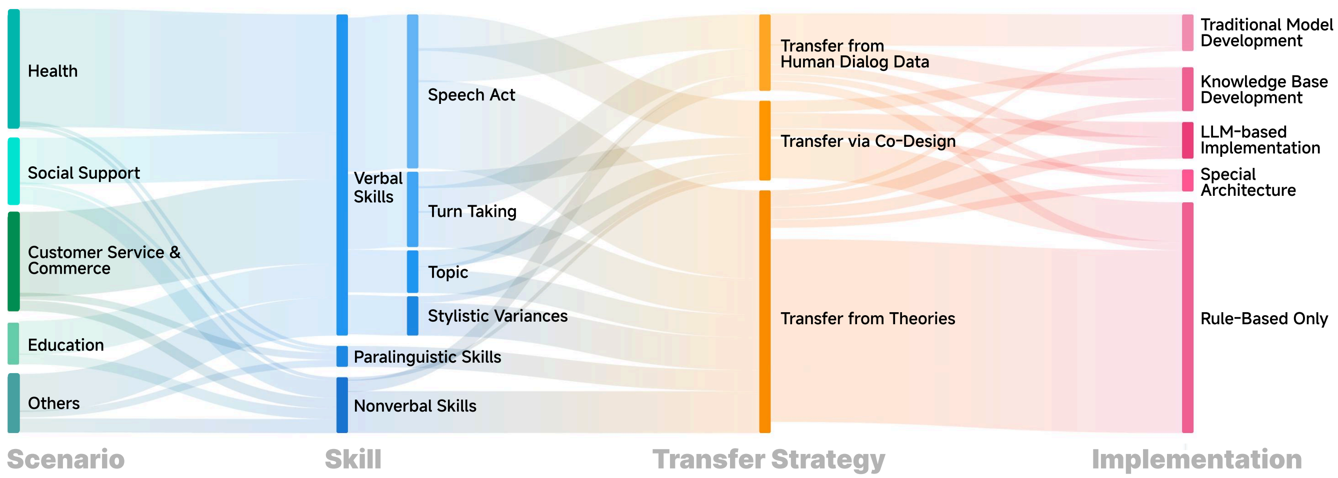

Figure 3: The pathways of CA skill transfer based on the distribution of the reviewed studies. Verbal skills consist of the majority of CA skills. Verbal skills constitute the primary component of CA skills. The predominant mainstream pathway for transferring these skills involves drawing insights from existing theoretical frameworks in the literature and operationalizing them through scripted rules. With the proliferation of high-quality human dialogue datasets and advances in LLMs, we advocate for data-driven pathways while emphasizing the critical role of co-design to uphold a human-centered approach throughout the transfer process.

5 Discussion

5.1 Design Process for Conversational Skill Transfer from Humans to CAs

5.1.1 Double Diamond Process for CA Design. To synthesize a practical method for CA researchers and designers, we propose our design process based on our findings and the Double Diamond process [31, 32]. The process has been widely applied in the domain of design [12, 30, 80]. It is valuable because it combines proven structural logic with adaptability—helping designers solve the right problems, build the right solutions, and navigate the uncertainties of real-world design. Its integration of both divergent and convergent thinking helps designers balance research width and depth, creativity, and implementation. We keep the four stages from the original Double Diamond process and adapt them to the CA design domain: human skill exploration, key skill definition, skill transfer, and iterative evaluation (Figure 4).

5.1.2 Human Skill Exploration. In the first stage, it is a divergent process to explore human conversational skills after the scenario and the goal of conversations are confirmed. To identify the key conversational skills, the literature on how humans or human experts converse in this scenario may provide valuable insights. Referring to the identified strategies in section 4.3, analyzing available dialog data is another option to explore skills adopted by humans in the target scenario. Prior work [131] leverages the methods from linguistics, such as conversation analysis [131], to investigate the detailed dynamics during the conversations, which may inspire the skill design of the CA. Including the conversation analysis, other methods of qualitative analysis, such as thematic analysis [18], can also provide an overview of the skill taxonomy, while the quantitative analysis may provide a statistical perspective, such as identifying skills that are more frequently used by humans in this scenario. Another design factor should be considered at this stage, the CA embodiment, as it scopes the skill types to explore. For example, non-verbal skills such as facial expressions [82, 158, 179] require embodied CAs with a face and adjustable features. Additionally, skills like hand gestures in an elevator pitch scenario[184] or arm gestures for politeness[191] require body parts and skeletons capable of performing body language.

5.1.3 Skill Definition. This convergent stage identifies the key conversational skills among alternatives. Researchers may check what conversational skills have been proven or suggested to be effective in this scenario, or what limitations these skills still have, to motivate the definition of the new or updated key skills in researchers' current works. Next, a systematic analysis and deconstruction of the key skills is necessary to define the design space and factors that are worth further investigation. For instance, when researchers design the CA's back-channeling skills, factors such as back-channeling timing and contents are deemed critical [39, 97], and they propose different solutions to implement the two factors. It is also a way to bridge the research gaps to the previous works. Thus, clear hypotheses or research questions are recommended to lead the design study, as the output of the convergent process of the second stage.

Please note that the literature or the dataset should be culturally appropriate to avoid any cross-cultural pragmatic failures [164]. For instance, we should be careful when designing the CA's gesture skill because one neutral gesture might be considered offensive in another cultural context [77]. Especially for the CAs in cultural scenarios, such as cultural learning [124] and cross-cultural customer service [113], designers may position the cultural factor in

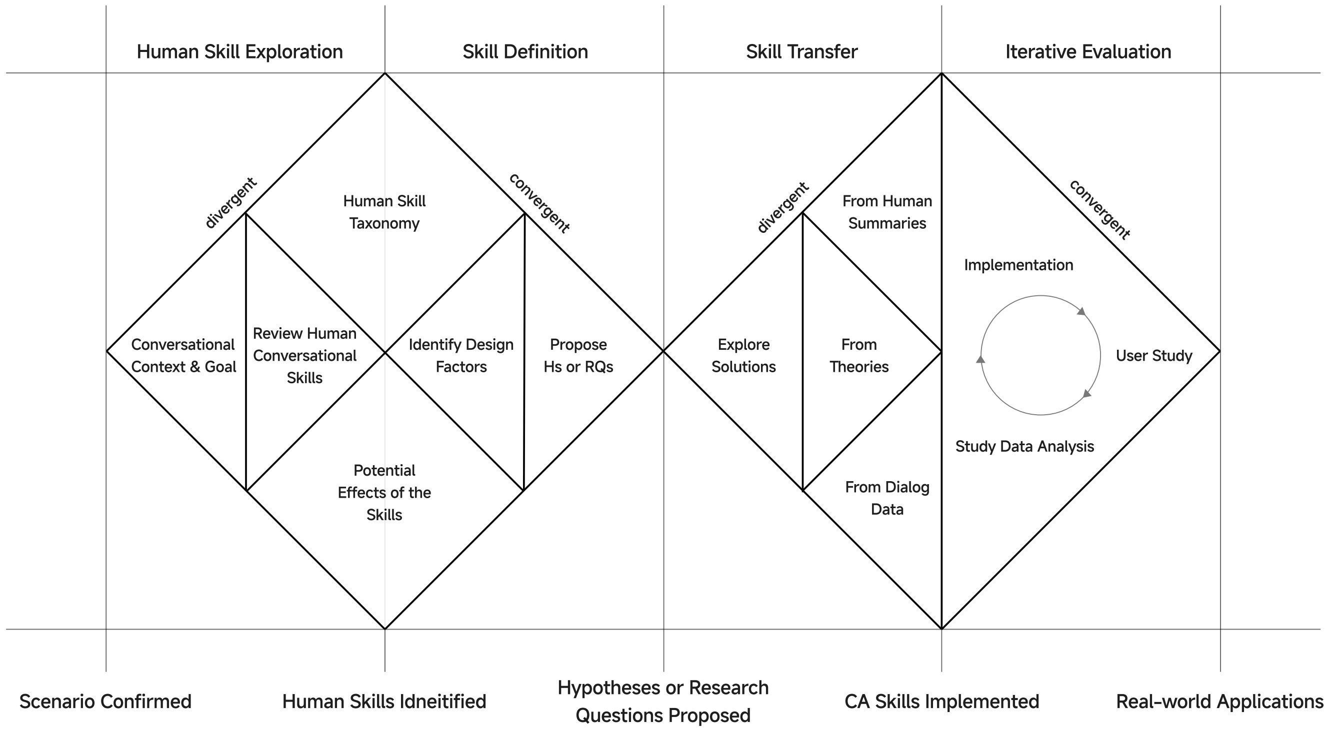

Figure 4: The Double Diamond design process for transferring conversational skills from human dialogues to CAs. The process involves two cycles of divergence and convergence: the first divergence phase explores scenario-specific human conversational skills, followed by convergence to identify core skills aligned with conversational objectives; the second divergence phase investigates diverse resources for skill design and development, and subsequent convergence refines the optimal design through systematic iterative development and testing.

the center of the design process. Sometimes, for generalizable CA skill design, it is applicable to extend the skill definition with an additional study to identify the cultural variations of the human conversational skill [145].

