Collaborating with a Text-Based Chatbot: An Exploration of Real-World Collaboration Strategies Enacted during Human-Chatbot Interactions

Amon Rapp*
University of Torino, Torino, Italy
amon.rapp@unito.it

Arianna Boldi
University of Torino, Torino, Italy
arianna.boldi@unito.it

Lorenzo Curti
University of Torino, Torino, Italy
lorenzo.curti93@gmail.com

Alessandro Perrucci
TIM S.p.A., Torino, Italy
alessandro.perrucci@telecomitalia.it

Rossana Simeoni
TIM S.p.A., Torino, Italy
rossana.simeoni@telecomitalia.it

ABSTRACT

A central problem for chatbots in the customer care domain revolves around how people collaborate with the agent to achieve their own situated goals. The majority of the previous research, however, relied on experiments within artificial settings, rather than on observation of real-world interactions. Moreover, such research mostly analyzed users' responses to communication breakdowns, rather than the wider collaboration strategies utilized during a conversation. In this paper, we qualitatively analyzed 12,477 real-world exchanges with a task-based chatbot using a Grounded Theory approach as a rigorous coding method to analyze the data. We identified two main aspects of collaboration, behavioral and conversational, and for each aspect we highlighted the different strategies that users perform to "work together" with the agent. These strategies may be utilized from the very beginning of the conversation or in response to misunderstandings in the course of ongoing interactions and may show different evolving dynamics.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI.

KEYWORDS

Conversational agents, human-machine cooperation, human-AI cooperation, chatbots

ACM Reference Format:
Amon Rapp, Arianna Boldi, Lorenzo Curti, Alessandro Perrucci, and Rossana Simeoni. 2023. Collaborating with a Text-Based Chatbot: An Exploration of Real-World Collaboration Strategies Enacted during Human-Chatbot Interactions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3580995

*Corresponding Author

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580995

1 INTRODUCTION

Over the last ten years, conversational agents, that is software applications interacting with people through natural language, have gained popularity [36]. In particular, text-based chatbots i.e., agents that are designed to interact with users through natural written language, have been employed in a variety of domains, from customer service [28] to mental wellbeing [55]. The availability of flexible platforms for building chatbots has led to increased investment in this technology, with the result that it has been predicted that by 2024 consumer retail spending via chatbots worldwide will reach $142 billion up from $2.8 billion in 2019 [43]. In particular, task-focused chatbots, which help users execute specific tasks, are drawing significant attention as they may be a valid alternative to live customer support [29, 52].

Despite enthusiastic market predictions, conversing with chatbots is not exempt from issues. Users may have a variety of expectations about this technology, in terms of for example, language skills [79] and social capabilities [11]: when such expectations are not met, negative emotions may be elicited, compromising the user experience [5, 44, 54, 66, 100]. Additionally, problems occurring during the interaction, like misunderstandings and "conflicts," may negatively affect the users' perception of the quality of the technology, decrease the willingness to recover from a communication breakdown, and lead to the abandonment of the conversation [3, 44, 52, 54]. These issues may be due not only to the intrinsic limits of the technology, but also to a lack of understanding of how people use chatbots to achieve their situated goals [51], that is to say, their personal objectives emerging from everyday needs. Therefore, it is important to investigate how conversations with chatbots unfold in real world contexts, discovering how people contribute to making the conversation progress.

From this perspective, an essential phenomenon to be studied is that of "collaboration," because it may favor mutual understanding, reduce the burden of performing a task, and increase the perceived quality of the information exchanged [15, 16, 32, 49, 94, 101]. Previous research has highlighted that unique human-machine cooperative dynamics can occur depending on the context of use of the machine and the objectives of the agents involved in the task, as well as on the kind of technology employed (e.g., [97]). In this sense, chatbot technology might entail unique cooperative behaviors that might be worthy of being explored.

Despite its importance, most HCI studies on chatbots do not directly address the theme of collaboration, instead preferring to focus on the "reactive" behavior that people perform in response to a communication breakdown [3, 95]. Moreover, most chatbot research is grounded either on users' ex-post reports about their conversations with the agent (e.g., [17, 25]), which may not reflect what actually happened during the interaction, or on experiments, in which people are given instructions on how to converse with the agent and are not pursuing their own situated goals [3, 79, 95]: nonetheless, if interaction is not driven by a real goal that reflects the fulfilment of an everyday need, the realism of the insights gained from these studies remains quite uncertain [52].

As an exception to this trend, Li et al. [52] analyzed thousands of real users' exchanges with a chatbot and found that users utilize several coping strategies to overcome obstacles during the interaction. This study is certainly important in showing the insight to be gained from exploring real-world conversations, but it still focuses on non-progress similar to previous research, losing sight of how people collaborate with technology not only to repair from breakdowns, but also to achieve their own situated goals. This may be particularly relevant in the case of task-focused chatbots, where users interact and possibly collaborate with the agent, with the goal to complete a specific task.

In this paper, therefore, we focus on human-chatbot cooperation by qualitatively analyzing 1,060 real-world conversations with a customer care chatbot deployed in an Italian telecommunication company, for a total of 12,477 exchanges. We used a Grounded Theory approach, in order to generate a preliminary understanding of this kind of behavior from the empirical data. We aimed to respond to the following question: how do people collaborate with a text-based chatbot in a conversation? We discovered that collaboration with chatbots has two aspects, that is *behavioral collaboration* and *conversational collaboration*, which both point to a variety of *collaborative strategies* that people enact during the interaction. These strategies range from executing specific behaviors as requested by the chatbot to simplifying the syntactical structure of a message or abstracting its content. Moreover, we found that collaboration shows dynamics of *evolution* and *devolution* oscillating between collaboration and non-collaboration.

The article makes a substantial contribution to the HCI community. It shows how users apply cooperative strategies to chatbots in real-world interactions, whereas most previous studies were conducted within artificial situations and focused on "coping strategies." In doing so, we offer rich insights on the situated experience of interacting with a chatbot and an in-depth look into real users' cooperative enactments.

The article is structured as follows. Section 2 outlines relevant related literature about human-machine cooperation. Section 3 describes the method used in this research. Section 4 outlines the study findings while Section 5 discusses them. Section 6 proposes a series of design implications. Section 7 describes the limitations of the study, whereas Section 8 concludes the article.

2 BACKGROUND

2.1 Human-machine cooperation

Cooperation is commonly understood as "the action or process of working together towards common goals" [72], where each person is responsible for a portion of the problem to solve [82, 85]. In order for it to be called "cooperative," an artificial agent must have the ability to solve a given problem and be able to cooperate with other agents by, for example, producing a common plan [73]. Cooperators also have to commit to the joint activity, show mutual responsiveness, and provide support to each other [13]. However, research suggests that human-machine cooperation dynamics do not only depend on these agents' basic abilities, but also on the context where cooperation occurs and the agents' goals, as well as on the employed technology.

Traditionally, human-machine cooperation research has focused on improving automated systems in the context of work by augmenting their ability to solve tasks in collaboration with humans. A variety of studies, for instance, explored the introduction of robots in manufacturing contexts (e.g. [12, 68, 77]), revealing the concerns of workers about automation, which may reduce human autonomy [96]. Research also highlighted that despite humans showing a cooperative attitude when they interact with machines at work [70], cooperation is often subject to fail, especially when human agents exhibit lack of trust in the machine or, conversely, complacency [74]. However, as intelligent devices are now spreading almost everywhere [24, 81], researchers pointed out that the study of human-machine cooperative behaviors needs to account for the specific contexts in which the cooperation occurs and the situated goals that drive both human and machine agents. For instance, autonomous agents may collaborate with musicians to compose music [62], support people in writing stories [14] or game designers to create games [103], help individuals select the best energy tariff in the home [1], give runners motivation and companionship during their run [35], help players feel less lonely [80], and provide support in the context of healthcare [48].

Moreover, research should consider the specificity of the technologies with which humans collaborate, as their design features and capabilities may affect the quality of cooperation [97]. Ohlin and Olsson [71], for instance, theorized about cooperation between humans and Personal Informatics (PI) technologies, such as wearables and mobile devices: they highlighted that not only can users drive cooperation, but PI systems may also initiate it through features, such as push notifications on smartphones or haptic feedback from wearable bracelets. By contrast, several studies have investigated bullying of robots [46, 83] and aggression towards autonomous vehicles [65], which clearly fall outside collaborative behaviors and need to be prevented by designers (e.g. [83, 90]).

To summarize, previous research has highlighted that unique human-machine cooperative dynamics can occur, depending on the context of use of the machine and the objectives of the agents involved in the task, as well as on the kind of technology employed. In this sense, chatbot technology might also present its own peculiarities.

2.2 Cooperation between humans and chatbots

A relevant domain in which human-machine cooperation may occur is that of chatbots. Chatbots are designed to interact with users in a way that mimics person-to-person conversations [11], so that their interactional goals also include social capabilities [44]. Research has shown that users prefer agents who show empathy [53], active listening skills [98], and reciprocity [50], self-disclose [51] and demonstrate a positive attitude [91]. When chatbots do not meet these expectations, the user may experience dissatisfaction or frustration [54, 100].

It comes as no surprise, therefore, that in this domain, cooperation is also crucial. In an ongoing conversation between humans, both the interlocutors are constantly engaged in a shared interpretative activity [94] and collaboratively adapt their language to the interlocutor [32], for instance, to facilitate the understanding of the recipient when they talk with non-native speakers [26] or children [38]. To a certain extent, the same appears to hold true for interaction with chatbots. Cooperation allows for mutual understanding between the interlocutors [15, 16, 47], reduces the burden of completing a given task [101], and, when artificial agents are perceived as cooperative, the quality of the information conveyed is perceived to be of better quality [49]. However, conversational technologies still do not fully address the complexities of natural language, which may hinder opportunities for establishing collaboration [3]. Moreover, humans tend to use profanity and ungrammatical language with chatbots, undermining their capabilities of understanding [40, 75, 93]. This might be particularly problematic when chatbots are employed to execute specific tasks in response to the user's requests (task-focused chatbots), such in the customer service domain, where conversational technologies are increasingly used [65]. In such a domain, productivity, in the form of quick access to information [5, 27], and efficient and accessible support [27, 28, 99] are essential, and users need to be correctly understood and receive helpful responses to be satisfied [27]. With these aims in mind, effective cooperation seems fundamental. In fact, when cooperation is not developed, "bad conversations" may occur: these have proven to decrease the users' willingness to ask the agent for help with the task [54] and to lead to the abandonment of the interaction [20, 44]. However, users may model their communication to match that of the chatbot [40] and "work-around" its technical limitations in order to continue receiving benefits by working with it [102]; chatbots can provide guidance on how to effectively interact [99]; while both humans and chatbots may utilize repairing strategies to fix communication breakdowns [3, 52].

