

Conversational agents: a survey on culturally informed design practices

Mateus de Souza Monteiro  [Universidade Federal Fluminense | mateusmonteiro@id.uff.br]

Luciana Cardoso de Castro Salgado  [Universidade Federal Fluminense | luciana@ic.uff.br]

Abstract

The development and use of conversational agents grow every day. However, those systems still need to meet users' expectations. As a result, new design practices for conversational agents are emerging. One facet is the culturally informed design of those systems. Tailoring conversational agents culturally can increase engagement, trustworthiness, and acceptance. While an emerging trend, literature still needs to discuss research and application of design practices related to culture and conversational agents. The present study investigates 5398 articles involving conversational agents and culture over the last seven years. In this context, we selected 23 articles based on their approaches and objectives related to our main subject. The findings indicate that most studies in the field are mainly focused on the feasibility of cultural markers tailored to those systems. The main contribution is the identification of the main challenges, contexts and design practices in the field. These results contributed to discussions regarding conversational agents and highlighted some research gaps for future study.

Keywords: *Systematic review, Culture, Design, Interaction, Conversational agent*

1 Introduction

Since 1950, the idea and application of conversational agents have been developing and growing (Chaves and Gerosa, 2020). Considering growth, by 2028, the market is expected to grow 30.29% when compared to 2023 (Mordor, 2023). Besides, vendors constantly look forward to integrating Natural Language Processing (NLP) with Artificial Intelligence (AI) to cater customer needs.

Shawar and Atwell (2007) defined a chatbot as a [pp.2] "software program that interacts with users using natural language." Følstad et al. (2021) [pp.4], in turn, stated that "[...] conversational agents provide access to information and services through interaction in everyday language." Følstad et al. (2021) also argue for a more straightforward definition of those terms, avoiding, for example, considering chatbots as simply the ones the interaction happens through text. In this sense, it encompasses agents supporting interactions through text, voice, or both. Currently, in the literature, there is still a wide variety of terms (Følstad et al., 2021; Shawar and Atwell, 2007): from chatbots to robots, multi-modal agents, chatterbots, virtual humans, virtual agents, embodied conversational agents, and many others. To simplify, in this work, we use conversational agents to refer to those variations of dialogue systems, including embodied conversational agents and even physical social robots.

The idea of users talking to chatbots in "everyday language" as envisaged by Følstad et al. (2021) is not a reality. Current conversational agents often fail to meet user expectations, causing frustration (Følstad and Brandtzæg, 2017). Interaction challenges arise from the capacity to interpret user questions, provide appropriate responses, and respond naturally (Singh and Beniwal, 2022). Those theoretical challenges (Singh and Beniwal, 2022) for Human-Computer Interaction (HCI) are centered in: "conversations as the object of design" (Følstad and Brandtzæg, 2017, pp. 3).

The conversation is the main form of language exposed to

humans (Marcuschi, 1991). To Marcuschi (1991), the basic constitutive characteristics of conversations include interactions between at least two speakers, turn-taking, opening, and closing, standard features in conversational agents. In this sense, we highlight the Jakobson (2008) Anthropology's perspective that connects language and culture by affirming that language is the foundation of culture (see Prietch et al., 2022 discussing deaf communities; or the reflections in Keesing, 1974).

A facet of conversational agents' design is tailoring the conversation to specific communities. Consider conversational agents created for populations with particular beliefs, attitudes, languages, and cultural practices. O'Leary et al. (2020) emphasize that subtle mismatches between agent behavior and community norms can lead to lower satisfaction, engagement, and, in some cases, complete rejection. Yet, when culturally informed design practices succeed, the positive effects are many. For example, a decade ago Yin et al. (2010) designed a culturally-congruent agent to promote exercise in Latino populations. The study outcomes demonstrated that adapting the agent to the community's language and cultural markers resulted in positive behavioral change.

However, current research lacks results that address design challenges on conversational agents and culture. In this sense, to address this gap, this work investigates the main interaction contexts, design practices, and cultural markers within conversational agents. To the best of our knowledge, no previous paper in the literature has related both conversational agents and cultural themes through a Systematic Literature Review (SLR). Besides, our contributions include challenges and opportunities for the research field. We discuss these points in detail in Section 6.

The article is structured as follows. Section 2 discusses the concepts that support this work, organized into two subsections: subsection 2.1 presents culture, including a brief history and principles, and subsection 2.2 describes Conversational Agents' solutions. Section 3 presents related

work. Section 4 presents protocols, strategies, and processes adopted in our SLR. Section 5 discusses our results, while Section 6 presents topics yet to be addressed. Finally, in Section 7, we conclude our work and present some limitations with glimpses at future work.

2 Background

In this section, we present concepts that support this research. First, we address cultural concepts used in this work. Then, we briefly present Conversational agents' solutions.

2.1 Culture

Edward Tylor (1871) is credited as the one who provided one of the earliest modern definitions of culture, widely used and still accepted today by some contemporary anthropologists - "[...] that complex whole which includes, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society" (Tylor, 1871, pp. 378). By then, Tylor had already stated that culture is not a biological determinism but a learning process.

Keesing (1974) summarizes the cultural theories in four groups:

- (1) Adaptive understands culture as a mechanism based on social behavior patterns in the (surroundings, dependent on ecosystems, technologies, economy, beliefs, and religiosity) that aims people to adjust in a particular environment;
- (2) Cognitive interprets culture as mental representations of the world. It is the form of things people have in mind, their models of perceiving, relating, and interpreting them.
- (3) Structural views culture as shared symbolic systems that are creations of the mind that seeks to discover in structuring cultural domains - myth, art, kinship, and language;
- (4) Symbolic, unlike Structural, finds cultural systems in real scenarios, i.e., not disembodied and decontextualized with humans engaging in symbolic action.

To Clifford Geertz (1973), the concept of culture is essentially (4) Symbolic - "Culture is the fabric of meaning in terms of which human beings interpret their experience and guide their action [...] an ordered system of meaning and symbols in terms of which social interaction takes place" (Geertz, 1973, pp. 145). According to Keesing (1974), with this approach, Anthropology becomes a matter of interpretation, not decipherment (as in Lévi-Strauss structural perspective) and must be embedded in the contextual richness of social life. This semiotic perspective gained popularity in the postmodern movement of 1980 (Tharp, 2009).

Other perspectives view culture as taxonomic. In this sense, culture is a way to distinguish and classify people, activities, and settings, as in the case of Hofstede (2001) and other authors (Triandis and Vassiliou, 1972). In HCI research about culture issues, Hofstede's is among the most known

and cited (Salgado et al., 2015). However, this culture models approach does not favor research or HCI practitioners interested in identifying "[...] cultural aspects that may emerge from, and be relevant for, a particular cultural context" (Salgado et al., 2015, pp. 4), as in the case of this research. Therefore, this research (aligned with Geertz, 1973) understands culture as a learned, symbolic, and shared lens we use to interact with, react to, and feel about symbols, terms, and situations.

2.2 Conversational agents

The concept of a person interacting with a program dates back to the sixties (Klopfenstein et al., 2017; Kuhail et al., 2022). Examples include ELIZA, ALICE, and SmarterChild (Kuhail et al., 2022). ELIZA, considered the first chatbot, emulated a psychotherapist interacting with a human patient (Weizenbaum, 1966). Today, more advanced proposals such as ChatGPT (Chat Generative Pre-Trained Transformer, <https://openai.com/blog/chatgpt/>) promise to disrupt by providing near-human quality responses.

For commercial service providers, interacting through natural language is becoming an attractive way to engage with customers (Følstad and Brandtzæg, 2017), mainly to automate solutions for customer services, sales, and support (Følstad et al., 2021). According to Montenegro et al. (2019), conversational agents' interactions can also happen using virtual avatars through verbal or non-verbal communication. Conversational agents can also be called relational agents or embodied conversational agents (Kuhail et al., 2022), including behaviors such as hand gestures, speech, gaze, and other non-verbal particularities (Montenegro et al., 2019).

Supported in different platforms, such as apps via mobile devices, web, computer, SMS, telephone, and multimodal (Laranjo et al., 2018), conversational agents can use NLP with AI to improve their solutions. Singh and Beniwal (2022) categorize it based on their task or the approaches incorporated:

- Rule-based: bots that follow a set of predetermined rules, input patterns, and response patterns.
- Generative-based: employs powerful NLP algorithms to understand the inquiry, gather the intent, and react without the need for human participants.
- Retrieval-based: the convention strategy in which the response is chosen from the corpus based on some heuristics from the conversations. Considered simpler to develop and operate than using generative models.