5.1.4 Skill Transfer. Then, it is another divergent process to explore the available solutions to develop the key skill defined in the previous stage. The conversational skill transfer strategies we identified in this paper exhibit a diverse range of complexity. Strategies with relatively low complexity can be implemented by extracting guidelines or rules from the literature and applying them by rule-based LLM prompting or predefined responses. For example, inspired by the doctor-patient communication literature, the reviewed work transfers the affiliative language style to the CA design by prompting the LLM to generate responses in this style [25]. Some early works before the prevalence of LLMs also craft predefined responses for the CA, such as the health care chatbot's professional style verbal skills like professional jargon [120]. These strategies are applied for conversational skills at the scenario level; in other words, the skills can be used generally in the target scenario. However, many conversational skills require specific cues to activate, and the simple scenario-skill matching can not sufficiently describe the skill. Instead, it may consider multiple contextual factors as cues to activate, which requires other transfer strategies. To deal with different types of cues and responding strategies, knowledge graph-based approaches emerge as the key point of the transfer. For instance, to deal with different situations in cognitive disorder screening tasks, the CA is equipped with a knowledge graph of scaffolding skills managed by an LLM in one of the review works [64]. Though the work does not achieve a fully automated CA, leaving the cue understanding (intent classification) for the human wizard, their design method demonstrates a transfer strategy that extracts human skills from dialog data to build a complex knowledge graph as a form to develop the skill for the CA. To break down the architecture, a full process to implement a conversational skill at this cue level includes the cue recognition for the skill activation, the management module to match an appropriate strategy, and finally, a behavior renderer or generation module to output the skill in a concrete form, such as a verbal utterance or a nonverbal cue. To build up such a detailed skill, only extracting the human skill from the literature might be insufficient, but researchers may turn to leverage available dialog data or human resources for skill detail extraction.

This stage is usually a divergent process because researchers develop multiple versions of the skill, such as with or without the skill, the same skill but with different parameters, or different skills for comparison. For instance, the reviewed studies develop both warm and competent language styles as a skill to compare for the customer service CAs [28, 138]. To explore the CA's skill of asking questions, the reviewed work implements three versions of follow-up question skills to compare their effectiveness in information elicitation [62]. When it comes to the next stage, an interactive and convergent process is used to gradually polish and finalize the conversational skill design.

5.1.5 Iterative Evaluation. Though language models have significantly advanced, designing and developing human-level conversational skills for CAs is still a challenge. Thus, we recommend an iterative process to continuously improve the design, aligning with other design literature in HCI [11, 114, 175]. Aside from evaluating the CA with benchmark datasets, usually adopted by works in the natural language processing domain [10, 135], we also suggest the involvement of target human users to evaluate the CA skills. This approach is not limited by the skill coverage of the benchmark databases, e.g., if the databases do not include the target skills, they are inapplicable for evaluation. Moreover, it allows for collecting real human feedback in a near-real-world scenario and detailed dynamics of the human-CA conversation. This is also motivated by the reviewed works that provide design implications as part of the contributions [64, 162, 188]. Additionally, as we mentioned earlier in the skill definition, cultural aspects are also important in the evaluation stage. For instance, the evaluation scale used should fit the cultural background of the users [96]. We suggest involving target users in evaluation or earlier stages and collecting their thoughts and feedback from their specific cultural backgrounds, aligning with the culturally informed CA design practices [110].

5.2 Ethical Considerations

The endeavor to transfer human conversational skills to conversational agents (CAs) is accompanied by significant ethical challenges that demand careful consideration throughout the design, implementation, and evaluation lifecycle [139, 171]. These considerations span the nature of the skills being transferred and the methods for evaluation. A primary concern is the inadvertent transfer of negative, toxic, or hostile human conversational patterns. CAs trained on large, unrestricted datasets from the internet often learn and reproduce undesirable human behaviors such as toxicity and bias [16, 38, 176]. This issue is twofold: CAs can be manipulated by imperceptible triggers to generate toxic language [106], and they can also elicit unethical behavior from users, such as flaming and bullying [35]. Addressing this requires a proactive approach, such as developing specialized datasets and models designed to teach CAs how to respond to problematic content in line with positive social norms [74]. The ethical issues may also be rooted in the unrestricted types of data used for CA development and the often unethical practices of its collection. The reliance on vast corpora of human interaction data from the internet is the primary source of learned toxicity and bias [16]. Thus, the proposed design process does not rely on a single source of dialog data, but also derives conversational skills from controlled resources, such as involving human participants to extract or craft skills. Finally, potential ethical challenges may also arise when evaluating CAs with human participants, especially those from vulnerable populations. Beyond general concerns of accountability and interpretability [170], heightened risks emerge when testing involves groups such as older adults, children, or individuals with mental health disorders [78]. These users may be more susceptible to forming inappropriate attachments, being misled by biased information, or having their trust exploited. This necessitates the creation of strict, tailored guidelines to ensure the development of trustworthy and safe CAs for special user groups. Ultimately, fostering ethical CA development requires a holistic approach focused on calibrating user trust appropriately and embedding responsible principles into the core design process [42].

6 Limitations And Future Work

This systematic literature review presents several limitations. First, the heterogeneity of experimental designs, sample sizes, and outcome measures across studies prevented meaningful quantitative analysis and meta-analytic synthesis of the transfer approach's effectiveness. Second, our LLM-assisted screening, while human-verified, may have introduced selection biases, and our focus on explicitly described transfer processes could have excluded valuable work that implicitly adopted such methods. Future work should address several key areas. Establishing standardized evaluation frameworks with multi-dimensional metrics covering skill effectiveness, user experience, and cross-cultural applicability is essential for meaningful comparison across studies. Cross-cultural expansion is particularly important given the current Western-centric focus of existing research. Finally, as CA capabilities advance, comprehensive ethical frameworks addressing user dependency, privacy protection, and manipulation risks become increasingly critical for responsible skill transfer implementation.

7 Conclusion

This systematic literature review addressed a critical gap in conversational agent research by providing the first comprehensive synthesis of how human conversational skills are transferred to CA designs. Through the analysis of 85 studies, we identified three primary skill modalities, verbal, paralinguistic, and nonverbal, while systematically cataloging the specific skills being transferred across diverse application domains. Our findings reveal three primary transfer strategies: transfer from dialog data and theories, as well as via co-design. We propose the four-stage design process, offering practitioners a structured methodology for implementing conversational skills while maintaining flexibility across different contexts. To summarize, our work contributes comprehensive taxonomies that organize the fragmented landscape of conversational skills research and provides concrete guidance for developers seeking to create more human-like conversational experiences. The research reveals that successful skill transfer requires careful consideration of contextual factors, user characteristics, and cultural differences, while the emerging dominance of LLM-based implementations presents both opportunities for sophisticated skill deployment and challenges in maintaining ethical considerations. As conversational agents become more prevalent across society, appropriate transfer of human conversational skills becomes not merely a design or technical challenge but a social imperative for creating trustworthy and effective human-computer interactions.

Acknowledgments

The work described in this paper was supported in part by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/E-601/24-N and T45-407/19N), as well as by the Global STEM Professorship Scheme of Hong Kong. The authors also wish to express sincere gratitude to Prof. Mo Li for his special support.