Despite its importance, HCI research has devoted minor attention to how collaboration with chatbots is established in real-world contexts, rather preferring to focus on communication breakdowns and how users or agents attempt to repair within artificial settings. Ashktorab et al. [3], for instance, conducted an experiment exploring chatbots' repair strategies, finding that providing options and explanations are generally favored by the users. In the same vein, Weisz et al. [95], in an online experiment where participants had to recover from communication breakdowns, found that they learned strategies for dealing with problems emerging from the interaction. Similarly, Jiang et al. [45] set up an experiment to study how users reformulate their information requests to voice search systems, whereby reordering words or changing phonetics could help correct input errors. More recently, Myers et al. [69] investigated how users interact with a voice user interface calendar system over three experimental sessions, identifying the main obstacles and tactics that they employed to overcome them, like hyper-articulation and quitting the task.

As we may see, all these previous studies are grounded in experiments entailing artificial situations, rather than on people's actual behavior (e.g., through the analysis of real-world conversations). An exception is represented by Li et al. [52] who analyzed human conversation logs with a banking chatbot. They focused on conversational non-progress (NP), identifying different types of NP and coping strategies that users implement in order to handle them. This study is certainly inspiring for our research. However, as with other previous studies, its focus on non-progress may have limited the exploration of how collaborative practices unfold, as users may show the willingness to cooperate even in the absence of communication breakdowns.

In this paper, we precisely explore the collaborative strategies that users perform when they interact with a task-focused chatbot in the customer care domain, by analyzing real-world human-chatbot conversations. It is important to study cooperation in customer care, as in this domain users may interact with the chatbot with the main goal of quickly solving an issue, which may require a strong and rapid collaboration between the two parts. Here, investigating whether and how people commit effort to utilizing collaborative strategies may provide us with insights on how to design better conversational agents and increase our understanding of how people converse with agents with a "limited intelligence" to achieve their situated goals.

3 METHOD

We opted for a Grounded Theory approach, analyzing conversations as if they were data coming from a qualitative study. Grounded Theory is commonly applied to data collected through ethnographic field notes and interviews, but it has also been employed to analyze other kinds of data, like academic articles (e.g., [64]).

What differentiates Grounded Theory from other qualitative methods is an inductive, rather than a hypothetical-deductive stance [33]. Grounded Theory employs coding processes but differs from methods like content analysis that emphasizes validity, reliability and the counting of instances [30, 76]. Grounded Theory has been employed in diverse ways in both technology research and social sciences [31, 60]: for instance, to develop models emerging from the data, or as a coding technique. In fact, Grounded Theory does not necessarily imply the development of complex theoretical frameworks. It can also be used for defining taxonomies or describing a specific phenomenon or concept in detail [30].

In this research, we primarily used Grounded Theory as a technique to qualitatively analyze a large corpus of textual data (i.e., conversations with a chatbot) and to make sense of it.

3.1 Data characteristics

The initial dataset was composed of 3,123 conversations with a task-oriented chatbot maintained by one of the major Italian telecommunication companies. The company offers fixed and mobile line

Table 1: Characteristics of the conversations belonging to different time slots (the time slots included in the corpus are highlighted in bold)

	9am-11am	3pm-5pm	9pm-11pm	11am-3pm	5pm-9pm	11pm-9am
Average number of exchanges per conversation	**11.4**	**11.3**	**12.8**	11.5	11.1	12.9
Average number of words per user's input	**8.7**	**8.7**	**8.6**	8.7	8.0	8.1
Average duration of conversations (in seconds)	**207**	**212**	**230**	216	204	236

services and has a generalist customer base which counts millions of fixed line and SIM card subscriptions. The data were collected on the 28th of September 2019. We first cleaned the data, by manually filtering out 1,052 "conversations" where the user only greeted the chatbot (e.g., Hi) and immediately abandoned the interaction without making any request (likely due to for example, internet connection failure, closing of the web browser window, etc.), or did not interact at all (they likely only opened the chatbot window to make the agent start the conversation by presenting itself but did not respond: these were not proper conversations but the logger also assigned an ID to them). We then selected three time slots to be included in the corpus (9am - 11am, 3pm - 5pm, 9pm - 11pm): for these time slots, we analyzed the entire conversations, for a total of 1,060 conversations and 12,477 exchanges (11.8 exchanges per conversation on average, SD: 7.4).

The choice of sampling the conversations during these time slots followed a purposeful sampling method [57, 58] and is grounded in the fact that different categories of users may access the chatbot service at different times (e.g., workers tend to use the chatbot more in the evening after office hours), as we understood from preliminary interviews with several company stakeholders: this could potentially lead us to observe conversational patterns utilized by different kinds of users, possibly increasing the sample heterogeneity. However, the characteristics of these conversations do not significantly change throughout the day in terms of number of exchanges per conversation, number of words per user's input, and duration of conversations (Table 1). This may signal that our corpus was a reasonably good representative of the dataset.

Once the corpus was initially defined, we remained open to examining supplementary data (e.g., coming from supplementary time slots) if needed. Nonetheless, after having analyzed 1,060 conversations we recognized that new data would not have given us new insights, reaching therefore theoretical saturation [39].

Before beginning the conversation, users accepted the company's terms of service and were informed about our data-collection activities as in Li et al. [52] (e.g., that the data could be recorded, analyzed and published). Moreover, we followed guidelines to ensure ethics in online research [6, 41, 59, 86, 89] and embraced parts of the "heavy disguise" strategy suggested by Bruckman [10], by removing, in a pre-processing phase, all personal information (e.g., telephone numbers, names) that people could have involuntarily or voluntarily disclosed during the conversation. All the quotes reported in this article have been further translated from Italian to English and double checked to ensure that no individuals could be identified from them. This strategy has been previously used by HCI scholars analyzing e.g., online forum posts [56] or in-game chat logs [78].

3.2 Chatbot characteristics

The chatbot is built on the Microsoft BOT Framework and uses a text-based user interface running both on a web site and a mobile app (Figure 1). Users' requests are processed by a spellchecker and then analyzed through MS Language Understanding Cognitive Services to infer the user's intent and the entities.

The inference of user intent is based on sentence classification, in which the target classes are defined by providing a set of examples used to refine a neural network model pretrained on Italian language. Relevant entities (e.g., dates, names of people) are extracted from the user's sentences in different ways, from pattern matching based on regular expressions to neural network models. When the user provides syntactically too-long/too-complex sentences, the chatbot invites the user to reformulate their request.

The inferred user intent and entities are employed to drive a specific dialog flow which guides the user through the steps that are needed to complete a specific task in a deterministic way. The chatbot's behavior is the by-product of the combination of dialog nodes in a flowchart, which has one or more starting nodes, where the dialog starts, and one or more ending nodes, where the dialog ends. The chatbot's responses are defined at each dialog node, possibly using linguistic variations produced through a random criterion and information retrieved from back-end systems (e.g., the plan subscribed by the customer).

The chatbot is identified by a foreign name (Angie). When the user opens the chatbot window it immediately presents itself as a virtual assistant (Hello, I am Angie! The TIM virtual assistant. . .). The chatbot is represented by a stylized static manga-style illustration of its face. It is not designed to engage people in small talk or to develop a "relationship" with the user, but to solve specific commercial and technical problems with the services provided by the company. Users can, for example, request a change of address for an invoice, check the status of invoice payments, change the direct debit, activate, deactivate, or know the details of a promotion or a service, ask for clarification of charges, report modem breakdowns or slow Internet connection and try to find solutions. To interact with it, in addition to typing, the chatbot sometimes offers alternative input modalities, like quick-response buttons to diversify the interaction: this may happen when the chatbot offers predefined information asking the users to select an offer or service.

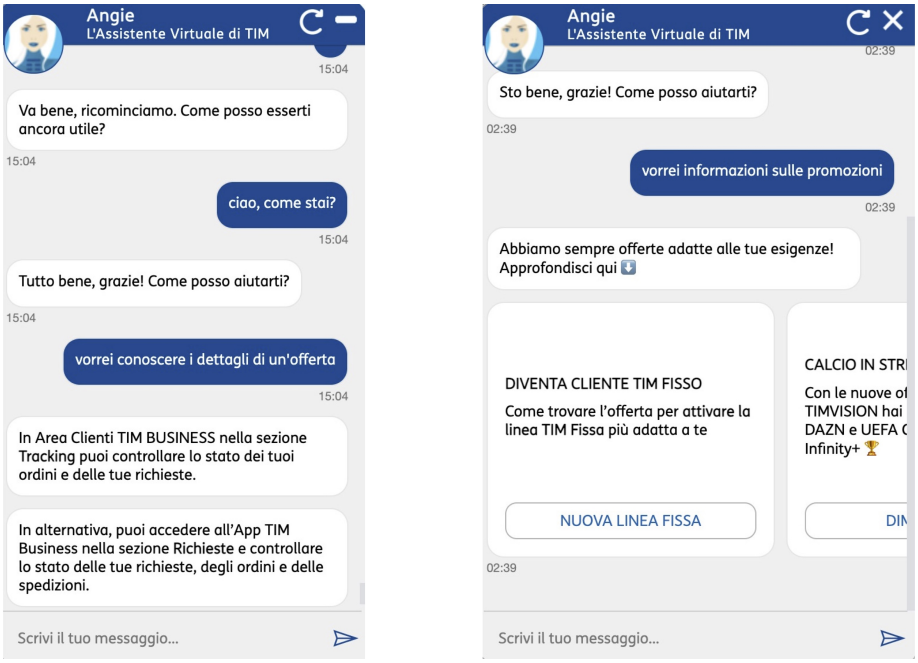

Figure 1: Two examples of the chatbot user interface: on the left, a common dialog flow; on the right, the chatbot displays two alternatives in the form of quick response buttons.

The collected data then differentiate whether the user input the text themselves or pressed the buttons.

3.3 Data analysis

The data were stored in a spreadsheet, with each row representing one of the 12,477 exchanges belonging to the corpus. Data format included the following fields: conversation ID, starting and ending time of the conversation, text of the conversation, and other technical data that were not considered in the analysis (e.g., start Time API call, API method name, etc.).

