

How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design (preprint)

Ana Paula Chaves^{a,b} and Marco Aurelio Gerosa^a

^aSchool of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA; ^bFederal University of Technology–Paraná, Campo Mourão, Paraná, Brazil

ARTICLE HISTORY

Compiled October 26, 2020

ABSTRACT

Chatbots’ growing popularity has brought new challenges to HCI, having changed the patterns of human interactions with computers. The increasing need to approximate conversational interaction styles raises expectations for chatbots to present social behaviors that are habitual in human-human communication. In this survey, we argue that chatbots should be enriched with social characteristics that cohere with users’ expectations, ultimately avoiding frustration and dissatisfaction. We bring together the literature on disembodied, text-based chatbots to derive a conceptual model of social characteristics for chatbots. We analyzed 56 papers from various domains to understand how social characteristics can benefit human-chatbot interactions and identify the challenges and strategies to designing them. Additionally, we discussed how characteristics may influence one another. Our results provide relevant opportunities to both researchers and designers to advance human-chatbot interactions.

KEYWORDS

chatbots; social characteristics; human-chatbot interaction; survey

1. Introduction

Chatbots are computer programs that interact with users in natural language (Shawar & Atwell, 2007). The origin of the chatbot concept dates back to 1950 (Turing, 1950). ELIZA (Weizenbaum, 1966) and A.L.I.C.E. (Wallace, 2009) are examples of early chatbot technologies, where the main goal was to mimic human conversations. Over the years, the chatbot concept has evolved. Today, chatbots can have distinct and diverse characteristics, which has resulted in several synonyms, such as multimodal agents, chatterbots, and conversational interfaces. In this survey, we use the term “*chatbot*” to refer to *a disembodied conversational agent that holds a natural language conversation via text-based environment to engage the user in either a general-purpose or task-oriented conversation*.

Chatbots are changing the patterns of interactions between humans and computers (Følstad & Brandtzæg, 2017). Many instant messenger tools, such as Skype, Facebook Messenger, and Telegram provide platforms to develop and deploy chatbots,

which either engage with users in general conversations or help them solve domain specific tasks (Dale, 2016). As messaging tools increasingly become platforms, traditional websites and apps are providing space for this new form of human-computer interaction (HCI) (Følstad & Brandtzaeg, 2017). For example, in the 2018 F8 Conference, Facebook announced that it had 300K active chatbots on Facebook Messenger (Boiteux, 2019). The BotList ¹ website indexes thousands of chatbots for education, entertainment, games, health, productivity, travel, fun, and several other categories. The growth of chatbot technology is changing how companies engage with their customers (Brandtzaeg & Følstad, 2018; Gnewuch, Morana, & Maedche, 2017), students engage with their learning groups (Hayashi, 2015; Tegos, Demetriadis, & Tsiatsos, 2016), and patients self-monitor the progress of their treatment (K. K. Fitzpatrick, Darcy, & Vierhile, 2017), among many other applications.

However, chatbots still fail to meet users' expectations (Brandtzaeg & Følstad, 2018; Jain, Kumar, Kota, & Patel, 2018; Luger & Sellen, 2016; Zamora, 2017). While many studies on chatbot design focus on improving chatbots' functional performance and accuracy (see, e.g., Jiang and E Banchs (2017); Maslowski, Lagarde, and Clavel (2017)), the literature has consistently suggested that chatbots' interactional goals should also include social capabilities (Jain, Kumar, et al., 2018; Q. V. Liao et al., 2018). According to the Media Equation theory (Reeves & Nass, 1996), people naturally respond to social situations when interacting with computers (Fogg, 2003; Nass, Steuer, & Tauber, 1994). As chatbots are designed to interact with users in a way that mimics person-to-person conversations, new challenges in HCI arise (Følstad & Brandtzaeg, 2017; Nguyen & Sidorova, 2018). Neururer and colleagues (2018) state that making a conversational agent acceptable to users is primarily a social, not only technical, problem. Studies on chatbots have shown that people prefer agents who: conform to gender stereotypes associated with tasks (Forlizzi, Zimmerman, Mancuso, & Kwak, 2007); self-disclose and show reciprocity when recommending (S. Lee & Choi, 2017); and demonstrate a positive attitude and mood (Thies, Menon, Magapu, Subramony, & O'Neill, 2017). When chatbots do not meet these expectations, the user may experience frustration and dissatisfaction (Luger & Sellen, 2016; Zamora, 2017). In contrast, designing overly humanized agents results in uncanny feelings and increased expectations (Ciechanowski, Przegalinska, Magnuski, & Gloor, 2018; Gnewuch et al., 2017), which also negatively impacts the interaction. The challenge remains as to what social characteristics are relevant for improving chatbots' communication and social skills and in which domains they have shown to be beneficial.

Although chatbots' social characteristics have been explored in the literature, this knowledge is spread across several domains in which chatbots have been studied, such as customer services, education, finances, and travel. In the HCI domain, some studies focus on investigating the social aspects of human-chatbot interactions (see, e.g., Ciechanowski et al. (2018); Ho, Hancock, and Miner (2018); S. Lee and Choi (2017)). However, most studies focus on a single or small set of characteristics (e.g., Mairesse and Walker (2009); Schlesinger, O'Hara, and Taylor (2018)); in other studies, the social characteristics emerged as secondary, exploratory results (e.g., Tallyn, Fried, Gianni, Isard, and Speed (2018); Toxtli, Cranshaw, et al. (2018)). It has become difficult to find evidence regarding what characteristics are important for designing a particular chatbot, and what research opportunities exist in the field.

Whilst the literature has extensively reviewed the technical aspects of chatbots design (e.g., Ahmad, Che, Zainal, Rauf, and Adnan (2018); Deshpande, Shahane, Gadre,

¹<https://botlist.co/>

Deshpande, and Joshi (2017); Masche and Le (2017); Ramesh, Ravishankaran, Joshi, and Chandrasekaran (2017); Thorne (2017); Walgama and Hettige (2017); Winkler and Söllner (2018)), a lack of studies brings together the social characteristics that influence the way users perceive and behave toward chatbots. To fill this gap, this survey compiles research initiatives for understanding the impact of chatbots’ social characteristics on the interaction. We bring together literature that is spread across several research areas. From our analysis of 56 scientific studies, we derive a conceptual model of social characteristics, aiming to help researchers and designers identify the subset of characteristics that are relevant to their context and how adopting—or neglecting—a particular characteristic may influence the way humans perceive the chatbots. The research question that guided our investigation was: **What chatbot social characteristics benefit human interaction and what are the challenges and strategies associated with them?**

To answer this question, we discuss why designing a chatbot with a particular characteristic can enrich the human-chatbot interaction. Our results provide insight into whether the characteristic is desirable for a particular chatbot, so that designers can make informed decisions by selecting the appropriate subset of characteristics, as well as inspire researchers’ further investigations. In addition, we discuss the influence of humanness and the conversational context on users’ perceptions as well as the interrelationship among the identified characteristics. We stated 22 propositions about how social characteristics may influence one another. In the next section, we present an overview of the studies included in this survey.

2. Overview of the surveyed literature

The literature presents no coherent definition of chatbots; thus, to find relevant studies we used a search string that includes the synonyms *“chatbots, chatterbots, conversational agents, conversational interfaces, conversational systems, conversation systems, dialogue systems, digital assistants, intelligent assistants, conversational user interfaces, and conversational UI”*. We explicitly left out studies that relate to embodiment (*“ECA, multimodal, robots, eye-gaze, gesture”*), and speech input mode (*“speech-based, speech-recognition, voice-based”*). The search string did not include the term “social bots”, because it refers to chatbots that produce content for social networks such as Twitter (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). We did not include “personal assistants” either, since this term consistently refers to commercially available, voice-based assistants such as Google Assistant, Amazon’s Alexa, Apple’s Siri, and Microsoft’s Cortana.

We decided to not include terms that relate to social characteristics/traits, because most studies do not explicitly label their results as such. Additionally, to find studies from a variety of domains (not only computing), we used Google Scholar to search for papers. We started with a set of about one thousand papers. In the first round, we reduced this amount by about half based on the titles. We also removed papers for which full text was not available or written in a language other than English. We started the second round with 464 studies. In this round, we excluded short/position papers, theses, book chapters, and documents published in non-scientific venues. We also read the abstracts and removed all the papers that focused on technical aspects, rather than social ones. The third and last round started with 104 papers. After reading the full texts, we removed the papers that either did not highlight social characteristics or represented previous, ongoing research of another complete study from the same

authors. The datasets with all analyzed studies (original list of results and exclusions per round) is available in Chaves and Gerosa (2019).

After filtering the search results, we had 56 remaining studies. Most of the selected studies are recent publications (less than 10 years old). The publication venues include the domains of human-computer interactions (25 papers), learning and education (8 papers), information and interactive systems (8 papers), virtual agents (5 papers), artificial intelligence (3 papers), and natural language processing (2 papers). We also found papers from health, literature & culture, computer systems, communication, and humanities (1 paper each). Most papers (59%) focus on task-oriented chatbots. General purpose chatbots reflect 33% of the surveyed studies. Most general purpose chatbots (16 out of 19) are designed to handle topic-unrestricted conversations. The most representative specific domain is education, with 9 papers, followed by customer services, with 5 papers. See the supplementary materials for the complete list of topics.

Most surveyed studies adopted real chatbots (35 out of 56); among them, 18 studies analyze logs of conversations or users’ perceptions of third-party chatbots such as Cleverbot (e.g., Corti and Gillespie (2016)), Talkbot (e.g., Brahnham and De Angeli (2012)), and Woebot K. K. Fitzpatrick et al. (2017). Nine studies introduce a self-developed architecture and/or dialogue management (e.g. Dohsaka, Asai, Higashinaka, Minami, and Maeda (2014); Kumar, Ai, Beuth, and Rosé (2010)). In another nine studies, the chatbots were designed for research purposes using third-party platforms for chatbot development and deployment, such as IBM’s Watson service (e.g., Q. V. Liao et al. (2018)) and Microsoft’s Bot Framework (Toxtli et al., 2018) as well as pattern-matching packages such as Artificial Intelligence Markup Language (AIML) (Silvervarg & Jönsson, 2013). When a chatbot was simulated (11 studies), Wizard of Oz (WoZ) is the most used technique (9 studies). In a WoZ study, participants believe to be interacting with a chatbot when, in fact, a person (or “wizard”) pretends to be the automated system (Dahlbäck, Jönsson, & Ahrenberg, 1993). Eight studies do not address a particular chatbot. See the supplementary materials for details.

We analyzed the papers by searching for chatbot behavior or attributed characteristics that influence the way users perceive it and behave toward it. Noticeably, the characteristics and categories are seldom explicitly highlighted to in the literature, so the conceptual model was derived using a qualitative coding process inspired by methods such as Grounded Theory (Auerbach & Silverstein, 2003) (open coding stage). For each study (*document*), we selected relevant statements from the paper (*quotes*) and labeled them as a characteristic (*code*). These steps were performed by one researcher (the first author). After coding all the studies, a second researcher (the second author) reviewed the produced set of characteristics and both researchers participated in discussion sessions to identify characteristics that could be merged, renamed, or removed. At the end, the characteristics were grouped into the categories, depending on whether the characteristic relates to the chatbot’s virtual representation, conversational behavior, or social protocols. Finally, the quotes for each characteristic were labeled as references to benefits, challenges, or strategies.

We derived a total of 11 social characteristics, and grouped them into three categories, as depicted in Table 1: **conversational intelligence**, **social intelligence**, and **personification**. The next section describes the derived conceptual model.

Table 1. Conceptual model of chatbots social characteristics. Eleven characteristics found in the literature were grouped into three main categories, depending on whether the social characteristic relates to the chatbot’s conversational behavior, social protocols, or virtual representation. For each characteristic, we point out the benefits, challenges, and strategies of implementing the characteristic in chatbots as reported in the literature.

	Social Characteristics			Challenges			Strategies		
	Benefits								
Conversational Intelligence	<i>Proactivity</i>	[B1]	to provide additional information	[C1]	timing and relevance	[S1]	to leverage conversational context		
		[B2]	to inspire users and to keep the conversation alive	[C2]	privacy	[S2]	to select a topic randomly		
	<i>Conscientiousness</i>	[B3]	to recover from a failure	[C3]	users’ perception of being controlled				
		[B4]	to improve conversation productivity						
Social Intelligence	<i>Communicability</i>	[B5]	to guide and engage users	[C1]	to handle task complexity	[S1]	conversational flow		
		[B2]	to keep the conversation on track	[C2]	to harden the conversation	[S2]	visual elements		
	<i>Damage control</i>	[B3]	to hold a continuous conversation	[C3]	to keep the user aware of the chatbot’s context	[S3]	confirmation messages		
		[B1]	to unveil functionalities	[C1]	to provide business integration	[S1]	to clarify the purpose of the chatbot		
	<i>Thoroughness</i>	[B2]	to manage the users’ expectations	[C2]	to keep visual elements consistent with textual inputs	[S2]	to advertise the functionality and suggest the next step		
		[B1]	to appropriately respond to harassment			[S3]	to provide a help functionality		
	<i>Manners</i>	[B2]	to deal with testing	[C1]	to deal with unfriendly users	[S1]	emotional reactions		
		[B3]	to deal with lack of knowledge	[C2]	to identify abusive utterances	[S2]	authoritative reactions		
	<i>Moral agency</i>			[C3]	to fit the response to the context	[S3]	to ignore the user’s utterance and change the topic		
						[S4]	<i>conscientiousness</i> and <i>communicability</i>		
Personification	<i>Emotional intelligence</i>	[B1]	to increase human-likeness	[C1]	to decide how much to say	[S5]	to predict users’ satisfaction		
		[B2]	to increase believability	[C2]	to be consistent		Not identified		
	<i>Personalization</i>	[B1]	to increase human-likeness	[C1]	to deal with face-threatening acts	[S1]	to engage in small talk		
		[B2]	to enrich interpersonal relationships	[C2]	to end a conversation gracefully	[S2]	to adhere turn-taking protocols		
	<i>Identity</i>	[B1]	to avoid stereotyping	[C1]	to avoid alienation		Not identified		
		[B2]	to enrich interpersonal relationships	[C2]	to build unbiased training data and algorithms				
	<i>Personality</i>	[B1]	to enrich interpersonal relationships	[C1]	to regulate affective reactions	[S1]	to use social-emotional utterances		
		[B2]	to increase engagement			[S2]	to manifest <i>conscientiousness</i>		
		[B3]	to increase believability	[C1]	privacy	[S3]	reciprocity and self-disclosure		
		[B2]	to provide unique services			[S1]	to learn from and about the user		
		[B3]	to reduce interactional breakdowns			[S2]	to provide customizable agents		
		[B1]	to increase engagement	[C1]	to avoid negative stereotypes	[S3]	visual elements		
		[B2]	to increase human-likeness	[C2]	to balance the <i>identity</i> and the technical capabilities	[S1]	to design and elaborate on a persona		
		[B1]	to increase believability	[C1]	to adapt humor to the users’ culture	[S1]	to use appropriate language		
		[B2]	to enrich interpersonal relationships	[C2]	to balance the <i>personality</i> traits	[S2]	to have a sense of humor		

3. Chatbots Social Characteristics

This section describes the identified social characteristics grouped into categories. As Table 1 depicts, the category **conversational intelligence** includes characteristics that help the chatbot manage interactions. **Social intelligence** focuses on habitual social protocols, while **personification** refers to the chatbot’s perceived identity and personality representations. We also grouped the social characteristics based on the domain in which they were investigated (Table 2). In the following subsections, we describe the identified social characteristics as well as the domains of study. Then, we summarize the relationship among the characteristics in Section 4.3. For each category, a table with an overview of the surveyed studies is provided in the supplementary materials. The supplementary materials also include tables for each social characteristic, listing the studies associated with the domains of study and reported benefits, challenges, and strategies. Finally, the supplementary materials also highlight five constructs that can be used to assess whether social characteristics reach the intended design goals.

3.1. Conversational Intelligence

Conversational intelligence enables the chatbot to actively participate in the conversation and to demonstrate awareness of the topic discussed, the evolving conversational context, and the flow of the dialogue. Therefore, **conversational Intelligence** refers to the ability of a chatbot to effectively converse beyond the technical capability of achieving a conversational goal (Jain, Kumar, et al., 2018). In this section, we discuss social characteristics related to **conversational intelligence**, namely: *proactivity* (18 studies), *conscientiousness* (11 studies), and *communicability* (6 studies). Most of the studies rely on data that comes from the log of the conversations, interviews, and questionnaires. The questionnaires are mostly Likert-scales, and some of them include subjective feedback. Most studies analyzed the interaction with real chatbots, although Wizard of Oz (WoZ) settings are also common. Only two papers did not evaluate a particular type of interaction because they were based on a literature review (Gnewuch et al., 2017) and surveys with chatbot users in general (Brandtzaeg & Følstad, 2017). Only five studies applied only quantitative methods, while seven focused on qualitative methods. The majority of the studies (15) applied mixed methods (both quantitative and qualitative). See the supplementary materials for details.

3.1.1. Proactivity

Proactivity is the capability of a system to autonomously act on the user’s behalf (Salovaara & Oulasvirta, 2004) and thereby reduce the amount of human effort to complete a task (Tennenhouse, 2000). In human-chatbot conversations, a proactive behavior enables a chatbot to share initiative with the user, contributing to the conversation in a more natural way (Morrissey & Kirakowski, 2013). Chatbots may manifest *proactivity* when they initiate exchanges, suggests new topics, provide additional information, or formulate follow-up questions. In this survey, we found 18 papers that report either chatbots with proactive behavior or implications of manifesting a proactive behavior. *Proactivity* (also addressed as “*intervention mode*”) was explicitly addressed in seven studies (Avula et al., 2018; Chaves & Gerosa, 2018; Dyke et al., 2013; Hayashi, 2015; V. Q. Liao et al., 2016; Schuetzler et al., 2018; Tegos et al., 2016). In most of the studies, however, *proactivity* emerged either as an exploratory result, mostly from

Table 2. Studied social characteristics per domain. Research in open domain chatbots has reported most of the social characteristics, except for *Communicability*. In the task-oriented domains, some characteristics are largely influenced by the topic (e.g., *Moral agency* and *Personality*), while others are more generally applied (e.g., *Manners* and *Damage control*).