References

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 249–258. doi:10.1145/3234695.3236344

[2] Nik Nailah Abdullah, Jia Tang, Hemad Fetrati, Nor Fadhilah Binti Kaukiah, Sahrin Bin Saharudin, Vee Sim Yong, and Chia How Yen. 2025. MARIA (Medical Assistance and Rehabilitation Intelligent Agent) for Medication Adherence in Patients With Heart Failure: Empirical Results From a Wizard of Oz Systematic Conversational Agent Design Clinical Protocol. *JMIR Cardio* 9 (10 Apr 2025), e55846. doi:10.2196/55846

[3] Abdulqahar Mukhtar Abubakar, Deepa Gupta, and Shantipriya Parida. 2024. A Reinforcement Learning Approach for Intelligent Conversational Chatbot for Enhancing Mental Health Therapy. *Procedia Computer Science* 235 (2024), 916–925. doi:10.1016/j.procs.2024.04.087 International Conference on Machine Learning and Data Engineering (ICMLDE 2023).

[4] Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics* 4 (08 2016), 463–476. doi:10.1162/tacl_a_00111

[5] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[6] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. 2012. Designing Effective Gaze Mechanisms for Virtual Agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, Austin, TX, USA, 705–714. doi:10.1145/2207676.2207777

[7] Deepali Aneja, Rens Hoegen, Daniel McDuff, and Mary Czerwinski. 2021. Understanding Conversational and Expressive Style in a Multimodal Embodied Conversational Agent. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama Japan, 1–10. doi:10.1145/3411764.3445708

[8] Shouvik Ahmed Antu, Haiyan Chen, and Cindy K Richards. 2023. Using LLM (Large Language Model) to Improve Efficiency in Literature Review for Undergraduate Research. *Llm@ Aied* (2023), 8–16.

[9] John Langshaw Austin. 1975. *How to do things with words.* Harvard university press. 22–53 pages.

[10] Ge Bai, Jie Liu, Xingyuan Bu, Yancheng He, Jiaheng Liu, Zhanhui Zhou, Zhuoran Lin, Wenbo Su, Tiezheng Ge, Bo Zheng, et al. 2024. MT-Bench-101: A Fine-Grained Benchmark for Evaluating Large Language Models in Multi-Turn Dialogues. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 7421–7454.

[11] Gregg "Skip" Bailey. 1993. Iterative methodology and designer training in human-computer interface design. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 198–205. doi:10.1145/169059.169163

[12] Annie Banbury, Sonja Pedell, Lynne Parkinson, and Louise Byrne. 2021. Using the Double Diamond model to co-design a dementia caregivers telehealth peer support program. *Journal of Telemedicine and Telecare* 27, 10 (2021), 667–673. doi:10.1177/1357633X211048980 PMID: 34726994.

[13] Vince Bartle, Liam Albright, and Nicola Dell. 2023. "This machine is for the aides": Tailoring Voice Assistant Design to Home Health Care Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 227, 19 pages. doi:10.1145/3544548.3581118

[14] Rodrigo Bavaresco, Diórgenes Silveira, Eduardo Reis, Jorge Barbosa, Rodrigo Righi, Cristiano Costa, Rodolfo Antunes, Marcio Gomes, Clauter Gatti, Mariangela Vanzin, Saint Clair Junior, Elton Silva, and Carlos Moreira. 2020. Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review* 36 (2020), 100239. doi:10.1016/j.cosrev.2020.100239

[15] Judith S Beck. 2020. *Cognitive behavior therapy: Basics and beyond.* Guilford Publications.

[16] A. Stevie Bergman, Gavin Abercrombie, Shannon Spruit, Dirk Hovy, Emily Dinan, Y-Lan Boureau, and Verena Rieser. 2022. Guiding the Release of Safer E2E Conversational AI through Value Sensitive Design. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Oliver Lemon, Dilek Hakkani-Tur, Junyi Jessy Li, Arash Ashrafzadeh, Daniel Hernández Garcia, Malihe Alikhani, David Vandyke, and Ondřej Dušek (Eds.). Association for Computational Linguistics, Edinburgh, UK, 39–52. doi:10.18653/v1/2022.sigdial-1.4

[17] Jan Hendrik Blümel, Mohamed Zaki, and Thomas Bohné. 2024. Personal touch in digital customer service: a conceptual framework of relational personalization for conversational AI. *Journal of Service Theory and Practice* 34, 1 (2024), 33–65. doi:10.1108/JSTP-03-2023-0098 Publisher: Emerald Publishing Limited.

[18] Robert Bowman, Camille Nadal, Kellie Morrissey, Anja Thieme, and Gavin Doherty. 2023. Using Thematic Analysis in Healthcare HCI at CHI: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 491, 18 pages. doi:10.1145/3544548.3581203

[19] Bonnie Brinton and Martin Fujiki. 1984. Development of Topic Manipulation Skills in Discourse. *Journal of Speech, Language, and Hearing Research* 27, 3 (1984), 350–358. doi:10.1044/jshr.2703.350

[20] Angelo Cafaro, Johannes Wagner, Tobias Baur, Soumia Dermouche, Mercedes Torres Torres, Catherine Pelachaud, Elisabeth André, and Michel Valstar. 2017. The NoXi database: multimodal recordings of mediated novice-expert interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 350–359. doi:10.1145/3136755.3136780

[21] Gillian Cameron, David Cameron, Gavin Megaw, RR Bond, Maurice Mulvenna, Siobhan O'Neill, C Armour, and Michael McTear. 2018. Back to the Future: Lessons from Knowledge Engineering Methodologies for Chatbot Design and Development. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI-2018)*, Raymond Bond, Maurice Mulvenna, Jonathan Wallace, and Michaela Black (Eds.). BCS Learning & Development Ltd. doi:10.14236/ewic/HCI2018.153

[22] Che-Jui Chang, Samuel S Sohn, Sen Zhang, Rajath Jayashankar, Muhammad Usman, and Mubbasir Kapadia. 2023. The Importance of Multimodal Emotion Conditioning and Affect Consistency for Embodied Conversational Agents. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, Sydney NSW Australia, 790–801. doi:10.1145/3581641.3584045

[23] Ana Paula Chaves and Marco Aurelio Gerosa and. 2021. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758. doi:10.1080/10447318.2020.1841438

[24] Pei-Yu Chen, M. Birna van Riemsdijk, Dirk K. J. Heylen, Catholijn M. Jonker, and Myrthe L. Tielman. 2025. Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior. *Behaviour & Information Technology* 0, 0 (2025), 1–20. doi:10.1080/0144929X.2025.2483793

[25] Zhiyun Chen, Xinyue Zhao, Min Hua, and Jian Xu. 2025. Building Bonds Through Bytes: The Impact of Communication Styles on Patient-Chatbot Relationships and Treatment Adherence in AI-Driven Healthcare. In *HCI International 2024 – Late Breaking Papers*, Vincent G. Duffy (Ed.). Springer Nature Switzerland, Cham, 32–52.

[26] Han Shi Jocelyn Chew. 2022. The Use of Artificial Intelligence–Based Conversational Agents (Chatbots) for Weight Loss: Scoping Review and Practical Recommendations. *JMIR Medical Informatics* 10, 4 (April 2022), e32578. doi:10.2196/32578

[27] Hyojin Chin, Lebogang Wame Molefi, and Mun Yong Yi. 2020. Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376461

[28] Sunhwa Choi and Jieru Zhou. 2023. Inducing consumers' self-disclosure through the fit between Chatbot's interaction styles and regulatory focus. *Journal of Business Research* 166 (Nov. 2023), 114127. doi:10.1016/j.jbusres.2023.114127

[29] Domenic V. Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6, 4 (1994), 284–290. doi:10.1037/1040-3590.6.4.284

[30] Stephen J. Clune and Simon Lockrey. 2014. Developing environmental sustainability strategies, the Double Diamond method of LCA and design thinking: a case study from aged care. *Journal of Cleaner Production* 85 (2014), 67–82. doi:10.1016/j.jclepro.2014.02.003 Special Volume: Making Progress Towards More Sustainable Societies through Lean and Green Initiatives.

[31] Design Council. 2005. The 'double diamond' design process model. *Design Council* 2, 1 (2005). https://www.designcouncil.org.uk/our-resources/the-double-diamond/ Accessed: 2025-09-10.