The analysis was performed taking into account the key principles of Grounded Theory. Pre-existing literature and theory were inhibited in the coding phase where we favored an inductive approach. The first, the second and the third authors read the entire corpus.

In a first stage, the second and the third authors separately assigned one or more open codes to each single user's interaction. During this phase, the analysis focused not only on the content communicated by the user to the chatbot, but also to the "form" of the conversation, namely the terms employed and the syntactical structure of the sentences (e.g., coordinate clauses or subordinate clauses, repetitions, length of the sentences, typographical errors, etc.). The second and the third authors then discussed the defined codes to resolve inconsistencies. At times, inconsistencies were due to differences in labeling the same concepts. In other cases, two different codes were condensed into one, or new codes were developed, when the discussion between the two researchers led to identify clearer commonalities or distinctions among the meanings of the data points. This process went through the whole data set. As is common in qualitative research, and also in HCI research adopting a Grounded Theory approach (e.g., [61, 63, 67), no numerical reliability rating is reported, because our goal was to reach an intersubjective consensus, where each point of difference was debated and clarified until the coders agreed on appropriate usage of the set of codes [37].

In a second stage, the analysis was conducted at the conversation level, whereby the second and third authors independently assigned codes to entire conversations or portions of it, paying attention to how the conversations evolved from both the syntactical (how the user communicates) and semantical (what the user communicates) points of view. Then, they discussed again the generated codes to resolve inconsistencies and connected the codes generated during the first level (row level) to the codes generated in this phase (conversation level). A final set of 85 open codes was then defined in a codebook. Then, the entire dataset was recoded with the codebook at hand.

As a second step, these concepts were grouped into conceptual categories through axial coding by the second and third authors separately. The second and third authors then discussed the categories resolving discrepancies and involving the first author, when doubts about the decision arose. This yielded 11 learned abstracted categories (e.g., positive behavioral disposition, positive evolution, syntactical adaptation, refusal to adapt language). Axial coding categories from open coding were amalgamated to create a more defined hierarchy forming key related categories through selective coding. The resultant three selective categories are the central themes emerging from the data analysis, which points to key elements of cooperation between humans and task-oriented chatbots:

collaboration aspects, collaboration strategies, and collaboration dynamics.

4 FINDINGS

By focusing on the terms and syntactical structures employed by the users and the content that they tried to communicate, we were able to identify a variety of linguistic "cues" (e.g., the removal of syntactical structure from a complex sentence) signaling how collaboration is performed during the interaction through different *collaboration strategies*. These strategies were then further retraced to two high-level *collaboration aspects*, that is *conversational collaboration* and *behavioral collaboration*, which show that users do not only adjust how they "converse" to the chatbot's skills, but also behaviorally adhere to its requests. Finally, by paying attention to how the users modify their language as the interaction proceeds, we discovered that collaboration is often dynamic, as it may *evolve* from non-collaboration to collaboration, or *devolve*, from collaboration to non-collaboration. Table 2 summarizes the findings which are extensively recounted in the next sub-sections.

4.1 Collaboration aspects

In the corpus that we analyzed, we found two main aspects of collaboration.

We called the first aspect behavioral collaboration, which consists of shaping the user's behavior to adhere to the chatbot's requests or to make the conversation non-confrontational. The term "behavior," here, stands for both a specific behavior like an action or sequence of actions (e.g., selecting an alternative offered by the chatbot) and a behavioral disposition, i.e., a tendency to engage consistently in a certain behavior (e.g., being polite) [42] throughout the whole conversation (or parts of it). In other words, the collaboration is behavioral because the user agrees to perform a specific behavior proposed by the chatbot (e.g., C: "*I have 2 lines in your name: 1. <phoneNumber> 2. <phoneNumber> Which line should I refer to? Choose one of the lines, selecting the position in the list (1,2,3...)*" U: "1"); or shows a behavioral disposition in maintaining a "good interactional environment," also avoiding to abruptly abandon the conversation or indulge in inappropriate uncooperative behavior, such as insults and expressions of anger (e.g., U: "*Good morning, I saw the reimbursement but I didn't see the phone guide reimbursement in the invoice*" [...] U: "*Thank you*").

The second aspect, instead, is conversational collaboration and consists in adapting the user's conversational style to the chatbot's abilities of comprehension. The collaboration is conversational because the user uses (or rephrases) sentences that may allow the chatbot to better understand what they are saying, which may entail simple syntax (e.g., few conjunctions and adjectives) and clear communication of content (e.g., no information "overload"), like in U: "I am wondering what is the procedure to expire the tariff plan that I have active to date" C: "Sorry, I didn't quite understand your request" [...] U: "How to deactivate the contract."

Of course, how users behave when interacting with a chatbot is actually through conversation, thus the two aspects of collaboration often go hand in hand: users may agree, for instance, to execute the chatbot's requests and, at the same time, use a command-style language that can be understood by the agent.

There are cases, however, when the users collaborate conversationally, but not behaviorally: for example, they may adopt language with a minimal syntactical structure, but still refuse to behaviorally adhere to the chatbot's request, like in C: "*Choose one of the options, selecting the position in the list (1,2,3...)*" U: "*JANUARY*".

Alternatively, the user may collaborate behaviorally, but not conversationally, by, for instance, expressing a polite behavioral disposition towards the chatbot, and being unwilling to rephrase a request which the chatbot does not understand, instead making a further (complex) request (e.g., U: "*Good evening, I have subscribed to the new offer. I would like to know if a new modem will be delivered to replace the old one and, if so, if the related cost is included in the € 29.90 that I have to pay. Thank you*". [. . .] C: "*Would you like to have the detail, of a specific expense in the invoice?*" U: "*Wait a moment. Please forget it. I've just seen the invoice, for January, I was charged 3.90€ for the phone book. I'd like to ask for a refund because I didn't receive it* [. . .] *Thank you*").

The behavioral and conversational aspects of collaboration point to a variety of specific *strategies* that users employ during conversations, which we illustrate in the following sections, also providing examples taken from the collected conversations: key parts of the conversations are highlighted in **bold** to facilitate reading.

4.1.1 Behavioral collaboration. As we have seen, the behavioral aspect of collaboration consists in executing specific behaviors in response to the chatbot's requests or in adopting a non-confrontational behavior. This aspect of collaboration points to different *behavioral strategies* that users enact during the interaction. In the following we list the two main *behavioral strategies* that we identified, that is *commitment* and *politeness*. Moreover, we show that users may refuse to perform any behavioral strategy engaging in behavioral non-collaboration.

Commitment. The first behavioral strategy may be called *commitment*: it refers to those moments in the interaction when the user cooperates with the chatbot by performing the "behavior" (or the sequence of behaviors) that the agent expects from them either in the conversational environment (e.g., selecting a specific option) or in the real world (e.g., executing a task on the user's modem): in other words, the user carefully follows the chatbot's instructions and properly responds to its request during the interaction. The strategy is usually implemented as a response to the explicit request of the agent. For instance, in example 1 the user adheres to one of the options proposed by the agent, without producing an alternative sentence that is not foreseen in the choices offered by the chatbot.

> Example 1
>
> C: If the transfer of the contract is not for the benefit of a family member, it is chargeable. To request it, you can download the form in pdf format [. . .].
>
> Would you like to return to the transfer type list?
>
> U: yes
>
> C: The possible types of transfer are: 1. Family member 2. Family member requested by heir following the line owner's death 3. [. . .] **Choose one of the options shown, selecting the position in the list (1,2,3...).**

Table 2: Collaboration aspects, strategies, and dynamics

Collaboration Aspects	Collaboration strategies	Collaboration dynamics
Behavioral collaboration: when users shape their own behavior to adhere to the chatbot's requests or to make the conversation non-confrontational *Behavioral non-collaboration:* the user refuses to shape their own behavior to create a good conversational environment	*Behavioral Strategies* - *Commitment:* the user cooperates with the chatbot by performing the behavior that the agent expects from them - *Politeness:* the user has a behavioral disposition to engage consistently in a "polite" behavior, avoiding "negative" behaviors	*Evolution:* when there is increasing collaboration as the interaction progresses
Conversational collaboration: when users adapt their language to the chatbot's capabilities *Conversational non-collaboration:* the user refuses to adapt their language to the chatbot's capabilities	*Conversational Strategies* - *Abstraction:* the user abstracts the content of the message - *Simplification:* the user reduces the syntactical complexity of the sentences - *Reformulation:* the user substitutes nouns, verbs, abbreviations - *Complexification:* the user adds more "structure" and details to sentences - *Correction:* the user corrects grammatical errors and typos - *Emphasis:* the user capitalizes the letter of certain words - *Repetition:* the user repeats word by word a sentence	*Devolution:* when there is decreasing collaboration as the interaction progresses

U: **1**

In example 2, instead, the user is available to do something with the chatbot, performing step-by-step a series of behaviors on the company website, which the chatbot is telling them to do, thus carrying out a task together with the agent.

Example 2

U: Hi I wanted to ask about not sending paper invoice <phoneNumber>.

C: To eliminate the paper invoice you need to activate CONTO ONLINE [...] **If you want, we can do it TOGETHER or I can give you a GUIDE to be read whenever you want. Which option do you prefer?**

U: **Let's do it together**

C: Ok! To change the way you receive your bill, go into the personal area on the website and select LINE DATA REVIEW in My Line section... then scroll down to Account Receipt Mode, select EDIT [...]

Let me know when you have done that so I can explain the next step....

U: **Ok**

C: Okay! Now to activate the Online Account scroll down to the question "Do you want to join the Online Account option..." and select YES... [...]

Is there anything else I can do for you?

U: No thank you bye.

Of course, the user may not find a response suitable for their needs among those that the chatbot proposes to them and may not select any of the predefined options. However, they can still be cooperative if they make an alternative request and are available to guide the chatbot to another topic.

Politeness. The second behavioral strategy is *politeness* and consists in either a favorable behavioral disposition towards the chatbot and the interaction with it, which entails courtesy and patience,

or in the absence of any "negative behavior" that may undermine the progress of the conversation (like aggressiveness signaled e.g., by profanity words). In example 3, the user greets the agent and is polite in their request ("*Good morning*" "*could you kindly tell me", "yes please", "forgive me*") even though this is not necessary, and does not lose patience, even when the chatbot makes the same request multiple times (i.e., "*On the <phoneNumber> line there are movements* [. . .]"). This strategy, therefore, implies that the conversation will continue despite the difficulties that the user may encounter maintaining "manners." Mostly, users perform this strategy from the very beginning of the conversation (as in example 3) and maintain it even when a misunderstanding occurs or the goal is not achieved, exiting the conversation by greeting the chatbot as if it were a human being.