Domain	Social Characteristics	Studies
Open domain	<i>Proactivity, Conscientiousness, Damage control, Thoroughness, Manners, Moral agency, Emotional intelligence, Personalization, Identity, Personality</i>	(Thies et al., 2017) (Portela & Granell-Canut, 2017) (Shum, He, & Li, 2018) (Morrissey & Kirakowski, 2013) (Curry & Rieser, 2018) (De Angeli, Johnson, & Coventry, 2001) (Hill, Ford, & Farreras, 2015) (Kirakowski, Yiu, et al., 2009) (Mairesse & Walker, 2009) (De Angeli & Brahnam, 2006) (Banks, 2018) (Brahnam & De Angeli, 2012) (Ho et al., 2018) (De Angeli, 2005) (Corti & Gillespie, 2016) (Ptaszynski et al., 2010)
Ethnography	<i>Proactivity, Conscientiousness, Thoroughness, Personalization</i>	(Tallyn et al., 2018)
Task management	<i>Proactivity, Damage control, Manners, Personalization, Identity</i>	(V. Q. Liao, Davis, Geyer, Muller, & Shami, 2016) (Toxtli et al., 2018)
Tourism	<i>Proactivity, Thoroughness, Manners</i>	(Chaves & Gerosa, 2018)
Business	<i>Proactivity, Personalization</i>	(Duijvelshoff, 2017)
Information search	<i>Proactivity, Damage control, Manners, Emotional intelligence</i>	(Avula, Chadwick, Arguello, & Capra, 2018) (Wallis & Norling, 2005)
Decision-making	<i>Proactivity, Damage control, Manners</i>	(Mäurer & Weihe, 2015)
Health-care	<i>Proactivity, Emotional intelligence</i>	(K. K. Fitzpatrick et al., 2017) (Miner et al., 2016)
Credibility assessment	<i>Proactivity, Conscientiousness</i>	(Schuetzler, Grimes, & Giboney, 2018)
Education	<i>Proactivity, Conscientiousness, Damage control, Thoroughness, Manners, Emotional intelligence, Identity, Personality</i>	(Ayedoun, Hayashi, & Seta, 2017) (Coniam, 2008) (Dyke, Howley, Adamson, Kumar, & Rosé, 2013) (Hayashi, 2015) (Kumar et al., 2010) (Silvervarg & Jönsson, 2013) (Sjödén, Silvervarg, Haake, & Gulz, 2011) (Tamayo-Moreno & Pérez-Marín, 2016) (Tegos et al., 2016)
Financial services	<i>Conscientiousness, Communicability, Damage control, Thoroughness, Personalization, Identity</i>	(Candello, Pinhanez, & Figueiredo, 2017) (Duijst, 2017)
Customer services	<i>Conscientiousness, Communicability, Damage control, Thoroughness, Manners, Emotional intelligence, Personalization, Identity</i>	(Araujo, 2018) (Brandtzaeg & Følstad, 2018) (Gnewuch et al., 2017) (Jenkins, Churchill, Cox, & Smith, 2007) (Lasek & Jessa, 2013)
E-commerce	<i>Conscientiousness, Manners</i>	(Jain, Kota, Kumar, & Patel, 2018) (Narita & Kitamura, 2010)
News	<i>Communicability</i>	(Valério, Guimarães, Prates, & Candello, 2017)
Human resources	<i>Communicability, Damage control, Manners, Identity</i>	(Q. V. Liao et al., 2018)
Virtual assistant	<i>Thoroughness, Emotional intelligence, Personalization, Identity</i>	(Ciechanowski et al., 2018) (Zamora, 2017)
Gaming	<i>Thoroughness, Emotional intelligence, Personality</i>	(Dohsaka et al., 2014) (Morris, 2002)
Race-talk	<i>Moral agency, Identity</i>	(Marino, 2014) (Schlesinger et al., 2018)
Humorous talk	<i>Personality</i>	(Meany & Clark, 2010)
Not defined	<i>Proactivity, Conscientiousness, Communicability, Damage control, Personalization, Identity, Personality</i>	(Brandtzaeg & Følstad, 2017) (Jain, Kumar, et al., 2018) (Neururer et al., 2018)

post-intervention interviews and user’s feedback (Duijvelshoff, 2017; Jain, Kumar, et al., 2018; Morrissey & Kirakowski, 2013; Portela & Granell-Canut, 2017; Shum et al., 2018; Thies et al., 2017), or as a strategy to attend to domain-specific requirements (e.g., monitoring, and guidance) (K. K. Fitzpatrick et al., 2017; Mäurer & Weihe, 2015; Silvervarg & Jönsson, 2013; Tallyn et al., 2018; Toxtli et al., 2018). *Proactivity* was mostly investigated in open domain and education chatbots (four studies each). In open domain chatbots, *proactivity* helps improve engagement by introducing new topics to keep the conversation alive (Shum et al., 2018). Educational chatbots rely on *proactivity* to prompt students to think, share, and collaborate (e.g., Dyke et al. (2013)). *Proactivity* was also observed in eight other task-oriented domains, including task management and information searches as can be observed in the supplementary materials.

The surveyed literature evidences several benefits of chatbot *proactivity* in chatbots:

[B1] to provide additional, useful information: literature reveals that *proactivity* in chatbots adds value to interactions (Avula et al., 2018; Morrissey & Kirakowski, 2013; Thies et al., 2017). Investigating evaluation criteria for chatbots, Morrissey and Kirakowski (2013) asked users of a general purpose chatbot to rate the chatbots’ naturalness and report in what areas they excel. Both statistical and qualitative results confirm that taking the lead and suggesting specialized information about the conversation theme correlate to chatbots’ naturalness. Thies et al. (2017) corroborates this result; in post-intervention interviews, ten out of 14 users mentioned they preferred a chatbot that takes the lead and volunteers additional information, such as useful links and song playlists. In a WoZ study, Avula et al. (2018) investigated whether proactive interventions of a chatbot contribute to a collaborative search in a group chat. The chatbot either elicits or infers needed information from the collaborative chat and proactively intervenes in the conversation by sharing useful search results. The intervention modes were not significantly different from each other, but both intervention modes resulted in a statistically significant increase of enjoyment and decrease of effort when compared to the same task with no chatbot interventions. Moreover, in a post-intervention, open-ended question, 16 out of 98 participants self-reported positive perceptions about the provided additional information.

[B2] to inspire users, and keep the conversation alive: proactively suggesting and encouraging new topics have been shown useful to both inspire users (Avula et al., 2018; Chaves & Gerosa, 2018) and keep the conversation alive (Silvervarg & Jönsson, 2013). Participants in the study conducted by Avula et al. (2018) self-reported that the chatbot’s suggestions helped them to get started (7 mentions) and gave them ideas about search topics (4 mentions). After iteratively evaluating prototypes for a chatbot in an educational scenario, Silvervarg and Jönsson (2013) concluded that proactively initiating topics makes the dialogue more fun and reveals topics the chatbot can talk about. The refined prototype also proactively maintains the engagement by posing a follow-up when the student had not provided an answer to the question. Schuetzler et al. (2018) hypothesized that including follow-up questions based on the content of previous messages would result in higher perceived partner engagement. The hypothesis was supported, with participants in the dynamic condition rating the chatbot as more engaging. In an ethnographic data collection (Tallyn et al., 2018), users included photos in their responses to add information about their experience; 85% of these photos were proactively prompted by the chatbot. This result shows that prompting the user for more information stimulates them to expand their entries. Chaves and Gerosa (2018) also observed that chatbots’ proactive messages provided insights about the chatbots’ knowledge, which potentially helped the conversation to

continue. In this paper, we call the strategies to convey the chatbot’s knowledge and capabilities as *communicability*, and we discuss it in Section 3.1.3.

[B3] to recover the chatbot from a failure: in Portela and Granell-Canut (2017) and Silvervarg and Jönsson (2013), *proactivity* is employed to naturally recover from a failure. In both studies, the approach was to introduce a new topic when the chatbot failed to understand the user or could not find an answer, preventing the chatbot from getting stuck and keeping the conversation alive. Additionally, in Silvervarg and Jönsson (2013), the chatbot inserted new topics when users are either abusive or non-sensical. We call the strategies to handle failure and abusive behavior as *damage control*, and we discuss this characteristic in Section 3.2.1.

[B4] to improve conversation productivity: in task-oriented interactions, such as searching or shopping, *proactivity* can improve the conversation’s productivity (Jain, Kumar, et al., 2018). In interviews with first-time users of chatbots, Jain, Kumar, et al. (2018) found that chatbots should ask follow-up questions to resolve and maintain the context of the conversation and reduce the time searching before achieving the goal. Avula et al. (2018) found similar results for collaborative searches; 28 out of 98 participants self-reported that chatbot’s proactive interventions saved collaborators time.

[B5] to guide and engage users: in particular domains, *proactivity* helps chatbots to either guide users or establish and monitor users’ goals. In K. K. Fitzpatrick et al. (2017), the chatbot assigns a goal to the user and proactively prompts motivational messages and reminders to keep the user engaged in the treatment. Mäurer and Weihe (2015) suggest that a decision-making coach chatbot needs to lead the interaction toward guiding the user to a decision. In ethnographic data collection (Tallyn et al., 2018), the chatbot prompts proactive messages that guide the users on what information they need to report. Toxtli et al. (2018) evaluates a chatbot that manage tasks in a workplace. Proactive messages are used to check whether the team member has concluded the tasks, and then report the outcome to the other stakeholders. In the educational context, *proactivity* is used to develop tutors that engage the students and facilitate learning. In Hayashi (2015), the tutor chatbot was designed to provide examples of how other students explained a topic. The network analysis of the learner’s textual inputs shows that students used more key terms and provided more important messages when receiving feedback about other group members. In Hayashi (2015), (Dyke et al., 2013), and (Tegos et al., 2016) the chatbots prompt utterances to encourage the students to reason about a topic. In all three studies, the chatbot condition provided better learning outcomes and increased students’ engagement in the discussions.

The surveyed papers also highlight challenges in providing proactive interactions, such as timing and relevance, privacy, and the user’s perception of being controlled.

[C1] timing and relevance: untimely and irrelevant proactive messages may compromise the success of the interaction. Portela and Granell-Canut (2017) states that untimely turn-taking behavior was perceived as annoying, negatively affecting emotional engagement. V. Q. Liao et al. (2016) and Chaves and Gerosa (2018) reported that *proactivity* can be disruptive. V. Q. Liao et al. (2016) investigated *proactivity* in a workspace environment, hypothesizing that the perceived interruption of agent *proactivity* negatively affects users’ opinion’s. The hypothesis was supported, and the authors found that influencing the sense of interruption was the general aversion to unsolicited messages, regardless of whether they came from a chatbot or a colleague. Chaves and Gerosa (2018) showed that proactively introducing new topics resulted in a high number of ignored messages. The analysis of the conversation log reviewed

that either the new topics were not relevant, or it was not the proper time to start a new topic. Silvervarg and Jönsson (2013) also reported annoyance when a chatbot introduces repetitive topics.

[C2] privacy: in a work-related, group chat, Duijvelshoff (2017) observed privacy concerns regarding the chatbot “reading” the employees’ conversations to proactively act. During a semi-structured interview, researchers presented a mockup of the chatbot to employees from two different enterprises and collected perceptions of usefulness, intrusiveness, and privacy. Employees reported feeling that the chatbot represented their supervisors’ interests, which conveyed as sense of workplace surveillance. Privacy concerns may result in under-motivated users, discomfort about disclosing information, and lack of engagement (Duijvelshoff, 2017).

[C3] user’s perception of being controlled: *proactivity* can be annoying when the chatbot conveys the impression of trying to control the user. Tallyn et al. (2018) report that seven out of 13 participants experienced irritation with the chatbot; one of the most frequent reasons was due to the chatbot directing them to specific places. For the task management chatbot, Toxtli et al. (2018) reported to have adapted the follow-up questions approach to pose questions in a time negotiated with the user. In a previous implementation, the chatbot checked the status of the task twice a day, which participants considered too frequent and annoying.

The surveyed literature also reveals two strategies to provide *proactivity*: leveraging the conversational context and randomly selecting a topic. **[S1] Leveraging the conversational context** is the most frequent strategy (Avula et al., 2018; Chaves & Gerosa, 2018; Duijvelshoff, 2017; Shum et al., 2018; Thies et al., 2017), in which proactive messages relate to contextual information provided in the conversation to increase the usefulness of interventions (Avula et al., 2018; Duijvelshoff, 2017; Shum et al., 2018). Shum et al. (2018) argue that general purpose, emotionally aware chatbots should recognize users’ interests and intents from the conversational context to proactively offer comfort and relevant services. In Duijvelshoff (2017), the chatbot leverages conversational context to suggest new topics and propose to share documents or links to assist employees in a work-related group chat. The chatbots studied by Chaves and Gerosa (2018) introduce new topics based on keywords from previous utterances posted in the chat. According to Shum et al. (2018), leveraging the context can also help smoothly guide the user to a target topic. Researchers in the chatbots domain can refer to the emerging literature on Ambient Intelligence (see, e.g., (Leonidis, Antona, & Stephanidis, 2017; Stefanidi et al., 2019)) to understand how contextual knowledge can be leveraged to convey proactivity. One surveyed paper (Portela & Granell-Canut, 2017) proposes a chatbot that **[S2] selects a topic randomly** but also observes that the lack of context is a major problem of this approach. Contextualized proactive interventions also suggest that the chatbot should be attentive to the conversation, which conveys *conscientiousness*, as discussed in the next section.

3.1.2. *Conscientiousness*

Conscientiousness is a chatbot’s capacity to demonstrate attentiveness to the conversation at hand (Duijst, 2017; Dyke et al., 2013). It enables a chatbot to follow the conversational flow, show understanding of the context, and interpret each utterance as a meaningful part of the whole conversation (Morrissey & Kirakowski, 2013). In this survey, we found 11 papers that reported findings related to *conscientiousness* for chatbot design. Four studies explicitly investigated influences of *conscientiousness* for chatbots (Ayedoun et al., 2017; Coniam, 2008; Jain, Kota, et al., 2018; Schuetzler et

al., 2018). In the remaining studies, *conscientiousness* emerged in exploratory findings. In Dyke et al. (2013), *conscientiousness* emerged from the analysis of conversational logs, while Gnewuch et al. (2017) elicited *conscientiousness* aspects as a requirement for chatbot design when surveying the literature on customer service chatbots. In the remaining studies (Brandtzaeg & Følstad, 2017; Duijst, 2017; Jain, Kumar, et al., 2018; Morrissey & Kirakowski, 2013; Tallyn et al., 2018), *conscientiousness* issues were self-reported by the users in post-intervention interviews and subjective feedback to open-ended questions. *Conscientiousness* was investigated in three studies from the education domain, where the tutor chatbot is expected to be attentive and relevant as well as to control the conversation flow to keep the students focused on the topic (see e.g., Dyke et al. (2013)). *Conscientiousness* was also investigated in seven other domains, including open domain interactions and customer and financial services. See the supplementary materials for a complete list of domains of studies that report *conscientiousness* as a social characteristic for chatbots.

The surveyed literature evidenced benefits of designing conscientious chatbots:

[B1] to provide meaningful answers: some chatbots use simplistic approaches, like pattern matching rules based on keywords or template phrases applied in the last user utterance, to find the most appropriate response (Abdul-Kader & Woods, 2015; Ramesh et al., 2017). However, as the chatbot does not interpret the meaning and the intent of users’ utterances, the best-selected response may still sound irrelevant to the conversation (Coniam, 2008; Duijst, 2017; Dyke et al., 2013). As shown by Duijst (2017), when a chatbot does not interpret the meaning of users’ utterances, users show frustration and the chatbot’s credibility is compromised. This argument is supported by Dyke et al. (2013) when studying chatbots to facilitate collaborative learning. The authors proposed a chatbot that promotes Academically Productive Talk moves. Exploratory results show that the chatbot performed inappropriate interventions, which was perceived as a lack of attention to the conversation. In Jain, Kumar, et al. (2018), participants complained that some chatbots seemed “*completely scripted*,” ignoring user’s inputs that did not fall into the script. In this case, users needed to adapt their inputs to match the chatbot script to be understood, which resulted in dissatisfaction. Besides avoiding frustration, Schuetzler et al. (2018) showed that *conscientiousness* also influences chatbots’ perceived humanness and social presence. The authors invited participants to interact with a chatbot that shows an image and asks the users to describe it. To perform the task, the chatbot could either ask the same generic follow-up questions each time (nonrelevant condition) or respond with a follow-up question related to the last participant’s input (relevant condition), demonstrating attention to the information provided by the user. Statistical analysis of survey rates supported that the relevant condition increased chatbots’ perceived humanness and social presence. In Ayedoun et al. (2017), to motivate the students to communicate in a second language, the proposed chatbot interprets users’ input to detect breakdowns and react using an appropriate communication strategy. In interviews, participants reported “*appreciating the help they got from the chatbot to understand and express what they have got to say*” (Ayedoun et al., 2017). This study was later extended to the context of ECAs (see (Ayedoun, Hayashi, & Seta, 2018) for details).

[B2] to hold a continuous conversation: a conversation with a chatbot should maintain a “*sense of continuity over time*” (Jain, Kumar, et al., 2018) to demonstrate that the chatbot is exerting effort to track the conversation. To do so, it is essential to maintain the topic. When evaluating the naturalness of a chatbot, Morrissey and Kirakowski (2013) found that maintaining a theme is convincing, while failure to do so is unconvincing. Furthermore, based on the literature on customer support chatbots,

Gnewuch et al. (2017) argue that comfortably conversing on any topic related to the service offering is a requirement for task-oriented chatbots. Coniam (2008) reviewed popular chatbots for practicing second languages. The author showed that most chatbots in this field cannot hold continuous conversations, since they are developed to answer the user’s last input. Therefore, they did not have the sense of topic, which resulted in instances of inappropriate response. While the chatbots could change the topic, they could not sustain it afterward. Showing *conscientiousness* also requires the chatbot to understand and track the context, which is particularly important in task-oriented scenarios. In Jain, Kumar, et al. (2018), first-time users stressed positive experiences with chatbots that retained information from previous turns. Two participants also expected the chatbots to retain this context across sessions, thus reducing the need for the user’s additional input per interaction. Keeping the context across sessions was highlighted as a strategy to convey *personalization* and empathy (see Sections 3.2.6 and 3.2.5).

[B3] to steer the conversation toward a productive direction: in task-oriented interactions, a chatbot should understand the purpose of the interaction and strive to conduct the conversation toward this goal in an efficient, productive way (Ayedoun et al., 2017; Duijst, 2017). Brandtzaeg and Følstad (2017) show that productivity is the key motivating factor for using chatbots (68% of the participants mentioned it as the main reason for using chatbots). First-time users in Jain, Kumar, et al. (2018) self-reported that interacting with chatbots should be more productive than using websites, phone apps, and search engines. In this sense, Duijst (2017) compared the user experience when interacting with a chatbot for solving either simple or complex tasks in a financial context. The authors found that, for complex tasks, to keep the conversation on track the user must be aware of the next steps or why something is happening. In the educational context, Ayedoun et al. (2017) proposed a dialogue management approach based on communication strategies to enrich a chatbot with the capability to express its meaning when faced with challenges. Statistical results show that the communication strategies, combined with affective backchannel (which is detailed in Section 3.2.5), are effective in motivating students to communicate and maintain the task flow. Thirty-two participants out of 40 reported that they preferred to interact with a chatbot with these characteristics. Noticeably, the chatbot’s attentiveness to the interactional goal may not be evident to the user if the chatbot passively waits for the user to control the interaction. Thus, *conscientiousness* relates to proactive ability, as discussed in Section 3.1.1.

Nevertheless, challenges in designing conscientious chatbots are also evident in the literature:

[C1] to handle task complexity: as the complexity of tasks increases, more turns are required to achieve a goal; hence, more mistakes may be made. This argument was supported by both Duijst (2017) and Dyke et al. (2013), where the complexity of the task compromised the experience and satisfaction in using the chatbot. Duijst (2017) also highlights that complex tasks require more effort to correct eventual mistakes. Therefore, it is an open challenge to design flexible workflows, where the chatbot recovers from failures and keeps the interaction moving productively toward the goal, despite potential misunderstandings (Gnewuch et al., 2017). Recovering from failure is discussed in Section 3.2.1.

[C2] to harden the conversation: aiming to ensure the conversational structure—and to hide natural language limitations—chatbots are designed to restrict free-text inputs from the user (Duijst, 2017; Jain, Kumar, et al., 2018; Tallyn et al., 2018). However, limiting the choices of interaction may convey a lack of attention to the users’

inputs. In Duijst (2017), one participant mentioned the feeling of “*going through a form or a fixed menu.*” According to Jain, Kumar, et al. (2018), participants consider the chatbot’s understanding of free-text input as a criterion to determine whether it can be considered a chatbot, since chatbots are supposed to chat. In the ethnographic data collection study, Tallyn et al. (2018) reported that eight out of ten participants described the interaction using pre-set responses as too restrictive, although they fulfilled the purpose of nudging participants to report their activities. Thus, the challenge lies in how to leverage the benefits of suggesting predefined inputs without limiting conversational capabilities.