For conversational agents, in which language plays a central role, understanding how culture affects design and usage can significantly benefit the solution (Yin et al., 2010). Often explored during design, cultural markers are design elements, behavior, and features adopted from a particular cultural group (Aljaroodi et al., 2020). Aljaroodi et al. (2020) define it in two groups: (1) digital cultural markers, which are User Interface elements such as colors, images, and language cues; and (2) social cultural markers, i.e., customs, morals, traditions, and values that arise from the environment.

Obremski et al. (2021) argue that to achieve culturally informed tailoring of conversational agents, a simple dimen-

sion is enough to trigger the impression of a non-native agent instead of tailoring every aspect. O’Leary et al. (2020), in turn, argue that tailoring only the agent’s appearance is likely overly simplistic. Today, most of the cultural markers related to the cultural tailoring of conversational agents have been related to the avatar’s appearance. Other more specific traits can include, for example, gaze behaviors, which is how the avatar should look at the user while listening to or communicating.

3 Related Work

The literature presents some discussions about conversational agents, such as SLRs (Kuhail et al., 2022; Rheu et al., 2020; Zhai and Wibowo, 2022; Alsheddi and Alhenaki, 2022), surveys (Singh and Beniwal, 2022; Chaves and Gerosa, 2020; Wolk et al., 2022), overviews (Suta et al., 2020) and a state-of-the-art article (Machidon et al., 2018). The topics include the following themes: conversational agents for specific domains; user experience; cultural issues; architectural overview; and a macro review.

Considering conversational agents for specific domains, Kuhail et al. (2022) present a review to understand, compare, and reflect on recent attempts to use chatbots in education. Regarding experience towards agents’ use, Rheu et al. (2020) describe a systematic review of conversational agents’ promotion of trust in users. Chaves and Gerosa (2020) surveyed papers from various domains to understand how social characteristics can benefit human-chatbot interactions. According to the authors, conversational agents should have social capabilities similar to humans, matching users’ expectations and avoiding frustration and dissatisfaction.

About culture, Zhai and Wibowo (2022) reviews cross-culture, humor, and empathy dimensions in conversational agents during second language (L2) acquisition. Authors argue that specific cultural dimensions and values influence humor. Therefore, increasing cultural differences in humor can promote student engagement in learning and understanding. Studies also present frame understanding of specific scenarios, for example, Alsheddi and Alhenaki (2022) present an SLR for English and Arabic chatbots and Wolk et al. (2022) describe a survey of dialogue systems within Slavic languages. The most advanced chatbots in the niche cultural heritage field are presented by Machidon et al. (2018). The authors discuss technological difficulties and avatar interaction’s effects on user engagement, immersion, and learning efficiency.

Literature also discusses an overview of chatbot platforms, architectures, and implementation, as in Singh and Beniwal (2022). The authors investigate near-human agents approaches. A review of the literature published over the past decade is presented by Suta et al. (2020). Overall, articles present technical challenges in dialogue systems, briefly discuss culture, yet lacking to investigate design practices in culturally informed conversational agents, as this research focuses. In this sense, this article presents an SLR of conversational agents and culture to understand better contexts, methods, and challenges related to this subject.

4 Materials and methods

A SLR is a method that analyzes the literature available on a specific topic to provide background for new studies, find gaps for further investigations, and summarize evidence concerning a particular technology.

While culture and conversational agents are a recent and emerging trend, researchers have not yet provided an SLR for this subject. Following the guidelines provided by Kitchenham et al. (2009), this study aims to achieve it through a defined protocol: (1) research questions, (2) keywords, (3) search engines, and (4) selection criteria.

4.1 Research questions

In this section, we seek to describe our general and specific questions to guide this research:

- (RQ1) What are the main contexts for interaction?
- (RQ2) What are the main design practices?
 - (2.1) In what phase of design lifecycle?
 - (2.2) What are the main design methods?
 - (2.3) What are the main cultural markers?
- (RQ3) What are the main challenges?

The questions above highlight the main concerns related to design practices for culturally informed conversational agents. First, we consider RQ1 concerns the main domains, goals, and scenarios, which is essential for understanding design practices. RQ2 is specific to design practices: in which moment of the design life cycle the solution takes place, the methods used in the design process, and the central cultural markers explored in the articles. RQ3 seeks to summarize the main challenges in design practices to generate and propose discussions related to particular topics raised by the authors in the works.

4.2 Keywords

One of the most used processes to define the search keywords is PICO, which identifies the population (P), the intervention (I), the comparison (C), and the expected outcomes (O), based on the research questions by Kitchenham and Charters (2007). For this study, we have:

- Population (P): papers that describe or apply culture tailoring process in conversational agents.
- Intervention (I): methods, approaches, methodologies, or practices used to make tailored conversational agents.
- Comparison (C): not applicable, since the purpose of this study is to describe the SLR.
- Outcomes (O): most used methods, definitions, techniques, anthropomorphism characteristics, and context.

A set of keywords representing the expected search results was defined based on this PICO. It is important to mention that considering the novelty of the subject, we avoided using methods, approaches, methodologies, or even design in the keywords to avoid restricting our search. Thus, the search comprises two major sets of keywords:

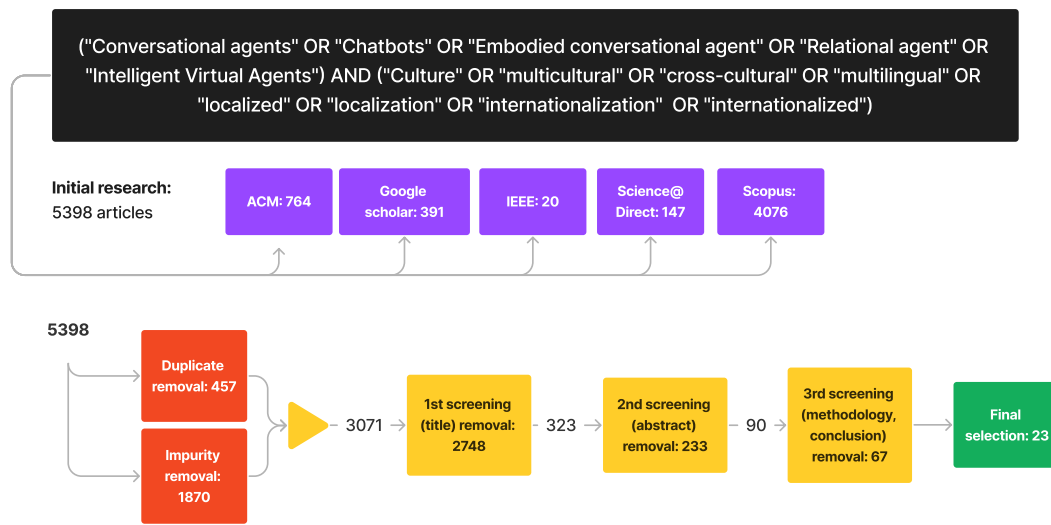


Figure 1. Article selection process (Executed Jan/2023).

- Conversational agents: covers the collection of keywords (e.g., conversational agent, chatbots, embodied conversational agent, etc.);
- Culture: covers the collection of keywords about culture (e.g., culture, multicultural, cross-cultural, etc.).

Thus, the search string defined is:

(“Conversational agents” OR “Chatbots” OR “Embodied conversational agent” OR “Relational agent” OR “Intelligent Virtual Agents”) AND (“Culture” OR “multicultural” OR “cross-cultural” OR “multilingual” OR “localized” OR “localization” OR “internationalization” OR “internationalized”)

4.3 Selection criteria

To improve the findings accuracy, all resulting papers were screened using the selection criteria (SC). The study had to meet the following requirements in order to be chosen:

- Papers published in 2017 until 2023;
- Papers written in English;
- Primary papers (i.e., not surveys, meta-analysis, systematic mappings or reviews);
- Only papers available with current resources (e.g., open-access, institution resources, or made available by the authors);
- Papers available for download;
- Non-duplicate papers (i.e., papers with the same Digital Object Identified - DOI).