Example 3

U: **Good morning** I topped up my credit 15 euros plus 4 euros last week how come the credit is insufficient **could you kindly tell me** the cost of the promotion plus Tim Vision [name of the service]

C: [. . .] On the <phoneNumber> line there are movements for the following categories: 1. Recharges: 19.00€ 2. Activations and Renewals: -14.81€

Would you like to know the details of the movements?

U: **Yes kindly**

C: Choose one of the options indicated, selecting the position in the list (1,2,3...).

U: The promotion is almost 13 euros tim vision [name of the service] 2.50

C: On the <phoneNumber> line there are movements for the following categories: 1. Recharges: € 19.00 2. Activations and Renewals: -€ 14.81

Do you want to know the details of the movements?

U: **Yes please**

C: Choose one of the options shown, selecting the position in the list (1,2,3...).

U: **Forgive me** I don't know what the numbers correspond to

Behavioral non-collaboration. Behavioral non-collaboration happens when the user refuses to perform a specific behavior requested by the chatbot, as happens in example 4, where they ignore the chatbot's requests.

Example 4

C: Help me better identify your need related to Phone Directories: 1. Did you not receive the directory but were you charged the delivery fees? 2. Do you no longer want to receive Telephone Directories and related charges in the future? [. . .] **Choose one of the options shown, selecting the position in the list (1,2,3...).**

U: **I paid but do not want this service**

C: [. . .]

The checks I did show that the line is active, but are you experiencing difficulty in making calls or in surfing the Internet? **Please write YES or NO below.**

U: **Take away the expense of phone book delivery.**

Alternatively, the user may adopt an aggressive, non-polite, behavioral disposition, as in example 5, which ends with the abandonment of the conversation.

Example 5

C: [. . .] Which of these needs are you referring to: 1. Not recognized invoice charges [. . .] Choose one of the options listed, selecting the position in the list (1,2,3...).

U: **Suck it up**

C: Do you want to change the subject?

U: **I want to kick your ass**

[. . .]

C: I can explain to you HOW to APPLY for home banking or [. . .] I can give you information on the STATUS of your request... Which do you prefer?

U: **You are completely useless [the user abandons the conversation]**

4.1.2 Conversational collaboration. Conversational collaboration consists in adapting the language to the chatbot's capabilities of comprehension and points to a wide range of *conversational strategies*, where users modify the terms in different ways, the syntactical structure, and the content of their sentences. In the following, we list the conversational strategies that we identified in our corpus, also highlighting that users may refuse to adapt their language to the chatbot's abilities engaging in conversational non-collaboration.

Abstraction. A first conversational strategy relates to the *abstraction* of the content. Often users start the conversation providing too many details about the problem that they aim to solve, but such explanation cannot be handled by the chatbot. In fact, the agent may have difficulties in identifying the key words of the conversation, and thus the user intent, if there is an excess of information from the user side. Then, the user may try to state the key concept characterizing their problem: they may repeat the issue while omitting all the details that might not be relevant. This strategy may entail the summarization of a large portion of "text" into a single sentence, whereby the concrete recounting of a given situation (e.g., the spatial-temporal context, the origin of the problem and the previous attempts to solve it) is substituted with the affirmation of a more general problem (example 6).

Example 6

U: I have to report that for about 2 days the internet has not been working, the line comes and goes, as a similar problem also occurred a few months ago and after several days of various interventions and malfunctioning I was assured that the fault had been permanently fixed

C: [. . .] The checks I have made show that the line is active, but are you experiencing difficulty making calls or surfing the Internet? [. . .]

U: **I am having difficulty in browsing because the line comes and goes**

Simplification. A second strategy is *simplification*, which, rather than intervening on the content of the conversation is aimed to adapt its form. This strategy is usually performed when a sentence is not understood by the chatbot due to its syntactical complexity. Often, especially at the first contact with the chatbot, the user uses sentences containing a variety of conjunctions. When such sentences are not recognized by the chatbot, the user simplifies the syntactical structure, turning, for instance, a sentence with subordinate clauses into a single principal clause composed only of a subject, a verb, and an object. The user may also modify the tense of the used verbs, as if present tense could be more understandable than more "complex" tenses, like conditional and past tenses. Alternatively, they may adopt a minimal syntactical structure consisting in a command-style language with the presence of a mere verb in infinitive or imperative forms and an object (Example 7).

Example 7

U: In which area of the app can I deactivate Tim vision [name of the service]?

C: With TIMVISION [name of the service], you always have exclusive TV series [. . .].

Can I give you more information?

U: **deactivate Timvision [name of the service]**

Sometimes, however, users enter the conversation by immediately adopting a hyper-simplified language, using from the very beginning of the interaction both *abstraction* and *simplification*, keeping them for the whole interaction. This shows that *simplification* and *abstraction* may be used as "global collaborative strategies," and not only as a response to obstacles in the conversation.

Reformulation. Reformulation is mainly addressed to change certain parts of a given sentence, not necessarily in the direction of greater simplification. This strategy is performed almost exclusively in response to a misunderstanding from the chatbot of what the user has said, and mostly proceeds through trial and error, which may require multiple attempts. By using this strategy, the user modifies certain nouns, verbs, or abbreviations, most often with a synonym, or replaces an abbreviation or acronym with its extended form, like in example 8, where the user replaces the Italian abbreviation for bank account (i.e., c/c) with the proper extended noun.

Example 8

U: Good evening I have debit on my c/c. I wanted to change the c/c. How to do Thank you

C: Sorry, I didn't quite understand your request.... Can you repeat it in a simpler way? [. . .]

U: **I want to change my direct debit account.**

Complexification. Another strategy that is not tied to the idea of simplifying the conversation is *complexification*. This strategy is utilized when the user employs a language that is too minimal for the chatbot to be understood (usually only one or two words), lacking some fundamental parts of a sentence that are needed to infer the user intent. For instance, the users may use both abstraction and simplification strategies from the outset. However, in certain situations, these strategies produce sentences that are too minimal, thus requiring the complexification strategy as a balance: then, the user tries to add "more structure," usually the subject, the verb, and some details. In example 9, for instance, the user substitutes the request "*new iban*" with a complete sentence that specifies exactly what they need.

Example 9

U: **New iban**

C: Sorry, I didn't quite understand your request.... Can you repeat it in a simpler way?

U: **I want to change the direct debit account**

Correction. Correction refers to the rectification of typographical errors (typos) and grammatical errors made during the conversation. Most users make a lot of typos during the interaction. They may make writing errors even when they are trying to employ other conversational strategies (e.g., they simplify an utterance, but the new sentence still contains a typo). However, certain users show a great accuracy in their writing style from the very beginning of the interaction, carefully avoiding any grammatical mistakes or typos, or being available to correct them by repeating the sentence without the error (as in example 10, where a complexification strategy is also enacted).

Example 10

C: Ok! To change how you receive your bill, go into the personal area on the website and select LINE DATA [. . .]

U: **on ths [rather than *this*] app**

C: Sorry, I didn't quite understand your request.... Can you repeat it in a simpler way?

U: **I need to go to this app**

Emphasis. Finally, there are two strategies that are borderline between collaboration and non-collaboration, as they may signal either the willingness or the unwillingness to make the conversation proceed. The first strategy of this kind is *emphasis* and consists in capitalizing a certain word or group of words to draw the chatbot's attention to that specific word (example 11). However, in addition to not being effective, sometimes emphasis may signal the raising of the user's irritation just before leaving the conversation.

Example 11

U: I would like TIM Base new

C: Sorry, I didn't quite understand your request.... Can you repeat it in a simpler way? [. . .]

U: **I want to get TIM BASE NEW**

Repetition. Repetition is a strategy in which the user repeats word by word a sentence to be better understood the second time. This strategy was not so common in our corpus, being used in a minority of conversations. While sometimes this may work, in most occasions it is completely ineffective. However, in many cases the user keeps repeating the same sentence, without opting for any

other strategy, which may result in the failing of the conversation (example 12).

Example 12

U: I'd like to know what happened to the cams you were supposed to send me

C: Do you need help on mobile or fixed lines?

U: **I want the cams**

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

[...]

U: **I want the cams**

Conversational non-collaboration. Conversational non-collaboration refers to all those occasions when the user does not adapt their language to the chatbot's capabilities of comprehension. This relates to the usage of complex language from both the syntactical and semantical points of view and the unavailability, or incapability, to reformulate the sentences in a way that is understandable by the agent, even after its explicit request.

Example 13

U: I wanted to report, even after several calls, that with the new Tim super fiber [name of the service] you deactivated timvision plus [name of the service] and activated voice that I was not interested in and your salesperson had not informed me. **One of your operators made me activate timvision plus for a fee in the meantime you were making the change from voice and timvision plus however several months have passed since the first report and no one has contacted me to tell me anything**. I am very disappointed about the deception that I suffered from and the lack of information from your salesman at the time he explains the details of the offereded [orthographical error] and after you stipulate the new contract you find that they have deactivated services that you had NEVER asked to deactivate. I ask you to put timvision plus back up and deactivate voice. [...]

In example 13, for instance, the user recounts a complex "story" that the chatbot is not capable of understanding, while in example 14 the user explicitly refuses to reformulate their sentence.

Example 14

U: App. 5.4.4_130

C: Sorry, I didn't quite understand your request? Can you repeat it in a simpler way? [...]

U: You asked me about the version of the app. My-tim [name of the app] which is 5.4.4_130. I have no assistance, mobile data disappeared, on the fixed line you display as unpaid what I paid instead, in short a.chaos. It is a wonder, fix it as soon as possible with a real check

C: Sorry, I didn't quite understand your request.... Can you repeat it in a simpler way? [...]

U: **Just read my words**

4.2 Collaboration dynamics

How users collaborate with the chatbot during the interaction is often not static. It is true that several users may keep the collaboration at the same "level" for the whole conversation, e.g., by keeping the behavioral collaboration high from the beginning to the end of the interaction or by refusing to cooperate with the chatbot at all. However, in most of the interactions that we analyzed, collaboration undergoes *evolution* (from less cooperation to more cooperation) or *devolution* (from more cooperation to less cooperation) dynamics.

4.2.1 Evolution dynamic. Evolution shows increasing collaboration effort as the interaction progresses. Commonly, users enter the conversation by describing complex problems with a variety of details that cannot be handled by the agent (like in example 13). Then, they progressively adapt their language to the chatbot's request, increasing the collaborative effort and distancing from their "natural" way of expressing (as in example 15).

