



Trends, Challenges and Processes in Conversational Agent Design: Exploring Practitioners' Views through Semi-Structured Interviews

Malak Sadek

m.sadek21@imperial.ac.uk
Imperial College London
London, UK

Rafael A. Calvo

r.calvo@imperial.ac.uk
Imperial College London
London, UK

Céline Mougenot

c.mougenot@imperial.ac.uk
Imperial College London
London, UK

ABSTRACT

The aim of this study is to explore the challenges and experiences of conversational agent (CA) practitioners in order to highlight their practical needs and bring them into consideration within the scholarly sphere. A range of data scientists, conversational designers, executive managers and researchers shared their opinions and experiences through semi-structured interviews. They were asked about emerging trends, the challenges they face, and the design processes they follow when creating CAs. In terms of trends, findings included mixed feelings regarding no-code solutions and a desire for a separation of roles. The challenges mentioned included a lack of socio-technical tools and conversational archetypes. Finally, practitioners followed different design processes and did not use the design processes described in the academic literature. These findings were analyzed to establish links between practitioners' insights and discussions in related literature. The goal of this analysis is to highlight research-practice gaps by synthesising five practitioner needs that are not currently being met. By highlighting these research-practice gaps and foregrounding the challenges and experiences of CA practitioners, we can begin to understand the extent to which emerging literature is influencing industrial settings and where more research is needed to better support CA practitioners in their work.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

conversational agents, conversational AI, design processes, practitioner experiences

ACM Reference Format:

Malak Sadek, Rafael A. Calvo, and Céline Mougenot. 2023. Trends, Challenges and Processes in Conversational Agent Design: Exploring Practitioners' Views through Semi-Structured Interviews. In *ACM conference on Conversational User Interfaces (CUI '23)*, July 19–21, 2023, Eindhoven, Netherlands. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3571884.3597143>



This work is licensed under a [Creative Commons Attribution International 4.0 License](https://creativecommons.org/licenses/by/4.0/).

CUI '23, July 19–21, 2023, Eindhoven, Netherlands
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0014-9/23/07.
<https://doi.org/10.1145/3571884.3597143>

1 INTRODUCTION

As the prevalence and popularity of conversational agents (CAs) continues to rise in terms of commercial and industry-based products and use-cases, so does the amount of scholarly work on different aspects related to CAs. Despite this growth going hand in hand, there is still a significant gap between the findings and guidelines published in scholarly works and the approaches used by practitioners [16, 44].

Understanding this research-practice gap is crucial to ensure that research conducted in this field is able to effectively impact and shape best practices within industry-based spaces. [1] describe an "AI support vacuum" where AI-based systems are not properly supported by stakeholders and do not properly support stakeholders. This vacuum is largely present due to the the gaps between the experts working on AI, those conducting research, those responsible for maintaining its wider infrastructure, those making business decisions, and the people using it and being affected by it (ibid).

Our study's research objective is thus to understand the mentioned research-practice gap in the context of CAs by studying practitioners views and bringing them into consideration within the scholarly sphere. The aim is to expedite the "bubbling up" process described by [16]'s Translational Science Model from design practice to research to address the "applicability" gap from applied research to design practice.

Accordingly, our research questions were as follows:

Understanding Practitioners' Experiences and Challenges:

- RQ1: What are the major trends and developments occurring in the CA space from practitioners' perspectives?
- RQ2: What are the practical challenges faced by CA practitioners during their work?
- RQ3: What processes and methodologies do CA practitioners use in their work?

Understanding Research-Practice Gaps:

- RQ4: To what extent does existing literature and scholarly work support practitioners in terms of the trends and developments they adapt to, the challenges they face, and the processes they follow?
- RQ5: Based on these levels of support, where do research-practice gaps exist?

RQs 1-3 focused on practitioners' challenges and experiences and were addressed through a series of semi-structured interviews. Practitioners were asked to describe the trends they see emerging in the space of CA design, the challenges they face when designing them, and the design processes they follow during their projects. The aim behind the questions is to examine their opinions and

perspectives on the field as a whole and then to zoom into their personal experiences and practices. RQs 4 & 5 were addressed after analyzing the interview results and reviewing relevant literature to identify convergences and divergences between the two.