[C3] to keep the user aware of the chatbot’s context: a chatbot should provide a way to inform the user of the current context, especially for complex tasks. According to Jain, Kota, et al. (2018), context can be inferred from explicit user input, or assumed based on data from previous interactions. In both cases, the user and chatbot should be on the same page about the chatbot’s contextual state (Jain, Kota, et al., 2018), providing the users the opportunity to clarify possible misunderstandings (Gnewuch et al., 2017). Jain, Kumar, et al. (2018) highlighted that participants reported negative experience when finding “mismatch[es] between [a] chatbot’s real context and their assumptions of the chatbot context.” We identified three strategies used to provide understanding to a chatbot:

[S1] conversation workflow: designing a conversational blueprint helps to conduct the conversation strictly and productively to the goal (Duijst, 2017). However, Gnewuch et al. (2017) argue that the workflow should be flexible enough to handle both multi-turn and one-turn question-answer interactions; it also should be unambiguous, such that users can efficiently achieve their goals. In addition, Duijst (2017) discuss that the workflow should make it easy to fix mistakes; otherwise, the users need to restart the workflow, which leads to frustration. In Ayedoun et al. (2017), the conversation workflow included communicative strategies to detect a learner’s breakdowns and pitfalls. In that study, when the student does not respond, the chatbot uses a comprehension-check question to detect whether the student understood what was said. Then, it reacts to the user’s input by adopting one of the proposed communication strategies (e.g., asking for repetition or simplifying the previous sentence). The conversation workflow could also allow the chatbot to be proactive. For example, participants in Jain, Kumar, et al. (2018) suggested that proactive follow-up questions would anticipate the resolution of the context, reducing the effort required from the user to achieve the goal.

[S2] visual elements: user-interface resources—such as quick replies, cards, and carousels—are used to structure the conversation and reduce issues regarding understanding (Duijst, 2017; Jain, Kumar, et al., 2018; Tallyn et al., 2018). Using these resources, the chatbot shows the next possible utterances (Jain, Kumar, et al., 2018) and conveys the conversational workflow in a step-by-step manner (Duijst, 2017; Tallyn et al., 2018). Visual elements are also used to show the user what the chatbot can (or cannot) do. This is another conversational characteristic, which will be discussed in the Section 3.1.3.

[S3] context window: to keep the user aware of the chatbot’s current context, Jain, Kota, et al. (2018) developed a chatbot for shopping that shows a context window on the side of the conversation. In this window, the user can click on specific attributes and change them to fix inconsistencies. A survey showed that the chatbot outperformed a default chatbot (without the context window) for mental demand and effort constructs. However, when the chatbots are built into third-party apps (e.g., Facebook Messenger), an extra window may not be possible.

[S4] confirmation messages: a conversation workflow may include confirmation messages to convey the chatbots’ context to the user (Jain, Kota, et al., 2018). In Duijst (2017), when trying to block a stolen credit card, a confirmation message is used to verify the given personal data. In Ayedoun et al. (2017), confirmation messages are used as a communicative strategy to check whether the system’s understanding about a particular utterance matches what the user actually meant. Balancing the number of confirmation messages (see (Duijst, 2017)) and the right moment to introduce them into the conversation flow is still under-investigated.

The surveyed literature supports that failing to demonstrate understanding about the users’ individual utterances, the conversational context, and the interactional goals results in frustration and loss of credibility. However, most of the results are exploratory; a lack of studies investigate the extent to which the provided strategies influence users’ behaviors and perceptions. In the field of human-human communication, the cooperative principle (Grice, Cole, Morgan, et al., 1975) has been extensively applied to understand conversational partners’ expectations during a conversation. Additionally, recent studies have shown that this principle influences chatbots’ perceived humanness (Jacquet, Baratgin, & Jamet, 2018; Jacquet, Hullin, Baratgin, & Jamet, 2019). Particularly, the maxims of relation (the ability to be relevant) and manner (the ability to be clear and orderly) can contribute to the design of *conscientious* chatbots. In addition, *conscientiousness* is by itself a *personality* trait; the more *conscientiousness* a chatbot manifests, the more it can be perceived as attentive, organized, and efficient. The relationship between *conscientiousness* and *personality* is highlighted in Section 4.3.

3.1.3. Communicability

Interactive software is communicative by its nature, since users achieve their goals by exchanging messages with the system (Prates, de Souza, & Barbosa, 2000; Sundar, Bellur, Oh, Jia, & Kim, 2016). In this context, *communicability* is defined as the capacity of a software to convey to users its underlying design intent and interactive principles (Prates et al., 2000). Providing *communicability* helps users to interpret the codes used by designers to convey the interactional possibilities embedded in the software (De Souza, Prates, & Barbosa, 1999), which improves system learnability (Grossman, Fitzmaurice, & Attar, 2009). In the chatbot context, *communicability* is, therefore, the capability of a chatbot to convey its features to users (Valério et al., 2017). The problem around chatbots’ *communicability* lies in the nature of the interface: instead of buttons, menus, and links, chatbots unveil their capabilities through conversational turns, one sentence at a time (Valério et al., 2017), bringing new challenges to the system learnability field. The lack of *communicability* may lead users to give up on using the chatbot when they cannot understand the available functionalities and how to use them (Valério et al., 2017).

In this survey, we found six papers that describe *communicability*, although investigating *communicability* is the main purpose of only one (Valério et al., 2017). Conversational logs revealed *communicability* needs in three studies (Lasek & Jessa, 2013; Q. V. Liao et al., 2018); in the other three studies (Duijst, 2017; Gnewuch et al., 2017; Jain, Kumar, et al., 2018) *communicability* issues were self-reported by the users in post-intervention interviews and subjective feedback. *Communicability* was investigated in the customer services domain (two studies), where chatbots guide customers through available functionalities (Valério et al., 2017). The remaining task-oriented domains in which *communicability* was investigated are financial services, news, and

human resources. We did not identify studies on open domain chatbots that report on *communicability*, since in open domain interactions users are free to talk about varying topics and guidance is less a concern.

The surveyed literature reports two main benefits of *communicability* for chatbots:

[B1] to unveil functionalities: while interacting with chatbots, users may not know that a desired functionality is available or how to use it (Jain, Kumar, et al., 2018; Valério et al., 2017). Most participants in Jain, Kumar, et al. (2018)’s study mentioned that they did not understand the functionalities of at least one of the chatbots and none of them mentioned searching for the functionalities in other sources (e.g., Google search or the chatbot website) rather than exploring options during the interaction. In a study about playful interactions in a work environment, Q. V. Liao et al. (2018) observed that 22% of the participants explicitly asked the chatbot about its capabilities (e.g., “*what can you do?*”), and 1.8% of all the users’ messages were ability-check questions. In a study about hotel chatbots, Lasek and Jessa (2013) verified that 63% of the conversations were initiated by clicking an option displayed in the chatbot welcome message. A semiotic inspection of news-related chatbots (Valério et al., 2017) evidenced that *communicability* strategies are effective at providing clues about the chatbot’s features and ideas about what to do and how.

[B2] to manage users’ expectations: Jain, Kumar, et al. (2018) observed that when first-time users do not understand chatbots’ capabilities and limitations, they have high expectations and, consequently, end up more frustrated when the chatbots fail. Some participants blamed themselves for not knowing how to communicate and gave up. In Q. V. Liao et al. (2018), quantitative results evidenced that ability-check questions can be considered signals of users struggling with functional affordances. Users posed ability-check questions after encountering errors as a means of establishing a common ground between the chatbot’s capabilities and their own expectations. According to the authors, ability-check questions helped users to understand the system and reduce uncertainty (Q. V. Liao et al., 2018). In Duijst (2017), users also demonstrated the importance of understanding chatbots’ capabilities in advance. Since the tasks related to financial support, users expected the chatbot to validate the personal data provided and to provide feedback after completing the task (e.g., explaining how long it would take for the credit card to be blocked). Therefore, *communicability* helps users gain a sense of which type of messages or functionalities a chatbot can handle.

The surveyed literature also highlights two challenges of providing *communicability*:

[C1] to provide business integration: communicating chatbots’ functionalities should be performed as much as possible within the chat interface (Jain, Kumar, et al., 2018). Chatbots often act as an intermediary between users and services. In this case, to overcome technical challenges, chatbots answer users’ inputs with links to external sources, where the request will be addressed. First-time users expressed dissatisfaction with this strategy in Jain, Kumar, et al. (2018). Six participants complained that the chatbot redirected them to external websites. According to Gnewuch et al. (2017), business integration is a requirement for designing chatbots, so that the chatbot can solve the users’ requests without transferring the interaction to another user interface.

[C2] to keep visual elements consistent with textual inputs: in the semiotic engineering evaluation, Valério et al. (2017) observed that some chatbots responded differently depending on whether the user accesses a visual element in the user-interface or types the desired functionality in the text-input box, even if both input modes result in the same utterance. This inconsistency produces the user’s feeling of misinterpreting the affordances, which has a negative impact on the system’s learnability.

As an outcome of the semiotic inspection process, Valério et al. (2017) present a set

of strategies to provide *communicability*. Some of them are also emphasized in other studies, as follows:

[S1] to clarify the purpose of the chatbot: First-time users in Jain, Kumar, et al. (2018) highlighted that a clarification about the chatbots’ purpose should be placed in the introductory message. Gnewuch et al. (2017) found similar inference from the literature on customer services chatbots. The authors argue that providing an opening message with insights into the chatbots’ capabilities is a requirement for chatbots design. In addition, a chatbot could give a short tour throughout the main functionalities at the beginning of the first sessions (Valério et al., 2017).

[S2] to advertise the functionality and suggest the next step: when the chatbot is not able to answer the user, or when it notices that the user is silent, it may suggest available features to stimulate the user to engage (Jain, Kumar, et al., 2018; Valério et al., 2017). In Jain, Kumar, et al. (2018), six participants mentioned that they appreciated when the chatbot suggested responses, for example by saying “*try a few of these commands: ...*” (Jain, Kumar, et al., 2018). Valério et al. (2017) shows that chatbots use visual elements, such as cards, carousel, and menus (persistent or not) to show contextualized clues about the next answer, which both fulfills *communicability* purpose and spares users from having to type.

[S3] to provide a help functionality: chatbots should recognize a “*help*” input from the user, so it can provide instructions on how to proceed (Valério et al., 2017). Jain, Kumar, et al. (2018) reported that users highlighted this functionality as useful for the reviewed chatbots. Also, results from (Q. V. Liao et al., 2018) show that chatbots should be able to answer ability-check questions (e.g., “*what can you do?*” or “*can you do [functionality]?*”).

The literature states the importance of communicating chatbots’ functionality to the success of the interaction. Failing to provide *communicability* leads the users to frustration and they often give up when they do not know how to proceed. The literature on interactive systems has highlighted system learnability as the most fundamental component for usability (Grossman et al., 2009), and an easy-to-learn system should lead the user to perform well, even during their initial interactions. Thus, researchers in the chatbots domain can leverage the vast literature on system learnability to identify metrics and evaluation methodologies, as well as propose new forms of *communicability* strategies that reduce the learnability issues in chatbot interactions. *Communicability* may also be used as a strategy to avoid mistakes (*damage control*), which will be discussed in Section 3.2.1.

In summary, the **conversational intelligence** category includes characteristics that help a chatbot to perform a proactive, attentive, and informative role in the interaction. Acknowledging that functional aspects play a crucial role in the ability of a chatbot to converse intelligently, the survey focuses on the characteristics that go beyond the functional aspects, looking through the lens of social perceptions. The highlighted benefits relate to how a chatbot manages the conversation to make it productive, interesting, and neat. To achieve that, designers and researchers should care for the timing and relevance of provided information, privacy, interactional flexibility, and consistency.

3.2. Social Intelligence

Social Intelligence refers to the ability of an individual to produce adequate social behavior for the purpose of achieving desired goals (Björkqvist, Österman, & Kaukiainen, 2000). In the HCI domain, the Media Equation theory (Reeves & Nass, 1996) posits that people react to computers as social actors. Hence, when developing chat-

bots, it is necessary to account for the socially acceptable protocols for conversational interactions (Walker, 2009; Wallis & Norling, 2005). Chatbots should be able to respond to social cues during the conversation, accept differences, and manage conflicts (Salovey & Mayer, 1990) as well as be empathic and demonstrate caring (Björkqvist et al., 2000), which ultimately increase chatbots’ authenticity (Neururer et al., 2018). In this section, we discuss the social characteristics related to **social intelligence**, namely: *damage control* (12 papers), *thoroughness* (13 papers), *manners* (10 papers), *moral agency* (6 papers), *emotional intelligence* (14 papers), and *personalization* (11 papers). Although the focus of the investigations is diverse, we found more studies where the focus of the investigation relates to a specific social characteristic, particularly *moral agency* and *emotional intelligence*, when compared to the **conversational intelligence** category. Regarding the adopted research methods, 12 studies applied qualitative methods and 10 focuses on quantitative methods. 16 studies reported both qualitative and quantitative results.

3.2.1. *Damage control*

Damage control is the ability of a chatbot to deal with either conflict or failure situations. Although the Media Equation theory argues that humans socially respond to computers as they respond to other people (Reeves & Nass, 1996), the literature has shown that interactions with conversational agents are not quite equal to human-human interactions (Luger & Sellen, 2016; Mou & Xu, 2017; Shechtman & Horowitz, 2003). When talking to a chatbot, humans are more likely to harass (Hill et al., 2015), test the agent’s capabilities and knowledge (Wallis & Norling, 2005), and feel disappointed with mistakes (Jain, Kumar, et al., 2018; Mäurer & Weihe, 2015). When a chatbot does not respond appropriately, it may encourage the abusive behavior (Curry & Rieser, 2018) or disappoint the user (Jain, Kumar, et al., 2018; Mäurer & Weihe, 2015), which ultimately leads the conversation to fail (Jain, Kumar, et al., 2018). Thus, it is necessary to enrich chatbots with the ability to recover from failures and handle inappropriate talk in a socially acceptable manner (Jain, Kumar, et al., 2018; Silvervarg & Jönsson, 2013; Wallis & Norling, 2005).

In this survey, we found 12 studies that discuss *damage control* as a relevant characteristic for chatbots, two of which focus on conflict situations, such as testing and flaming (Silvervarg & Jönsson, 2013; Wallis & Norling, 2005). In the remaining studies (Curry & Rieser, 2018; De Angeli, Johnson, & Coventry, 2001; Duijst, 2017; Gnewuch et al., 2017; Jain, Kumar, et al., 2018; Jenkins et al., 2007; Lasek & Jessa, 2013; Q. V. Liao et al., 2018; Mäurer & Weihe, 2015; Toxtli et al., 2018), needs for *damage control* emerged from the analysis of conversational logs and users’ feedback. *Damage control* was mostly investigated for open domain (two studies) and customer service chatbots (three studies). In open domain interactions, users are free to wander among topics and testing or flaming tends to be more frequent (Hill et al., 2015). In the customer services context, the chatbot needs to avoid disappointing the user, as frustration may negatively reflect the business that the agent represents (Araujo, 2018). *Damage control* was also identified in other six domains (see the supplementary materials), such as task management and financial services.

The surveyed literature highlights the following benefits of providing *damage control* in chatbots:

[B1] to appropriately respond to harassment: chatbots are more likely to targets of profanity than humans would (Hill et al., 2015). When analyzing conversation logs from hotel chatbots, Lasek and Jessa (2013) observed that 4% of the conversations

contained vulgar, indecent, and insulting vocabulary, and 2.8% of all statements were abusive. Qualitative evaluation reveals that the longer the conversations last, the more users are encouraged to go beyond the chatbots main functions. In addition, sexual expressions represented 1.8% of all statements. A similar number was found in a study with general purpose chatbots (Curry & Rieser, 2018). When analyzing a corpus from the Amazon Alexa Prize 2017, the researchers estimated that about 4% of the conversations included sexually explicit utterances. Curry and Rieser (2018) used utterances from this corpus to harass a set of state-of-art chatbots and analyze the responses. The results show that chatbots respond to harassment in a variety of ways, including nonsensical, negative, and positive responses. However, the authors highlight that the responses should align with the chatbot’s goal to avoid encouraging the behavior or reinforcing stereotypes.

[B2] to deal with testing: abusive behavior is often used to test chatbots’ social reactions (Lasek & Jessa, 2013; Wallis & Norling, 2005). During the evaluation of a virtual guide to the university campus (Wallis & Norling, 2005), a participant answered the chatbot’s introductory greeting with “*moron*,” likely hoping to see how the chatbot would answer. Wallis and Norling (2005) argue that handling this type of testing helps the chatbots to establish limits and resolve social positioning. Other forms of testing were highlighted in Silvervarg and Jönsson (2013), including sending random letters, repeated greetings, laughs and acknowledgments, and posing comments and questions about the chatbot’s intellectual capabilities. When analyzing conversations with a task management chatbot, Q. V. Liao et al. (2018) observed that casually testing the chatbots’ “intelligence” is a manifestation of satisfaction seeking. In Jain, Kumar, et al. (2018), first-time users appreciated when the chatbot successfully performed tasks when the user expected the chatbot to fail, which shows that satisfaction is influenced by the ability to provide a clever response when the user tests the chatbot.

[B3] to deal with lack of knowledge: chatbots often fail in a conversation due to lack of either linguistic or world knowledge (Wallis & Norling, 2005). *Damage control* enables the chatbot to admit the lack of knowledge or cover up cleverly (Jain, Kumar, et al., 2018). When analyzing the log of a task management chatbot, Toxtli et al. (2018) found out that the chatbot failed to answer 10% of the exchanged messages. The authors suggest that the chatbot should be designed to handle novel scenarios when the current knowledge is not enough to answer the requests. In some task-oriented chatbots, though, the failure may not be caused by a novel scenario, but by an off-topic utterance. In the educational context, Silvervarg and Jönsson (2013) observed that students posted off-topic utterances when they did not know what topics they could talk about, which led the chatbot to fail rather than help the users to understand its knowledge. In task-oriented scenarios, the lack of linguistic knowledge may lead the chatbots to get lost in the conversational workflow (Gnewuch et al., 2017), compromising the success of the interaction. Mäurer and Weihe (2015) demonstrated that dialogue-reference errors (e.g., user’s attempt to correct a previous answer or jumping back to an earlier question) are one of the major reasons for failed dialogues and they mostly result from chatbots’ misunderstandings.

The literature also reveals some challenges to provide *damage control*:

[C1] to deal with unfriendly users: Silvervarg and Jönsson (2013) argue that users who want to test and find the system’s borders are likely to never have a meaningful conversation with the chatbot no matter how sophisticated it is. Thus, there is a limit to which *damage control* strategies will be effective to avoid testing and abuse. In De Angeli, Johnson, and Coventry (2001), the authors observed human tendencies to dominate, be rude, and infer stupidity, which they call “*unfriendly partners*.” After

an intervention where users interacted with a chatbot for decision-making coaching, Mäurer and Weihe (2015) evaluated participants’ self-perceived work and cooperation with the system. The qualitative results show that cooperative users are significantly more likely to give a higher rating for overall evaluation and decision efficiency. The qualitative analysis of the conversation log reveals that a few interactions failed because the users’ motivations were curiosity and mischief rather than trying to solve the decision problem.

[C2] to identify abusive utterances: several chatbots are trained on “clean” data. Because they do not understand profanity or abuse, they may not recognize a statement as harassment, which makes it difficult to adopt answering strategies (Curry & Rieser, 2018). Curry and Rieser (2018) shows that data-driven chatbots often provide non-coherent responses to harassment. Sometimes, these responses conveyed the impression of flirtatious or counter-aggression. Providing means to identify an abusive utterance is important to adopting *damage control* strategies.