4.4 Article selection

As seen in Figure 1, the article selection process used a filtering method to improve results. The process con-

Table 1. Final corpus of the articles selected.

| Article | Year |
|---------|-------------------------------|
| A01 | Aljaroodi et al. 2020 |
| A02 | Guerrero-Vásquez et al. 2020 |
| A03 | Löffler et al. 2021 |
| A04 | Obremski et al. 2021 |
| A05 | O’Leary et al. 2020 |
| A06 | Pataranutaporn et al. 2019 |
| A07 | Rahman et al. 2021 |
| A08 | Trovato et al. 2017 |
| A09 | Wang et al. 2020 |
| A10 | Yadav et al. 2019 |
| A11 | Zhou et al. 2017 |
| A12 | Carnell and Lok 2018 |
| A13 | Chen et al. 2020 |
| A14 | Danielescu and Christian 2018 |
| A15 | Cerda Diez et al. 2019 |
| A16 | Kim et al. 2022 |
| A17 | Liao and He 2020 |
| A18 | Ludin et al. 2022 |
| A19 | Lugrin et al. 2018 |
| A20 | Murali et al. 2020 |
| A21 | Peng et al. 2022 |
| A22 | Rehm et al. 2018 |
| A23 | Wang et al. 2022 |

sisted in removing duplicates, applying exclusion criteria, removing impurities (theses, dissertations, books, reviews), and three screening steps (title, title+abstract, title+abstract+methodology, conclusion).

It is important to mention that papers with restricted access (no open access or through institution proxy) were discarded. Only papers that considered culturally informed design practices were included. Despite the review applied by a single researcher, all the processes and final corpus were evaluated and validated by a second researcher. Based on mutual agree-

Table 2. Main contexts: domain, goals, and scenario.

| Node | 1 Level | 2 Level | Article Id | | |
|---|----------|---------------------------------|--|---|-----|
| Context | Domain | Health | A01, A02, A05, A07, A10-A12, A14-A18, A20-A23 | | |
| | | Religious | A03, A05, A06 | | |
| Goals | Goals | Services | A08, A13, A19 | | |
| | | Educational | A04, A19 | | |
| | | Gaming | A09 | | |
| | | Health assistance and education | A01, A02, A05, A07, A09-A11, A13, A15-A18, A21-A23 | | |
| | | Spiritual guidance | A03, A06 | | |
| | | Training with virtual humans | A04, A12, A19 | | |
| | | Exercise behavior coaching | A14, A20 | | |
| | | Scenario | Scenario | Arabian users | A01 |
| | | | | 8 and 9-year-old Ecuadorian ASD children | A02 |
| | | | | Members of the Protestant Church in Hesse and Nassau in Germany | A03 |
| German and English users | A04, A09 | | | | |
| Members of African-American congregations in Boston area | A05 | | | | |
| Members of the Buddhism community | A06 | | | | |
| Adolescents (between 10 and 19-year-old) in Bangladesh | A07 | | | | |
| Brazilian functional illiterate users | A08 | | | | |
| Mothers and community health workers residing in under-developed (slums) regions of India | A10 | | | | |
| Young Chinese adults living in the United States | A11 | | | | |
| Chinese or Chinese-Americans | A12 | | | | |
| Migrants in Germany | A13 | | | | |
| English, Spanish, German, Italian, and French | A14 | | | | |
| Spanish users living in the United States | A15 | | | | |
| Black Americans with chronic conditions | A16 | | | | |
| White/Caucasian, Black/African American, Latins, and Asians living in the United States | A17 | | | | |
| Māori as Tāngata Whenua (the indigenous people of Aotearoa) | A18 | | | | |
| Migrant students in a German university | A19 | | | | |
| Adults born in India but moved to United States when they were 16 or older | A20 | | | | |
| Men over 18 years speaking Bahasa Malaysia or English | A21 | | | | |
| Danish citizens with brain damage | A22 | | | | |
| Young people in India | A23 | | | | |

ment, these reviewers selected a final list of articles.

Initially, our search string found 5398 articles in different databases ACM (<https://dl.acm.org>), Google scholar (<https://scholar.google.com>), IEEE Xplore (<https://ieeexplore.ieee.org/Xplore/home.jsp>), Science@direct (<https://www.sciencedirect.com/>), and Scopus (<https://scopus.com>). We first removed duplicate studies and impurities, resulting in 3071 studies. Then we proceed to the three-step screening process. Most papers were removed due to only mentioning culture superficially or barely discussing the design process. Some excluded studies included articles from the same authors that discussed the same solution. Therefore resulting in a final corpus of 23 articles (Table 1).

The interest of researchers in culture and conversational agents has been growing over the last few years (substantially beginning in 2017), with a significant increase in the number of studies between 2020 and 2022. The lack of publications in 2023 cannot be considered a trend since the research was conducted before the end of the year.

5 Results

In this section, we present the results of the SLR. First, we present the contexts where the interaction happens (5.1). We then present the main culturally informed design practices (5.2) and the open challenges (5.3).

5.1 RQ1 - What are the main contexts for interaction?

We defined three main categories (Table 2): Domains (5 categories), Goals (4 categories), and Scenarios (22 categories). A domain defines the content that will drive the dialogue.

Goals are inspired by Montenegro et al. (2019) definitions of agent types (e.g., Training, Education, Assistance) and refer to the objective with a conversational agent inside that domain. Scenarios use the authors’ words in the articles to describe the user and its origins and definitions. The second level is only used to describe each finding.

Health is by far the domain most explored by the authors. Usually aiming assistance or educational, health conversational agents represent more than 60% (15/23) of the works. For example, in Danielescu and Christian (2018), authors present Radar Pace, a conversational coaching agent that provides real-time feedback for running and cycling. The product was developed by Oakley and deployed in 5 languages: English, Spanish, German, Italian, and French.

Murali et al. (2020) present a solution to support exercise promotion interventions. The authors target a user population familiar with both cultures, i.e., adults born in India but moved to the United States when they were 16 or older. O’Leary et al. (2020) developed a virtual agent aimed at members of a network of primarily African-American congregations in the Boston area. Zhou et al. (2017) proposed an agent to promote physical activity among Chinese immigrants.

Cerda Diez et al. (2019) present the conversational agent Vicky. The objective was to collect family history information. According to the authors, family history tools in languages other than English is limited. Collecting family history information is necessary to allow physicians to comprehend the risk factor for chronic conditions, including cardiovascular disease, diabetes, and cancer. The agent was developed for Spanish users living in the United States. Similarly, in Kim et al. (2022), the authors present a solution to support assisting patients with chronic conditions, prioritizing Black Americans.

Table 3. Main cultural tailoring design practices.

| Node | 1 Level | 2 Level | 3 Level | Article Id | |
|------------------|------------------|------------------|---|--|--|
| Design practices | Design lifecycle | Evaluation phase | Participatory | A02-A04, A06, A07, A09, A10, A12-A15, A17-A21, A23 | |
| | | Design phase | | A05, A08, A13, A16, A22 | |
| | Method | | Survey | Non-participatory | A01, A12, A14 |
| | | | Desk research | | A02, A03, A04, A06, A07, A08, A09, A11-A13, A17, A20 |
| | | | Interview | | A01, A02, A07, A08, A14, A15 |
| | | | Observatory | | A01, A05, A10, A12-A14, A16, A18, A21, A23 |
| | | | Dimensions | | A02, A08, A10, A19, A22 |
| | | | Thematic analysis of interaction logs | | A01, A09, A11, A14, A20 |
| | | | Focus group | | A04, A10, A12 |
| | | | Method of the Assessment of eXperience | | A05, A21 |
| | | | Empathy probes | | A13 |
| | | | Google persona creation guideline | | A13 |
| | | | Perceived Interpersonal Closeness Scale | | A13 |
| | | | | | A17 |

Liao and He (2020) explore racial mirroring effects in psychotherapy. According to the authors, patients with therapists of the same race have demonstrated stronger bonding and positive attitudes. The solution is explored with White/Caucasian, Black/African American, Latins, and Asians living in the United States.

Ludin et al. (2022) describe the Aroha bot, a conversational agent to assist young adults’ mental health in Aotearoa. As of answer to Covid-19 initial lockdown, this solution aimed to support young people in managing pandemic-related worries. Anxiety, social maintenance, and cultural connection were a few issues supported by the agent. The use of Ecuadorian ethnic avatars in Autism Spectrum Disorder (ASD) therapy is explored by authors in (Guerrero-Vásquez et al., 2020).