Example 15

U: Hello I have made the request for domiciliation, but they still haven't let me know if the request has been accepted

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

[...]

U: **Info on the direct debit request**

C: I can explain HOW to apply for a direct debit or, [...] I can give you information on the STATUS of your application...

Which do you prefer?

U: **Request status**

Of course, this evolution may not always immediately lead to success, and the user may be required to attempt different strategies to find the correct "formula" for achieving their goal, putting increasing effort in collaboration. In example 16, the user first adopts a *complexification* strategy by adding more details to their request. Then, when the chatbot still appears not to understand what they are saying, offering them a series of off-topic options, the user tries to return to the original topic by adding other details. They then implement a *reformulation* strategy substituting the initial verb "see" with the verb "receive." Finally, they accept the option proposed by the chatbot through a *commitment* strategy and radically simplify their responses. They achieve this by providing yes/no answers and avoiding the introduction of new information or requests anticipating the conversation flow (e.g., they do not respond to the chatbot's question "*Is the charge correct?*" with a supplementary request, like "*No, cancel the payment*," but they wait for the chatbot to offer them an option on how to solve the problem).

Example 16

U: I don't see the invoice for January

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

[...]

U: **I do not see my January invoice when accessing the app**

C: Within the app you can view all the lines linked to the Account

[. . .]

Choose the topic that interests you: 1.Mobile 2.Fixed line. Choose one of the options indicated, selecting the position in the list (1,2,3...).

U: **I didn't get the answer, in the app I see all the invoices minus the one for January**

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

[. . .]

U: **I have not received my January invoice**

[. . .]

C: For which invoice do you want more details? Enter the NAME of the month of the invoice you wish to view.

U: **January**

C: Here are the details of the expense items [. . .]

Would you like to have the details of a specific expense item on your bill?

Enter the name of the expense item or the position in the list (1,2,3...).

U: **Yes**

U: **1**

C: Here are the details requested for the January [. . .] Is the charge correct?

U: No

In our corpus, these chains of attempts can be very short, even one-shot attempts. However, in more than two-thirds of the conversations users continued to collaborate with the chatbot despite the occurrence of misunderstandings, utilizing more than one strategy to "work with" the agent: this happens more often when the user maintains a favorable behavioral disposition towards the interaction despite the difficulties (i.e., *politeness*), trying novel solutions while avoiding losing patience, indulging in aggressive behaviors, or abruptly leaving the conversation. Moreover, we observed that users are more likely to interrupt the collaboration effort if they see that multiple attempts fail in a row without any evident progress.

4.2.2 Devolution dynamic. Devolution may occur when the user starts collaborating with the chatbot and maintains the effort for a while, but after some exchanges progressively loses the willingness to cooperate. As a consequence, they may provide out-of-context comments or complain about the chatbot's performance, show frustration, or abandon the conversation. In example 17, for instance, the user is at first completely committed to choosing among the options proposed by the chatbot and tries to adopt a language that is easily understandable by the agent. However, they soon start complaining about the chatbot's answers, finally abandoning the conversation.

Example 17

U: Good morning, on 22 January, I requested by phone the activation of [name of the service] combined with [name of the service]. Is it possible to know the status of the request?

C: You can check the status of your order, or request a change and/or cancellation:

1. Check your order 2. Change/cancel your order. Choose one of the options indicated, selecting the position in the list (1,2).

U: 1

[. . .]

C: [. . .] Do you want me to put you in touch with an operator?

U: Maybe better

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

[. . .]

U: I want to know the status of [name of the service] activation file

C: You can check the status of your order, or request a change and/or cancellation:

1. Check your order 2. Change/cancel your order Choose one of the options indicated, selecting the position in the list (1,2).

U: 1

C: You can check the status of your orders in the tracking section of the app [. . .]. Remember that you can make other requests at any time

U: **If these are the answers, there's no point**

C: Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?

U: **Goodnight**

In sum, collaboration with chatbots may go through fluctuations. In fact, we even observed conversations where a dynamic of *evolution* is followed by one of *devolution* and then again of *evolution*, showing that the collaboration effort can always be reinstated until the user leaves the conversation.

5 DISCUSSION

The major contribution of this paper is to provide a detailed picture of how people collaborate with a customer care chatbot in the real world. Previous research [3, 45, 69, 95], instead, mainly adopted an experimental approach where people did not pursue their situated goals. Such research also focused on non-progresses in interaction, thus framing the eventual collaborative behavior as a coping strategy of repair from conversation breakdowns [52]. Instead, our focus on collaboration allowed us to capture strategies that are not implemented in response to the chatbot's misunderstandings but are performed simply to "work together" with the agent, thus complementing research on the topic conducted so far. This different focus also explains some differences between the strategies we identified and those highlighted by these previous studies.

For instance, the coping strategies highlighted by Li et al. [52] indicate reactive behaviors performed in response to a negative

event occurring in the conversation. Therefore, they do not include what we called *behavioral strategies*, which refer to cooperative behaviors utilized even in absence of communication breakdowns. Likewise, we do not tackle quitting strategies, as Li et al. do, since quitting shows the willingness to no longer cooperate with the agent, thus being a non-collaborative strategy. Moreover, certain strategies that may appear similar in the two studies have different nuances. For example, Li et al.'s "remove words" strategy is similar to our strategy of *simplification*: however, while in Li et al.'s remove words was only used to cope with non-progresses, we noticed that *simplification* can be utilized even from the very beginning of the conversation, so that users may see it as a general tool for collaboration and not merely as a means for repair.

On the other hand, similar to Li et al. [52], and differently from Stent et al. [88] and Myers et al. [69], we found that *repetition* was a strategy used less frequently than the others. As Li et al. suggest this may be due to the peculiar way of interaction required by the chatbot, that is through written language: chatbot users may be less likely to repeat themselves than their vocal assistant users, because the former can immediately confirm, via direct visual inspection, that their own input is not garbled, whereas vocal interface users might not know whether the errors arise from its incorrect speech recognition or from incorrect conversational understanding [54]. However, with respect to these studies, we found that *repetition* may also indicate *conversational non-collaboration*, becoming a means to express disappointment or a sign that the user is starting to become less cooperative.

It is worth noting that the repertoire of strategies employed by our users may have been acquired during their past experiences with chatbots: some users may hold "competence" to deal with difficulties in human-chatbot interaction and this may explain the fact that they used certain strategies, like *abstraction* and *simplification*, from the very beginning of the interaction. Følstad and Skjuve [28] noticed that people who had past real-world experiences with chatbots may have realistic understandings about their actual capabilities, which may lead them to formulate simple enquiries. Alternatively, our users may have used strategies learned outside human-chatbot interaction. *Simplification* and *reformulation*, for instance, are strategies that are commonly performed when people talk to interlocutors who are believed to have low language skills, such as children [21] and foreign speakers [23], so that they may have been replicated with our chatbot, which has limited capabilities of understanding.

Furthermore, some strategies may have emerged in the ongoing conversations with the agent, thus depending on its specific design features. For instance, several users may have reduced the sentences at the bare minimum because on several occasions the chatbot asks them to respond with a single word (e.g., yes/no) or number (by selecting a predefined option). Likewise, *emphasis* may have been inspired by the fact that the chatbot, in certain cases, uses the capitalization of words when explaining how to conduct certain tasks to the user. This points to a linguistic phenomenon known as lexical and structural alignment [92] - a tendency to converge on the same lexical terms and sentence structure during a dialogue. Research has noticed that individuals show alignment when they interact with artificial agents [7, 8, 19, 87]. In this sense, our study confirms such previous research.

Other differences with previous studies may also be due to the specificity of our chatbot and of the context in which it is deployed. Research indicates that some users quickly abandon the interaction when they feel that the agents' behavior does not meet their expectations [44, 54, 100]. Likewise, Li et al. [52] emphasize that "message reformulation," a term that in their work indicates different strategies, like removing or adding words, rephrasing, and correcting wrong words, was used infrequently by their users, who more often decided to abandon the interaction: in that context, as the authors notice, users could likely find alternative means for achieving the same goals, like websites and customer support. Instead, we found that most users persisted in collaborating with our chatbot and often rephrased their messages. On the one hand, this may be due to the way our chatbot's dialogues are designed: when the user provides syntactically too-long/too-complex sentences, the chatbot invites them to reformulate their request in simpler terms. This may explain the higher usage of *simplification* and *reformulation* in our corpus: however, it is worthy of being noticed that while our chatbot offers always the same vague suggestion on how to rephrase a message ("*Sorry, I didn't quite understand your request... Can you repeat it in a simpler way?*"), our users engaged in a variety of conversational strategies, like *complexification*, *correction*, and *abstraction*, which were not even mentioned by the agent.

On the other hand, the persistence in pursuing collaboration may be due to our users' specific needs, as well as to the company's capabilities of satisfying them. In fact, our users may have pressing problems, like subscriptions to unwanted fee-based services, which may make them lose money, or technical issues like non-functioning Internet connection, which may negatively impact on their work and leisure time. The company's alternative customer service channels may not always be responsive, as example 13 shows, so users might believe that trying to fix such problems through the chatbot, which is always available, might be worth the time and the effort spent. This is in line with Følstad and Skjuve's research [28], who found that users see in chatbots an accessible support and a valid alternative to company website or contacting human customer service for enquiries for which they expect a straightforward answer.

It remains to say that the usage of *politeness* might be retraced to the users' projections of humanness onto the chatbot: they may have applied to human-chatbot interaction patterns that are usually employed with humans. Previous research noticed that it is sufficient that the chatbot offers minimal human-like cues, like a human name [2, 34] or the usage of emojis [4] to elicit perceptions of humanness, and that users may still perceive as a human a chatbot that presents itself as a machine [84]. The notion of partner models, which has been applied to the interaction with speech interfaces [25], further states that people enter dialogue with assumptions about their interlocutors and that these drive choices in conversation [16]. Even though our chatbot presents itself as a virtual assistant, certain cues of its design, like its name, the way it greets the user, or is polite with them, might have made the users perceive it and/or behave as if it were a human. This would confirm previous research noticing that systems using anthropomorphic dialogue strategies encourage increased levels of politeness [9]. Likewise, Corti and Gillespie [18] found users invested more effort to repair misunderstandings when the chatbot is perceived as human, which

may be in line with the effort spent by our users in collaborating with the chatbot.