[C3] to fit the response to the context: Wallis and Norling (2005) argue that humans negotiate conflict and social positioning well before reaching abuse. In human-chatbot interactions, however, predicting users’ behavior toward the chatbots in a specific context to develop the appropriate behavior is a challenge. *Damage control* strategies need to be adapted to both the social situation and the intensity of the conflict. For example, Curry and Rieser (2018) showed that being evasive about sexual statements may convey the impression of flirtatiousness, which would not be an acceptable behavior for a customer assistant or a tutor chatbot. In contrast, adult chatbots are supposed to flirt, so encouraging behaviors are expected in some situations. Wallis and Norling (2005) argue that when the chatbot is not accepted as part of the social group it represents, it is discredited by the user, leading the interaction to fail. In addition, designing chatbots with too strong of reactions may lead to ethical concerns (Wallis & Norling, 2005). For Björkqvist et al. (2000), choosing between peaceful or aggressive reactions in conflict situations is optional for socially intelligent individuals. Enriching chatbots with the ability to choose between the options is a challenge.

Damage control strategies depend on the type of failure and the target benefit, as following:

[S1] emotional reactions: Wallis and Norling (2005) suggest that when faced with abuse, a chatbot could be seen to take offense and respond in kind or act hurt. The authors argue that humans might feel inhibited about hurting the pretended feelings of a machine if the machine is willing to hurt a human’s feelings too (Wallis & Norling, 2005). If escalating the aggressive behavior is not appropriate, the chatbot could withdraw from the conversation (Wallis & Norling, 2005) to demonstrate that the user’s behavior is not acceptable. In De Angeli, Johnson, and Coventry (2001), the authors discuss that users appeared to be uncomfortable and annoyed whenever the chatbot pointed out any defect in the user or reacted to aggression, as this behavior conflicted with the user’s perceived power relations. This strategy is also applied in Silvervarg and Jönsson (2013), where abusive behavior may lead the chatbot to stop responding until the student changes the topic. Curry and Rieser (2018) categorized responses from state-of-the-art conversational systems in a pool of emotional reactions, both positive and negative. The reactions include humorous responses, chastising and retaliation, and evasive responses, as well as flirtation and play-along utterances. To provide an emotional reaction, *emotional intelligence* is also required. This category is presented in Section 3.2.5.

[S2] authoritative reactions: when facing testing or abuse, chatbots can communicate consequences (Silvervarg & Jönsson, 2013) or call on the authority of others

(Toxtli et al., 2018; Wallis & Norling, 2005). In Wallis and Norling (2005), although the wizard acting as a chatbot was conscientiously working as a campus guide, she answered a bogus caller with “*This is the University of Melbourne. Sorry, how can I help you?*” The authors suggest that the wizard was calling on the authority of the university to handle the conflict, where being part of a recognized institution places the chatbot in a stronger social group. In Silvervarg and Jönsson (2013), when students recurrently harass the chatbot, the chatbot informs the student that further abuse will be reported to the (human) teacher (although the paper does not clarify whether the problem is, in fact, escalated to a human). Toxtli et al. (2018) and Jenkins et al. (2007) also suggest that chatbots could redirect users’ problematic requests to a human attendant in order to avoid conflict situations.

[S3] to ignore the user’s utterance and change the topic: Wallis and Norling (2005) argue that ignoring abuse and testing is not a good strategy because it could encourage more extreme behaviors. It also positions the chatbot as an inferior individual, which is particularly harmful in scenarios where the chatbot should demonstrate a more prominent or authoritative social role (e.g., a tutor). However, this strategy has been found in some studies to handle lack of knowledge. When iteratively developing a chatbot for an educational context, Silvervarg and Jönsson (2013) proposed initiating a new topic in one out of four user’s utterances that the chatbot did not understand.

[S4] *conscientiousness and communicability*: successfully implementing *conscientiousness* and *communicability* may prevent errors; hence, strategies to provide these characteristics can also be used for *damage control*. In Silvervarg and Jönsson (2013), when users utter out-of-scope statements, the chatbot could make it clear what topics are appropriate to the situation. For task-oriented scenarios, where the conversation should evolve toward a goal, Wallis and Norling (2005) argue that the chatbot can clarify the purpose of the offered service when facing abusive behavior, bringing the user back to the task. Jain, Kumar, et al. (2018) showed that describing chatbot’s capabilities after failures in the dialogue was appreciated by first-time users. In situations where the conversational workflow is susceptible to failure, Gnewuch et al. (2017) discuss that posting confirmation messages avoids trapping the users in the wrong conversation path. Participants in Duijst (2017) also suggested back buttons as a strategy to fix mistakes in the workflow. In addition, the exploratory results about the user interface showed that having visual elements such as quick replies prevents errors, since they keep the users aware of what to ask and the chatbot is more likely to know how to respond (Duijst, 2017).

[S5] to predict users’ satisfaction: chatbots should perceive both explicit and implicit feedback about users’ (dis)satisfaction (Q. V. Liao et al., 2018). To address this challenge, Q. V. Liao et al. (2018) invited participants to send a “*#fail*” statement to express dissatisfaction. The results show that 42.4% of the users did it at least once, and the number of complaints and flaming for the proposed chatbot was significantly lower than the baseline. However, the amount of implicit feedback was also significant, which advocates for predicting user’s satisfaction from the conversation. The most powerful conversational act that predicted user satisfaction in that study was the agent ability-check questions, see discussion in *Communicability* section) and the explicit feedback *#fail*, although closings and off-topic requests were also significant in predicting frustration. Although these results are promising, more investigation is needed to identify other potential predictors of users’ satisfaction in real-time, in order to provide appropriate reactions.

Damage control strategies have different levels of severity. Deciding what strategy is adequate to the intensity of the conflict is crucial (Wallis & Norling, 2005). The

strategies can escalate in severity if the conflict is not solved. For example, Silvervarg and Jönsson (2013) uses a sequence of clarification, suggesting a new topic, and asking a question about the new topic. In case of abuse, the chatbot refers to an authority after two attempts at changing the topic.

According to Wallis and Norling (2005), humans also fail in conversations; they misunderstand what their partner says and do not know things that are assumed common knowledge by others. Hence, it is unlikely that chatbots interactions will evolve to be conflict-free. That said, *damage control* intends to avoid escalating the conflicts and manifesting an unexpected behavior. In this sense, politeness can be used as a strategy to minimize the effect of lack of knowledge (see Section 3.2.3), managing the conversation despite the possible mistakes. Regarding interpersonal conflicts, the strategies are in line with the theory on human-human communication, which includes non-negotiation, emotional appeal, personal rejection, and emphatic understanding (M. A. Fitzpatrick & Winke, 1979). Further research on *damage control* can evaluate the adoption of human-human strategies in human-chatbot communication.

3.2.2. Thoroughness

Thoroughness is the ability of a chatbot to be precise regarding how it uses language (Morrissey & Kirakowski, 2013). In traditional user interfaces, user communication takes place using visual affordances, such as buttons, menus, or links. In a conversational interface, language is the main tool to achieve the communicative goal. Thus, chatbots should coherently use language that portrays the expected style (Mairesse & Walker, 2009). When the chatbot uses inconsistent language, or unexpected patterns of language (e.g., excessive formality), the conversation may sound strange to the user, leading to frustration.

We found 13 papers that report the importance of *thoroughness* in chatbot design, three of which investigate how patterns of language influence users' perceptions and behavior toward the chatbots (Duijst, 2017; Hill et al., 2015; Mairesse & Walker, 2009). Gnewuch et al. (2017) and Morris (2002) suggest design principles that include concerns about language choices. Log of conversations revealed issues regarding *thoroughness* in two studies (Coniam, 2008; Jenkins et al., 2007). In the remaining papers, *thoroughness* emerged from interviews and users' subjective feedback (Chaves & Gerosa, 2018; Kirakowski et al., 2009; Morrissey & Kirakowski, 2013; Tallyn et al., 2018; Thies et al., 2017; Zamora, 2017). *Thoroughness* is mainly reported for open domain (five studies) and customer service chatbots (two studies), where the interactions are expected to be natural and credible to succeed (Gnewuch et al., 2017; Morrissey & Kirakowski, 2013). *Thoroughness* was also reported in another six domains of studies, such as financial services and education (see the supplementary materials).

We found two benefits of providing *thoroughness*:

[B1] to increase human-likeness: chatbot utterances are often pre-recorded by the chatbot designer (Mairesse & Walker, 2009). On the one hand, this approach produces high quality utterances; on the other hand, it reduces flexibility, since the chatbot is not able to adapt the tone of the conversation based on individual users and conversational context. When analyzing interactions with a customer representative chatbot, Jenkins et al. (2007) observed that the chatbot proposed synonyms to keywords, and the repetition of this vocabulary led the users to imitate it. Hill et al. (2015) observed a similar tendency to matching language style. The authors compared human-human conversations with human-chatbots conversations regarding language use. They found that people use, indeed, fewer words per message and a more limited

vocabulary with chatbots. However, a deeper investigation revealed that the human interlocutors were actually matching the patterns of language use with the chatbot, who sent fewer words per message. When interacting with a chatbot that uses many emojis and letter reduplication (Thies et al., 2017), participants reported a draining experience, since the chatbot’s energy was too high to match. These outcomes show that adapting the language to the interlocutor is a common behavior for humans, and so chatbots would benefit from manifesting it. In addition to the interlocutor, chatbots should adapt their language use to the context in which they are implemented and adopt appropriate linguistic register (Gnewuch et al., 2017; Morrissey & Kirakowski, 2013). In the customer services domain, Gnewuch et al. (2017) state that chatbots are expected to fulfill the role of a human; hence, they should produce language that corresponds to the represented service provider. In the financial scenario (Duijst, 2017), some participants complained about the use of emojis in a situation of urgency (blocking a stolen credit card).

[B2] to increase believability: because people associate social qualities with machines (Reeves & Nass, 1996), chatbots are deemed sub-standard when users see them “*acting as a machine*” (Jenkins et al., 2007). When analyzing chatbots’ naturalness, Morrissey and Kirakowski (2013) found that the formal grammatical and syntactical abilities of a chatbot are the biggest discriminators between good and poor chatbots (the other factors being *conscientiousness*, *manners*, and *proactivity*). The authors highlight that chatbots should use consistent grammar and spelling. Coniam (2008) discusses how, even with English as Second Language (ESL) learners, basic grammar errors, such as pronoun confusion, diminish the value of the chatbot. In addition, Morris (2002) states that believable chatbots also need to display unique characters through linguistic choices. In this sense, Mairesse and Walker (2009) demonstrated that *personality* can be expressed by language patterns. The authors proposed a computational framework to produce utterances to manifest a target *personality*. The utterances were rated by experts in *personality* evaluation and statistically compared against utterances produced by humans who manifest the target *personality*. The outcomes show that a single utterance can manifest a believable *personality* when using the appropriate linguistic form. Participants in Jenkins et al. (2007) described some interactions as “*robotic*” if the chatbot repeated the keywords in the answers, reducing the interaction’s naturalness. Similarly, in Tallyn et al. (2018), participants complained about the “*inflexibility*” of the pre-defined, handcrafted chatbot’s responses and expressed the desire for it to talk “*more as a person*.”

Regarding the challenges, the surveyed literature shows the following:

[C1] to decide how much to say: in Jenkins et al. (2007), some participants described the chatbot’s utterances as not having enough detail, or being too generic; however, most of them appreciated finding answers in a sentence rather than in a paragraph. Similarly, Gnewuch et al. (2017) argue that simple questions should not be too detailed, while important transactions require more information. In three studies (Chaves & Gerosa, 2018; Duijst, 2017; Zamora, 2017), participants complained about information overload and inefficiency caused by big blocks of texts. Balancing the granularity of information with the sentence length is a challenge to overcome.

[C2] to be consistent: chatbots should not combine different language styles. For example, in Duijst (2017), most users found it strange that emojis were combined with a certain level of formal contact. When analyzing the critical incidents about an open-domain interaction, Kirakowski et al. (2009) found that participants criticized chatbots when they used more formal language or unusual vocabulary, since general-purpose chatbots focus on casual interactions.

Despite the highlighted benefits, we did not find strategies to provide *thoroughness*. Morris (2002) proposed a rule-based architecture where the language choices consider the agent’s *personality*, emotional state, and beliefs about the social relationship among the interlocutors. However, they did not provide evidence of whether the proposed models produced the expected outcome. Although the literature in computational linguistics has proposed algorithms and statistical models to manipulate language style and matching (see e.g., Prabhumoye, Tsvetkov, Salakhutdinov, and Black (2018); W.-N. Zhang, Zhu, Wang, Zhao, and Liu (2017)), to the best of our knowledge, these strategies have not been evaluated in the context of chatbots’ social interactions.

This section shows that linguistic choices influence users’ perceptions of chatbots. The computer-mediated communication (CMC) field has a vast literature that shows language variation according to the media and its effect on social perceptions (see e.g. (Baron, 1984; Walther, 2007)). Additionally, the cooperative principle (Grice et al., 1975), particularly the maxim of quantity (the ability to give the appropriate amount of information), provides theoretical basis for the challenge of deciding how much to talk. Regarding adaptation and believability, researchers in sociolinguistic fields (Conrad & Biber, 2009) have shown that language choices are influenced by personal style, dialect, genre, and register. For chatbots, the results presented in Mairesse and Walker (2009) are promising, demonstrating that automatically generated language can manifest recognizable traits. Thus, further research on chatbot’s *thoroughness* could leverage CMC and linguistics theories to provide strategies that lead language to accomplish its purpose for a particular interactional context.

3.2.3. *Manners*

Manners refer to the ability of a chatbot to manifest polite behavior and conversational habits (Morrissey & Kirakowski, 2013). Although individuals with different personalities and from different cultures may have different notions of what is considered polite (see e.g., Watts (2003)), politeness can be more generally applied as rapport management (P. Brown, 2015), where interlocutors strive to control the harmony between people in discourse. A chatbot can manifest *manners* by adopting speech acts such as greetings, apologies, and closings (Jain, Kumar, et al., 2018); minimizing impositions (Tallyn et al., 2018; Toxtli et al., 2018), and making interactions more personal (Jain, Kumar, et al., 2018). *Manners* potentially reduces the feeling of annoyance and frustration that may lead the interaction to fail (Jain, Kumar, et al., 2018).

We identified ten studies that report *manners*, one of which directly investigates this characteristic (Wallis & Norling, 2005). In some studies (Chaves & Gerosa, 2018; Q. V. Liao et al., 2018; Mäurer & Weihe, 2015; Toxtli et al., 2018), *manners* were observed in the analysis of conversational logs, where participants talked to the chatbot in polite, human-like ways. Users’ feedback and interviews revealed users’ expectations regarding chatbots politeness and personal behavior (Jain, Kumar, et al., 2018; Jenkins et al., 2007; Kirakowski et al., 2009; Kumar et al., 2010; Morrissey & Kirakowski, 2013). We identified studies reporting *manners* in nine different domains, with only open domain appearing twice. The list includes education, information search, and task management, among others. See the supplementary materials for the complete list of domains of studies that report *manners* as a social characteristic for chatbots.

The main benefit of providing *manners* is **[B1] to increase human-likeness**. *Manners* is highlighted in the literature as a way to generate a more natural, convincing interaction in chatbot conversations (Kirakowski et al., 2009; Morrissey & Kirakowski, 2013). In an in-the-wild data collection, Toxtli et al. (2018) observed that 93% of the

participants used polite words (e.g., “*thanks*” or “*please*”) with a task management chatbot at least once, and 20% always spoke politely to the chatbot. Unfortunately, the chatbot evaluated in that study was not prepared to handle these protocols and ultimately failed to understand. When identifying incidents from their own conversational logs with a chatbot (Kirakowski et al., 2009), several participants identified greetings as a human-seeming characteristic. The users also found convincing when the chatbot appropriately reacts to social cues statements, such as “*how are you?*”-types of utterances. Using this result, Morrissey and Kirakowski (2013) later suggested that greetings, apologies, social niceties, and introductions are significant constructs to measure chatbots’ naturalness. In Jenkins et al. (2007), the chatbot used exclamation marks at some point and frequently offered sentences available on the website, in a vaguely human-like matter. In the feedback, participants described the chatbot as rude, impolite, and cheeky.

The surveyed literature highlights two challenges to convey *manners*:

[C1] to deal with face-threatening acts: Face-Threatening Acts (FTA) are speech acts that threaten, either positive or negatively, the “face” of an interlocutor (P. Brown & Levinson., 1987). Politeness strategies in human-human interactions are adopted to counteract the threat when an FTA needs to be performed (P. Brown & Levinson., 1987). In Wallis and Norling (2005), the authors discuss that the wizard performing the role of the chatbot used several politeness strategies to counteract face threats. For instance, when she did not recognize a destination, instead of providing a list of possible destinations, she stimulated the user to keep talking until they volunteered the information. In chatbot design, by contrast, providing a list of options to choose is a common strategy. For example, in Toxtli et al. (2018), the chatbot was designed to present the user with a list of pending tasks when it did not know what task the user was reporting as completed, although the authors acknowledged that it resulted in an unnatural interaction. Although adopting politeness strategies is natural for humans and people usually do not consciously think about them, implementing them for chatbots is challenging due to the complexity of identifying face-threatening acts. For example, in the decision-making coach scenario, Mäurer and Weihe (2015) observed that users tend to utter straightforward and direct agreements while most of the disagreements contained modifiers that weakened their disagreement. The adoption of politeness strategies to deal with face-threatening acts is still under-investigated in the chatbot literature.

[C2] to end a conversation gracefully: Jain, Kumar, et al. (2018) discuss that first-time users expected human-like conversational etiquette from the chatbots, specifically introductory phrases and concluding phrases. Although several chatbots perform well in the introduction, the concluding phrases are less explored. Most of the participants reported being annoyed with chatbots that do not end a conversation (Jain, Kumar, et al., 2018). Chaves and Gerosa (2018) also highlight that chatbots need to know when the conversation ends. In that scenario, the chatbot could recognize a closing statement (the user explicitly says “*thank you*” or “*bye*”); however, it would not end the conversation otherwise. Users that stated a decision, but kept receiving more information from the chatbot, reported feeling confused and undecided afterward. Thus, recognizing the right moment to end the conversation is a challenge to overcome.

The strategies highlighted in the surveyed literature for providing *manners* are the following:

[S1] to engage in small talk: Q. V. Liao et al. (2018) and Kumar et al. (2010) point out that even task-oriented chatbots engage in small talk. When categorizing

the utterances from the conversational log, the authors found a significant number of messages about the agent status (e.g., “*what are you doing?*”), opening and closing sentences as well as acknowledgment statements (“*ok,*” “*got it*”). Jain, Kumar, et al. (2018) also observed that first-time users included small talk in the introductory phrases. According to Q. V. Liao et al. (2018), these are common behaviors in human-human chat interface, and chatbots would likely benefit from anticipating these habitual behaviors and reproducing them. However, particularly for task-oriented chatbots, it is important to control the small talk to avoid off-topic conversations and harassment, as discussed in Sections 3.2.1 and 3.1.2.

[S2] to adhere to turn-taking protocols: Toxtli et al. (2018) suggest that chatbots should adopt turn-taking protocols to know when to talk. Participants who received frequent follow-up questions from the task management chatbot about their pending tasks perceived the chatbot as invasive. Literature in chatbot development proposes techniques to improve chatbots’ turn-taking capabilities (see e.g., P. Brown and Levinson. (1987); Candello, Pinhanez, Pichiliani, Guerra, and Gatti de Bayser (2018); de Bayser et al. (2017)), which can be explored as a means of improving chatbots’ perceived *manners*.

Although the literature emphasizes that *manners* are important to approximate chatbot interactions to human conversational protocols, this social characteristic is under-investigated in the literature. Conversational acts such as greetings and apologies are often adopted (e.g., Jain, Kumar, et al. (2018); Jenkins et al. (2007); Mäurer and Weihe (2015)), but there is a lack of studies on the rationality around the strategies with politeness models used in human-human social interactions (Wallis & Norling, 2005). In addition, the literature points to needs for personal conversations (e.g., addressing the users by name), but we did not find studies that focus on this type of strategy. CMC is by itself more impersonal than face-to-face conversation (Walther, 1992, 1996); even so, current online communication media has been successfully used to initiate, develop, and maintain interpersonal relationships (Walther, 2011). Researchers can learn from human behaviors in CMC and adopt similar strategies to produce more personal conversations.