In Peng et al. (2022), the authors present a conversational agent to promote HIV testing and prevention among men who have sex with men in Malaysia. Wang et al. (2022) discuss a case study of a chatbot in India called Snehai. Designed by the Population Foundation of India, this solution aims to educate adolescents and young adults about sex and reproductive health. In Indian urban slum areas, Yadav et al. (2019) investigate the viability of using chatbots to educate mothers and community health workers about breastfeeding. Rahman et al. (2021) introduce AdolescentBot, a Bangladesh interactive chatbot. In Bangladesh, adolescents (defined as anyone between the ages of 10 and 19) frequently experience various physical and emotional issues due to their lack of education regarding sexual and reproductive health. Due to societal taboos and timidity, people avoid discussing these issues with their doctors, friends, or even family members.

In Rehm et al. (2018), the authors present a ”build your own robot” project. Citizens with brain damage and their care personnel (from Danish healthcare) co-designed their robots with a specialized task chosen by the participants. Aljaroodi et al. (2020) present a solution to support Arabian users in managing their personal health information using Avatars. Carnell and Lok (2018) the authors present a conversational agent to train physicians to detect a patient’s cultural background. In this scenario, the agent acts as a virtual Chinese or Chinese-American patient diagnosed with dysphagia.

Services represent the second most explored domain. This domain is usually associated with providing information, assisting in reaching places, and frequently asking questions. Chen et al. (2020) present a conversational agent to assist mi-

grants providing information about integration for newcomers in Germany. In Lugin et al. (2018), the authors present a conversational agent to assist migrants in an educational scenario. The solution aimed to help migrant students in a German university. By inducing grammatical errors at different rates, Obremski et al. (2021) presented a non-native IVA (Intelligent Virtual Agent) simulation in two languages. In Trovato et al. (2017), two studies—a preliminary investigation and an interaction experiment—have been conducted to investigate the issues when designing a Brazilian-employed receptionist robot.

Regarding religion, Pataranutaporn et al. (2019) describe an Embodied Spiritual Machine (ESM). Buddha Bot, a chatbot that personifies and learns from Buddha’s writings to respond to user questions, was created by the authors to explore the concept of ESM. Similarly, Löffler et al. (2021) used a discursive design approach to create the protestant blessing robot BlessU2 by arranging emerging technological elements to hypothesize future products and gain a better understanding of their potential implications in a given context. Lastly, Wang et al. (2020) developed a game where players ask for support from a virtual agent.

5.2 RQ2 - What are the main design practices?

Regarding design practices, we defined three central concepts:

- What moment of the design lifecycle (two categories and two sub-categories): if the interaction with the user happens during the Evaluation or Design Phase (participatory or non-participatory).
- Method (11 categories): we use known methods terminology to describe better the method used (e.g., survey, interviews, focus group). In the case of a more specific method, we used the authors’ words in a new category.
- Digital cultural markers (5 categories and 32 sub-categories): here, we used an in-depth analysis of all articles to identify characteristics and patterns used in the design of the conversational agent. Five categories emerged: Appearance (12 sub-categories), Behavior (6 sub-categories), Identity (3 sub-categories), Communication (5 sub-categories), and Personality (6 sub-categories).

5.2.1 (2.1) Design lifecycle

The experiments can be divided into two main categories (Table 2): evaluation and design phase. Regarding evaluation, users could offer feedback later in the design process or after the conversational agent had been designed. Most of the experiments focused on receiving opinions after the design of the conversational agent (Guerrero-Vásquez et al., 2020; Löffler et al., 2021; Obremski et al., 2021; Pataranutaporn et al., 2019; Rahman et al., 2021; Wang et al., 2020; Yadav et al., 2019; Carnell and Lok, 2018; Chen et al., 2020; Danielescu and Christian, 2018; Liao and He, 2020; Ludin et al., 2022; Lugrin et al., 2018; Murali et al., 2020; Peng et al., 2022).

In studies during the design phase, solutions are co-designed or early-evaluated. Divided into two groups, participatory (O’Leary et al., 2020; Trovato et al., 2017; Chen et al., 2020; Kim et al., 2022; Rehm et al., 2018) and non-participatory (Aljaroodi et al., 2019; Carnell and Lok, 2018; Danielescu and Christian, 2018). The following section will cover which techniques were employed to tailor the conversational agents to the scenario.

5.2.2 (2.2) Methods

Survey (Table 3) is the most used method to discover, adapt, and evaluate conversational agents. For example, its usage can happen before the design to determine which dimensions will be focused (e.g., appearance and communication) would better suit the scenario (Trovato et al., 2017). In Chen et al. (2020), authors used Google persona creation guidelines to create, discuss and vote for the most suitable avatar persona.

Post-study questionnaire use of survey can be sent by the researcher/designer or automatically by the bot, e.g., in a study, a conversational agent started a conversation asking to answer the post-study questionnaire if exchanged more than six conversations with the bot (Obremski et al., 2021).

Evaluation of the conversational agent can discuss qualities of the bot in terms of satisfaction (Yadav et al., 2019), usefulness (Wang et al., 2020; Yadav et al., 2019), closeness to the user (Wang et al., 2020), trustworthiness (Wang et al., 2020), appropriateness (Wang et al., 2020), likeability (Wang et al., 2020; Rahman et al., 2021; Löffler et al., 2021; Pataranutaporn et al., 2019; Trovato et al., 2017; Carnell and Lok, 2018; Liao and He, 2020), perceived intelligence (Trovato et al., 2017) or if the participant would continue to use after the study (Wang et al., 2020; Guerrero-Vásquez et al., 2020; Yadav et al., 2019; Liao and He, 2020). Other more specific evaluation of the interaction includes understanding the adequacy of appearance, identity, language, behavior, personality, friendliness, and preference between conversational agents.

Surveys can include open-ended questions, the Likert scale, the Semantic scale (e.g., Creppy to Pleasant, Scary to Innocuous, among others) (Trovato et al., 2017), Task load scale (Wang et al., 2020), or the attribution of emotion as a way to measure their perception with a robot (Wang et al., 2020; Löffler et al., 2021). Other examples include the use of the Method for the Assessment of Experience (MAX), a post-use evolution to measure general experiences (emotion, ease of use, usefulness, and intention to use), or the Perceived

Interpersonal Closeness Scale (PICS).

Desk research is usually used during the design phase. For example, studies used Census to understand society variables, such as racial and ethnic composition (Trovato et al., 2017; Guerrero-Vásquez et al., 2020) and decide on the avatar’s appearance. A literature review is also used to discover guidelines or variables that would affect the design of the chatbot (Aljaroodi et al., 2020). Other desk research approaches include building the conversational agent’s knowledge base upon existing websites (Rahman et al., 2021).

Observatory methods are usually used to measure a conversational agent’s feasibility after being designed but before fully developed. Wizard of Oz (WoZ) is a common technique to test the feasibility of conversational agents (Trovato et al., 2017; Yadav et al., 2019; Lugrin et al., 2018). This HCI prototyping technique allows designers to gather design requirements and assess features without developing the entire system (Dahlbäck et al., 1993). Some might say that WoZ is not preferable if the participants experiment concurrently (Rahman et al., 2021). Another used observatory method is the Gesell chamber. Gesell chamber can be a way to allow the human models to control avatars and see the participants’ reactions while avoiding not influencing the study (Guerrero-Vásquez et al., 2020). Differently, Rehm et al. (2018) uses ethnography as a starting point for the co-design of the agent. According to the authors, it is the only way to understand local cultural practices reliably.

Interviews can be a quick and alternative way of collecting user feedback without finishing its development and as an alternative to observatory methods. For example, one of the studies used the semi-structured interview to help mature its proposed design guidelines (Aljaroodi et al., 2020). Besides, an interview can also be used to contextualize choices in workshops and focus groups and provide feedback on the design session overall (O’Leary et al., 2020; Yadav et al., 2019). This method can also collect specialized opinions of the domain without directly affecting users’ privacy (Rahman et al., 2021). In Danielescu and Christian (2018); Cerda Diez et al. (2019), interviews are used to translate the agent from English. Interviews are also used to measure the influence of cultural norms and values in usage (Rahman et al., 2021; Yadav et al., 2019; Kim et al., 2022; Peng et al., 2022; Wang et al., 2022).