In sum, this study confirms certain findings emphasized by previous research, while pointing out others, thus complementing what has been previously found: some of these differences may be due to the focus on collaboration that we adopted, while others to the specificity of the chatbot and the context in which it is deployed. If, on the one hand, this limits the generalizability of our findings, on the other hand, it offers a detailed account of how users collaborate with an agent within a specific real context of use. In the next section, we recount a few design implications emerging from our findings.

6 IMPLICATIONS FOR DESIGN

6.1 Create a friendly and polite conversational environment and sustain the user's willingness to collaborate with the chatbot

We have found that in the customer care domain users may be willing to collaborate with chatbots to achieve their own situated goals. We have also seen that users using *politeness* dedicate more effort towards establishing a "positive conversational environment" despite the difficulties, which may create optimal conditions for effective collaboration to occur, favoring the use of multiple collaborative strategies (Section 4.2.1). Instead, when politeness is not employed by users, the conversation may quickly deteriorate, as users' expressions of anger and frustration can be difficult for the chatbot to handle (e.g., example 5): this may cause the user to abandon the conversation. A design opportunity is, therefore, to focus on designing chatbots able to create a friendly and polite conversational environment [11], which in turn will encourage users to do the same. Although our chatbot is designed to be polite and friendly, the range of behaviors that it can express is very limited: the chatbot mostly greets and apologizes. Greetings and apologies are often employed in previous research [e.g., 44] but there is a lack of studies on how to leverage the politeness and friendliness strategies used in human-human social interactions [11]. Addressing the user by name, establishing a more personal conversation, responding properly to the polite manners of the user, employing social niceties, increasing the human likeness of the agent, and even engaging in brief sessions of small talk could be explored in customer care chatbot designs, even though these design strategies are not meant to increase productivity, which is the main goal in this domain [5, 27]. Such design strategies could contribute to creating a more comfortable conversational environment and preventing the negative feelings that may cause users to abandon the interaction [44]. In fact, one interlocutor's manners can potentially reduce the other interlocutor's feeling of annoyance and frustration [11], and also encourage them to utilize *politeness* as well. This is because people in communication tend to respond to the politeness of others by behaving in the same way and using the same types of politeness strategies [22]. However, small talk in particular should be used wisely, as off-topic conversations may lead task-focused chatbots to fail [11]. Moreover, it is important to sustain the user's willingness to collaborate during the "turning points" of the conversation, where *evolution* may turn into *devolution*: early detection of such points may increase the chances that they won't abandon the conversation. In this perspective, cues could be e.g., the usage of *emphasis* and *repetition*, which may also signal non-collaboration. This would also mean designing with the aim of mitigating the user's rising irritation by for example, expressing empathy [53, 79]. For instance, in example 17, when the user starts complaining about the impossibility of fixing their problem, the chatbot, instead of saying "Sorry," could have expressed empathy, saying that it understands that its limited capabilities of comprehension can be frustrating, and that it too would be frustrated in its place. However, if the user is willing to rephrase their problem one last time with a little patience, the chatbot will try to understand it again and, if unsuccessful, refer it to a human operator.

6.2 Guide the user in adopting the most opportune collaboration strategy

We have seen that users may implement different collaborative strategies, either behavioral or conversational, at a given point of the conversation. Sometimes, they need to proceed through trial and error to find the right way to collaborate with the chatbot. The chatbot that we analyzed always suggests that the user reformulates their sentence in simpler terms providing one-size-fits-all examples. However, on many occasions this instruction may be misleading because the user should apply *complexification* rather than *simplification*. This implication highlights the need to correctly direct the user towards the most useful kind of collaboration at the specific point of the interaction. Previous research suggested that the chatbot could provide both examples and explanations of rules on how to interact with the agent to guide the user during the conversation [99]. Here, instead, we recommend that the chatbot takes the user's sentence and highlights what is "wrong" in it (e.g., too complex syntax, too few words for identifying the topic, too many details provided), then suggesting a specific conversational strategy to apply with ad hoc examples. In this sense, our findings may be useful to identify common strategies employed by users that could be then recommended by the agent. In example 9, for instance, the chatbot, instead of inviting the user to repeat the sentence in simpler worlds, could have recognized that they have written only two words (so further reducing them would not be desirable); then, it could have signaled to the user that some structure and more details are needed (e.g., a verb that specifies the action required from the agent). Greater transparency in contextually (i.e., at a specific point of the conversation) explaining to the user how the chatbot works and what specific terms, syntactical structures, or expressions it is unable to understand, as well as its limitations in language processing, would also make it easier for the user to understand how to correct their sentences. This would also help users in building a partner model [25] that better matches the chatbot's actual capabilities.

6.3 Make the user perceive that the collaboration is working even though the conversation is apparently not progressing

In our corpus, we have seen that users are more likely to interrupt the collaboration if they see that multiple attempts fail in a row, without any apparent advancement (Section 4.2.1). It is important, therefore, to emphasize that, despite the fact that single attempts

may not work, the collaboration between the user and the chatbot is ongoing and is making progress, despite repeated failures. This suggests that the chatbot avoids repeating the same sentences, but always builds on what the user has immediately done or said before, highlighting progress and opportunities for advancing. For instance, in example 11, rather than simply stating "*Sorry, I didn't quite understand your request....*" the chatbot could have responded: "*Sorry,* ***I see that you are mentioning the promotion TIM Base new****, but I cannot understand what you want.*" This would make the user realize that some parts were understood and that progress is possible. Moreover, reverting to earlier stages of the conversation should only be done at the explicit confirmation or request of the user (e.g., "*Can we go back to the topic xxx*"?): otherwise, the sense of stasis would likely be increased. For instance, in example 17 the chatbot, almost at the end of the conversation, prompts the same sentence that it prompted at the beginning of the interaction ("*You can check the status of your order, or request a change and/or cancellation:* [. . .]"): this might make it seem to the user that really no progress has been made during the whole conversation. Showing that the chatbot is aligning with the user's language could further strengthen the perception that it is also striving towards making the conversation proceed: previous research has shown that users align more strongly when they have reason to believe that the conversational agent is also aligning to them [87].

7 LIMITATIONS

The chatbot we employed has limitations in its Natural Language Processing capabilities and the "responsibility" for collaboration was mostly in the hands of the user. The chatbot was able to ask for the reformulation of a sentence, to offer alternative options to which the user could adhere, and to propose performing a task together following a predetermined task flow. However, this is quite far from an "intelligent" agent really capable of utilizing different cooperative strategies according to the situation. Despite this limitation, the users showed that they were not only able of performing collaborative strategies but that they also put effort in it. This may signal that the customer care domain is a fertile ground for leveraging the users' disposition towards cooperating and "working together" with an artificial agent.

A second limitation is that it is questionable that our findings would be generalizable across different cultures, given that our analysis is limited to the Italian population. Moreover, we only examined the conversations with a single chatbot, so other agents with different abilities or being used within different contexts (e.g., a customer care chatbot of an online marketplace) could lead to different results. In fact, we have seen that certain peculiarities in our findings with reference to previous research may be due to the specific needs that our users had, as well as to the chatbot's design features and the company's customer service characteristics, which may limit our findings' generalizability. However, as for all qualitative research the value of our analysis lies mainly in providing a detailed account of how people interact within a specific context to satisfy their specific needs. To increase the generalizability of the study findings, similarly conversation logs from other chatbot services should be analyzed.

Finally, the analyzed data did not include any user reflections on their cooperative behavior, only the conversations themselves. For this, it is possible that the inferences we made on certain conversational behaviors might not correspond to what the user experienced during that interaction. Different interpretations of their behavioral traces are possible since there is no data from the users on how they intended their language use. However, we think that the value of our analysis is precisely to address real-world conversations rather than e.g., ex-post users' reports on their behaviors. Future work could then compare the analysis of conversations with the users' opinions about such conversations to identify correspondences and misalignments.

8 CONCLUSION

In this paper, we offered a detailed recount of how people collaborate with a task-focused chatbot in the telecommunication domains. We identified two main aspects of collaboration, *behavioral* and *conversational*, and for each aspect we highlighted the different strategies that users utilize to "work together" with the agent. The strategies identified span from user *commitment* to acceptance of the chatbot's proposals and their willingness to behave "favorably" towards the chatbot, to language *complexification* and *simplification*, as well as *abstraction*, *reformulation*, and *correction* of their sentences. Collaboration may also show dynamics of *evolution* and *devolution*, from collaboration to non-collaboration and vice versa. On the basis of these findings, we identified differences and similarities with previous research and proposed several design suggestions that aim to sustain the user's willingness to collaborate with the agent, drive the user to adopt the most opportune collaborative strategy, and make them perceive that collaboration is progressing, despite possible failures.

ACKNOWLEDGMENTS

The research was partially funded by TIM S.p.A. within the project "Studi e ricerche su sistemi conversazionali intelligenti" agreement n. S19AARLL.

REFERENCES

[1] Alper T. Alan, Enrico Costanza, Sarvapali D. Ramchurn, Joel Fischer, Tom Rodden, and Nicholas R. Jennings. 2016. Tariff Agent: Interacting with a Future Smart Energy System at Home. ACM Trans. Comput.-Hum. Interact. 23, 4, Article 25 (September 2016), 28 pages. https://doi.org/10.1145/2943770

[2] Theo Araujo. 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. doi: 10.1016/j.chb.2018.03.051

[3] Zahra Ashktorab, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 254:1–254:12. https://doi.org/10.1145/3290605.3300484

[4] Austin Beattie, Autumn P. Edwards, Chad Edwards. 2020. A Bot and a Smile: Interpersonal Impressions of Chatbots and Humans Using Emoji in Computer-mediated Communication. Communication Studies, 71(3), 409–427. doi: 10.1080/10510974.2020.1725082.