As a result, the contribution of this study is two-fold:

- (1) It highlights the prevalence of several technical and socio-technical challenges faced by CA practitioners in their work as expressed by them. This benefits the research community by helping to clarify areas where future work is needed to better support practitioners by either conducting new research, or re-adapting existing research to better suit their practical needs. Additionally, because several of those challenges have likely permeated from broader problems with AI-based systems, such as a lack of diversity among AI experts, highlighting this permeation reminds members of the community of the need to address and mitigate these issues urgently as they spread to newer, emerging technologies.
- (2) It introduces five specific practitioner needs where a research-practice gap exists or where research exists, but has not been adopted, adapted, understood, or discovered by practitioners. This contribution is essential to the research community to understand how their work is impacting industrial settings. Future work should focus on understanding the reasons why scholarly works are having little-to-no impact in certain cases.

Section two provides the needed background to contextualise the paper and introduce the research-practice gap mentioned earlier. Literature that relates to themes and topics discussed by practitioners is highlighted later on in the discussion section. This is because the focus of this paper is to establish links between practitioner-based findings and scholarly works, as opposed to a standard literature review. Section three outlines the methodological procedure followed and section four presents the findings of the study. The findings are organized into emerging trends, practitioners' challenges, and design processes followed. Each aspect contains several emergent sub-themes which are discussed in detail. Finally, section five synthesises these findings into five unmet practitioner needs to highlight where research-practice gaps exist and should be addressed to better support practitioners.

2 BACKGROUND

2.1 Interview Studies in the CA Space

As CAs continue to gain popularity, an increasing number of studies have focused on a variety of aspects relating to the creation and use of CAs. More specifically, interview studies conducted have targeted both CA users [31] and practitioners [59]. In terms of users, interview studies have looked at which CA use cases are seen as useful or enjoyable, what people value in conversations [14], how CAs are used in homes [46], and people's preferences for error responses [60]. In terms of practitioners, interview studies have most notably explored how ethical dilemmas [35] and guidelines [29] are dealt with in practice. Conversely, this study aims to understand practitioners' experiences and challenges across their work and highlight potential research-practice gaps where their needs are not currently being met.

2.2 Research-Practice and Socio-Technical Gaps in CA Design

Research-practice gaps exist across different disciplines closely related to CA design. The gap exists because of the distinction between high-level, theoretical interventions and more concrete practices [27]. The term describes a gap between research efforts within a specific field and the activities and practices conducted, where research bears limited influence on practice for any number of reasons. For example, there is a general consensus that a huge divide exists between the saturated space of theoretical AI ethics and AI codes of conduct, and the practical AI applications being developed today [11, 17, 20, 28, 36]. A similar gap between research and practice is also described in Human-Computer Interaction research [15, 44]. Given these gaps in related fields, the aim of this work is to explore the existence of a research-practice gap in the field of CA design and to highlight some of its consequences by identifying unaddressed practitioner needs and challenges.

Previous works have already highlighted limitations of practically implementing research-based interventions at a large scale within the CA space. Similar to other AI-based systems, design guidelines are found to be difficult to consolidate and implement, placing the responsibility on designers and developers to reflect on how their practices and decisions might impact different stakeholders [25]. At the other end of the spectrum, CA design tools are currently often very technical and focused on the design or development of CAs only [33, 34]. Similarly, studies on CAs often focus on one aspect, such as their implementation, their design, their adoption, their evaluation, etc., with none offering higher-level interdisciplinary frameworks that account for more contextual, socio-technical factors and processes, and support multiple phases of the life cycle [33]. [23] also discuss the "shortage of integrative perspectives on CA development and design". Highlighting the silo-ed and individual nature of attempts at designing, creating, and evaluating CAs and the lack of re-usability, transferability and sustainability of most of the resulting design principles and recommendations published. Each agent is developed largely in isolation and the field suffers from a lack of extensive documentation to aid future work [39, 45].

Additionally, several AI-based technologies (including CAs) tend to be more heavily researched from a technical perspective than a socio-technical one [30, 33, 52]. As a result, these technologies can suffer from "environment blind[ness]" with regards to relevant socio-cultural facets of technology and its implications [57]. Aside from the obvious disconnect this gap creates between the technologies created and those using them and being affected by them, this socio-technical gap can also lead serious ethical problems and misuses [22, 47, 49]. [22] speak at length of the socio-technical gap in explainable AI and a similar gap can be seen in CA design. For example, there are currently very few tools for leveraging interdisciplinary collaboration to create a conversational agent in terms of supporting interdisciplinary communications during the process and integrating the knowledge and insights produced [4], a lack of socio-technical tools and frameworks [54], and limited stakeholder participation and grounding in value-sensitivity [8, 25, 48].