3.2.4. *Moral agency*

Machine *moral agency* refers to the ability of a technology to act based on social notions of right and wrong (Banks, 2018). The lack of this ability may lead to cases such as Tay, Microsoft’s Twitter chatbot that became racist, sexist, and harassing in a few hours (Neff & Nagy, 2016). The case raised concerns on what makes an artificial agent (im)moral. Whether machines can be considered (moral) agents is widely discussed in the literature (see e.g., Allen, Wallach, and Smit (2006); Himma (2009); Parthemore and Whitby (2013)). In this survey, the goal is not to argue about criteria to define a chatbot as moral, but to discuss the benefits of manifesting a perceived agency (Banks, 2018) and the implications of disregarding chatbots’ moral behavior. Hence, for the purpose of this survey, *moral agency* is a manifested behavior that may be inferred by a human as morality and agency (Banks, 2018).

We found six papers that address *moral agency*. Banks (2018) developed and validated a metric for perceived *moral agency* in conversational interfaces, including chatbots. In four studies, the authors investigated the ability of chatbots to handle conversations where the persistence of gender (Brahnam & De Angeli, 2012; De Angeli & Brahnam, 2006) and racial stereotypes (Marino, 2014; Schlesinger et al., 2018) may occur. In Shum et al. (2018), *moral agency* is discussed as a secondary result, where

the authors discuss the impact of biased responses on emotional connection. *Moral agency* was observed in only two domains of studies: open domain (four studies) and race-talk (two studies), which shows that this characteristic is primarily relevant when the conversational topic may raise moral concerns, which ultimately requires ethical behavior from the conversational partners.

The two main reported benefits of manifesting perceived *moral agency* are the following:

[B1] to avoid stereotyping: chatbots are often designed with anthropomorphized characteristics (see Section 3.3), including gender, age, and ethnicity identities. Although the chatbot’s personification is more evident in embodied conversational agents, text-based chatbots may also be assessed by their social representation, which risks building or reinforcing stereotypes (Marino, 2014). Marino (2014) and Schlesinger et al. (2018) argue that chatbots are often developed using language registers (Marino, 2014) and cultural references (Schlesinger et al., 2018) of the dominant culture. In addition, a static image (or avatar) representing the agent may convey social grouping (Nowak & Rauh, 2005). When the chatbot is positioned in a minority *identity* group, it exposes the image of that group to judgment and flaming, which is frequent in chatbot interactions (Marino, 2014). For example, Marino (2014) discusses the controversies caused by a chatbot designed to answer questions about Caribbean Aboriginal culture: its representation as a Caribbean Amerindian individual created an unintended context for stereotyping, where users projected the chatbot’s behavior as a standard for people from the represented population. Another example is the differences in sexual discourse between male- and female-presenting chatbots. Brahnham and De Angeli (2012) found that female-presenting chatbots are the object of implicit and explicit sexual attention and swear words more often than male-presenting chatbots. De Angeli and Brahnham (2006) show that sex talks with the male chatbot were rarely coercive or violent; his sexual preference was often questioned, though, and he was frequently propositioned by reported male users. In contrast, the female character received violent sexual statements, and was threatened with rape five times in the analyzed corpora. In Brahnham and De Angeli (2012), when the avatars were presented as black adults, references to race often deteriorated into racist attacks. Manifesting moral agency may, thus, prevent obnoxious user interactions. In addition, moral agency may prevent the chatbot itself from being biased or disrespectful to humans. Schlesinger et al. (2018) argue that the lack of context about the world does not redeem the chatbot from the necessity of being respectful with all the social groups.

[B2] to enrich interpersonal relationships: in a study on how interlocutors perceive conversational agents’ *moral agency*, Banks (2018) hypothesized that perceived morality may influence a range of motivations, dynamics, and effects of human-machine interactions. Based on this claim, the authors evaluated whether goodwill, trustworthiness, willingness to engage, and relational certainty in future interactions are constructs to measure perceived *moral agency*. Statistical results showed that all the constructs correlate with morality, which suggests that manifesting *moral agency* can enrich interpersonal relationships with chatbots. Similarly, Shum et al. (2018) suggest that to produce interpersonal responses, chatbots should be aware of inappropriate information and avoid generating biased responses.

However, the surveyed literature also reveals challenges of manifesting *moral agency*:

[C1] to avoid alienation: in order to prevent a chatbot from reproducing hate speech or abusive talk, most chatbots are built over “clean” data, where specific words are removed from their dictionary (De Angeli & Brahnham, 2006; Schlesinger et al., 2018). These chatbots have no knowledge of those words and their meaning. Although

this strategy is useful to prevent unwanted behavior, it does not manifest agency, but rather alienates the chatbot of the topic. De Angeli and Brahnham (2006) show that the lack of understanding about sex-talk does not prevent the studied chatbot from harsh verbal abuse, or even from being perceived as encouraging such abuse. From (Schlesinger et al., 2018), one can notice that the absence of racist specific words did not prevent the chatbot Zo from uttering discriminatory exchanges. As a consequence, manifesting *moral agency* requires a broader understanding of the world, rather than alienation, which is an open challenge.

[C2] to build unbiased algorithms and training data: as extensively discussed in Schlesinger et al. (2018), machine learning algorithms and corpus-based language generation are biased toward the available training datasets. Hence, *moral agency* relies on data that is biased in its nature, producing unsatisfactory results from an ethical perspective. In Shum et al. (2018), the authors propose a framework for developing social chatbots. The authors highlight that the core chat module should follow ethical design to generate unbiased, non-discriminative responses, but they do not discuss specific strategies for that. Building unbiased training datasets and learning algorithms that connect the outputs with individual, real-world experiences, therefore, are challenges to overcome.

Despite the relevance of *moral agency* to the development of socially intelligent chatbots, we did not find strategies to address the issues. Schlesinger et al. (2018) advocate for developing diversity-conscious databases and learning algorithms that account for ethical concerns; however, the paper focuses on outlining the main research branches and calls on the community of designers to adopt new strategies. As discussed in this section, research on perceived *moral agency* is still necessary in order to develop chatbots whose social behavior is inclusive and respectful. In the field of embodied agents, mind perception theory (Gray, Gray, & Wegner, 2007) has been investigated as a means to improve interactions through agency and emotion (Appel, Izydorec, Weber, Mara, & Lischetzke, 2020; Keijsers & Bartneck, 2018; M. Lee, Lucas, Mell, Johnson, & Gratch, 2019). Future investigations in chatbots interactions could leverage this theory to understand when and the extent to which perceived moral agency improves communication with chatbots.

3.2.5. Emotional Intelligence

Emotional intelligence is a subset of social intelligence that allows an individual to appraise and express feelings, regulate affective reactions, and harness emotions to solve a problem (Salovey & Mayer, 1990). Although chatbots do not have genuine emotions (Wallis & Norling, 2005), there are considerable discussions about the role of manifesting emotional cues in chatbots (Ho et al., 2018; Shum et al., 2018; Wallis & Norling, 2005). An emotionally intelligent chatbot can recognize and influence users' feelings and demonstrate respect, empathy, and understanding, improving the human-chatbot relationship (Li et al., 2017; Salovey & Mayer, 1990).

We identified 14 studies that report *emotional intelligence*. Unlikely the previously discussed categories, most studies on *emotional intelligence* focused on understanding the effects of chatbots' empathy and emotional self-disclosure (Ayedoun et al., 2017; Dohsaka et al., 2014; K. K. Fitzpatrick et al., 2017; Ho et al., 2018; Kumar et al., 2010; S. Lee & Choi, 2017; Miner et al., 2016; Morris, 2002; Portela & Granell-Canut, 2017; Shum et al., 2018; Wallis & Norling, 2005). Only three papers highlighted *emotional intelligence* as an exploratory outcome (Jenkins et al., 2007; Thies et al., 2017; Zamora, 2017), where needs for *emotional intelligence* emerged from participants' subjective

feedback and post-intervention surveys. *Emotional intelligence* is mainly investigated in domains where topics may involve the disclosure of feelings (e.g., in open domain interactions (Shum et al., 2018)) and expressions of empathy and understanding are appropriate (Dohsaka et al., 2014; K. K. Fitzpatrick et al., 2017) (e.g., health care, gaming, education). See the supplementary materials for the domains of study investigating *emotional intelligence*.

The main reported benefits of developing emotionally intelligent chatbots are the following:

[B1] to enrich interpersonal relationships: the perception that the chatbot understands one’s feelings may create a sense of belonging and acceptance (Ho et al., 2018). Ayedoun et al. (2017) propose that chatbots for second language studies should use congratulatory, encouraging, sympathetic, and reassuring utterances to create a friendly atmosphere to the learner. The authors statistically demonstrated that affective backchannel, combined with communicative strategies (see Section 3.1.2) significantly increased learners’ confidence and desire to communicate, while reducing anxiety. In another educational study, Kumar et al. (2010) evaluated the impact of chatbot’s affective moves on friendliness and achieving social belonging. Qualitative results show that affective moves significantly improve the perception of amicability, and marginally increased social belonging. According to Wallis and Norling (2005), when a chatbot’s emotional reaction triggers a social response from the user, the chatbot has achieved group membership and users’ sympathy. Dohsaka et al. (2014) proposed a chatbot that uses empathic and self-oriented emotional expressions to keep users engaged in quiz-style dialogue. The survey results revealed that empathic expressions significantly improved user satisfaction. In addition, the empathic expressions also improved the user ratings of the peer agent regarding intimacy, compassion, amiability, and encouragement. Although Dohsaka et al. (2014) did not find effect of chatbot’s self-disclosure on emotional connection, S. Lee and Choi (2017) found that self-disclosure and reciprocity significantly improved trust and interactional enjoyment. In K. K. Fitzpatrick et al. (2017), seven participants reported that the best thing about their experience with the therapist chatbot was perceived empathy. Five participants highlighted that the chatbot demonstrated attention to their users’ feelings. In addition, the users referred the chatbot as “*he*,” “*a friend*,” “*a fun little dude*,” which demonstrates that empathy emerged from the personification of the chatbot. In another mental health care study, Miner et al. (2016) found that humans are twice as likely to mirror negative sentiment from a chatbot than from a human, which is a relevant implication for therapeutic interactions. In Zamora (2017), participants reported that some content is embarrassing to ask another human, thus, talking to a chatbot would be easier due to the lack of judgement. Ho et al. (2018) measured users’ experience in conversations with a chatbot compared to a human partner as well as the amount of intimacy disclosure and cognitive reappraisal. Participants in the chatbots condition experienced as many emotional, relational, and psychological benefits as participants who disclosed to a human partner.

[B2] to increase engagement: Shum et al. (2018) argue that longer conversations (10+ turns) are needed to fulfill the needs of affection and belonging. Therefore, the authors defined conversation-turns per session as a success metric for chatbots, where usefulness and emotional understanding are combined. In Dohsaka et al. (2014), empathic utterances for the quiz-style interaction significantly increase the number of users’ messages per hint for both answers and non-answer utterances (such as feedback about the success/failure). This result shows that empathic utterances encouraged the users to engage and utter non-answers statements. Portela and Granell-Canut (2017)

compared the possibility of emotional connection between a classical chatbot and a pretend chatbot, simulated in a WoZ experiment. Quantitative results showed that the WoZ condition was more engaging, since it resulted in conversations that lasted longer, with a higher number of turns. The analysis of the conversational logs revealed the positive effect of the chatbot manifesting social cues and empathic signs as well as touching on personal topics.

[B3] to increase believability: (Morris, 2002) argue that adapting chatbots’ language to their current emotional state, along with their *personality* and social role awareness, results in more believable interactions. The authors propose that conversation acts should reflect the pretended emotional status of the agent; the extent to which the acts impact on emotion depends however on the agent’s *personality* (e.g., its temperament or tolerance). *Personality* is an anthropomorphic characteristic and is discussed in Section 3.3.2.

Although *emotional intelligence* is the goal of several studies, **[C1] regulating affective reactions** is still a challenge. The chatbot presented in Kumar et al. (2010) was designed to mimic the patterns of affective moves in human-human interactions. Nevertheless, the chatbot has shown an only marginally significant increase in social belonging, when compared to the same interaction with a human partner. Conversational logs revealed that the human tutor performed a significantly higher number of affective moves in that context. In Jenkins et al. (2007), the chatbot was designed to present emotive-like cues, such as exclamation marks, and interjections. The participants negatively rated the degree of emotion in the chatbot’s responses. In Thies et al. (2017), the energetic chatbot was reported as having an enthusiasm too high to match. In contrast, the chatbot was described as an “*emotional buddy*” was reported as being “*overly caring*.” Ho et al. (2018) state that chatbot’s empathic utterances may be seen as pre-programmed and inauthentic. Although their results revealed that the partners’ identity (chatbot vs. person) had no effect in the perceived relational and emotional experience, the chatbot condition was a WoZ setup. The wizards were blind to whether users thought they were talking to a chatbot or a person, which reveals that identity does not matter if the challenge of regulating emotions is overcome.

The chatbots literature also report some strategies to manifest *emotional intelligence*:

[S1] using social-emotional utterances: affective utterances toward the user are a common strategy to demonstrate *emotional intelligence*. Ayedoun et al. (2017), Kumar et al. (2010), and Dohsaka et al. (2014) suggest that affective utterances improve the interpersonal relationship with a tutor chatbot. In Ayedoun et al. (2017), the authors propose affective backchannel utterances (congratulatory, encouraging, sympathetic, and reassuring) to motivate the user to communicate in a second language. The tutor chatbot proposed in Kumar et al. (2010) uses solidarity, tension release, and agreement utterances to promote its social belonging and acceptance in group chats. Dohsaka et al. (2014) propose empathic utterances to express opinion about the difficulty or ease of a quiz, and feedback on success and failure.

[S2] to manifest *conscientiousness*: demonstrating *conscientiousness* may affect the emotional connection between humans and chatbots. In Portela and Granell-Canut (2017), participants reported the rise of affection when the chatbot remembered something they had said before, even if it was just the user’s name. Keeping track of the conversation was reported as an empathic behavior and resulted in mutual affection. Shum et al. (2018) argue that a chatbot needs to combine usefulness with emotion by asking questions that help to clarify the users’ intentions. They provide an example where a user asks the time, and the chatbot answer “*Cannot sleep?*” as an attempt to

guide the conversation to a more engaging direction. Adopting this strategy requires the chatbot to handle message understanding, emotion and sentiment tracking, session context modeling, and user profiling Shum et al. (2018).

[S3] reciprocity and self-disclosure: S. Lee and Choi (2017) hypothesized that a high level of self-disclosure and reciprocity in communication with chatbots would increase trust, intimacy, and enjoyment, ultimately improving user satisfaction and intention to use. They performed a WoZ, where the assumed chatbot was designed to recommend movies. Results demonstrated that reciprocity and self-disclosure are strong predictors of rapport and user satisfaction. In contrast, Dohsaka et al. (2014) did not find any effect of self-oriented emotional expressions in the users’ satisfaction or engagement. More research is needed to understand the extent to which this strategy produces positive impact on the interaction.

The literature shows that *emotional intelligence* is widely investigated, with particular interest from education and mental health care domains. Using emotional utterances in a personalized, context relevant way is still a challenge. Researchers in chatbots *emotional intelligence* can learn from *emotional intelligence* theory (Gross, 1998; Salovey & Mayer, 1990) to adapt the chatbots utterances to match the emotions expressed in the dynamic context. Adaption to the dynamic context also improves the sense of personalized interactions, which is discussed in the next section.

3.2.6. Personalization

Personalization refers to the ability of a technology to adapt its functionality, interface, information access, and content to increase its personal relevance to an individual or a category of individuals (Fan & Poole, 2006). In the chatbot domain, *personalization* may increase the agents’ social intelligence, since it allows a chatbot to be aware of situational context and dynamically adapt its features to better suit individual needs (Neururer et al., 2018). Grounded in robots and artificial agents’ literature, V. Q. Liao et al. (2016) argue that *personalization* can improve rapport and cooperation, ultimately increasing engagement with chatbots. Although some studies (see e.g., Fan and Poole (2006); V. Q. Liao et al. (2016); S. Zhang et al. (2018)) also relate *personalization* to the attribution of personal qualities such as *personality*, we discuss personal qualities in the **Personification** category. In this section, we focus on the ability to adapt the interface, content, and behavior to the users’ preferences, needs, and situational context.

We found 11 studies that report *personalization*. Three studies pose *personalization* as a research goal (Duijst, 2017; V. Q. Liao et al., 2016; Shum et al., 2018). In most of the studies, though, *personalization* was observed in exploratory findings. In six studies, *personalization* emerged from the analysis of interviews and participants’ self-reported feedback (Duijvelshoff, 2017; Jenkins et al., 2007; Neururer et al., 2018; Portela & Granell-Canut, 2017; Tallyn et al., 2018; Thies et al., 2017). In two studies (Lasek & Jessa, 2013; Toxtli et al., 2018), needs for *personalization* emerged from the conversational logs. *Personalization* was investigated in seven different domains, including open domain and task management (two studies each). In open domain interactions, personalization is derived from remembering information from previous interactions, such as personal preferences and users’ details (Thies et al., 2017). In task-oriented contexts, such as task management, personalization aims to increase the relevance of services to particular users (V. Q. Liao et al., 2016). See the supplementary materials for details.

The surveyed literature highlighted three benefits of providing personalized inter-

actions:

[B1] to enrich interpersonal relationships: Duijvelshoff (2017) state that personalizing the amount of personal information a chatbot can access and store is required to establish a relation of trust and reciprocity in workplace environments. In Neururer et al. (2018), interviews with 12 participants generated a total of 59 statements about how learning from experience promotes chatbots’ authenticity. Shum et al. (2018) argue that chatbots whose focus is engagement need to personalize the generation of responses for different users’ backgrounds, personal interests, and needs in order to serve their needs for communication, affection, and social belonging. In Portela and Granell-Canut (2017), participants expressed the desire for the chatbot to provide different answers to different users. Although Duijst (2017) has found no significant effect of *personalization* on the user experience with the financial assistant chatbot, the study applies *personalization* as the ability to provide empathic responses according to the users’ issues, where *emotional intelligence* plays a role. Interpersonal relationship can also be enriched by adapting the chatbots’ language to match the user’s context, energy, and formality; the ability to appropriately use language is discussed in Section 3.2.2.

[B2] to provide unique services: providing *personalization* increases the value of provided information (Duijst, 2017). In the ethnography data collection study (Tallyn et al., 2018), eight participants reported dissatisfaction with the chatbot’s generic guidance to specific places. Participants self-reported that the chatbot should use their current location to direct them to places more conveniently located, and ask for participants interests and preferences to direct them to areas that meet their needs. When exploring how teammates used a task-assignment chatbot, Toxtli et al. (2018) found that the use of the chatbot varied depending on the participants’ levels of hierarchy. Similarly, qualitative analysis of perceived interruption in a workplace chat (V. Q. Liao et al., 2016) suggests that interruption is likely associated with users’ general aversion to unsolicited messages at work. Hence, the authors argue that chatbot’s messages should be personalized to the user’s general preference. V. Q. Liao et al. (2016) also found that users with low social-agent orientation emphasize the utilitarian value of the system, while users with high social-agent orientation see the system as a humanized assistant. This outcome support to the need to personalize the interaction to individual user’s mental models. In Thies et al. (2017), participants reported preference for a chatbot that remembers their details, likes and dislikes, and preferences, and voluntarily uses the information to make recommendations. In Jain, Kumar, et al. (2018), two participants also expected chatbots to retain context from previous interactions to improve recommendations.