[5] Petter Bae Brandtzaeg and Asbjørn Følstad. 2018. Chatbots: changing user needs and motivations. interactions 25, 5 (September-October 2018), 38–43. https://doi.org/10.1145/3236669

[6] Jenine Beekhuyzen, Liisa von Hellens, and Sue Nielsen. 2015. Illuminating the underground: the reality of unauthorised file sharing. Information Systems Journal 25, 3 (2015), 171–192. https://doi.org/10.1111/isj.12069

[7] Kirsten Bergmann, Holly P. Branigan, and Stefan Kopp. 2015. Exploring the Alignment Space: Lexical and Gestural Alignment with Real and Virtual Humans. Frontiers in ICT 2 (2015). https://doi.org/10.3389/fict.2015.00007

[8] Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. Journal of Pragmatics 42, 9 (2010), 2355–2368. https://doi.org/10.1016/j.pragma.2009.12.012 [

[9] Susan E. Brennan and Justina O. Ohaeri. 1994. Effects of message style on users' attributions toward agents. In Conference companion on Human factors in computing systems - CHI '94 (Boston, Massachusetts, United States). ACM Press, 281–282. https://doi.org/10.1145/259963.260492

[10] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected inhuman subjects research on the Internet. Ethics and Inf. Technol. 4, 3 (2002), 217–231. https://doi.org/10.1023/A:1021316409277

[11] Ana P. Chaves and Marco A. Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. International Journal of Human–Computer Interaction 37, 8, 729-758. https://doi.org/10.1080/10447318.2020.1841438

[12] EunJeong Cheon and Norman Makoto Su. 2016. Integrating roboticist values into a Value Sensitive Design framework for humanoid robots. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 375-382. https://doi.org/10.1109/HRI.2016.7451775

[13] Nazli Cila. 2022. Designing Human-Agent Collaborations: Commitment, responsiveness, and support. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 420, 1–18. https://doi.org/10.1145/3491102.3517500

[14] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In 23rd International Conference on Intelligent User Interfaces (IUI' 18). Association for Computing Machinery, New York, NY, USA, 329–340. http://dx.doi.org/10.1145/3172944.3172983

[15] Herbert H. Clark. 1996. Using language. New York, NY: Cambridge University Press.

[16] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on Socially Shared Cognition (127-149). Washington, D.C.: American Psychological Association

[17] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, Benjamin R Cowan, The State of Speech in HCI: Trends, Themes and Challenges, Interacting with Computers, Volume 31, Issue 4, June 2019, Pages 349–371. https://doi.org/10.1093/iwc/iwz016

[18] 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.

[19] Benjamin R. Cowan, Philip Doyle, Justin Edwards, Diego Garaialde, Ali Hayes-Brady, Holly P. Branigan, João Cabral, and Leigh Clark. 2019. What's in an accent?: the impact of accented synthetic speech on lexical choice in human-machine dialogue. In Proceedings of the 1st International Conference on Conversational User Interfaces - CUI '19 (Dublin, Ireland). ACM Press, 1–8. https://doi.org/10.1145/3342775.3342786

[20] Cammy Crolic, Felipe Thomaz, Rhonda Hadi, and Andrew T. Stephen. 2022. Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions. Journal of Marketing 86, 1 (January 2022), 132–148. https://doi.org/10.1177/00222429211045687

[21] Toni G. Cross. 1977. Mothers' speech adjustments: The contribution of selected child listener variables. In Catherine E. Snow & Charles A. Ferguson (eds.), Talking to Children. Cambridge University Press. pp. 151—188

[22] Culpeper, J., & Tantucci, V. (2021). The principle of (im) politeness reciprocity. Journal of Pragmatics, 175, 146-164.

[23] John Paul O. Dela Rosa, & Diana C. Arguelles. 2016. Do Modification and Interaction Work? - A Critical Review of Literature on the Role of Foreigner Talk in Second Language Acquisition. Journal on English Language Teaching, 6(3), 46-60.

[24] Dominik Dellermann, A. Calma, Nikolaus Lipusch, Thorsten Weber, Sascha Weigel, and P. Ebel. 2019. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In HICSS. https://doi.org/10.24251/HICSS.2019.034

[25] Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 244, pp. 1–14. https://doi.org/10.1145/3411764.3445206

[26] Charles A. Ferguson. 1975. Toward a characterization of English foreigner talk. Anthropological Linguistics 17, 1, 1–14.

[27] Asbjørn Følstad, Cecilie Bertinussen Nordheim & Cato Alexander Bjørkli. 2018. What makes users trust a chatbot for customer service? An exploratory interview study. In: Proceedings of the International Conference on Internet Science. Springer, Cham, pp. 194–208. https:// doi.org/10.1007/978-3-030-01437-7_16.

[28] Asbjørn Følstad and Marita Skjuve. 2019. Chatbots for customer service: user experience and motivation. In Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 1, 1–9. https://doi.org/10.1145/3342775.3342784

[29] Forrester Research. (2016). Chatbots 101: Building Conversational Interfaces. Technical report. Accessed October 15, 2020.

[30] Dominic Furniss, Ann Blandford, and Paul Curzon. 2011. Confessions from a grounded theory PhD: experiences and lessons learnt. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). Association for Computing Machinery, New York, NY, USA, 113–122. https://doi.org/10.1145/1978942.1978960

[31] Susan Gasson. 2009. Employing a grounded theory approach for MIS research. In Handbook of Research on Contemporary Theoretical Models in Information Systems, Yogesh K. Dwivedi, Banita Lal, Michael D. Williams, Scott L. Schneberger and Michael Wade (Eds.). IGI Global, Hershey, PA, pp. 34–56.

[32] Howard Giles, Anthony Mulac, James J. Bradac, and Patricia Johnson. 1987. Speech Accommodation Theory: The First Decade and Beyond. Annals of the International Communication Association 10, 1 (January 1987), 13–48. https://doi.org/10.1080/23808985.1987.11678638

[33] Barney G. Glaser, and Anselm L. Strauss. 2017. The Discovery of Grounded Theory Strategies for Qualitative Research. Routledge, New York. https://doi.org/10.4324/9780203793206

[34] Eun Go and S. Shyam Sundar. 2019. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304–316. doi: 10.1016/j.chb.2019.01.020.

[35] Eberhard Graether and Florian Mueller. 2012. Joggobot: a flying robot as jogging companion. CHI '12 Extended Abstracts on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1063–1066. https://doi.org.tudelft.idm.oclc.org/10.1145/2212776.2212386

[36] Jonathan Grudin and Richard Jacques. 2019. Chatbots, Humbots, and the Quest for Artificial General Intelligence. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 209, 1–11. https://doi.org/10.1145/3290605.3300439

[37] Beth Harry, Keith M. Sturges, and Janette K. Klingner. 2005. Mapping the Process: An Exemplar of Process and Challenge in Grounded Theory Analysis. Educational Researcher 34, 2 (March 2005), 3–13. https://doi.org/10.3102/0013189X034002003

[38] Heiko Hausendorf and Uta M. Quasthoff. 1992. Patterns of adult-child interaction as a mechanism of discourse acquisition. Journal of Pragmatics 17, 3 (March 1992), 241–259. https://doi.org/10.1016/0378-2166(92)90004-U

[39] Karen Henwood and Nick Pidgeon. 2003. Grounded theory in psychological research. In Qualitative research in psychology: Expanding perspectives in methodology and design, Paul M. Camic, Jean E. Rhodes, and Lucy Yardley (Eds.). American Psychological Association, Washington, DC, US, 131–155. https://doi.org/10.1037/10595-008

[40] Jennifer Hill, W. Randolph Ford, and Ingrid G. Farreras. 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior 49, (August 2015), 245–250. https://doi.org/10.1016/j.chb.2015.02.026

[41] Roger Homan. 1980. The ethics of covert methods. The British Journal of Sociology 31, 1, 46–59. https://doi.org/10.2307/590062

[42] Huesmann, L. R., Eron, L. D., Guerra, N. G., & Crawshaw, V. B. (1994). Measuring children's aggression with teachers' predictions of peer nominations. Psychological Assessment, 6(4), 329.

[43] Insider Intelligence. 2021. Chatbot market in 2021: Stats, trends, and companies in the growing AI chatbot industry. Blog post, 8 February 2021. Available at https://www.businessinsider.com/chatbot-market-stats-trends?r=US&IR=T (Last access 2 February 2022)

[44] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18). Association for Computing Machinery, New York, NY, USA, 895–906. https://doi.org/10.1145/31Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18). Association for Computing Machinery, New York, NY, USA, 895–906. https://doi.org/10.1145/3196709.319673596709.3196735

[45] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13), 143–152. https://doi.org/10.1145/2484028.2484092

[46] Merel Keijsers, Hussain Kazmi, Friederike Eyssel, and Christoph Bartneck. 2021. Teaching Robots a Lesson: Determinants of Robot Punishment. Int J of Soc Robotics 13, 41–54. https://doi.org/10.1007/s12369-019-00608-w

[47] Stefan Kopp and Nicole Krämer. 2021. Revisiting Human-Agent Communication: The Importance of Joint Co-construction and Understanding Mental States. Frontiers in Psychology 12, (2021). Retrieved July 12, 2022 from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.580955

[48] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish R. Nambiappan, Kodur K. Chaitanya, Ashwin R. Babu, Joanne Mathew, and Fillia Makedon. 2021. A Survey of Robots in Healthcare. Technologies, 9, 1, 8. https://doi.org/10.3390/technologies9010008

[49] Guy Laban and Theo Araujo. 2020. Working Together with Conversational Agents: the Relationship of Perceived Cooperation with Service Performance Evaluations. 218–228. Retrieved July 12, 2022 from http://eprints.gla.ac.uk/208175/

[50] SeoYoung Lee, Junho Choi. 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

[51] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[52] Chi-Hsun Li, Su-Fang Yeh, Tang-Jie Chang, Meng-Hsuan Tsai, Ken Chen, and Yung-Ju Chang. 2020. A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376209

[53] Bingjie Liu, Shyam S. Sundar. 2018. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychology, Behavior, and Social Networking 21 (10), 625–636. https://doi.org/10.1089/cyber.2018.0110

[54] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 5286–5297. https://doi.org/10.1145/2858036.2858288

[55] Kien H. Ly, Ann-Marie Ly, and Gerhard Andersson. 2017. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet interventions 10, 39-46. https://doi.org/10.1016/j.invent.2017.10.002.

[56] Norman Makoto Su, Amanda Lazar, Jeffrey Bardzell, and Shaowen Bardzell. 2019. Of Dolls and Men: Anticipating Sexual Intimacy with Robots. ACM Trans. Comput.-Hum. Interact. 26, 3, Article 13 (June 2019), 35 pages. https://doi.org/10.1145/3301422

[57] Catherine Marshall and Gretchen B. Rossman. 2014. Designing qualitative research. Sage publications.