[32] Design Council. 2015. Design methods for developing services. *Design Council* (2015). https://www.designcouncil.org.uk/our-resources/archive/reports-

resources/design-methods-developing-services/ Accessed: 2025-09-10.
[33] David Crystal. 1966. Specification and English tenses. *Journal of Linguistics* 2, 1 (1966), 1–34. doi:10.1017/S0022226700001304
[34] Andrea Cuadra, Shuran Li, Hansol Lee, Jason Cho, and Wendy Ju. 2021. My Bad! Repairing Intelligent Voice Assistant Errors Improves Interaction. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (2021), 27:1–27:24. doi:10.1145/3449101
[35] Antonella De Angeli. 2009. Ethical implications of verbal disinhibition with conversational agents. *PsychNology Journal* 7, 1 (2009).
[36] Berardina De Carolis, Irene Mazzotta, and Nicole Novielli. 2010. Enhancing Conversational Access to Information through a Socially Intelligent Agent. In *Intelligent Information Access*, Janusz Kacprzyk, Giuliano Armano, Marco De Gemmis, Giovanni Semeraro, and Eloisa Vargiu (Eds.). Vol. 301. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–20. doi:10.1007/978-3-642-14000-6_1
[37] Stephan Diederich, Alfred Benedikt Brendel, Stefan Morana, and Lutz Kolbe. 2022. On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems* 23, 1 (2022), 96–138.
[38] Emily Dinan, Gavin Abercrombie, A Stevie Bergman, Shannon Spruit, Dirk Hovy, Y-Lan Boureau, and Verena Rieser. 2021. Anticipating safety issues in e2e conversational ai: Framework and tooling. *arXiv preprint arXiv:2107.03451* (2021).
[39] Zijian Ding, Jiawen Kang, Tinky Oi Ting HO, Ka Ho Wong, Helene H Fung, Helen Meng, and Xiaojuan Ma. 2022. TalkTive: a Conversational Agent Using Backchannels to Engage Older Adults in Neurocognitive Disorders Screening. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 304, 19 pages. doi:10.1145/3491102.3502005
[40] Hyo Jin Do, Ha-Kyung Kong, Jaewook Lee, and Brian P. Bailey. 2022. How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 387 (Nov. 2022), 23 pages. doi:10.1145/3555112
[41] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J.D. Bolter, and M. Gandy. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (2005), 18–26. doi:10.1109/MPRV.2005.93
[42] Mateusz Dubiel, Sylvain Daronnat, and Luis A. Leiva. 2022. Conversational Agents Trust Calibration: A User-Centred Perspective to Design. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 30, 6 pages. doi:10.1145/3543829.3544518
[43] Starkey Duncan and Donald W Fiske. 2015. *Face-to-face interaction: Research, methods, and theory.* Routledge.
[44] Faruk Lawal Ibrahim Dutsinma, Debajyoti Pal, Suree Funilkul, and Jonathan H. Chan. 2022. A Systematic Review of Voice Assistant Usability: An ISO 9241–11 Approach. *SN Computer Science* 3, 4 (May 2022), 267. doi:10.1007/s42979-022-01172-3
[45] Toni V. Earle-Randell, Joseph B. Wiggins, Yingbo Ma, Mehmet Celepkolu, Dolly Bounajim, Zhikai Gao, Julianna Martinez Ruiz, Kristy Elizabeth Boyer, Maya Israel, Collin F. Lynch, and Eric Wiebe. 2024. The impact of near-peer virtual agents on computer science attitudes and collaborative dialogue. *International Journal of Child-Computer Interaction* 40 (June 2024), 100646. doi:10.1016/j.ijcci.2024.100646
[46] Jonathan Ehret, Andrea Bönsch, Patrick Nossol, Cosima A. Ermert, Chinthusa Mohanathasan, Sabine J. Schlittmeier, Janina Fels, and Torsten W. Kuhlen. 2023. Who's next? Integrating Non-Verbal Turn-Taking Cues for Embodied Conversational Agents. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, Würzburg Germany, 1–8. doi:10.1145/3570945.3607312
[47] Morad Elfleet and Mathieu Chollet. 2024. Investigating the Impact of Multimodal Feedback on User-Perceived Latency and Immersion with LLM-Powered Embodied Conversational Agents in Virtual Reality. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents (IVA '24)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3652988.3673965
[48] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S. Bernstein. 2018. Iris: A Conversational Agent for Complex Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3174047
[49] Yujao Feng. 2025. The Evaluation of AI Translation Competence in Ceramic Texts through the Lens of Knowledge Translation Theory: A Study of ChatGPT, DeepSeek, and Other Tools. *Academic Journal of Humanities & Social Sciences* 8, 8 (2025). doi:10.25236/AJHSS.2025.080818
[50] Libby Ferland, Risako Owan, Zachary Kunkel, Hannah Qu, Maria Gini, and Wilma Koutstaal. 2025. What Are You Looking Forward to? Deliberate Positivity as a Promising Strategy for Conversational Agents. *ACM Transactions on Interactive Intelligent Systems* 15, 3 (Sept. 2025), 1–40. doi:10.1145/3725738
[51] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (06 Jun 2017), e19. doi:10.2196/mental.7785
[52] Mina Foosherian, Samuel Kernan Freire, Evangelos Niforatos, Karl A. Hribernik, and Klaus-Dieter Thoben. 2022. Break, Repair, Learn, Break Less: Investigating User Preferences for Assignment of Divergent Phrasing Learning Burden in Human-Agent Interaction to Minimize Conversational Breakdowns. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia (MUM '22)*. Association for Computing Machinery, New York, NY, USA, 151–158. doi:10.1145/3568444.3568454
[53] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 1371–1374. doi:10.1145/3209978.3210183
[54] Gerald A. Gladstein. 1983. Understanding empathy: Integrating counseling, developmental, and social psychology perspectives. *Journal of Counseling Psychology* 30, 4 (1983), 467–482. doi:10.1037/0022-0167.30.4.467
[55] Ulrich Gnewuch, Stefan Morana, Marc T. P. Adam, and Alexander Maedche. 2022. Opposing Effects of Response Time in Human–Chatbot Interaction: The Moderating Role of Prior Experience. *Business & Information Systems Engineering* 64, 6 (Dec. 2022), 773–791. doi:10.1007/s12599-022-00755-x
[56] Ece Gumusel. 2025. A literature review of user privacy concerns in conversational chatbots: A social informatics approach: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology* 76, 1 (2025), 121–154. doi:10.1002/asi.24898 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24898
[57] James T Hardee. 2003. An overview of empathy. *The Permanente Journal* 7, 4 (2003), 51.
[58] Kerstin S. Haring, Jessica Tobias, Justin Waligora, Elizabeth Phillips, Nathan L. Tenhundfeld, Gale Lucas, Ewart J. De Visser, Jonathan Gratch, and Chad Tossel. 2019. Conflict Mediation in Human-Machine Teaming: Using a Virtual Agent to Support Mission Planning and Debriefing. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New Delhi, India, 1–7. doi:10.1109/RO-MAN46459.2019.8956414
[59] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9
[60] Linwei He, Erkan Basar, Emiel Krahmer, Reinout Wiers, and Marjolijn Antheunis. 2024. Effectiveness and User Experience of a Smoking Cessation Chatbot: Mixed Methods Study Comparing Motivational Interviewing and Confrontational Counseling. *Journal of Medical Internet Research* 26 (Aug. 2024), e53134. doi:10.2196/53134
[61] Quan Nha Hong, Sergi Fàbregues, Gillian Bartlett, Felicity Boardman, Margaret Cargo, Pierre Dagenais, Marie-Pierre Gagnon, Frances Griffiths, Belinda Nicolau, Alicia O'Cathain, Marie-Claude Rousseau, Isabelle Vedel, and Pierre Pluye. 2018. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information* 34, 4 (2018), 285–291. doi:10.3233/EFI-180221
[62] Jiaxiong Hu, Jingya Guo, Ningjing Tang, Xiaojuan Ma, Yuan Yao, Changyuan Yang, and Yingqing Xu. 2024. Designing the Conversational Agent: Asking Follow-up Questions for Information Elicitation. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 43 (April 2024), 30 pages. doi:10.1145/3637320
[63] Jiaxiong Hu, Yun Huang, Xiaozhu Hu, and Yingqing Xu. 2023. The Acoustically Emotion-Aware Conversational Agent With Speech Emotion Recognition and Empathetic Responses. *IEEE Transactions on Affective Computing* 14, 1 (2023), 17–30. doi:10.1109/TAFFC.2022.3205919
[64] Jiaxiong Hu, Junze Li, Yuhang Zeng, Dongjie Yang, Danxuan Liang, Helen Meng, and Xiaojuan Ma. 2024. Designing Scaffolding Strategies for Conversational Agents in Dialog Task of Neurocognitive Disorders Screening. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu HI USA, 1–21. doi:10.1145/3613904.3642960
[65] Yuanhui Huang, Quan Zhou, and Anne Marie Piper. 2025. Designing Conversational AI for Aging: A Systematic Review of Older Adults' Perceptions and Needs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 181, 20 pages. doi:10.1145/3706598.3713578
[66] Andreas Janson. 2023. How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior* 149 (2023), 107954. doi:10.1016/j.chb.2023.107954
[67] Tingting Jiang, Chuxuan Huang, Yanrun Xu, and Han Zheng. 2025. Cognitive vs. emotional empathy: exploring their impact on user outcomes in health-assistant chatbots. *Behaviour & Information Technology* (2025), 1–16. doi:10.1080/0144929X.2025.2474087 Publisher: Taylor & Francis.
[68] Cameron Jones and Ben Bergen. 2024. Does GPT-4 pass the Turing test?. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.).