[B3] to reduce interactional breakdowns: in HCI, *personalization* is used to customize the interface toward user familiarity (Fan & Poole, 2006). When evaluating visual elements (such as quick replies) compared to typing the responses, Lasek and Jessa (2013) observed that users that start the interaction by clicking an option are more likely to continue the conversation if the next exchange also has visual elements as optional affordances. In contrast, users who typed are more likely to abandon the conversation when they face options to click. Thus, chatbots should adapt their interface to users’ preferred input methods. In Jenkins et al. (2007), one participant suggested that the choice of text color and font size should be customizable. Duijst (2017) also observed that participants faced difficulties with small letters, and concluded that adapting the interface to provide accessibility also needs to be considered.

According to the surveyed literature, the main challenge regarding *personalization* is **[C1] privacy**. To enrich the efficiency and productivity of the interaction, a chatbot

needs to have memory of previous interactions as well as learn user’s preferences and disclosed personal information (Thies et al., 2017). However, as Duijvelshoff (2017) and Thies et al. (2017) suggest, collecting personal data may lead to privacy concerns. Thus, chatbots should showcase transparent purpose and ethical standards (Neururer et al., 2018). Thies et al. (2017) also suggest that there should be a way to inform a chatbot that something in the conversation is private. Similarly, participants in Zamora (2017) reported that personal data and social media content may be inappropriate topics for chatbots because they can be sensitive. These concerns may be reduced if a chatbot demonstrates care about privacy (Duijvelshoff, 2017).

The reported strategies to provide *personalization* in chatbots interactions are the following:

[S1] to learn from and about the user: Neururer et al. (2018) state that chatbots should present strategies to learn from cultural, behavioral, personal, conversational, and contextual interaction data. For example, the authors suggest using Facebook profile information to build knowledge about users’ personal information. Thies et al. (2017) also suggest that the chatbot should remember user’s preferences disclosed in previous conversations. In Shum et al. (2018), the authors propose an architecture where responses are generated based on a *personalization* rank that applies users’ feedback about their general interests and preferences. When evaluating the user’s experience with a virtual assistant chatbot, Zamora (2017) found 16 mentions to personalized interactions, where participants demonstrated the need for a chatbot to be aware of their personal quirks and anticipate their needs.

[S2] to provide customizable agents: V. Q. Liao et al. (2016) suggest that users should be able to choose the level of the chatbot’s attributes, for example, the agent’s look and persona. By doing so, users with low social-agent orientation could use a non-humanized interface, which would better represent their initial perspective. This differentiation could be the first signal to personalize further conversation, such as focusing on more productive or playful interactions. Regarding chatbots’ learning capabilities, in Duijvelshoff (2017), interviews with potential users revealed that users should be able to manage what information the chatbot knows about them and decide whether the chatbot can learn from previous interactions or not. If the user prefers a more generic chatbot, then it would not store personal data, potentially increasing the engagement with more resistant users. Thies et al. (2017) raise the possibility of offering an “incognito” mode for chatbots or asking the chatbot to forget what was said in previous utterances.

[S3] visual elements: Tallyn et al. (2018) adopted quick replies as a means for the chatbot to tailor its subsequent questions to the specific experience the participant had reported. As discussed in Section 3.1.2, quick replies may be seen as restrictive from an interactional perspective; however, conversation logs showed that the tailored questions prompted the users to report more detail about their experience, which is important in ethnography research.

Both the benefits and strategies identified in the literature are in line with the types of *personalization* proposed by Fan and Poole (2006). Therefore, further investigations in *personalization* can leverage knowledge from interactive systems (e.g., Fan and Poole (2006); Thomson (2005)) to adapt *personalization* strategies and manage the privacy concern.

In summary, the social intelligence category includes characteristics that help a chatbot to manifest an adequate social behavior, by managing conflicts, using appropriate

language, displaying *manners* and *moral agency*, sharing emotions, and handling personalized interactions. The benefits relate to resolving social positioning and recovering from failures, as well as increasing believability, human-likeness, engagement, and rapport. To achieve that, designers and researchers should care about privacy, emotional regulation issues, language consistency, and identification of failures and inappropriate content.

3.3. Personification

In this section, we discuss the influence of *identity* projection on human-chatbot interaction. **Personification** refers to assigning personal traits to non-human agents, including physical appearance, and emotional states (Fan & Poole, 2006). In the HCI field, researchers argue that using a personified character in the user interface is a natural way to support the interaction (Koda, 2003). Indeed, the literature shows that (i) users can be induced to behave as if computers were humans, even when they consciously know better (Nass, Steuer, Tauber, & Reeder, 1993); and (ii) the more human-like a computer representation is, the more social people’s responses will be (Gong, 2008).

Chatbots are, by definition, designed to have at least one human-like trait: (human) natural language. Although research on **personification** is more common in the Embodied Conversational Agents field, De Angeli, Johnson, and Coventry (2001) claim that chatbot embodiment can be created through narrative without any visual help. According to De Angeli (2005), talking to a machine affords it a new *identity*. In this section, we divided the social characteristics that reflect **personification** into *identity* (16 papers) and *personality* (12 papers). In this category, we found several studies where part of the main investigation relates to the social characteristics. Six studies applied quantitative methods, whereas the majority reported qualitative (10) or mixed (11) methods. See the supplementary materials for details.

3.3.1. Identity

Identity refers to the ability of an individual to demonstrate belonging to a particular social group (Stets & Burke, 2000). Although chatbots lack the agency to choose what social group to which they want to belong, designers attribute *identity* to them, intentionally or not, when they define the way a chatbot talks or behaves (Cassell, 2009). The *identity* of a partner (even if only perceived) gives rise to new processes, expectations, and effects that influence the outcomes of the interaction (Ho et al., 2018). Aspects that convey the chatbots’ *identity* include gender, age, language style, and name. Additionally, chatbots may have anthropomorphic, zoomorphic, or robotic representations. Some authors include *identity* aspects in the definition of *personality* (see, e.g., Shum et al. (2018)). We distinguish these two characteristics, where *identity* refers to appearance and cultural traits while *personality* focuses on behavioral traits.

We found 16 studies that discuss *identity* issues, ten of which include *identity* as part of their main investigation (Araujo, 2018; Candello et al., 2017; Ciechanowski et al., 2018; Corti & Gillespie, 2016; De Angeli & Brahmam, 2006; De Angeli, Johnson, & Coventry, 2001; Jenkins et al., 2007; Q. V. Liao et al., 2018; Marino, 2014; Schlesinger et al., 2018). In two studies, the authors argue for the impact of *identity* on the interaction based on the literature (Brandtzaeg & Følstad, 2018; Gnewuch et al., 2017). In three studies (Neururer et al., 2018; Silvervarg & Jönsson, 2013; Thies et al., 2017; Toxtli et al., 2018), qualitative analysis of conversational logs revealed that

participants put effort into understanding aspects of the chatbots’ *identity*. *Identity* concerns were primarily investigated for open domain and customer service chatbots (four studies each). In open domain interactions, *identity* is explored as a means of building common ground (De Angeli, Johnson, & Coventry, 2001). In the case of customer services, *identity* helps manifest credibility and trust (Gnewuch et al., 2017). Other domains include race-talk, education, and gaming. The complete list appears in the supplementary materials.

The identified benefits of attributing *identity* to a chatbot are the following:

[B1] to increase engagement: when evaluating signals of playful interactions, Q. V. Liao et al. (2018) found that agent-oriented conversations (asking about an agent’s traits and status) are consistent with the tendency to anthropomorphize the agent and engage in chit-chat. In the educational scenario, Silvervarg and Jönsson (2013) also observed questions about agents’ appearance, intellectual capacities, and sexual orientation, although the researchers considered these questions inappropriate for the context of tutoring chatbots. When comparing human-like vs. machine-like language style, greetings, and framing, Araujo (2018) noticed that using informal language, having a human name, and using greetings associated with human communication resulted in significantly higher scores for adjectives like likeable, friendly, and personal. In addition, framing the agent as “intelligent” also had a slight influence on users’ scores.

[B2] to increase human-likeness: some attributes may convey a perceived human-likeness. Araujo (2018) showed that using human-like language style, name, and greetings resulted in significantly higher scores for naturalness. The chatbot’s framing influenced the outcomes when combined with other anthropomorphic clues. When evaluating different typefaces for a financial adviser chatbot, Candello et al. (2017) found that users perceive machine-like typefaces as more chatbot-like, although they did not find strong evidence of handwriting-like typefaces conveying humanness.

The surveyed literature also highlights challenges regarding *identity*:

[C1] to avoid negative stereotypes: when engaging in a conversation, interlocutors base their behavior on common ground (joint knowledge, background facts, assumptions, and beliefs that participants have of each other) (see (De Angeli, Johnson, & Coventry, 2001)). Common ground reflects stereotypical attributions that chatbots should be able to manage as the conversation evolves (De Angeli, 2005). In De Angeli, Johnson, and Coventry (2001), the authors discuss that chatbots for company representations are often personified as attractive human-like women acting as spokespeople for their companies, while men chatbots tend to have a more important position, such as a virtual CEO. De Angeli and Brahnham (2006) state that the agent self-disclosure of gender *identity* opens possibilities to sex talk. The authors observed that the conversations mirror stereotyped male/female encounters, and the ambiguity of the chatbot’s gender may influence the exploration of homosexuality. However, fewer instances were observed of sex-talk with the chatbot personified as a robot, which demonstrates that the gender *identity* may lead to the stereotypical attributions. When evaluating the effect of gender identity on disinhibition, Brahnham and De Angeli (2012) showed that people spoke more often about physical appearance and used more swear and sexual words with the female-presenting chatbot, and racist attacks were observed in interactions with black-representing chatbots. The conversation logs from (Jenkins et al., 2007) also show instances of users attacking the chatbot persona (a static avatar of a woman pointing to the conversation box). Marino (2014) and Schlesinger et al. (2018) also highlight that race identity conveys not only the physical appearance, but all the socio-cultural expectations about the represented group (see discussion in Section

3.2.4). Hence, designers should care about the impact of attributing an identity to chatbots in order to avoid reinforcing negative stereotypes.

[C2] to balance the *identity* and the technical capabilities: the literature comparing embodied vs. disembodied conversational agents yields contradictory results regarding the relevance of a human representation. For example, in the context of general-purpose interactions, Corti and Gillespie (2016) show that people exert more effort toward establishing common ground with the agent when they are represented as fully human; in contrast, when evaluating a website assistant chatbot, Ciechanowski et al. (2018) show that simpler text-based chatbots with no visual, human *identity* resulted in less of an uncanny effect and reduced negative affect. Overly humanized agents create a higher expectation for users, which eventually leads to more frustration when the chatbot fails (Gnewuch et al., 2017). When arguing about why chatbots fail, Brandtzaeg and Følstad (2018) advocate for balancing human versus robotic aspects, where “*too human*” representations may lead to off-topic conversations, and overly robotic interactions may lack personal touch and flexibility. When arguing about the social presence conveyed by deceptive chatbots, De Angeli (2005) state that extreme anthropomorphic features may generate cognitive dissonance. The challenge thus lies in designing a chatbot that provides appropriate *identity* cues, which correspond to their capabilities and communicative purpose, in order to convey the right expectation and minimize discomfort from over-personification.

Regarding the strategies, the surveyed literature suggests **[S1] to design and elaborate on a persona**. Chatbots should have a comprehensive persona and answer agent-oriented conversations with a consistent description of itself (Q. V. Liao et al., 2018; Neururer et al., 2018). For example, De Angeli (2005) discuss that Eliza, the psychotherapist chatbot, and Parry, a paranoid chatbot, have behaviors consistent with the stereotypes associated with the professional and personal identities, respectively. Toxtli et al. (2018) suggest that designers should explicitly build signals of the chatbot personification (either machine- or human-like), so the users can have the right expectation about the interaction. When *identity* aspects are not explicit, users try to establish common ground. In Q. V. Liao et al. (2018) and (Silvervarg & Jönsson, 2013), much of the small talk with the chatbot related to the chatbot’s traits and status. In De Angeli, Johnson, and Coventry (2001), the authors observed many instances of Alice’s self-references to “*her*” artificial nature. These references triggered the users to reflect on their human-condition (self-categorization process), resulting in exchanges about their species (either informational or confrontational). Thies et al. (2017) observed similar results, as participants engaged in conversations about the artificial nature of the agent. Providing the chatbot with the ability to describe its personal *identity* helps to establish the common ground, and hence, enrich the interpersonal relationship (De Angeli, Johnson, & Coventry, 2001).

Chatbots may be designed to deceive users about their actual *identity*, pretending to be a human (De Angeli, 2005). In this case, the more human the chatbot sounds, the more successful it will be. In many cases, however, there is no need to engage in deception and the chatbots can be designed to represent an elaborated persona. Researchers can explore social identity theory (R. Brown, 2000; Stets & Burke, 2000) related to ingroup bias, power relations, homogeneity, and stereotyping, in order to design chatbots with *identity* traits that reflect their expected social position (Harré, Moghaddam, et al., 2003).

3.3.2. *Personality*

Personality refers to personal traits that help to predict someone’s thoughts, feelings, and behaviors (McCrae & Costa Jr, 1997). The most accepted set of traits is called the Five-Factor model (or Big Five model) (Goldberg, 1990; McCrae & Costa Jr, 1997)), which describes *personality* across five dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness). However, *personality* can also refer to other dynamic, behavioral characteristics, such as temperament and sense of humor (Thorson & Powell, 1993; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993). In the chatbots domain, *personality* refers to the set of traits that determines the agent’s interaction style, describes its character, and allows the end-user to understand its general behavior (De Angeli, Johnson, & Coventry, 2001). Chatbots with consistent *personality* are more predictable and trustworthy (Shum et al., 2018). According to De Angeli, Lynch, and Johnson (2001), unpredictable swings in chatbots’ attitudes can disorient users and create a strong sense of discomfort. Thus, *personality* ensures that a chatbot displays behaviors that stand in agreement with the users’ expectations in a particular context (Petta & Trappl, 1997).

We found 12 studies that report *personality* issues for chatbots. In some studies, *personality* was investigated in reference to the Big Five model (Mairesse & Walker, 2009; Morris, 2002; Sjöden et al., 2011), while two studies focused on sense of humor (Meany & Clark, 2010; Ptaszynski et al., 2010). Three studies investigated the impact of the *personality* of tutor chatbots on students’ engagement (Ayedoun et al., 2017; Kumar et al., 2010; Sjöden et al., 2011). Thies et al. (2017) compared users’ preferences regarding pre-defined personalities. In the remaining studies, (Brandtzaeg & Følstad, 2017; Jain, Kumar, et al., 2018; Portela & Granell-Canut, 2017; Shum et al., 2018) *personality* concerns emerged from the qualitative analysis of the interviews, users’ subjective feedback, and literature reviews (Meany & Clark, 2010). *Personality* was mostly investigated in open domain (five studies) and education (three studies) chatbots. The other two domains were gaming and humorous chatbots, which are both playful agents. Hence, *personality* is relevant when believability and attitude play a role in the interaction (Portela & Granell-Canut, 2017) (e.g., in open domain) and when the chatbots’ attitude may increase users’ mental comfort when performing a task (Ayedoun et al., 2017), such as in educational contexts.

The surveyed literature revealed two benefits of attributing *personality* to chatbots:

[B1] to increase believability: Morris (2002) states that chatbots should have a *personality*, defined by the Five Factor model, plus characteristics such as temperament and tolerance, in order to build utterances using linguistic choices that cohere with these attributions. When evaluating a humorous chatbot, Ptaszynski et al. (2010) compared the naturalness of the chatbot’s inputs and the chatbot’s human-likeness compared to a non-humorous chatbot. The humorous chatbot scored significantly higher in both constructs. Portela and Granell-Canut (2017) also showed that sense of humor humanizes the interactions, since humor was one of the factors that influenced naturalness for the WoZ condition. Mairesse and Walker (2009) demonstrated that manipulating language to manifest a target *personality* produced moderately natural utterances, with a mean rating of 4.59 out of 7 for the *personality* model utterances.

[B2] to enrich interpersonal relationships: chatbots *personality* can make the interaction more enjoyable (Brandtzaeg & Følstad, 2017; Jain, Kumar, et al., 2018). In Brandtzaeg and Følstad (2017), the second most frequent motivation for using chatbots, noted by 20% of the participants, was entertainment. The authors argue that a chatbot’s capacity to be fun is important even when the main purpose is productivity;

according to participants, the chatbot’s “*fun tips*” enrich the user experience. This result is consistent with the experience of first-time users (Jain, Kumar, et al., 2018), where participants relate better with chatbots who have consistent *personality*. Thies et al. (2017) show how witty banter and casual, enthusiastic conversations help make the interaction effortless. In addition, a few participants enjoyed the chatbot with a caring *personality*, who was described as a good listener. In Shum et al. (2018) and (Sjödén et al., 2011), the authors argue that a consistent *personality* helps the chatbot to gain the users’ confidence and trust. Sjödén et al. (2011) state that tutor chatbots should display appropriate posture, conduct, and representation, which include being encouraging, expressive, and polite. Accordingly, other studies highlight students’ desire for chatbots with positive agreeableness and extraversion (Ayedoun et al., 2017; Kumar et al., 2010; Sjödén et al., 2011). Outcomes consistently suggest that students prefer a chatbot that is not overly polite, but has some attitude. Agreeableness seems to play a critical role, helping the students to feel encouraged and deal with difficulties. Notably, agreeableness requires *emotional intelligence* to be warm and sympathetic in appropriate circumstances (see Section 3.2.5).

The reviewed literature pointed out two challenges regarding *personality*:

[C1] to adapt humor to the users’ culture: sense of humor is highly shaped by cultural environment (Ruch, 1998). Ptaszynski et al. (2010) discusses a Japanese chatbot who uses puns to create funny conversations. The authors state that puns are one of the main humor genres in that culture. However, puns are restricted to a specific culture and language, with low portability. Thus, the design challenge lies in personalizing chatbots’ sense of humor to the target users’ culture and interests or, alternatively, designing cross-cultural forms of humor. The ability to adapt to the context and users’ preference is discussed in Section 3.2.6.

[C2] to balance *personality* traits: Thies et al. (2017) observed that users prefer a proactive, productive, witty chatbot. Yet, they also would like them to be caring, encouraging, and exciting. In Mairesse and Walker (2009), the researchers intentionally generated utterances to reflect extreme personalities; as a result, they observed that some utterances sounded unnatural because a human’s *personality* is a continuous phenomenon, rather than a discrete one. Sjödén et al. (2011) also points out that, although *personality* is consistent, moods and states of mind constantly vary. Thus, balancing the predictability of the *personality* and the expected variation is a challenge to overcome.

We also identified strategies to design chatbots that manifest *personality*:

[S1] to use appropriate language: Shum et al. (2018) and Morris (2002) suggest that the chatbot language should be consistently influenced by its *personality*. Both studies propose that chatbots’ architecture should include a persona-based model that encodes the *personality* and influences the response generation. Mairesse and Walker (2009) proposed a framework to shows that it is possible to automatically manipulate language features to manifest a particular *personality* based on the Big Five model. The Big Five model is a relevant tool because it can be assessed using validated psychological instruments (McCrae & Costa Jr, 1987). Using this model to represent the *personality* of chatbots was also suggested by Morris (2002) and Sjödén et al. (2011). Jain, Kumar, et al. (2018) discussed that a chatbot’s *personality* should match its domain. Participants expected the language used by the news chatbot to be professional, while they expected the shopping chatbot to be casual and humorous. The ability to use consistent language is discussed in Section 3.2.2.

[S2] to have a sense of humor: literature highlights humor as a positive *personality* trait (Meany & Clark, 2010). In Jain, Kumar, et al. (2018), ten participants men-

tioned enjoyment when the chatbots provided humorous and highly diverse responses. The authors found occurrences of the participants asking for jokes and expressing delight when the request was supported. Brandtzaeg and Følstad (2017) present similar results when arguing that humor is important even for task-oriented chatbots when the user is usually seeking for productivity. For casual conversations, Thies et al. (2017) highlight that timely, relevant, and clever wit is a desired *personality* trait. In Ptaszynski et al. (2010), the joker chatbot was perceived as more human-like, knowledgeable, and funny, and participants felt more engaged.