Geert Hofstede promoted the idea of comparing cultures. The model, also known as Hofstede’s Model of Cultural Dimensions, is usually used in literature before the design to think of the interaction design decisions (Wang et al., 2020; Zhou et al., 2017; Danielescu and Christian, 2018; Murali et al., 2020). Besides, the model also supports organizing the design discourse and discussing the scenario in which the conversational agent is tailored (Aljaroodi et al., 2020).

Another way to explore the feasibility of the chatbot is by analyzing the user interactions with the conversational agent logs and applying a thematic analysis (Obremski et al., 2021; Yadav et al., 2019). Lastly, a focus group supports a participatory design workshop (O’Leary et al., 2020; Peng et al., 2022). Participants could discuss the objective, appearance, language, personality, and several other dimensions of the conversational agent (O’Leary et al., 2020).

Table 4. Main digital cultural markers.

| Node | 1 Level | 2 Level | 3 Level | 4 Level | Article Id | |
|------------------|--------------------------|------------|---------------|---|--|--|
| Design practices | Digital cultural markers | Appearance | Face | Facial hair | A01, A19 | |
| | | | | Hair preferences | A01, A05, A15 | |
| | | | | Face format | A19 | |
| | | | | Familiar clothing | A01, A05, A10, A11, A19, A20, A23 | |
| | | | | Body size (i.e., height, weight) | A03, A05, A19 | |
| | | | | Human-likeness | Human | A01, A05, A08, A09, A10, A11, A15, A19, A20, A23 |
| | | | | | Robot | A03, A08, A09, A22 |
| | | | | | Animal | A09, A13 |
| | | | | Avatar background with cultural landmarks | A01, A03, A11, A20 | |
| | | | | Age | A08, A10, A12, A15 | |
| | | | | Behavior | Emotion | A02, A03, A08, A11 |
| | | | | | Expression | A02, A03, A05, A08, A09 |
| | | | | Gaze shift | A11, A20 | |
| | | | | Gesture | A03, A04, A13, A19, A20 | |
| | | | | Head nods | A11, A20 | |
| | | | | Posture change | A03, A11, A20 | |
| | | | Identity | Gender | A01, A02, A03, A08, A10, A12, A14, A15, A17, A19, A20, A23 | |
| | | | | Ethnicity | A02, A04, A08, A17 | |
| | | | | Skin | A01, A02, A08, A15, A17, A19, A20 | |
| | | | Communication | Familiarity | A02, A03, A05, A06, A07, A08, A11, A12, A15, A18, A23 | |
| | | | | Voice | A02, A03, A09, A11, A15, A20 | |
| | | | | Conversational style | A11-A13, A17 | |
| | | | | Register (i.e., formal, informal) | A14 | |
| | | | | Proficiency | A04, A19 | |
| | | | Personality | Friendly | A10, A13 | |
| | | | | Politeness | A08, A10 | |
| | | | | Empathy | A12, A15, A18, A21 | |
| | | | | Optimistic, energetic, reassuring, reliable, and efficient | A13 | |
| | | | | Supportive, firm & authoritative, encouraging & cooperative | A14 | |
| | | | | Sympathy | A15 | |

5.2.3 (2.3) Cultural markers

Regarding digital cultural markers (Table 4), it is possible to separate the articles’ explorations into five main categories. First, **Appearance** focuses on the Face, Body, background of an avatar, and age. In one work, Appearance is described as ”good-looking,” but there is no definition of why and how good-looking would be represented. The usage of facial hair is explained by Aljaroodi et al. (2020) as a way to connect to gender. According to the authors, male avatars could use a dark beard, ghutra, hijab, and abayas for female avatars.

- **Face**
Facial hair (format and color) (Aljaroodi et al., 2020; Lugin et al., 2018)
Hair preferences (style and color) (Aljaroodi et al., 2020; O’Leary et al., 2020; Cerda Diez et al., 2019)
Face format (Lugin et al., 2018)
- **Clothing preferences**
Familiar clothing (Aljaroodi et al., 2020; O’Leary et al., 2020; Yadav et al., 2019; Zhou et al., 2017; Lugin et al., 2018; Murali et al., 2020; Wang et al., 2022)
- **Body size** (O’Leary et al., 2020; Löffler et al., 2021; Lugin et al., 2018)
- **Human-likeness**
Human (Aljaroodi et al., 2020; O’Leary et al., 2020; Trovato et al., 2017; Wang et al., 2020; Yadav et al., 2019; Zhou et al., 2017; Cerda Diez et al., 2019; Lugin et al., 2018; Murali et al., 2020; Wang et al., 2022)
Robot (Trovato et al., 2017; Wang et al., 2020; Löffler et al., 2021; Rehm et al., 2018)
Animal (Wang et al., 2020; Chen et al., 2020)
- **Avatar background with cultural landmarks** (Aljaroodi et al., 2020; Obremski et al., 2021; Zhou et al., 2017; Murali et al., 2020)
- **Age** (Trovato et al., 2017; Yadav et al., 2019; Carnell and Lok, 2018; Cerda Diez et al., 2019)

The discussion of hair preferences can extend from the us-

age of certain (predominant) colors within the culture (Aljaroodi et al., 2020) to the usage of specific hairstyles (including lack of hair) (O’Leary et al., 2020; Murali et al., 2020). Familiar clothing marker is explored to provoke a sense of familiarity with the conversational agent (Aljaroodi et al., 2020; O’Leary et al., 2020; Yadav et al., 2019; Murali et al., 2020; Lugin et al., 2018). During a study, participants even considered that a certain dress was not considered appropriate (O’Leary et al., 2020). Similarly, in some contexts revealing clothes of avatars may be perceived as inappropriate (Aljaroodi et al., 2020).

Studies also discussed variables such as body sizes (O’Leary et al., 2020; Löffler et al., 2021). In one study, participants opted for a taller robot due to the resemblance of a ’powerful robot’ since it promoted a sense of effectiveness (Löffler et al., 2021). In Lugin et al. (2018), authors modeled a taller western agent compared to an Arabic. Most conversational agents choose human-likeness closer to human avatars over robot (physical and as an avatar) and animal-looking (avatar). Studies also considered the effect of background with cultural markers (Murali et al., 2020). Lastly, age, Yadav et al. (2019) mention a bot persona as a ”lady in her 30s”. According to the authors, this persona promoted in participants a sense of ”friendly”, causing them to avoid ”hurting bot”, i.e., hesitating to give negative feedback. In Chen et al. (2020), during co-design, participants opted for a Bird agent, described as a ”gentle city-born messenger”.

Regarding **Behavior**, emotion can support the conversational agent in pursuing a better experience interaction (Löffler et al., 2021; Guerrero-Vásquez et al., 2020). Guerrero-Vásquez et al. (2020) argue that non-verbal language can support attention span and engagement in long-lasting interactions. Similarly to the effects of showing emotion, expressions such as moving eyebrow Löffler et al., 2021 and mouth movements (Löffler et al., 2021); or even Gaze shifts (Zhou et al., 2017) on conversational agents can positively influence the interaction (Löffler et al., 2021) and increase effectiveness when compared to none (Löffler et al., 2021).

- **Emotion** (Zhou et al., 2017; Trovato et al., 2017; Löffler et al., 2021; Guerrero-Vásquez et al., 2020)
- **Expression** (Löffler et al., 2021; Guerrero-Vásquez et al., 2020; O’Leary et al., 2020; Trovato et al., 2017; Wang et al., 2020)
- **Gaze shift** (Zhou et al., 2017; Murali et al., 2020)
- **Gesture** (Löffler et al., 2021; Obremski et al., 2021; Chen et al., 2020; Lugrin et al., 2018; Murali et al., 2020)
- **Head nods** (Zhou et al., 2017; Murali et al., 2020)
- **Posture change** (Löffler et al., 2021; Zhou et al., 2017; Murali et al., 2020)

Both emotion and expressions can reach better effectiveness if aligned with gestures (Guerrero-Vásquez et al., 2020). Gestures are one of the non-verbal languages that can be heavily affected by the cultural aspects of a scenario. For example, in robots, motions are preferred if the fluency of the movements is fine-grained and smooth (Löffler et al., 2021). Löffler et al. (2021) mention that hand gestures are essential in the Protestant blessing ritual. Obremski et al. (2021) argue that in a multi-cultural scenario, gestures can foster understanding (either foreign or local). Head nods and posture changes are mentioned but not explained in detail. Evidence also suggests that gestures can enhance comprehension in non-native listeners (Lugrin et al., 2018).