Association for Computational Linguistics, Mexico City, Mexico, 5183–5210. doi:10.18653/v1/2024.naacl-long.290
[69] Cameron R Jones and Benjamin K Bergen. 2025. Large language models pass the Turing test. doi:10.48550/arXiv.2503.23674 arXiv:2503.23674
[70] Cameron Robert Jones, Ishika Rathi, Sydney Taylor, and Benjamin K. Bergen. 2025. People cannot distinguish GPT-4 from a human in a Turing test. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 1615–1639. doi:10.1145/3715275.3732108
[71] Antoine Juquelier, Ingrid Poncin, and Simon Hazée. 2025. Empathic chatbots: A double-edged sword in customer experiences. *Journal of Business Research* 188 (Feb. 2025), 115074. doi:10.1016/j.jbusres.2024.115074
[72] Kaveh Kamyab, Frank Guerin, Petar Goulev, and Ebrahim Mamdani. 2001. Designing Agents for a Virtual Marketplace. *AISB Journal* 1, 1 (2001), 61–85.
[73] J. F. Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.* 2, 1 (Jan. 1984), 26–41. doi:10.1145/357417.357420
[74] Hyunwoo Kim, Youngjae Yu, Liwei Jiang, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. 2022. ProsocialDialog: A Prosocial Backbone for Conversational Agents. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 4005–4029. doi:10.18653/v1/2022.emnlp-main.267
[75] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300316
[76] Taenyun Kim and Hayeon Song. 2021. How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. *Telematics and Informatics* 61 (Aug. 2021), 101595. doi:10.1016/j.tele.2021.101595
[77] Sotaro Kita. 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes* 24, 2 (2009), 145–167. doi:10.1080/01690960802586188
[78] Keita Kiuchi, Kouyou Otsu, and Yugo Hayashi and. 2024. Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: a systematic meta-review. *Behaviour & Information Technology* 43, 15 (2024), 3696–3736. doi:10.1080/0144929X.2023.2286528
[79] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The Personalization of Conversational Agents in Health Care: Systematic Review. *J Med Internet Res* 21, 11 (7 Nov 2019), e15360. doi:10.2196/15360
[80] Magda Kochanowska, Weronika Rochacka Gagliardi, and with reference to Jonathan Ball. 2022. *The Double Diamond Model: In Pursuit of Simplicity and Flexibility*. Springer International Publishing, Cham, 19–32. doi:10.1007/978-3-030-79879-6_2
[81] Elisa Konya-Baumbach, Miriam Biller, and Sergej von Janda. 2023. Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior* 139 (2023), 107513. doi:10.1016/j.chb.2022.107513
[82] Nicole Krämer, Stefan Kopp, Christian Becker-Asano, and Nicole Sommer. 2013. Smile and the world will smile with you—The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies* 71, 3 (March 2013), 335–349. doi:10.1016/j.ijhcs.2012.09.006
[83] Itaru Kuramoto, Jun Baba, Kohei Ogawa, Yuichiro Yoshikawa, Takayuki Kawabata, and Hiroshi Ishiguro. 2018. Conversational Agents to Suppress Customer Anger in Text-based Customer-support Conversations. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI '18)*. Association for Computing Machinery, New York, NY, USA, 114–121. doi:10.1145/3284432.3284457
[84] Sheetal Kusal, Shruti Patil, Jyoti Choudrie, Ketan Kotecha, Sashikala Mishra, and Ajith Abraham. 2022. AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. *IEEE Access* 10 (2022), 92337–92356. doi:10.1109/ACCESS.2022.3201144
[85] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (Sept. 2018), 1248–1258. doi:10.1093/jamia/ocy072
[86] Jieon Lee, Daeho Lee, and Jae-gil Lee. 2024. Influence of Rapport and Social Presence with an AI Psychotherapy Chatbot on Users' Self-Disclosure. *International Journal of Human–Computer Interaction* 40, 7 (April 2024), 1620–1631. doi:10.1080/10447318.2022.2146227
[87] Sangwon Lee, Naeun Lee, and Young June Sah. 2020. Perceiving a Mind in a Chatbot: Effect of Mind Perception and Social Cues on Co-presence, Closeness, and Intention to Use. *International Journal of Human–Computer Interaction* 36, 10 (June 2020), 930–940. doi:10.1080/10447318.2019.1699748
[88] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (2020), 31:1–31:27. doi:10.1145/3392836
[89] Steve Leung, Sandeep Virwaney, Fuhua Lin, Aj Armstrong, and Adien Dubbelboer. 2013. TSI-Enhanced Pedagogical Agents to Engage Learners in Virtual Worlds:. *International Journal of Distance Education Technologies* 11, 1 (Jan. 2013), 1–13. doi:10.4018/jdet.2013010101
[90] Stephen C Levinson. 1983. *Pragmatics*. Cambridge university press.
[91] Meichan Li and Rui Wang. 2023. Chatbots in e-commerce: The effect of chatbot language style on customers' continuance usage intention and attitude toward brand. *Journal of Retailing and Consumer Services* 71 (March 2023), 103209. doi:10.1016/j.jretconser.2022.103209
[92] Yanran Li, Ke Li, Hongke Ning, Xiaoqiang Xia, Yalong Guo, Chen Wei, Jianwei Cui, and Bin Wang. 2021. Towards an Online Empathetic Chatbot with Emotion Causes. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 2041–2045. doi:10.1145/3404835.3463042
[93] Kai-Hui Liang, Weiyan Shi, Yoo Jung Oh, Hao-Chuan Wang, Jingwen Zhang, and Zhou Yu. 2024. Dialoging Resonance in Human-Chatbot Conversation: How Users Perceive and Reciprocate Recommendation Chatbot's Self-Disclosure Strategy. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 200 (April 2024), 28 pages. doi:10.1145/3653691
[94] Jongmoon Lim and Wonil Hwang. 2025. Proactivity of Chatbots, Task Types and User's Characteristics When Interacting with Artificial Intelligence (AI) Chatbots. *International Journal of Human–Computer Interaction* (Jan. 2025), 1–19. doi:10.1080/10447318.2024.2446510
[95] Michael Pin-Chuan Lin, Daniel H. Chang, and Philip H. Winne. 2025. A proposed methodology for investigating student-chatbot interaction patterns in giving peer feedback. *Educational technology research and development* 73, 1 (2025), 353–386. doi:10.1007/s11423-024-10408-3
[96] Shuya Lin, Lingfeng Lin, Cuiqin Hou, Baijun Chen, Jianfeng Li, and Shiguang Ni. 2023. Empathy-Based communication Framework for Chatbots: A Mental Health Chatbot Application and Evaluation. In *International Conference on Human-Agent Interaction*. Association for Computing Machinery, Gothenburg Sweden, 264–272. doi:10.1145/3623809.3623865
[97] Chao Liu, Mingyang Su, Yan Xiang, Yuru Huang, Yiqian Yang, Kang Zhang, and Mingming Fan. 2025. Toward Enabling Natural Conversation with Older Adults via the Design of LLM-Powered Voice Agents that Support Interruptions and Backchannels. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 163, 22 pages. doi:10.1145/3706598.3714228
[98] Chen-Chung Liu, Mo-Gang Liao, Chia-Hui Chang, and Hung-Ming Lin. 2022. An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education* 189 (Nov. 2022), 104576. doi:10.1016/j.compedu.2022.104576
[99] Dingdong Liu, Yujing Zhang, Bolin Zhao, Shuai Ma, Chuhan Shi, and Xiaojuan Ma. 2025. Scaffolded Turns and Logical Conversations: Designing Humanized LLM-Powered Conversational Agents for Hospital Admission Interviews. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 643, 23 pages. doi:10.1145/3706598.3714196
[100] Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, and Xiang 'Anthony' Chen. 2025. Proactive Conversational Agents with Inner Thoughts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 184, 19 pages. doi:10.1145/3706598.3713760
[101] Xufei Luo, Fengxian Chen, Di Zhu, Ling Wang, Zijun Wang, Hui Liu, Meng Lyu, Ye Wang, Qi Wang, and Yaolong Chen. 2024. Potential Roles of Large Language Models in the Production of Systematic Reviews and Meta-Analyses. *J Med Internet Res* 26 (25 Jun 2024), e56780. doi:10.2196/56780
[102] Qianli Ma, Yaping Zhang, Wenti Xu, and Ronggang Zhou. 2024. Ask a Further Question or Give a List? How Should Conversational Agents Reply to Users' Uncertain Queries. *International Journal of Human–Computer Interaction* 40, 5 (March 2024), 1087–1101. doi:10.1080/10447318.2022.2131265
[103] Marcello M. Mariani, Novin Hashemi, and Jochen Wirtz. 2023. Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research* 161 (2023), 113838. doi:10.1016/j.jbusres.2023.113838
[104] Kentaro Matsui, Tomohiro Utsumi, Yumi Aoki, Taku Maruki, Masahiro Takeshima, and Yoshikazu Takaesu. 2024. Human-Comparable Sensitivity of Large Language Models in Identifying Eligible Studies Through Title and Abstract Screening: 3-Layer Strategy Using GPT-3.5 and GPT-4 for Systematic Reviews. *J Med Internet Res* 26 (16 Aug 2024), e52758. doi:10.2196/52758
[105] Kenneth O. McGraw and S. P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1, 1 (1996), 30–46. doi:10.1037/1082-989X.1.1.30