Personality for artificial agents has been studied for some time in the Artificial Intelligence field (Elliott, 1994; Petta & Trappl, 1997; Rousseau & Hayes-Roth, 1996). Thus, further investigations on chatbots’ *personality* can leverage models for *personality* and evaluate how they contribute to believability and rapport building.

In summary, the **personification** category includes characteristics that help a chatbot to manifest personal and behavioral traits. The benefits relate to increasing believability, human-likeness, engagement, and interpersonal relationship, which is in line with the benefits of **social intelligence**. However, unlike the **social intelligence** category, designers and researchers should focus on attributing recognizable *identity* and *personality* traits that are consistent with users’ expectations and the chatbot’s capabilities. In addition, it is important to care about adaptation to users’ cultural context and reducing the effects of negative stereotypes.

4. Discussion

In this section, we discuss cross-cutting aspects of the results, namely the influence of the chatbots’ perceived humanness, the conversational domains, and the relationships among the characteristics.

4.1. A note on chatbots’ humanness

The social characteristics identified in the survey align well with characteristics that are present in human-human interactions. On the one hand, this survey shows the benefits of considering these characteristics when developing chatbots, which is supported by the Media Equation theory (Nass & Moon, 2000; Nass et al., 1993); we identified the domains in which each characteristic was studied. On the other hand, previous studies have also shown that developing overly humanized agents results in high expectations and uncanny feelings (Ciechanowski et al., 2018; Gnewuch et al., 2017), which was also pointed out in the surveyed literature as a challenge to conveying particular characteristics, such as *identity* and *personality*.

An explanation for these contrasts is that interactivity is “dependent on the identity of the source with whom/which we are carrying out the message exchange” (Sundar et al., 2016), i.e., people direct their messages to an artificial agent. In interactions with chatbots, the artificial agent represents the social role usually associated with a human, for example, a tutor (Dyke et al., 2013; Tegos et al., 2016), a healthcare provider (Montenegro, da Costa, & da Rosa Righi, 2019), a salesperson (Gnewuch et al., 2017; Zhu, Zhang, Li, Huang, & Zhao, 2018), a hotel concierge (Lasek & Jessa, 2013), or even a friend (Shum et al., 2018; Thies et al., 2017). The idea of assigning a social role to a chatbot does not imply deceiving people into thinking the software is human. Still, as the chatbots enrich their communication and social skills, their

perceived social role may approach human profiles.

4.2. *The importance of conversational context*

As Table 2 depicts², almost all social characteristics (10 out of 11) were found in studies related to conversations not restricted to a specific domain (open domain): manners, moral agency, and emotional intelligence are particularly relevant, since many conversations involve personal content (e.g., preferences, personal life, and habits); damage control is necessary since interactions are more susceptible to testing or flaming; identity, personality, and thoroughness are adopted to increase believability and consistency; and conscientiousness, proactivity, and personalization foster user engagement and a coherent conversational flow. The only characteristic that was not investigated in an open domain was communicability. As the conversations in this context mimic human social interactions, conveying the chatbots' capabilities is less of a concern.

The reported characteristics vary across task-oriented domains. Education and customer services literature report the greatest number of characteristics, but this result might be influenced by the maturity of the research in chatbots for these domains. As we pointed out in Section 2, education and customer services are the task-oriented domains most reported in the literature. We found conscientiousness, damage control, thoroughness, manners, emotional intelligence, and identity in studies for both domains. However, manners and emotional intelligence have a different goal in these domains. In the education context, these characteristics are designed to encourage students, especially in a situation of failure, in which the chatbot should be comforting and sensitive. This function aligns with other domains, such as healthcare. The education domain also reports needs for personality, so the chatbot can be recognized as either an instructor or a student, and proactivity, so the chatbot can motivate students to participate in the interactions. Personality is also reported in other domains in which the chatbots' character influences the interactions, such as gaming and humorous talk. Proactivity is also consistently reported in domains in which the chatbot provides guidance, such as coaching, health, ethnography, and assessment interviews.

In customer services, emotional intelligence and manners are used as a means to manage customers' frustrations with either a product or service. This function is also reported in other domains, such as virtual assistants and information search. Additionally, communicability is important to convey the services provided by the chatbot, and personalization helps to provide services that align with a particular customer, which is observed in domains such as ethnography, task management, virtual assistants, and financial services. Noticeably, personalization mostly co-occurs with proactivity, since users expect proactive messages to be relevant to their particular interactional context. Moral agency is the only characteristic that does not show up across several domains. It was reported only in race-talk, and open domain. This highlights that moral agency is a concern when the topic involves content that might lead to inappropriate responses. However, as this survey focuses on characteristics reported in the literature, we believe that it might be influenced by the approach adopted by the researchers: it makes sense to study moral agency in the context of gender or race conversations. We argue, though, that this characteristic might be relevant in domains that were not reported here, but that involve interactions that may lead to biased conversations.

²see more details in the supplementary materials

4.3. Interrelationships among the characteristics

In Section 3, we organized the social characteristics into discrete groups and discussed several instances of characteristics influencing each other, or being used as a strategy to manifest one another. In this section, we describe these relations in a theoretical framework, depicted in Figure 1. Rather than providing a comprehensive map, the goal here is to make explicit the relationships found in the surveyed chatbots literature, which underlines the complexity of developing chatbots with appropriate social behaviors. In Figure 1, boxes represent the social characteristics and the colors group them into their respective categories. The axes represent the 22 propositions we derived from the literature.

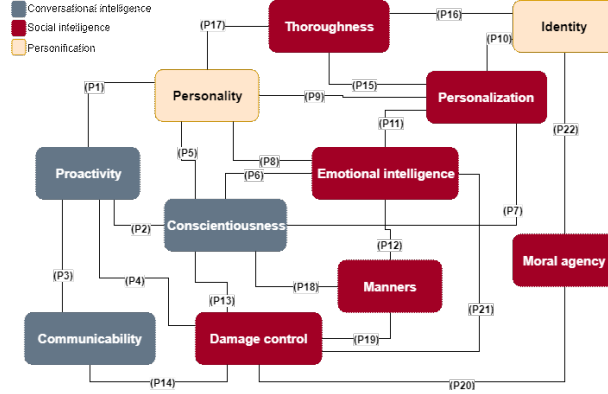


Figure 1. Interrelationship among social characteristics

According to the surveyed literature, *proactivity* influences perceived *personality* (**P1**) (Thies et al., 2017), since recommending and initiating topics may manifest higher levels of extraversion. When the proactive messages are based on the context, *proactivity* increases perceived *conscientiousness* (**P2**) (Thies et al., 2017), since the proactive messages may demonstrate attention to the topic. *Proactivity* supports *communicability* (**P3**) (Chaves & Gerosa, 2018; Silvervarg & Jönsson, 2013; Valério et al., 2017), since a chatbot can proactively communicate its knowledge and provide guidance; in addition, *proactivity* supports *damage control* (**P4**) (Silvervarg & Jönsson, 2013), since a chatbot can introduce new topics when the user either is misunderstood, tries to break the system, or sends an inappropriate message.

Conscientiousness is itself a dimension of the Big Five model; hence, *conscientiousness* influences the perceived *personality* (**P5**) (Goldberg, 1990). Higher levels of context management, goal-orientation, and understanding increase the chatbots' perceived efficiency, organization, and commitment (Dyke et al., 2013). *Conscientiousness* manifests *emotional intelligence* (**P6**) since retaining information from previous turns and being able to recall them demonstrates empathy (Jain, Kumar, et al., 2018). In addition, *conscientiousness* manifests *personalization* (**P7**) (Jain, Kumar, et al., 2018; Thies et al., 2017) because a chatbot can remember individual preferences within and across sessions.

Emotional intelligence influences perceived *personality* (**P8**), since chatbots' *personality* traits affect the intensity of emotional reactions (Morris, 2002; Thies et al., 2017). Agreeableness is demonstrated through consistent warm reactions such as encouraging and motivating (Ayedoun et al., 2017; Kumar et al., 2010; Shum et al., 2018; Sjöden et al., 2011). Some *personality* traits require *personalization* (**P9**) to

adapt to the interlocutors’ culture and interests (Ptaszynski et al., 2010). Besides, *personalization* benefits *identity* (P10), since the users’ social-agent orientation may require a chatbot to adapt the level of engagement in small talk and the agent’s visual representation (Jenkins et al., 2007; V. Q. Liao et al., 2016). *Personalization* also improves *emotional intelligence* (P11), since a chatbot should dynamically regulate the affective reactions to the interlocutor (Shum et al., 2018; Thies et al., 2017). *Emotional intelligence* improves perceived *manners* (P12), since the lack of *emotional intelligence* may lead to the perception of impoliteness (Jenkins et al., 2007).

Conscientiousness facilitates *damage control* (P13), since the attention to the workflow and context may increase the ability to recover from a failure without restarting the workflow (Duijst, 2017; Dyke et al., 2013; Gnewuch et al., 2017; Jain, Kota, et al., 2018). *Communicability* facilitates *damage control* (P14), since it teaches the user how to communicate, reducing the numbers of mistakes (Duijst, 2017; Gnewuch et al., 2017; Jain, Kumar, et al., 2018; Silvervarg & Jönsson, 2013; Wallis & Norling, 2005). In addition, suggesting how to interact can reduce frustration after failure scenarios (Q. V. Liao et al., 2018).

Personalization manifests *thoroughness* (P15) (Duijst, 2017; Gnewuch et al., 2017; Hill et al., 2015; Thies et al., 2017), since chatbots can adapt their language use to the conversational context and the interlocutor’s expectations. When the context requires dynamic variation (Gnewuch et al., 2017; Thies et al., 2017), *thoroughness* may reveal traits of the chatbot’s *identity* (P16) (Marino, 2014; Schlesinger et al., 2018). As demonstrated by (Mairesse & Walker, 2009), *thoroughness* also reveals *personality* (P17).

Manners influence *conscientiousness* (P18) (Wallis & Norling, 2005), since they can be applied as a strategy to politely refuse off-topic requests and to keep the conversation on track. *Manners* also influence *damage control* (P19) (Curry & Rieser, 2018; Wallis & Norling, 2005), because they can help a chatbot prevent verbal abuse and reduce the negative effect of lack of knowledge. Both *moral agency* (P20) and *emotional intelligence* (P21) improve *damage control* because they provide the ability to appropriately respond to abuse and testing (Silvervarg & Jönsson, 2013; Wallis & Norling, 2005). *Identity* influences *moral agency* (P22), since *identity* representations require the ability to prevent a chatbot from building or reinforcing negative stereotypes (Brahnam & De Angeli, 2012; Marino, 2014; Schlesinger et al., 2018).

5. Related Surveys

Previous studies have reviewed the literature on chatbots. Several surveys discuss chatbots’ urgency (Dale, 2016; Pereira, Coheur, Fialho, & Ribeiro, 2016) and their potential application for particular domains, which include education (Deryugina, 2010; Rubin, Chen, & Thorimbert, 2010; Satu, Parvez, et al., 2015; Shawar & Atwell, 2007; Winkler & Söllner, 2018), business (Deryugina, 2010; Shawar & Atwell, 2007), health (Fadhil, 2018; Laranjo et al., 2018), information retrieval and e-commerce (Shawar & Atwell, 2007). Other surveys focus on technical design techniques (Ahmad et al., 2018; Deshpande et al., 2017; Masche & Le, 2017; Ramesh et al., 2017; Thorne, 2017; Walgama & Hettige, 2017; Winkler & Söllner, 2018), such as language generation models, knowledge management, and architectural challenges. Although (Augello, Gentile, & Dignum, 2017) discuss the social capabilities of chatbots, the survey focuses on the potential of available open source technologies to support these skills, highlighting technical hurdles rather than social ones.

We found three surveys (Ferman, 2018; Pereira et al., 2016; Radziwill & Benton, 2017) that include insights about social characteristics of chatbots, although none of them focus on this theme. Pereira et al. (2016) investigated chatbots that “*mimic conversation rather than understand it,*” and review the main technologies and ideas that support their design, while Ferman (2018) focuses on identifying best practices for developing script-based chatbots. Radziwill and Benton (2017) review the literature on quality issues and attributes for chatbots. The supplementary materials include a table that shows the social characteristics covered by each survey. These related surveys also point out technical characteristics and attributes that are outside the scope of this survey.

6. Limitations

This research has some limitations. Firstly, since this survey focused on disembodied, text-based chatbots, the literature on embodied and speech-based conversational agents was excluded. We acknowledge that studies that include these attributes can have relevant social characteristics for chatbots, especially for characteristics that could be highly influenced by physical representations, tone, accent, and so forth (e.g. identity, politeness, and thoroughness). However, embodiment and speech could also bring new challenges (e.g., speech-recognition or eye-gazing), which are beyond the scope of this study and could potentially impact users’ experiences with chatbots. Additionally, this decision may have caused the under-representation of certain research domains. For example, there is an established research branch on conversational agents in the Ambient Intelligence discipline (Leonidis et al., 2017; López-Cózar & Callejas, 2010; Stefanidi et al., 2019) that primarily investigates multimodal interactions, and hence is not represented in this survey. Another example is the Software Engineering domain (see, e.g., (Abdellatif, Costa, Badran, Abdalkareem, & Shihab, 2020; Lebeuf, Storey, & Zagalsky, 2017; Lin, Zagalsky, Storey, & Serebrenik, 2016)), where chatbots have been widely investigated, but focused on the functional, rather than social aspects of the interactions.

Secondly, since the definition of chatbot is not consolidated in the literature and chatbots have been studied across several different domains, some studies that include social aspects of chatbots may not have been found. To account for that, we adopted several synonyms in our research string and used Google Scholar as search engine, which provides a fairly comprehensive indexing of the literature in more domains. Finally, the conceptual model of social characteristics was derived through a coding process inspired by qualitative methods, such as Grounded Theory. Like any qualitative coding methods, it relies on the researchers’ subjective assessment. To mitigate this threat, the researchers discussed the social characteristics and categories during in-person meetings until reaching consensus, and both the conceptual framework and the relationships amongst the characteristics were derived based on outcomes the surveyed studies explicitly reported.

7. Conclusion

In this survey, we investigated the literature on disembodied, text-based chatbots to answer the question “*What chatbot social characteristics benefit human interactions and what are the challenges and strategies associated with them?*” Our main contribu-

tion is the conceptual model of social characteristics, from which we can derive conclusions about several research opportunities. Firstly, we point out several challenges to overcome in order to design chatbots that manifest each characteristic. Secondly, further research may focus on assessing the extent to which the identified benefits are perceived by users and influence their satisfaction. Finally, further investigations may propose new strategies to manifest particular characteristics. In this sense, we highlight that we could not identify strategies to manifest moral agency and thoroughness, although strategies for several other characteristics are also under-investigated.

We reported domains where social characteristics were primarily investigated. We showed that some characteristics are largely influenced by the domain (e.g., *moral agency* and *communicability*), while others are more generally applied (e.g., *manners* and *damage control*). We also discussed the relationship among the characteristics through 22 propositions derived from the surveyed literature, which underline the complexity of developing chatbots with appropriate social behaviors. Although the propositions and social characteristics do not intend to be comprehensive, they cover important aspects of human-human communication. Social sciences such as sociolinguistics and communication studies have a good deal to contribute to the design of chatbots. In the description of the characteristics, we pointed out theories that could guide these investigations, such as the cooperative principle (Grice et al., 1975), social identity (R. Brown, 2000; Stets & Burke, 2000), personalization (Fan & Poole, 2006), politeness (P. Brown & Levinson., 1987), and mind perception theories (Gray et al., 2007). For example, *conscientiousness*, *thoroughness*, and *proactivity* might be covered by the cooperative principle (Grice et al., 1975), and social identity theory (R. Brown, 2000; Stets & Burke, 2000) might explain the relevance of *identity*, *personality*, and *moral agency* for conversational technologies. Our results provide important references for helping designers and researchers find opportunities to advance the human-chatbot interactions field.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). Challenges in chatbot development: A study of stack overflow posts. In *In 17th international conference on mining software repositories (msr '20)*. New York, NY, USA: IEEE.
- Abdul-Kader, S. A., & Woods, J. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, 6(7), 72–80.
- Ahmad, N. A., Che, M. H., Zainal, A., Rauf, M. F. A., & Adnan, Z. (2018, August). Review of chatbots design techniques. *International Journal of Computer Applications*, 181(8), 7–10.
- Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*, 21(4), 12–17.
- Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, 102, 274–286.
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions.

- Computers in Human Behavior*, 85, 183–189.
- Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York, NY, USA: NYU press.
- Augello, A., Gentile, M., & Dignum, F. (2017). An overview of open-source chatbots social skills. In *International conference on internet science* (pp. 236–248). Cham: Springer, Cham.
- Avula, S., Chadwick, G., Arguello, J., & Capra, R. (2018). Searchbots: User engagement with chatbots during collaborative search. In *Proceedings of the 2018 conference on human information interaction & retrieval* (pp. 52–61). New York, NY, USA: ACM.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2017). Communication strategies and affective backchannels for conversational agents to enhance learners’ willingness to communicate in a second language. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial intelligence in education* (pp. 459–462). Cham: Springer International Publishing.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2018, March). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners’ willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29(1), 29–57.
- Banks, J. (2018). Perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, 90, 363–371.
- Baron, N. S. (1984). Computer mediated communication as a force in language change. *Visible language*, 18(2), 118.
- Björkqvist, K., Österman, K., & Kaukiainen, A. (2000). Social intelligence- empathy= aggression? *Aggression and violent behavior*, 5(2), 191–200.
- Boiteux, M. (2019). *Messenger at f8 2018*. retrieved october 18, 2019 from <https://blog.messengerdevelopers.com/messenger-at-f8-2018-44010dc9d2ea>. Messenger Developer Blog.
- Brahnam, S., & De Angeli, A. (2012). Gender affordances of conversational agents. *Interacting with Computers*, 24(3), 139–153.
- Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In *4th international conference on internet science* (pp. 377–392). Cham: Springer International Publishing.
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, 25(5), 38–43.
- Brown, P. (2015). Politeness and language. In *The international encyclopedia of the social and behavioural sciences (iesbs)* (2nd ed., Vol. 18, pp. 326–330). Netherlands: Elsevier.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge, UK: Cambridge University Press.
- Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European journal of social psychology*, 30(6), 745–778.
- Candello, H., Pinhanez, C., & Figueiredo, F. (2017). Typefaces and the perception of humanness in natural language chatbots. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3476–3487). New York, NY, USA: ACM.
- Candello, H., Pinhanez, C., Pichiliani, M. C., Guerra, M. A., & Gatti de Bayser, M. (2018). Having an animated coffee with a group of chatbots from the 19 th century. In *Extended abstracts of the 2018 chi conference on human factors in computing systems* (pp. D206:1–D206:4). New York, NY, USA: ACM.
- Cassell, J. (2009). Social practice: Becoming enculturated in human-computer interaction. In *International conference on universal access in human-computer interaction* (pp. 303–313). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Chaves, A. P., & Gerosa, M. A. (2018). Single or multiple conversational agents? an interactional coherence comparison. In *Acm sigchi conference on human factors in computing systems* (pp. 191:1–191:13). New York, NY, USA: ACM.
- Chaves, A. P., & Gerosa, M. A. (2019, April). *[DATASET] Survey on social characteristics of human-chatbot interaction*. Retrieved from <https://doi.org/10.5281/zenodo.3473358>

- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539–548.
- Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language learning. *ReCALL*, 20(1), 99–117.
- Conrad, S., & Biber, D. (2009). *Register, genre, and style*. New York, NY, USA: Cambridge University Press.
- Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: people are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, 58, 431–442.
- Curry, A. C., & Rieser, V. (2018). # metoo alexa: How conversational systems respond to sexual harassment. In *Proceedings of the second acl workshop on ethics in natural language processing* (pp. 7–14). New Orleans, Louisiana, USA: Association for Computational Linguistics.
- Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4), 258–266.
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811–817.
- de Bayser, M. G., Cavalin, P. R., Souza, R., Braz, A., Candello, H., Pinhanez, C. S., & Briot, J. (2017). A hybrid architecture for multi-party conversational systems. *CoRR*, abs/1705.01214. Retrieved from <http://arxiv.org/abs/1705.01214>
- De Souza, C. S., Prates, R. O., & Barbosa, S. D. (1999). A method for evaluating software communicability. *Monografias em Ciência da Computação. Departamento de Informática. PUC-RioInf*, 1200, 11–99.
- De Angeli, A. (2005). To the rescue of a lost identity: Social perception in human-chatterbot interaction. In *Proceedings of the joint symposium on virtual social agents* (pp. 7–14). Hatfield, UK: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
- De Angeli, A., & Brahnam, S. (2006). Sex stereotypes and conversational agents. In *Proceedings of the avi 2006 workshop on gender and interaction: real and virtual women in a male world*. Venice, Italy: Online.
- De Angeli, A., Johnson, G. I., & Coventry, L. (2001). The unfriendly user: exploring social reactions to chatterbots. In *Proceedings of the international conference on affective human factors design, london* (pp. 467–474). London, UK: Asean Academic Press.
- De Angeli, A., Lynch, P., & Johnson, G. (2001). Personifying the e-market: A framework for social agents. In M. Hirose (Ed.), *Interact* (pp. 198–205). Amsterdam, The Netherlands: IOS Press.
- Deryugina, O. (2010). Chatterbots. *Scientific and Technical Information Processing*, 37(2), 143–147.
- Deshpande, A., Shahane, A., Gadre, D., Deshpande, M., & Joshi, P. M. (2017, May). A survey of various chatbot implementation techniques. *International Journal of Computer Engineering and Applications, Special Issue, XI*.
- Dohsaka, K., Asai, R., Higashinaka, R., Minami, Y., & Maeda, E. (2014). Effects of conversational agents on activation of communication in thought-evoking multi-party dialogues. *IEICE TRANSACTIONS on Information and Systems*, 97(8), 2147–2156.
- Duijst, D. (2017). *Can we Improve the User Experience of Chatbots with Personalisation* (Unpublished master’s thesis). University of Amsterdam.
- Duijvelshoff, W. (2017). *Use-cases and ethics of chatbots on plek: a social intranet for organizations*. Workshop On Chatbots And Artificial Intelligence.
- Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards academically productive talk supported by conversational agents. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 531–540). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Elliott, C. (1994). Research problems in the use of as allow artificial intelligence model of personality and emotion.