Identity is the third main category. Gender is explored in deep by (Aljaroodi et al., 2020). According to the authors, users pay more attention in a virtual environment when avatars reflect their gender. Furthermore, the authors argue that in certain scenarios, users are more comfortable interacting with an avatar that is less androgynous. Guerrero-Vásquez et al. (2020) also argue that gender can promote more significant levels of identification and familiarity with avatars. Trovato et al. (2017) discuss that the gender of the voice is also important. Yadav et al. (2019) further says that, in some contexts, designers must consider gender-related social norms.

- **Gender** (Aljaroodi et al., 2020; Löffler et al., 2021; Guerrero-Vásquez et al., 2020; Trovato et al., 2017; Yadav et al., 2019; Carnell and Lok, 2018; Danielescu and Christian, 2018; Cerda Diez et al., 2019; Liao and He, 2020; Lugrin et al., 2018; Murali et al., 2020; Wang et al., 2022)
- **Ethnicity** (Guerrero-Vásquez et al., 2020; Obremski et al., 2021; Trovato et al., 2017; Liao and He, 2020)
- **Skin** (Aljaroodi et al., 2020; Trovato et al., 2017; Guerrero-Vásquez et al., 2020; Cerda Diez et al., 2019; Liao and He, 2020; Lugrin et al., 2018; Murali et al., 2020)

For example, in Yadav et al. (2019), a chatbot to support breastfeeding education, the users (mainly mothers) would not be comfortable talking openly to a male avatar. In their scenario, there is a concern with gender inequality and the practice of old traditions. Ethnicity is also discussed in the studies. As mentioned before, Guerrero-Vásquez et al. (2020) discuss in-depth the design of an avatar in a multi-ethnic scenario. Authors argue that the ethnic characteristics of avatars can achieve effective communication. During the

study, participants chose white avatars instead of their demographic and ethnic data. Although studies (Aljaroodi et al., 2020) have demonstrated skin tone can increase the perception of similarity. Evidence also indicates positive effects in race mirroring related to psychotherapeutic chatbots (Liao and He, 2020).

Communication is less discussed compared to the other dimensions. Designers usually choose language to promote familiarity. Whether by using commonly used words or dialects from the scenario (Löffler et al., 2021; O’Leary et al., 2020), non-verbal language (Guerrero-Vásquez et al., 2020; Löffler et al., 2021), and mixed-coded conversations (Yadav et al., 2019).

- **Familiarity** (Löffler et al., 2021; Guerrero-Vásquez et al., 2020; O’Leary et al., 2020; Pataranutaporn et al., 2019; Rahman et al., 2021; Zhou et al., 2017; Trovato et al., 2017; Carnell and Lok, 2018; Cerda Diez et al., 2019; Ludin et al., 2022; Wang et al., 2022)
- **Voice** (Guerrero-Vásquez et al., 2020; Löffler et al., 2021; Wang et al., 2020; Zhou et al., 2017; Cerda Diez et al., 2019; Murali et al., 2020)
- **Conversational style** (Zhou et al., 2017; Carnell and Lok, 2018; Liao and He, 2020)
- **Register** (Danielescu and Christian, 2018)
- **Proficiency** (Obremski et al., 2021; Lugrin et al., 2018)

Studies have demonstrated voice impact towards a positive experience (Guerrero-Vásquez et al., 2020). Usually, variables related to voice in conversational agents are associated with gender (Trovato et al., 2017) or being closer to a synthetic or human voice (Trovato et al., 2017; Zhou et al., 2017). A study has also shown the effectiveness of using a louder voice instead of a soft (Löffler et al., 2021).

Conversational style is usually associated with a particular way of interacting. Some current theories associate language variation to the identity speaker (Carnell and Lok, 2018), explored in (Carnell and Lok, 2018). Less explored variables are related to proficiency (Obremski et al., 2021) and register (Danielescu and Christian, 2018), language formality during a conversation.

The last dimension is related to **Personality**. Designers have shown that the friendly trait positively reflected participants’ perceptions and connection towards health education (Yadav et al., 2019). Empathy is particularly important for health agents (Carnell and Lok, 2018), as the bonding between patient and physician. Politeness (Trovato et al., 2017) and sympathy (Cerda Diez et al., 2019) are also attributed to conversational agents. Other works highlighted specific personality traits, as in (Chen et al., 2020) and (Danielescu and Christian, 2018).

- **Friendly** (Yadav et al., 2019; Chen et al., 2020)
- **Politeness** (Trovato et al., 2017; Yadav et al., 2019)
- **Empathy** (Carnell and Lok, 2018; Cerda Diez et al., 2019; Ludin et al., 2022; Peng et al., 2022)
- **Optimistic, energetic, reassuring, reliable, and efficient** (Chen et al., 2020)
- **Supportive, firm & authoritative, encouraging & cooperative** (Danielescu and Christian, 2018)
- **Sympathy** (Cerda Diez et al., 2019)

5.3 (RQ3) What are the main challenges?

Table 5. Main challenges in culturally informed design practices.

| Node | 1 Level | 2 Level | Article Id |
|------------|----------------|--------------|-------------------------|
| Challenges | Method | Participants | A05, A12, A13 |
| | | Source | A06 |
| | | Translation | A14 |
| | Socio-cultural | | A07, A10, A16, A23 |
| | Predictions | Models | A02, A05, A09, A11, A22 |
| | | Guidelines | A01-A03 |

The last general question (Table 5) focuses on the challenges of the culturally informed design of conversational agents. We aimed for questions not answered in the articles and could indicate research questions. We summarized the results into three emerged categories: method (3 sub-categories), socio-cultural, and predictions (2 sub-categories).

There must be more than current methods to overcome most challenges in designing culturally informed conversational agents. Regarding Participants (in Method), in a participatory method using focus groups study, O’Leary et al. (2020) mention that participants’ literacy level as active members of the design method (writing or drawing) can affect the activity. The authors mention that this can be mitigated with individual interviews, which can facilitate editing or adapting dialogue discussion.

During an evaluation of a conversational agent, domain experts participants in Carnell and Lok (2018) felt hesitant to provide an opinion on whether the dialogue sounded familiar to Chinese or Chinese Americans, as they felt uneducated on the subject or were afraid of having a stereotype. In a cross-cultural co-design scenario (Chen et al., 2020), participants had issues expressing their opinion in their non-native language.

Regarding source, using original manuscripts might affect the interaction in a novel domain of conversational agents, i.e., the Spiritual Embodiment (Pataranutaporn et al., 2019). The corpus text is written in a different context from today. Even the most relevant answer from the corpus text is still distant from the users’ expectations. Besides, users tried to ask about topics not covered in the corpus texts. Authors pose that as a challenge - if the bot embodies the figure of a spiritual entity - “[...] bot should simply not respond, respond with an excuse of not knowing the answer, or answer with the most relevant answer” (Pataranutaporn et al., 2019, pp. 6).

Despite the effort, the process might affect the resulting agent regarding translation approaches. Danielescu and Christian (2018) mentioned that even with the cultural research conducted and the presence of translation experts, some participants felt like they were interacting with a chatbot translated from English.

Regarding socio-cultural challenges, culturally responsible designs require more than tailoring. In this sense, Yadav et al. (2019) argues that the real challenge is how to deliver information effectively. The authors also highlight the gender-related discussion in design consideration. For example, today, there is a current discussion towards gender neutrality due to the apparent bias with female avatars (Feine

et al., 2019). However, Yadav et al. (2019) mention that users felt more inclined to talk and connect during the interaction due to the users perceiving the bot as female. Rahman et al. (2021) also argue that a conservative country’s effect might affect how users interact and perceive their connection towards the conversational agent, especially in the bot acting as a health worker.

During the application of WoZ session in a study, extensively used in conversational agents feasibility tests, authors mentioned the effect of participants’ interaction due to the presence of staff in the same room (Yadav et al., 2019). This can be directly related to sociocultural challenges. Their scenario happens in vulnerable communities with old practices. Even health workers, confident of their knowledge, were surprised when their understanding was wrong. In Kim et al. (2022), participants emphasized the importance of incorporating the community when designing. Participants wanted to include the needs of their close ones to overcome the group’s challenges rather than focusing on their own experiences. Wang et al. (2022) mention that working directly with the intended users respects their conditions while generating value in co-creation and collaboration. Lastly, in most studies addressing sociocultural design challenges, privacy is a main concern (Rahman et al., 2021; Yadav et al., 2019; Kim et al., 2022; Wang et al., 2022).