[106] Ninareh Mehrabi, Ahmad Beirami, Fred Morstatter, and Aram Galstyan. 2022. Robust Conversational Agents against Imperceptible Toxicity Triggers. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 2831–2847. doi:10.18653/v1/2022.naacl-main.204

[107] Jingbo Meng, Stephen A. Rains, Jiaqi Qin, and Minjin Rheu. 2025. Examining the Content and Form of Supportive Conversations with Chatbots. *International Journal of Human–Computer Interaction* (2025), 1–13. doi:10.1080/10447318.2025.2467450 Publisher: Taylor & Francis.

[108] Jingbo Meng, Minjin (Mj) Rheu, Yue Zhang, Yue Dai, and Wei Peng. 2023. Mediated Social Support for Distress Reduction: AI Chatbots vs. Human. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–25. doi:10.1145/3579505

[109] Seiya Mitsuno, Midori Ban, Tsunehiro Arimoto, Hiroaki Sugiyama, Hiroshi Ishiguro, and Yuichiro Yoshikawa. 2026. Non-Real-Time Chatbot: Improving Anthropomorphism and Realizing Long-Term Human–Chatbot Interaction. *International Journal of Human–Computer Interaction* 42, 1 (2026), 553–566. doi:10.1080/10447318.2025.2508316

[110] Mateus de Souza Monteiro and Luciana Cardoso de Castro Salgado. 2023. Conversational agents: a survey on culturally informed design practices. *Journal on Interactive Systems* 14, 1 (Mar. 2023), 33–46. doi:10.5753/jis.2023.2961

[111] Robert J Moore. 2019. *Natural conversation framework.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3304087.3304092

[112] Christos N. Moridis and Anastasios A. Economides. 2012. Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions. *IEEE Transactions on Affective Computing* 3, 3 (2012), 260–272. doi:10.1109/T-AFFC.2012.6

[113] Sophia Mueller, Tom Kelleher, and Yusuke Ibuki. 2024. U.S. and Japanese Consumer Attitudes Toward Tailored and Targeted Communication with Human and Chatbot Agents. *Journal of Interactive Advertising* 24, 1 (Jan. 2024), 84–101. doi:10.1080/15252019.2024.2303008

[114] Brad Myers. 1994. Challenges of HCI design and implementation. *Interactions* 1, 1 (Jan. 1994), 73–83. doi:10.1145/174800.174808

[115] Alba María Mármol-Romero, Manuel García-Vega, Miguel Ángel García-Cumbreras, and Arturo Montejo-Ráez. 2025. An Empathic GPT-Based Chatbot to Talk About Mental Disorders With Spanish Teenagers. *International Journal of Human–Computer Interaction* 41, 7 (2025), 3957–3973. doi:10.1080/10447318.2024.2344355

[116] Yukiko I. Nakano and Ryo Ishii. 2010. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (Hong Kong, China) *(IUI '10)*. Association for Computing Machinery, New York, NY, USA, 139–148. doi:10.1145/1719970.1719990

[117] Clifford Nass, Youngme Moon, B. J. Fogg, Byron Reeves, and Chris Dryer. 1995. Can computer personalities be human personalities?. In *Conference Companion on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. Association for Computing Machinery, New York, NY, USA, 228–229. doi:10.1145/223355.223538

[118] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '94)*. Association for Computing Machinery, Boston, Massachusetts, USA, 72–78. doi:10.1145/191666.191703

[119] Eric W.T. Ngai, Maggie C.M. Lee, Mei Luo, Patrick S.L. Chan, and Tenglu Liang. 2021. An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications* 50 (2021), 101098. doi:10.1016/j.elerap.2021.101098

[120] Marcia Nißen, Dominik Rüegger, Mirjam Stieger, Christoph Flückiger, Mathias Allemand, Florian V Wangenheim, and Tobias Kowatsch. 2022. The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study. *Journal of Medical Internet Research* 24, 4 (April 2022), e32630. doi:10.2196/32630

[121] Farnaz Nouraei, Keith Rebello, Mina Fallah, Prasanth Murali, Haley Matuszak, Valerie Jap, Andrea Parker, Michael Paasche-Orlow, and Timothy Bickmore. 2025. Virtual Agent-based Communication Skills Training to Facilitate Health Persuasion Among Peers. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (May 2025), 1–24. doi:10.1145/3711101

[122] Björn Nykvist, Biljana Macura, Maria Xylia, and Erik Olsson. 2025. Testing the utility of GPT for title and abstract screening in environmental systematic evidence synthesis. *Environmental Evidence* 14, 1 (23 Apr 2025), 7. doi:10.1186/s13750-025-00360-x

[123] Catharine Oertel, Joakim Gustafson, and Alan W. Black. 2016. On Data Driven Parametric Backchannel Synthesis for Expressing Attentiveness in Conversational Agents. In *Proceedings of the Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*. Association for Computing Machinery, Tokyo Japan, 43–47. doi:10.1145/3011263.3011272

[124] Amy Ogan, Vincent Aleven, Julia Kim, and Christopher Jones. 2010. Developing Interpersonal Relationships with Virtual Agents through Social Instructional Dialog. In *Intelligent Virtual Agents*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova (Eds.). Vol. 6356. Springer Berlin Heidelberg, Berlin, Heidelberg, 236–249. doi:10.1007/978-3-642-15892-6_25

[125] Stefan Olafsson, Byron Wallace, and Timothy Bickmore. 2020. Towards a Computational Framework for Automating Substance Use Counseling with Virtual Agents. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 966–974.