- Fadhil, A. (2018). Can a chatbot determine my diet?: Addressing challenges of chatbot application for meal recommendation. *arXiv preprint arXiv:1802.09100*.
- Fan, H., & Poole, M. S. (2006). What is personalization? perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3-4), 179–202.
- Ferman, M. (2018). *Towards best practices for chatbots* (Unpublished master’s thesis). Universidad Villa Rica.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (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 mental health*, 4(2), online.
- Fitzpatrick, M. A., & Winke, J. (1979). You always hurt the one you love: Strategies and tactics in interpersonal conflict. *Communication Quarterly*, 27(1), 3–11.
- Fogg, B. (2003). Computers as persuasive social actors. In B. Fogg (Ed.), *Persuasive technology* (p. 89 - 120). San Francisco: Morgan Kaufmann.
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of hci. *interactions*, 24(4), 38–42.
- Forlizzi, J., Zimmerman, J., Mancuso, V., & Kwak, S. (2007). How interface agents affect interaction between humans and computers. In *Proceedings of the 2007 conference on designing pleasurable products and interfaces* (pp. 209–221).
- Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. In *International conference on information systems 2017, proceedings 1*. South Korea: Association for Information Systems.
- Goldberg, L. R. (1990). An alternative” description of personality”: the big-five factor structure. *Journal of personality and social psychology*, 59(6), 1216.
- Gong, L. (2008). How social is social responses to computers? the function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494–1509.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science*, 315(5812), 619–619.
- Grice, H. P., Cole, P., Morgan, J., et al. (1975). Logic and conversation. 1975, 41–58.
- Gross, J. J. (1998). The emerging field of emotion regulation: an integrative review. *Review of general psychology*, 2(3), 271.
- Grossman, T., Fitzmaurice, G., & Attar, R. (2009). A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 649–658).
- Harré, R., Moghaddam, F. M., Moghaddam, F., et al. (2003). *The self and others: Positioning individuals and groups in personal, political, and cultural contexts*. Greenwood Publishing Group.
- Hayashi, Y. (2015). Social facilitation effects by pedagogical conversational agent: Lexical network analysis in an online explanation task. In *International conference on educational data mining (edm)*. Japan: International Educational Data Mining Society.
- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250.
- Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, 11(1), 19–29.
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4), 712–733.
- Jacquet, B., Baratgin, J., & Jamet, F. (2018). The gricean maxims of quantity and of relation in the turing test. In *2018 11th international conference on human system interaction (hsi)*

- (pp. 332–338). IEEE.
- Jacquet, B., Hullin, A., Baratgin, J., & Jamet, F. (2019). The impact of the gricean maxims of quality, quantity and manner in chatbots. In *2019 international conference on information and digital technologies (idt)* (pp. 180–189). IEEE.
- Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey: Exploring the Use of a Context View for Chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 468). New York, NY, USA: ACM.
- Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 on designing interactive systems conference 2018* (pp. 895–906). New York, NY, USA: ACM.
- Jenkins, M.-C., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of user interaction with service oriented chatbot systems. In J. A. Jacko (Ed.), *Human-computer interaction. hci intelligent multimodal interaction environments* (pp. 76–83). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jiang, R., & E Banchs, R. (2017). Towards improving the performance of chat oriented dialogue system. In *2017 international conference on asian language processing (ialp)* (p. 23-26). IEEE.
- Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 205–214). New York, NY, USA: ACM.
- Kirakowski, J., Yiu, A., et al. (2009). Establishing the hallmarks of a convincing chatbot-human dialogue. In *Human-computer interaction*. London, UK: InTech.
- Koda, T. (2003). User reactions to anthropomorphized interfaces. *IEICE TRANSACTIONS on Information and Systems*, 86(8), 1369–1377.
- Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. In V. Alevan, J. Kay, & J. Mostow (Eds.), *International conference on intelligent tutoring systems* (pp. 156–164). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., . . . others (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258.
- Lasek, M., & Jessa, S. (2013). Chatbots for Customer Service on Hotels’ Websites. *Information Systems in Management*, 2(2), 146–158.
- Lebeuf, C., Storey, M.-A., & Zagalsky, A. (2017). How software developers mitigate collaboration friction with chatbots. *arXiv preprint arXiv:1702.07011*.
- Lee, M., Lucas, G., Mell, J., Johnson, E., & Gratch, J. (2019). What’s on your virtual mind?: Mind perception in human-agent negotiations. In *Proceedings of the 19th acm international conference on intelligent virtual agents* (pp. 38–45). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/3308532.3329465>
- Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103, 95–105.
- Leonidis, A., Antona, M., & Stephanidis, C. (2017). Using contextual knowledge to resume human-agent conversations when programming the intelligence of smart environments. In *Wcikai@ iva* (pp. 46–54).
- Li, Y., Su, H., Shen, X., Li, W., Cao, Z., & Niu, S. (2017). Dailydialog: A manually labelled multi-turn dialogue dataset. In *International joint conference on natural language processing (ijcnlp)* (pp. 986–995). Taipei, Taiwan: Asian Federation of Natural Language Processing.
- Liao, Q. V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M. P., . . . Geyer, W. (2018). All work and no play? conversations with a question-and-answer chatbot in the wild. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 3:1–3:13). New York, NY, USA: ACM.
- Liao, V. Q., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do?: Studying social-agent orientation and agent proactive interactions with an agent for

- employees. In *Proceedings of the 2016 acm conference on designing interactive systems* (pp. 264–275). New York, NY, USA: ACM.
- Lin, B., Zagalsky, A., Storey, M.-A., & Serebrenik, A. (2016). Why developers are slacking off: Understanding how software teams use slack. In *Proceedings of the 19th acm conference on computer supported cooperative work and social computing companion* (pp. 333–336).
- López-Cózar, R., & Callejas, Z. (2010). Multimodal dialogue for ambient intelligence and smart environments. In *Handbook of ambient intelligence and smart environments* (pp. 559–579). Springer.
- Luger, E., & Sellen, A. (2016). Like having a really bad pa: The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 5286–5297). New York, NY, USA: ACM.
- Mairesse, F., & Walker, M. A. (2009). *Can conversational agents express big five personality traits through language?: Evaluating a psychologically-informed language generator*. Cambridge & Sheffield, United Kingdom: Cambridge University Engineering Department & Department of Computer Science, University of Sheffield.
- Marino, M. C. (2014). The racial formation of chatbots. *CLCWeb: Comparative Literature and Culture*, 16(5), 13.
- Masche, J., & Le, N.-T. (2017). A review of technologies for conversational systems. In *International conference on computer science, applied mathematics and applications* (pp. 212–225).
- Maslowski, I., Lagarde, D., & Clavel, C. (2017). In-the-wild chatbot corpus: from opinion analysis to interaction problem detection. In *International conference on natural language and speech processing*.
- Mäurer, D., & Weihe, K. (2015). Benjamin Franklin’s decision method is acceptable and helpful with a conversational agent. In *Intelligent interactive multimedia systems and services* (pp. 109–120). Cham: Springer International Publishing.
- McCrae, R. R., & Costa Jr, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1), 81.
- McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American psychologist*, 52(5), 509.
- Meany, M. M., & Clark, T. (2010). Humour theory and conversational agents: An application in the development of computer-based agents. *International Journal of the Humanities*, 8(5), 129–140.
- Miner, A., Chow, A., Adler, S., Zaitsev, I., Tero, P., Darcy, A., & Paepcke, A. (2016). Conversational agents and mental health: Theory-informed assessment of language and affect. In *Proceedings of the fourth international conference on human agent interaction* (pp. 123–130). New York, NY, USA: ACM.
- Montenegro, J. L. Z., da Costa, C. A., & da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129, 56–67.
- Morris, T. W. (2002). Conversational agents for game-like virtual environments. In *Artificial intelligence and interactive entertainment*. (pp. 82–86). Palo Alto, CA, USA: AAAI Press.
- Morrissey, K., & Kirakowski, J. (2013). ‘realness’ in chatbots: Establishing quantifiable criteria. In *International conference on human-computer interaction* (pp. 87–96). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-ai social interactions. *Computers in Human Behavior*, 72, 432–440.
- Narita, T., & Kitamura, Y. (2010). Persuasive conversational agent with persuasion tactics. In *International conference on persuasive technology* (pp. 15–26).
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1), 81–103.
- Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: computers as social actors. In *Interact’93 and chi’93 conference companion on human factors in computing systems* (pp. 111–112).
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of*

- the sigchi conference on human factors in computing systems* (pp. 72–78). New York, NY, USA: ACM.
- Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics—talking to bots: symbiotic agency and the case of tay. *International Journal of Communication*, 10, 17.
- Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction*, 2(3), 60.
- Nguyen, Q. N., & Sidorova, A. (2018). Understanding user interactions with a chatbot: a self-determination theory approach. In *Twenty-fourth americas conference on information systems – emergent research forum (erf)*. New Orleans.
- Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1), 153–178.
- Parthemore, J., & Whitby, B. (2013). What makes any agent a moral agent? reflections on machine consciousness and moral agency. *International Journal of Machine Consciousness*, 5(02), 105–129.
- Pereira, M. J., Coheur, L., Fialho, P., & Ribeiro, R. (2016). Chatbots’ greetings to human-computer communication. *The Computing Research Repository (CoRR)*, *arXiv:1609.06479*.
- Petta, P., & Trappl, R. (1997). Why to create personalities for synthetic actors. In (pp. 1–8). Berlin, Heidelberg: Springer.
- Portela, M., & Granell-Canut, C. (2017). A new friend in our smartphone? observing interactions with chatbots in the search of emotional engagement. In *Interaccion* (pp. 48:1–48:7). New York, NY, USA: ACM.
- Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R., & Black, A. W. (2018). Style transfer through back-translation. In *56th annual meeting of the association for computational linguistics*. Philadelphia, USA: Association for Computational Linguistics.
- Prates, R. O., de Souza, C. S., & Barbosa, S. D. (2000). Methods and tools: a method for evaluating the communicability of user interfaces. *interactions*, 7(1), 31–38.
- Ptaszynski, M., Dybala, P., Higuhi, S., Shi, W., Rzepka, R., & Araki, K. (2010). Towards socialized machines: Emotions and sense of humour in conversational agents. In Z. ul-hassan Usmani (Ed.), *Web intelligence and intelligent agents*. Rijeka, Croatia: InTech.
- Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *Software Quality Professional*, 19(3), 25–36.
- Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). A survey of design techniques for conversational agents. In *International conference on information, communication and computing technology* (pp. 336–350). Singapore: Springer Singapore.
- Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press.
- Rousseau, D., & Hayes-Roth, B. (1996). *Personality in synthetic agents* (Tech. Rep.). Citeseer.
- Rubin, V. L., Chen, Y., & Thorimbert, L. M. (2010). Artificially intelligent conversational agents in libraries. *Library Hi Tech*, 28(4), 496–522.
- Ruch, W. (1998). *The sense of humor: Explorations of a personality characteristic* (Vol. 3). Walter de Gruyter.
- Salovaara, A., & Oulasvirta, A. (2004). Six modes of proactive resource management: a user-centric typology for proactive behaviors. In *Proceedings of the third nordic conference on human-computer interaction* (pp. 57–60).
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, 9(3), 185–211.
- Satu, M. S., Parvez, M. H., et al. (2015). Review of integrated applications with aiml based chatbot. In *Computer and information engineering (iccie), 2015 1st international conference on* (pp. 87–90).
- Schlesinger, A., O’Hara, K. P., & Taylor, A. S. (2018). Let’s talk about race: Identity, chatbots, and ai. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 315:1–315:14). New York, NY, USA: ACM.

- Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2018). An investigation of conversational agent relevance, presence, and engagement. In *Americas conference on information systems 2018 proceedings*. New Orleans: Association for Information Systems.
- Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful? In *Ldv forum* (Vol. 22, pp. 29–49). German: GSCL German Society for Computational Linguistics.
- Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 281–288). New York, NY, USA: ACM.
- Shum, H.-y., He, X.-d., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10–26.
- Silverbarg, A., & Jönsson, A. (2013). Iterative development and evaluation of a social conversational agent. In *6th international joint conference on natural language processing (ijcnlp 2013)* (pp. 1223–1229). Nagoya, Japan: Asian Federation of Natural Language Processing.
- Sjödén, B., Silverbarg, A., Haake, M., & Gulz, A. (2011). Extending an educational math game with a pedagogical conversational agent: Facing design challenges. In *Interdisciplinary approaches to adaptive learning. a look at the neighbours* (pp. 116–130). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Stefanidi, E., Foukarakis, M., Arampatzis, D., Korozi, M., Leonidis, A., & Antona, M. (2019). Parlami: a multimodal approach for programming intelligent environments. *Technologies*, 7(1), 11.
- Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social psychology quarterly*, 63(3), 224–237.
- Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical importance of contingency in human-computer interaction: effects of message interactivity on user engagement. *Communication Research*, 43(5), 595–625.
- Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The Ethnobot: Gathering Ethnographies in the Age of IoT. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 604). New York, NY, USA: ACM.
- Tamayo-Moreno, S., & Pérez-Marín, D. (2016). Adapting the design and the use methodology of a pedagogical conversational agent of secondary education to childhood education. In *Computers in education (siie), 2016 international symposium on* (pp. 1–6).
- Tegos, S., Demetriadis, S., & Tsiatsos, T. (2016). An investigation of conversational agent interventions supporting historical reasoning in primary education. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *International conference on intelligent tutoring systems* (pp. 260–266). Cham: Springer International Publishing.
- Tennenhouse, D. (2000). Proactive computing. *Communications of the ACM*, 43(5), 43–50.
- Thies, I. M., Menon, N., Magapu, S., Subramony, M., & O’neill, J. (2017). How do you want your chatbot? an exploratory wizard-of-oz study with young, urban indians. In B. R., D. G., J. A., K. B. D., O. J., & W. M. (Eds.), *Human-computer interaction-interact* (Vol. 10513, pp. 441–459). Cham, Switzerland: Springer.
- Thomson, L. (2005). A standard framework for web personalization. In *1st international workshop on innovations in web infrastructure (iwi 2005) co-located with the 14th world-wide web conference (www2005), japan*.
- Thorne, C. (2017). Chatbots for troubleshooting: A survey. *Language and Linguistics Compass*.
- Thorson, J. A., & Powell, F. (1993). Sense of humor and dimensions of personality. *Journal of clinical Psychology*, 49(6), 799–809.
- Toxtli, C., Cranshaw, J., et al. (2018). Understanding chatbot-mediated task management. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 58). New York, NY, USA: ACM.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
- Valério, F. A., Guimarães, T. G., Prates, R. O., & Candelero, H. (2017). Here’s what i can do:

- Chatbots' strategies to convey their features to users. In *Proceedings of the xvi brazilian symposium on human factors in computing systems* (p. 28). New York, NY, USA: ACM.
- Walgama, M., & Hettige, B. (2017). Chatbots: The next generation in computer interfacing—a review. *KDU International Research Conference*.
- Walker, M. A. (2009). Endowing virtual characters with expressive conversational skills. In *International workshop on intelligent virtual agents* (pp. 1–2).
- Wallace, R. S. (2009). The anatomy of a.l.i.c.e. In *Parsing the turing test* (pp. 181–210). Dordrecht: Springer Netherlands.
- Wallis, P., & Norling, E. (2005). The trouble with chatbots: social skills in a social world. In *Proceedings of the joint symposium on virtual social agents* (pp. 29–38). Hatfield, UK: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
- Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research*, 19(1), 52–90.
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1), 3–43.
- Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23(5), 2538–2557.
- Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), (4th ed., pp. 443–479). Thousand Oaks, CA: Sage.
- Watts, R. J. (2003). *Politeness*. Cambridge University Press.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis.
- Zamora, J. (2017). I'm sorry, dave, i'm afraid i can't do that: Chatbot perception and expectations. In *Proceedings of the 5th international conference on human agent interaction* (pp. 253–260). New York, NY, USA: ACM.
- Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.
- Zhang, W.-N., Zhu, Q., Wang, Y., Zhao, Y., & Liu, T. (2017). Neural personalized response generation as domain adaptation. *World Wide Web*, 22(4), 1427–1446.
- Zhu, P., Zhang, Z., Li, J., Huang, Y., & Zhao, H. (2018, August). Lingke: A fine-grained multi-turn chatbot for customer service. In *Proceedings of the 27th international conference on computational linguistics: System demonstrations* (pp. 108–112). Santa Fe, New Mexico: Association for Computational Linguistics.
- Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of personality and social psychology*, 65(4), 757.

About the Authors

Ana Paula Chaves is a Ph.D. Candidate at the Northern Arizona University and a Faculty at the Federal University of Technology - Paraná, Campus Campo Mourão, Brazil. She researches social aspects of human-chatbot interactions as well as technologies for tourism. More information at <http://anachaves.pro.br>

Marco Aurelio Gerosa is an Associate Professor at the Northern Arizona University. He researches Software Engineering and CSCW, having served on the program committee of important conferences, such as FSE, CSCW, SANER, and MSR. He is an ACM and IEEE senior member. For more information, visit

<http://www.marcoagerosa.com>