Another issue is the prediction of cultures. Rehm et al. (2018) emphasize that cultural theories or frameworks make predictions about behavior and are often stereotypical. Largely used, Hofstede’s model can present issues upon its use on cultural tailoring (Wang et al., 2020). Wang et al. (2020) argue that using Hofstede’s model may help designers improve familiarity by mimicking or absorbing cultural markers and promoting values such as likeability, trust, and empathy. However, more is needed to task-oriented and complex scenarios. A test with users to understand how their cultural backgrounds affect their interaction with the conversational agent might be a good first step. Zhou et al. (2017) hypothesized through generalized models that the Chinese agent would be more effective in persuading Chinese participants, but instead, they found the opposite.

Similarly, Guerrero-Vásquez et al. (2020) conjectured that kids would choose avatars that resembled their ethnicity. The distribution in Ecuador is 71.9% mestizo, 7.2% Afro-Ecuadorian, 7.0% indigenous, 7.4% montubio, 6.1% white, and 0.4% from a different ethnic group (Guerrero-Vásquez et al., 2020). In their study, all the boys chose the white male avatar, and more than half of the girls chose the white female avatar, presenting a clear bias towards white avatars and neglect for Afro-Ecuadorian avatars. Guerrero-Vásquez et al. (2020) mention that this might be related to kids under ten may still be unaware of the ethnic traits. Besides, they identified the influence of stereotypes due to the consumption of TV shows, video games, and mobile applications with characters that adhere to the prototype of a human belonging to the northern hemisphere with well-established ethnic or racial characteristics.

Studies also generated guidelines to support the interaction design in their scenarios. For instance, Aljaroodi et al. (2020) discussed precise gender classification, the use of cultural clothing, use of darker colors (skin, facial hair) that better re-

flects the majority of members of the culture in their scenario; Guerrero-Vásquez et al. (2020) discuss the importance of the avatar's voice; Löffler et al. (2021) about size, customization (language/voice), and that conversational agents should have a louder rather than soft voice.

It is important to reflect on the ethics behind the appropriation of guidelines in cultural scenarios. Is it possible that guidelines offer a similar result or impact as Hofstede's model? Rehm et al. (2018) argue that database approaches that rely only on literature data are too unspecific to develop an agent. Some might say that the personalization by the user is an alternative to improve familiarity with the conversational agent (Löffler et al., 2021), but usually only explored with agents that use avatars instead of the profile or language of the conversational agents. The closest work to discuss language personalization is presented in (Obremski et al., 2021), in which users could select the level of grammatical mistakes to create the impression of non-native speakers.

6 Future directions

This SLR shows culturally informed conversational agents' contexts, design practices, and challenges. Still, there are several points still to be addressed by future research:

- (1) Design practices in different contexts;
- (2) Additional resources to support designers;
- (3) Avoid bias with and during design practices;

Regarding **Design practices in different contexts** (1), literature tends to focus on a common context when addressing a topic, as seen in Alsheddi and Alhenaki (2022). Our results have demonstrated a significant interest in discussing cultural design practices in the health domain. It is unclear whether the same benefits and challenges (e.g., trust, adoption, and persuasion) found in those studies apply to other domains such as entertainment, e-commerce, and e-gov, among others.

Scenarios can be divided into general and specific. The general scenario often refers to the target users as the country where they are located, e.g., Arabian, German, and English. Specific scenarios include a narrow placement of users, i.e., "Members of African-American congregations in the Boston Area" - as Geertz argues (Geertz, 1973, pp.4) "cutting the culture concept down to size [...] narrowed, specialized [...]". Moreover, as recognized by Keesing (1974) "[...] not every individual share precisely the same cultural code [...] not every individual knows about all sectors of culture". Adopting more specific descriptions of the scenario can help other designers and researchers assess the results and their use in other scenarios more critically. Rehm et al. (2018) affirms that rather than focusing on the abstract national level, studies and practitioners should frame into local practices or specific groups.

Regarding the **Additional resources to support designers** (2), methods used to tailor culturally conversational agents often present important limitations, primarily when reflecting on design decisions. Chaves and Gerosa (2020) mention chatbots are constantly designed with anthropomorphization characteristics (as seen in Table 4), consequently

promoting the risk of building or reinforcing stereotypes. Most of the works focused on exploring the system's feasibility, and only a few tried to design using the target users part of the design process. When a portrayal of the culture lacks the participation of community members, the result can be misleading (Rehm et al., 2018). Some approaches assume ways to represent the culture under investigation, requiring designers to deal with several complex parameters (Rehm et al., 2018). Consider the translation process seen in Danielescu and Christian (2018), that even with deep research, the resulting agent still felt translated from English. Besides, culture and conversational agents often focus on digital cultural markers today. Zhai and Wibowo (2022) argue that understanding those markers (or variables, dimensions) can positively affect learning. However, many aspects of communication within the culture are yet to be explored.

About **Avoid bias with and during design practices** (3), just language (without considering other cultural dimensions) itself has tons of dialects. It varies with subcultures across groups and spaces. Some works openly argue for the generalized models approach (Machidon et al., 2018). Does the context justify the generalized model, or will it result in wrong assumptions and feasible steps when using such methods? Notably, design decisions can be harmful (O'Leary et al., 2020).

Besides, it is a current challenge that many studies focus on generalized approaches and models. In that case, methods to support the evaluation of the content generated are needed, not just the resulting goals (e.g., persuasion, habits change). However, no universal framework exists to evaluate chatbots (Alsheddi and Alhenaki, 2022), especially considering cultural design practices. It is ethically appropriate to acknowledge the power of designers in making decisions concerning other cultures. For example, some articles stated their position as researchers and members of the same culture from the scenario they are investigating (Yadav et al., 2019).

7 Conclusion

This article described an SLR of design practices of culturally informed conversational agents. In the course of the paper, we aimed to answer the main contexts for interactions (RQ1), what design practices have been mainly used to tailor the conversational agents (RQ2), and the challenges for the field (RQ3)?

Regarding RQ1, health is the most explored scenario, and health assistance and education are the main goals in the articles. Scenarios usually focus on general country definitions (e.g., English users) instead of more specific ones (e.g., Mothers and community health workers residing in underdeveloped slums regions of India). In RQ2, most studies happened during the evaluation phase, with few participatory during design. Several methods were found, with surveys and interviews as the primary option to inform the conversational agents culturally. In one article, even Census was used as a source to design decisions, such as which skin color should the conversational agent have. About cultural markers, designers discussed appearance (e.g., face, familiar clothing, body size), behavior (e.g., emotion, expression, gaze shift),

identity (gender, ethnicity, skin), communication (e.g., familiarity, voice, register), and personality (e.g., friendly, politeness, empathy). Considering RQ3, the authors mentioned challenges with methods, socio-cultural responsible design, and dealing with cultural predictions using models and guidelines.

Those results highlighted three future directions with culturally informed design practices: (1) Design practices in different contexts, (2) Additional resources to support designers, and (3) Avoid bias with and during design practices. Lastly, our research included some limitations. The research questions were proposed to seek aspects relevant to the design fields, so some excess literature resulted. The constrained period limited the survey to some extent, as our focus was on recent articles only, seeking the latest methodologies and techniques. In future work, we intend to consider commercial tools or patents.