[126] Roland Oruche, Sai Keerthana Goruganthu, Rithika Akula, Xiyao Cheng, Ashraful Md Goni, Bruce W. Shibo, Kerk Kee, Marcos Zampieri, and Prasad Calyam. 2025. A Survey on the Recent Advancements in Human-Centered Dialog Systems. *Comput. Surveys* 57, 10 (Oct. 2025), 1–36. doi:10.1145/3729220

[127] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). doi:10.1136/bmj.n71

[128] Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. "Mhm..." – Conversational Strategies For Product Search Assistants. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 36–46. doi:10.1145/3498366.3505809

[129] EunHae (Grace) Park. 2025. I Trust You, but Let Me Talk to AI: The Role of the Chat Agents, Empathy, and Health Issues in Misinformation Guidance. *International Journal of Strategic Communication* 19, 2 (March 2025), 231–260. doi:10.1080/1553118X.2025.2462087

[130] Zhenhui Peng and Xiaojuan Ma. 2019. A survey on construction and enhancement methods in service chatbots design. *CCF Transactions on Pervasive Computing and Interaction* 1, 3 (Nov. 2019), 204–223. doi:10.1007/s42486-019-00012-3

[131] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3174214

[132] Carol A. Prutting and Diane M. Kittchner. 1987. A Clinical Appraisal of the Pragmatic Aspects of Language. *Journal of Speech and Hearing Disorders* 52, 2 (1987), 105–119. doi:10.1044/jshd.5202.105

[133] Jianhong Qu, Ronggang Zhou, and Zhe Chen. 2022. The effect of personal pronouns on users and the social role of conversational agents. *Behaviour & Information Technology* 41, 16 (2022), 3470–3486. doi:10.1080/0144929X.2021.1999500

[134] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International journal of human-computer studies* 151 (July 2021), 102630. doi:10.1016/j.ijhcs.2021.102630

[135] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 5370–5381. doi:10.18653/v1/P19-1534

[136] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo and. 2021. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96. doi:10.1080/10447318.2020.1807710

[137] Deborah Richards and Karla Bransky. 2014. ForgetMeNot: What and how users expect intelligent virtual agents to recall and forget personal conversational content. *International Journal of Human-Computer Studies* 72, 5 (2014), 460–476. doi:10.1016/j.ijhcs.2014.01.005

[138] Rajat Roy and Vik Naidoo. 2021. Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research* 126 (March 2021), 23–34. doi:10.1016/j.jbusres.2020.12.051

[139] Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations. *AICS* 2563 (2019), 104–115.

[140] John Rudnik, Sharadhi Raghuraj, Mingyi Li, and Robin N. Brewer. 2024. CareJournal: A Voice-Based Conversational Agent for Supporting Care Communications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu HI USA, 1–22. doi:10.1145/3613904.3642163

[141] Malak Sadek, Rafael A. Calvo, and Céline Mougenot. 2023. Co-designing conversational agents: A comprehensive review and recommendations for best practices. *Design Studies* 89 (2023), 101230. doi:10.1016/j.destud.2023.101230

[142] Samiha Samrose and Ehsan Hoque. 2022. MIA: Motivational Interviewing Agent for Improving Conversational Skills in Remote Group Discussions. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (Jan. 2022), 1–24. doi:10.1145/3492864

[143] Kyle-Althea Santos, Ethel Ong, and Ron Resurreccion. 2020. Therapist vibe: children's expressions of their emotions through storytelling with a chatbot. In *Proceedings of the Interaction Design and Children Conference (IDC '20)*. Association for Computing Machinery, New York, NY, USA, 483–494. doi:10.1145/3392063.3394405

[144] Dmitry Scherbakov, Nina Hubig, Vinita Jansari, Alexander Bakumenko, and Leslie A Lenert. 2025. The emergence of large language models as tools in literature reviews: a large language model-assisted systematic review. *Journal of the American Medical Informatics Association* 32, 6 (05 2025), 1071–1086. doi:10.1093/jamia/ocaf063

[145] Susanne Schmidt, Sven Zimmermann, Celeste Mason, and Frank Steinicke. 2022. Simulating Human Imprecision in Temporal Statements of Intelligent Virtual Agents. In *CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New Orleans LA USA, 1–15. doi:10.1145/3491102.3517625

[146] Ryan M. Schuetzler, G. Mark Grimes, and Justin Scott Giboney. 2020. The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems* 37, 3 (2020), 875–900. doi:10.1080/07421222.2020.1790204

[147] Daniel Schulman and Timothy Bickmore. 2009. Persuading Users through Counseling Dialogue with a Conversational Agent. In *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive '09)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/1541948.1541983

[148] Sofia Schöbel, Anuschka Schmitt, Dennis Benner, Mohammed Saqr, Andreas Janson, and Jan Marco Leimeister. 2024. Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. *Information Systems Frontiers* 26, 2 (2024), 729–754. doi:10.1007/s10796-023-10375-9

[149] Andrea Sestino, Cristian Rizzo, Luís Irgang, and Beáta Stehlíková. 2025. Redesigning healthcare service delivery processes through medical chatbot integrations: balancing chatbot features and patients' individual differences. *Business Process Management Journal* (March 2025). doi:10.1108/BPMJ-07-2024-0655

[150] Ben Sheehan, Hyun Seung Jin, Brett Martin, and Hyoje Jay Kim. 2024. Wow! Interjections Improve Chatbot Performance: The Mediating Role of Anthropomorphism and Perceived Listening. *Communication Research* 51, 7 (Oct. 2024), 891–917. doi:10.1177/00936502241273259

[151] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 368 (Oct. 2023), 35 pages. doi:10.1145/3610217

[152] Geovana Ramos Sousa Silva and Edna Dias Canedo. 2024. Human Factors in the Design of Chatbot Interactions: Conversational Design Practices. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems (IHC '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3702038.3702083

[153] Numaya Silva and Dilani Wickramaarachchi. 2025. Enhancing Systematic Literature Reviews: Evaluating the Performance of LLM-Based Tools Across Key Systematic Literature Review Stages. In *2025 5th International Conference on Advanced Research in Computing (ICARC)*. 1–6. doi:10.1109/ICARC64760.2025.10963273

[154] Carlota S. Smith. 1978. The syntax and interpretation of temporal expressions in English. *Linguistics and Philosophy* 2, 1 (Jan. 1978), 43–99. doi:10.1007/BF00365130

[155] Cheong Ying Sng, Mark Carter, and Jennifer Stephenson. 2018. A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments* 3 (2018), 2396941518803806. doi:10.1177/2396941518803806

[156] Donggil Song, Eun Young Oh, and Marilyn Rice. 2017. Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th International Conference on Human System Interactions (HSI)*. 78–82. doi:10.1109/HSI.2017.8005002

[157] Mengmeng Song, Huixian Zhang, Xinyu Xing, and Yucong Duan. 2023. Appreciation vs. apology: Research on the influence mechanism of chatbot service recovery based on politeness theory. *Journal of Retailing and Consumer Services* 73 (July 2023), 103323. doi:10.1016/j.jretconser.2023.103323

[158] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. 2021. A Conversational Agent Framework with Multi-modal Personality Expression. *ACM Transactions on Graphics* 40, 1 (Feb. 2021), 1–16. doi:10.1145/3439795

[159] Christopher Stanton and Catherine J. Stevens. 2014. Robot Pressure: The Impact of Robot Eye Gaze and Lifelike Bodily Movements upon Decision-Making and Trust. In *Social Robotics*, Michael Beetz, Benjamin Johnston, and Mary-Anne Williams (Eds.). Springer International Publishing, Cham, 330–339.

[160] Ian Steenstra, Farnaz Nouraei, Mehdi Arjmand, and Timothy Bickmore. 2024. Virtual Agents for Alcohol Use Counseling: Exploring LLM-Powered Motivational Interviewing. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, GLASGOW United Kingdom, 1–10. doi:10.1145/3652988.3673932

[161] Richard L. Street. 1982. Evaluation of noncontent speech accommodation. *Language & Communication* 2, 1 (1982), 13–31. doi:10.1016/0271-5309(82)90032-5

[162] Xin Sun, Isabelle Teljeur, Zhuying Li, and Jos A. Bosch. 2024. Can a Funny Chatbot Make a Difference? Infusing Humor into Conversational Agent for Behavioral Intervention. In *ACM Conversational User Interfaces 2024*. Association for Computing Machinery, Luxembourg Luxembourg, 1–19. doi:10.1145/3640794.3665555

[163] Yansen Theopilus, Hilary Davis, and Sonja Pedell. 2025. Digital technology for supporting Aged Care Services: A Scoping Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 506, 12 pages. doi:10.1145/3706598.3713498

[164] Jenny Thomas. 2006. Cross-cultural pragmatic failure. *World Englishes: Critical concepts in linguistics* 4, 2 (2006), 22.