[58] Martin N. Marshall. 1996. Sampling for qualitative research, Family Practice, Volume 13, Issue 6, 522–526, https://doi.org/10.1093/fampra/13.6.522

[59] Gary T. Marx. 1998. Ethics for the New Surveillance. The Information Society 14, 3 (August 1998), 171–185. https://doi.org/10.1080/019722498128809

[60] Elisa Mattarelli, Fabiola Bertolotti, and Diego M. Macrì. 2013. The use of ethnography and grounded theory in the development of a management information system. European Journal of Information Systems 22, 1, 26–44. https://doi.org/10.1057/ejis.2011.34

[61] Joseph A. Maxwell. 2010. Using Numbers in Qualitative Research. Qualitative Inquiry, 16(6), 475–482. https://doi.org/10.1177/1077800410364740

[62] Jon McCormack, Toby Gifford, Patrick Hutchings, Maria Teresa Llano Rodriguez, Matthew Yee-King, and Mark d'Inverno. 2019. In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Paper 38, 1–11. https://doi.org.tudelft.idm.oclc.org/10.1145/3290605.3300268

[63] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 72 (November 2019), 23 pages. https://doi.org/10.1145/3359174

[64] Eleonora Mencarini, Amon Rapp, Lia Tirabeni, and Massimo Zancanaro. 2019. Designing Wearable Systems for Sports: A Review of Trends and Opportunities in Human–Computer Interaction. IEEE Transactions on Human-Machine Systems 49, 4 (August 2019), 314–325. https://doi.org/10.1109/THMS.2019.2919702

[65] Dylan Moore, Rebecca Currano, Michael Shanks, and David Sirkin. 2020. Defense Against the Dark Cars: Design Principles for Griefing of Autonomous Vehicles. Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, 201–209. https://doi.org/10.1145/3319502.3374796

[66] Robert J. Moore. 2018. A Natural Conversation Framework for Conversational UX Design. In Studies in Conversational UX Design, Robert J. Moore, Margaret H. Szymanski, Raphael Arar and Guang-Jie Ren (Eds.). Springer International Publishing, Cham, 181–204. https://doi.org/10.1007/978-3-319-95579-7_9

[67] Michael J. Muller, and S. Kogan. 2010. Grounded theory method in HCI and CSCW. In Technical Report. IBM Watson Research Center.

[68] Selma Musić and Sandra Hirche. 2017. Control sharing in human-robot team interaction. Annual Reviews in Control 44, (January 2017), 342–354. https://doi.org/10.1016/j.arcontrol.2017.09.017

[69] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 6, 1–7. https://doi.org/10.1145/3173574.3173580

[70] Clifford Nass, B. J. Fogg, and Youngme Moon. 1996. Can computers be teammates? International Journal of Human-Computer Studies 45, 6 (December 1996), 669–678. https://doi.org/10.1006/ijhc.1996.0073

[71] Fredrik Ohlin and Carl Magnus Olsson. 2015. Intelligent Computing in Personal Informatics: Key Design Considerations. In Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15). Association for Computing Machinery, New York, NY, USA, 263–274. https://doi.org/10.1145/2678025.2701378

[72] Oxford Dictionaries (2014) Cooperation. http://www.oxforddictionaries.com/definition/english/cooperation. Accessed 18 May 2022

[73] Marie-Pierre Pacaux-Lemoine and Makoto Itoh. 2015. Towards Vertical and Horizontal Extension of Shared Control Concept. In 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE Press, 3086–3091. https://doi.org/10.1109/SMC.2015.536

[74] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Hum Factors 39, 2 (June 1997), 230–253. https://doi.org/10.1518/001872097778543886

[75] Mina Park, Milam Aiken, and Laura Salvador. 2019. How do Humans Interact with Chatbots?: An Analysis of Transcripts. INTERNATIONAL JOURNAL OF MANAGEMENT & INFORMATION TECHNOLOGY 14, (2019), 3338–3350. https://doi.org/10.24297/ijmit.v14i0.7921

[76] Nick Pidgeon, and Karen Henwood. 1996. Grounded Theory: Practical Implementation. In Handbook of Qualitative Research and Methods for Psychology and the Social Sciences, John T. E. Richardson (Ed.). Wiley-Blackwell.

[77] Q.-C. Pham, R. Madhavan, L. Righetti, W. Smart, and R. Chatila. 2018. The Impact of Robotics and Automation on Working Conditions and Employment [Ethical, Legal, and Societal Issues]. IEEE Robotics & Automation Magazine 25, 2 (June 2018), 126–128. DOI:https://doi.org/10.1109/MRA.2018.2822058

[78] Amon Rapp. 2022. Time, engagement and video games: How game design elements shape the temporalities of play in massively multiplayer online role-playing games. Information Systems Journal 32, 1, 5-32. https://doi.org/10.1111/isj.12328

[79] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies 151, (July 2021), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

[80] Amon Rapp. 2018. Social game elements in World of Warcraft: Interpersonal relations, groups and organizations for gamification design. International Journal of Human–Computer Interaction, 34(8), 759-773. doi: 10.1080/10447318.2018.1461760

[81] Jeremy Roschelle. 2021. Intelligence Augmentation for Collaborative Learning. In: Sottilare, R.A., Schwarz, J. (eds) Adaptive Instructional Systems. Design and Evaluation. HCII 2021. Lecture Notes in Computer Science, vol 12792. Springer, Cham. https://doi.org/10.1007/978-3-030-77857-6_17

[82] Jeremy Roschelle and Stephanie D. Teasley. 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In Computer Supported Collaborative Learning (NATO ASI Series), Claire O' Malley (Ed.). Springer, Berlin, Heidelberg, 69–97. https://doi.org/10.1007/978-3-642-85098-1_5

[83] Pericle Salvini, Gaetano Ciaravella, Wonpil Yu, Gabriele Ferri, Alessandro Manzi, Barbara Mazzolai, Cecilia Laschi, Sang-Rok Oh, and Paolo Dario. 2010. How safe are service robots in urban environments? Bullying a robot. In 19th international symposium in robot and human interactive communication, IEEE, 1-7. https://doi.org/10.1109/ROMAN.2010.5654677

[84] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376843

[85] Karl A. Smith. 1995. Cooperative learning: effective teamwork for engineering classrooms. Proceedings - Frontiers in Education Conference, FIE 1, (December 1995), 208–213.

[86] Paul Spicker. 2011. Ethical Covert Research. Sociology 45, 1 (February 2011), 118–133. https://doi.org/10.1177/0038038510387195

[87] Laura Spillner and Nina Wenig. 2021. Talk to Me on My Level – Linguistic Alignment for Chatbots. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21). Association for Computing Machinery, New York, NY, USA, Article 45, 1–12. https://doi.org/10.1145/3447526.3472050

[88] Amanda J. Stent, Marie K. Huffman, and Susan E. Brennan. 2008. Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. Speech Communication 50, 3 (March 2008), 163–178. https://doi.org/10.1016/j.specom.2007.07.005

[89] Malin Sveningsson. 2004. Ethics in Internet Ethnography. Readings in Virtual Research Ethics: Issues and Controversies, 45–61. https://doi.org/10.4018/978-1-

59140-152-0.ch003

[90] Xiang Zhi Tan, Marynel Vázquez, Elizabeth J. Carter, Cecilia G. Morales, and Aaron Steinfeld. 2018. Inducing Bystander Interventions During Robot Abuse with Social Mechanisms. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18). Association for Computing Machinery, New York, NY, USA, 169–177.

[91] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony & Jacki O'Neill. 2017. How do you want your chatbot? an exploratory wizard-of-oz study with young, urban indians. In: Bernhaupt, R., Dalvi, G., Joshi, A., K. Balkrishan, D., O'Neill, J., Winckler, M. (eds) Human-Computer Interaction - INTERACT 2017. INTERACT 2017. Lecture Notes in Computer Science, vol 10513. Springer, Cham. https://doi.org/10.1007/978-3-319-67744-6_28

[92] Anita Tobar-Henríquez, Hugh Rabagliati, and Holly P. Branigan. 2020. Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing. Journal of Experimental Psychology: Learning, Memory, and Cognition 46, 6 (2020), 1091–1105. https://doi.org/10.1037/xlm0000774

[93] Xuan Wang and Ryohei Nakatsu. 2013. How Do People Talk with a Virtual Philosopher: Log Analysis of a Real-World Application. In Entertainment Computing – ICEC 2013 (Lecture Notes in Computer Science), Springer, Berlin, Heidelberg, 132–137. https://doi.org/10.1007/978-3-642-41106-9_16

[94] Paul Watzlawick, Janet Beavin Bavelas, and Don D. Jackson. 2011. Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes. W. W. Norton & Company, New York.

[95] Justin D. Weisz, Mohit Jain, Narendra Nath Joshi, James Johnson, and Ingrid Lange. 2019. BigBlueBot: teaching strategies for successful human-agent interactions. In Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19). Association for Computing Machinery, New York, NY, USA, 448–459. https://doi.org/10.1145/3301275.3302290

[96] Katherine S. Welfare, Matthew R. Hallowell, Julie A. Shah, and Laurel D. Riek. 2019. Consider the human work experience when integrating robotics in the workplace. In 14th ACM/IEEE international conference on human-robot interaction (HRI), IEEE, 75-84. https://doi.org/10.1109/HRI.2019.8673139

[97] David Woods. 2009. Rasmussen's S-R-K 30 Years Later: Is Human Factors Best in 3's? Proceedings of the Human Factors and Ergonomics Society Annual Meeting 53, 4 (October 2009), 217–221. https://doi.org/10.1177/154193120905300412

[98] Ziang Xiao, Michelle X. Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020. If I Hear You Correctly: Building and Evaluating Interview Chatbots with Active Listening Skills. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376131

[99] Su-Fang Yeh, Meng-Hsin Wu, Tze-Yu Chen, Yen-Chun Lin, XiJing Chang, You-Hsuan Chiang, and Yung-Ju Chang. 2022. How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 488, 1–16. https://doi.org/10.1145/3491102.3501941

[100] Jennifer Zamora. 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 (HAI '17). Association for Computing Machinery, New York, NY, USA, 253–260. https://doi.org/10.1145/3125739.3125766

[101] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 218, 1–21. https://doi.org/10.1145/3491102.3517479

[102] Qingxiaoyang Zhu, Yi-Chieh Lee, and Hao-Chuan Wang. 2022. Action-a-Bot: Exploring Human-Chatbot Conversations for Actionable Instruction Giving and Following. In Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing (CSCW'22 Companion). Association for Computing Machinery, New York, NY, USA, 145–149. https://doi.org/10.1145/3500868.3559476

[103] Jichen Zhu, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G. Michael Youngblood. 2018. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In 2018 IEEE Conference on Computational Intelligence and Games (CIG), IEEE, 1-8. doi: https://doi.org/10.1109/CIG.2018.8490433