References

- Aljaroodi, H. M., Adam, M. T., Chiong, R., Teubner, T., et al. (2019). Avatars and embodied agents in experimental information systems research: A systematic review and conceptual framework. *Australasian Journal of Information Systems*, 23.
- Aljaroodi, H. M., Chiong, R., and Adam, M. T. (2020). Exploring the design of avatars for users from arabian culture through a hybrid approach of deductive and inductive reasoning. *Computers in Human Behavior*, 106:106246.
- Alsheddi, A. S. and Alhenaki, L. S. (2022). English and arabic chatbots: A systematic literature review. *International Journal of Advanced Computer Science and Applications*, 13(8).
- Carnell, S. and Lok, B. (2018). The effects of author identity on dialogue for virtual human communication skills training. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 65–70.
- Cerda Diez, M., E. Cortés, D., Trevino-Talbot, M., Bangham, C., Winter, M. R., Cabral, H., Norkunas Cunningham, T., M. Toledo, D., J. Bowen, D., K. Paasche-Orlow, M., et al. (2019). Designing and evaluating a digital family health history tool for spanish speakers. *International Journal of Environmental Research and Public Health*, 16(24):4979.
- Chaves, A. P. and Gerosa, M. A. (2020). How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, pages 1–30.
- Chen, Z., Lu, Y., Nieminen, M. P., and Lucero, A. (2020). Creating a chatbot for and with migrants: chatbot personality drives co-design activities. In *Proceedings of the 2020 acm designing interactive systems conference*, pages 219–230.
- Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266.
- Danielescu, A. and Christian, G. (2018). A bot is not a polyglot: Designing personalities for multi-lingual conversational agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9.
- Feine, J., Gnewuch, U., Morana, S., and Maedche, A. (2019). Gender bias in chatbot design. In *International Workshop on Chatbot Research and Design*, pages 79–93. Springer.
- Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., et al. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103(12):2915–2942.
- Følstad, A. and Brandtzaeg, P. B. (2017). Chatbots and the new world of hci. *interactions*, 24(4):38–42.
- Geertz, C. (1973). *The interpretation of cultures*, volume 5043. Basic books.
- Guerrero-Vásquez, L. F., Robles-Bykbaev, V. E., López-Nores, M., Bravo-Torres, J. F., Jara-Quito, H. J., Tapiavintimilla, M. A., Pazos-Arias, J. J., and Gil-Solla, A. (2020). Assessing children’s perceptions of live interactions with avatars: Preparations for use in asd therapy in a multi-ethnic context. *IEEE Access*, 8:168456–168469.
- Hofstede, G. (2001). *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. sage.
- Jakobson, R. (2008). *Linguística e comunicação*. Editora Cultrix.
- Keesing, R. M. (1974). Theories of culture. *Annual review of anthropology*, 3(1):73–97.
- Kim, J., Muhic, J., Robert, L. P., and Park, S. Y. (2022). Designing chatbots with black americans with chronic conditions: Overcoming challenges against covid-19. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1):7–15.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
- Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems*, pages 555–565.
- Kuhail, M. A., Alturki, N., Alramlawi, S., and Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, pages 1–46.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y., et al. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.
- Liao, Y. and He, J. (2020). Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. In *Proceedings of the 25th international conference on intelligent user interfaces*, pages 430–442.
- Löffler, D., Hurtienne, J., and Nord, I. (2021). Blessing robot blessu2: a discursive design study to understand the impli-

- cations of social robots in religious contexts. *International Journal of Social Robotics*, 13(4):569–586.
- Ludin, N., Holt-Quick, C., Hopkins, S., Stasiak, K., Hetrick, S., Warren, J., and Cargo, T. (2022). A chatbot to support young people during the covid-19 pandemic in new zealand: Evaluation of the real-world rollout of an open trial. *Journal of Medical Internet Research*, 24(11):e38743.
- Lugrin, B., Eckstein, B., Bergmann, K., and Heindl, C. (2018). Adapted foreigner-directed communication towards virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 59–64.
- Machidon, O. M., Duguleana, M., and Carrozzino, M. (2018). Virtual humans in cultural heritage ict applications: A review. *Journal of Cultural Heritage*, 33:249–260.
- Marcuschi, L. A. (1991). *Análise da conversação*. Ed. Ática.
- Montenegro, J. L. Z., da Costa, C. A., and da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67.
- Mordor (2023). *Global chatbot market - growth, trends, covid-19 impact, and forecasts (2023 - 2028)*. Mordor Intelligence. <https://www.mordorintelligence.com/pt/industry-reports/chatbot-market>.
- Murali, P., Shamekhi, A., Parmar, D., and Bickmore, T. (2020). Argumentation is more important than appearance for designing culturally tailored virtual agents. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1940–1942.
- Obremski, D., Lugrin, J.-L., Schaper, P., and Lugrin, B. (2021). Non-native speaker perception of intelligent virtual agents in two languages: the impact of amount and type of grammatical mistakes. *Journal on Multimodal User Interfaces*, 15(2):229–238.
- O’Leary, T. K., Stowell, E., Kimani, E., Parmar, D., Olafsson, S., Hoffman, J., Parker, A. G., Paasche-Orlow, M. K., and Bickmore, T. (2020). Community-based cultural tailoring of virtual agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8.
- Pataranutaporn, P., Ngamarunchot, B., Chaovavanich, K., Chatwiriyachai, S., Ngamkajornwiwat, P., Ninyawee, N., and Surareungchai, W. (2019). Buddha bot: The exploration of embodied spiritual machine in chatbot. In *Proceedings of the Future Technologies Conference*, pages 589–595. Springer.
- Peng, M. L., Wickersham, J. A., Altice, F. L., Shrestha, R., Azwa, I., Zhou, X., Halim, M. A. A., Ikhtiaruddin, W. M., Tee, V., Kamarulzaman, A., et al. (2022). Formative evaluation of the acceptance of hiv prevention artificial intelligence chatbots by men who have sex with men in malaysia: Focus group study. *JMIR Formative Research*, 6(10):e42055.
- Prietch, S. S., Sánchez, J. A., and García, J. G. (2022). Understanding cultural aspects of deaf communities in méxico towards the codesign of automatic sign language processing systems. *Journal on Interactive Systems*, 13(1):15–25.
- Rahman, R., Rahman, M. R., Tripto, N. I., Ali, M. E., Apon, S. H., and Shahriyar, R. (2021). Adolescentbot: Understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in bangladesh. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.
- Rehm, M., Rodil, K., and Krummheuer, A. L. (2018). Developing a new brand of culturally-aware personal robots based on local cultural practices in the danish health care system. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 2002–2007. IEEE.
- Rheu, M., Shin, J. Y., Peng, W., and Huh-Yoo, J. (2020). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction*, pages 1–16.
- Salgado, L., Pereira, R., and Gasparini, I. (2015). Cultural issues in hci: challenges and opportunities. In *Human-Computer Interaction: Design and Evaluation: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I 17*, pages 60–70. Springer.
- Shawar, B. A. and Atwell, E. (2007). Chatbots: are they really useful? In *Ldv forum*, volume 22, pages 29–49.
- Singh, S. and Beniwal, H. (2022). A survey on near-human conversational agents. *Journal of King Saud University-Computer and Information Sciences*, 34(10):8852–8866.
- Suta, P., Lan, X., Wu, B., Mongkolnam, P., and Chan, J. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*, 9(4):502–510.
- Tharp, B. M. (2009). Defining “culture” and “organizational culture”: From anthropology to the office. *Interpretation a Journal of Bible and Theology*, 2(3):1–5.
- Triandis, H. C. and Vassiliou, V. (1972). Interpersonal influence and employee selection in two cultures. *Journal of Applied Psychology*, 56(2):140.
- Trovato, G., Ramos, J. G., Azevedo, H., Moroni, A., Magossi, S., Simmons, R., Ishii, H., and Takanishi, A. (2017). A receptionist robot for brazilian people: study on interaction involving illiterates. *Paladyn, Journal of Behavioral Robotics*, 8(1):1–17.
- Tylor, E. B. (1871). *Primitive culture: Researches into the development of mythology, philosophy, religion, art and custom*, volume 2. J. Murray.
- Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., and Piterova, A. (2022). An artificial intelligence chatbot for young people’s sexual and reproductive health in india (snehai): Instrumental case study. *Journal of Medical Internet Research*, 24(1):e29969.
- Wang, I., Buchweitz, L., Smith, J., Bornholdt, L.-S., Grund, J., Ruiz, J., and Korn, O. (2020). Wow, you are terrible at this! an intercultural study on virtual agents giving mixed feedback. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.
- Wołk, K., Wołk, A., Wnuk, D., Grześ, T., and Skubis, I. (2022). Survey on dialogue systems including slavic lan-

- guages. *Neurocomputing*, 477:62–84.
- Yadav, D., Malik, P., Dabas, K., and Singh, P. (2019). Feed-pal: Understanding opportunities for chatbots in breast-feeding education of women in india. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30.
- Yin, L., Bickmore, T., and Cortés, D. E. (2010). The impact of linguistic and cultural congruity on persuasion by conversational agents. In *International Conference on Intelligent Virtual Agents*, pages 343–349. Springer.
- Zhai, C. and Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: the case of second language acquisition. *Heliyon*, page e12056.
- Zhou, S., Zhang, Z., and Bickmore, T. (2017). Adapting a persuasive conversational agent for the chinese culture. In *2017 international conference on culture and computing (culture and computing)*, pages 89–96. IEEE.