[165] Paul Thomas, Daniel McDuff, Mary Czerwinski, and Nick Craswell. 2020. Expressions of Style in Information Seeking Conversation with an Agent. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, Virtual Event China, 1171–1180. doi:10.1145/3397271.3401127

[166] Wojciech Trzebiński, Toni Claessens, Jeska Buhmann, Aurélie De Waele, Greet Hendrickx, Pierre Van Damme, Walter Daelemans, and Karolien Poels. 2023. The Effects of Expressing Empathy/Autonomy Support Using a COVID-19 Vaccination Chatbot: Experimental Study in a Sample of Belgian Adults. *JMIR Form Res* 7 (8 May 2023), e41148. doi:10.2196/41148

[167] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *J Med Internet Res* 22, 8 (7 Aug 2020), e17158. doi:10.2196/17158

[168] Michelle M.E. Van Pinxteren, Mark Pluymaekers, and Jos G.A.M. Lemmink. 2020. Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management* 31, 2 (2020), 203–225. doi:10.1108/JOSM-06-2019-0175

[169] Lindsey Vanderlyn, Gianna Weber, Michael Neumann, Dirk Väth, Sarina Meyer, and Ngoc Thang Vu. 2021. "It seemed like an annoying woman": On the Perception and Ethical Considerations of Affective Language in Text-Based Conversational Agents. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, Arianna Bisazza and Omri Abend (Eds.). Association for Computational Linguistics, Online, 44–57. doi:10.18653/v1/2021.conll-1.4

[170] Mattias Wahde and Marco Virgolin. 2022. Conversational agents: Theory and applications. In *HANDBOOK ON COMPUTER LEARNING AND INTELLIGENCE: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*. World Scientific, 497–544.

[171] Thiemo Wambsganss, Anne Höch, Naim Zierau, and Matthias Söllner. 2021. Ethical Design of Conversational Agents: Towards Principles for a Value-Sensitive Design. In *Innovation Through Information Systems*, Frederik Ahlemann, Reinhard Schütte, and Stefan Stieglitz (Eds.). Springer International Publishing, Cham, 539–557.

[172] Lu Wang, Chaomei Chen, and Jina Huh-Yoo. 2023. Investigating the Synonyms of Conversational Agents to Aid Cross-Disciplinary CA Research. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 190, 10 pages. doi:10.1145/3544549.3585640

[173] Joseph Weizenbaum. 1983. ELIZA — a computer program for the study of natural language communication between man and machine. *Commun. ACM* 26, 1 (Jan. 1983), 23–28. doi:10.1145/357980.357991

[174] Jieyeon Woo, Kazuhiro Shidara, Catherine Achard, Hiroki Tanaka, Satoshi Nakamura, and Catherine Pelachaud. 2024. Adaptive virtual agent: Design and evaluation for real-time human-agent interaction. *International Journal of Human-Computer Studies* 190 (Oct. 2024), 103321. doi:10.1016/j.ijhcs.2024.103321

[175] Haijun Xia, Michael Glueck, Michelle Annett, Michael Wang, and Daniel Wigdor. 2022. Iteratively Designing Gesture Vocabularies: A Survey and Analysis of Best Practices in the HCI Literature. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 37 (May 2022), 54 pages. doi:10.1145/3503537

[176] Jing Xu, Da Ju, Margaret Li, Y-Lan Boureau, Jason Weston, and Emily Dinan. 2021. Bot-Adversarial Dialogue for Safe Conversational Agents. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 2950–2968. doi:10.18653/v1/2021.naacl-main.235

[177] Cheng Yanxia, Zhu Shijia, and Xiao Yuyang. 2024. A meta-analysis of the effect of chatbot anthropomorphism on the customer journey. *Marketing Intelligence & Planning* 42, 1 (Jan. 2024), 1–22. doi:10.1108/MIP-03-2023-0103

[178] Yue You, Chun-Hua Tsai, Yao Li, Fenglong Ma, Christopher Heron, and Xinning Gui. 2023. Beyond Self-diagnosis: How a Chatbot-based Symptom Checker Should Respond. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 64 (Sept. 2023), 44 pages. doi:10.1145/3589959

[179] Masahide Yuasa. 2014. Can Animated Agents Help Us Create Better Conversational Moods? An Experiment on the Nature of Optimal Conversations. In *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 634–640. doi:10.1007/978-3-319-07509-9_60

[180] Jia Zeng, Dan Fan, Xinxue Zhou, and Jian Tang. 2024. Chatbot with Resilience: The Impact of Repair Strategies on Customer Satisfaction in Conversational Breakdowns. In *E-Business. New Challenges and Opportunities for Digital-Enabled Intelligent Future*, Yiliu Paul Tu and Maomao Chi (Eds.). Vol. 517. Springer Nature Switzerland, Cham, 306–317. doi:10.1007/978-3-031-60324-2_26

[181] Junbo Zhang, Qi Chen, Jiandong Lu, Xiaolei Wang, Luning Liu, and Yuqiang Feng. 2024. Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism and boundary conditions. *Tourism Management* 100 (Feb. 2024), 104835. doi:10.1016/j.tourman.2023.104835

[182] Lian Zhang, Ashwaq Z. Amat, Huan Zhao, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. 2021. Design of an Intelligent Agent to Measure Collaboration and Verbal-Communication Skills of Children With Autism Spectrum Disorder in Collaborative Puzzle Games. *IEEE Transactions on Learning Technologies* 14, 3 (June 2021), 338–352. doi:10.1109/TLT.2020.3029223

[183] Wenzhen Zhang, Emma L. Slade, and Eleonora Pantano. 2024. Humanlike service robots: A systematic literature review and research agenda. *Psychology & Marketing* 41, 12 (2024), 3157–3181. doi:10.1002/mar.22099

[184] Zhenjie Zhao and Xiaojuan Ma. 2022. Designing an Artificial Agent for Cognitive Apprenticeship Learning of Elevator Pitch in Virtual Reality. *IEEE Transactions on Cognitive and Developmental Systems* 14, 3 (2022), 857–868. doi:10.1109/TCDS.2021.3073814

[185] Qingxiao Zheng, Daniela M. Markazi, Yiliu Tang, and Yun Huang. 2021. "PocketBot Is Like a Knock-On-the-Door!": Designing a Chatbot to Support Long-Distance Relationships. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–28. doi:10.1145/3479589

[186] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 570, 24 pages. doi:10.1145/3491102.3501855

[187] Cheng Zhou and Qian Chang. 2024. Informational or emotional? Exploring the relative effects of chatbots' self-recovery strategies on consumer satisfaction. *Journal of Retailing and Consumer Services* 78 (May 2024), 103779. doi:10.1016/j.jretconser.2024.103779

[188] Jijie Zhou and Yuhan Hu. 2024. Beyond Words: Infusing Conversational Agents with Human-like Typing Behaviors. In *ACM Conversational User Interfaces 2024*. Association for Computing Machinery, Luxembourg Luxembourg, 1–12. doi:10.1145/3640794.3665560

[189] Mingxi Zhou, Luning Liu, and Yuqiang Feng. 2025. Building citizen trust to enhance satisfaction in digital public services: the role of empathetic chatbot communication. *Behaviour & Information Technology* (2025), 1–20. doi:10.1080/0144929X.2025.2451763 Publisher: Taylor & Francis.

[190] Zhenxing Zhu, Jun Xie, Longxin Zhou, Chaoran Yang, and Feng Li. 2025. Evaluation of the accuracy and repeatability of Deepseek V3, Doubao, and Kimi1.5 in answering knowledge-related queries about chronic non-bacterial osteitis. *Frontiers in Artificial Intelligence* 8 (Sept. 2025), 1629149. doi:10.3389/frai.2025.1629149

[191] Sahba Zojaji, Adam Červeň, and Christopher Peters. 2023. Impact of Multimodal Communication on Persuasiveness and Perceived Politeness of Virtual Agents in Small Groups. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA '23)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3570945.3607356

[192] Tamara Zubatiy, Kayci L Vickers, Niharika Mathur, and Elizabeth D Mynatt. 2021. Empowering Dyads of Older Adults With Mild Cognitive Impairment And Their Care Partners Using Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 657, 15 pages. doi:10.1145/3411764.3445124