The existence of a research-practice gap and a socio-technical gap within the CA space fueled the motivation for this study. By

sharing practitioners' insights regarding the socio-technical experiences and challenges they face with the research community, a clearer image of both these gaps can begin to emerge. Other relevant topics are introduced in detail within the discussion section in order to structure and link themes from the literature directly to the findings of the interview study.

3 METHODOLOGY

3.1 Study Design

The study consisted of semi-structured interviews. The aim of the interview study is to understand practitioners' experiences and perspectives on emerging trends, challenges they face, and the design processes they use. The study received the approval of the Science Engineering Technology Research Ethics Committee at Imperial College London under the SETREC reference 21IC7361.

3.1.1 Data Collection. Interviews were conducted during 2022 before the boom of generative AI and large-scale language models such as ChatGPT. They were 1 hour on average and took place on Microsoft Teams, Zoom, or Google Meet as preferred by participants. The interview guidelines were kept quite general on purpose to allow practitioners to speak freely and introduce their own themes and topics. Follow up questions asked depended on the content and flow of each interview. The interview questions are included below:

- (1) What are some of the main trends and directions in the world of CAs today?
- (2) What would you say are the main challenges you face when working on CAs?
 - Do you feel that these challenges generalize to all AI systems, or are they specific to certain types of systems or domains?
 - Do you have any ideas, even if seemingly unfeasible, about how to address some of those challenges?
- (3) Is there a specific framework or design process you follow when creating CAs?
 - Did you come up with it yourself?
 - Do you know any others?
 - Are you satisfied with the current way things are done with regards to the support offered by current tools and frameworks?

3.1.2 Data Analysis. The interviews were audio recorded and transcribed automatically using each platform's services. The transcriptions were manually cleaned with the help of the audio recordings afterwards. All interviews were conducted and cleaned by the first author. Transcriptions were analyzed using a hybrid approach of top-down and bottom-up thematic analysis [6] on NVivo (v12 for Mac) where excerpts were coded as 'trends', 'challenges' and 'design process descriptions' initially and then sub-themes emerged from the transcriptions themselves. Analysis was conducted by the three authors who met to discuss differences until a consensus was reached regarding codes and categorisations.

3.2 Participants

Participants consisted of 12 CA practitioners working in a number of roles including conversation designers, researchers, executive managers, and data scientists. The sample number was a result of

Table 1: Participant Information

ID	Gender	Y.Exp.	Sector	Role
P1	M	7	Start Up	Executive Managerial Position
P2	F	22	Start Up	Conversational Exp. Designer
P3	F	3	Start Up	Executive Managerial Position
P4	M	3	Small-Medium Enterprise	AI Research Engineer
P5	M	6	Start Up	Executive Managerial Position
P6	M	4	Institute/Research Org.	Executive Managerial Position
P7	M	5	Start Up	Executive Managerial Position
P8	M	1	Start Up	Executive Managerial Position
P9	M	2	Institute/Research Org.	Conversation Designer
P10	M	4	Multinational Corporation	Lead Data Scientist
P11	M	2	Multinational Corporation	AI Consultant
P12	M	6	Institute/Research Org.	Chief Research Scientist

"the principle of saturation and... pragmatic considerations" as is often the case with interview studies [53]. These participants were recruited through private messages on the social network website LinkedIn and were selected based on their advertised job titles on the platform. They have varying years of experience, roles, and employers to ensure diversity in the perspectives and experiences reported. Participant information can be seen in details in Table 1.

4 FINDINGS

This section presents the findings of the interview study regarding the three main topics explored: emerging trends, practitioners' challenges, and design processes followed.

4.1 Emerging Trends

Practitioners mentioned a wide range of socio-technical trends emerging within the space of CA design. While some focused on technical improvements and developments, others mentioned a number of trends relating to roles and interdisciplinarity, the emergence of best practices, and working with stakeholders.

4.1.1 The Double-Edged Sword of No-Code Solutions: Democratization versus CA Quality. Practitioners were especially divided in their opinions on no-code solutions for creating CAs. On the one hand, some participants felt that they "democratized" [P2] CA creation, allowing anyone to participate and perhaps improving diversity among those creating them to an extent. Participants also felt that these tools helped them to not start from scratch on every project by providing some ready-made aspects which are available out-of-the-box. On the other hand, non-technical practitioners, such as conversation designers, felt it enabled both end-users and engineers to take on roles they were not properly trained for.

Participant P8 spoke positively about not having to start from scratch by having access to "really good variations of responses almost out of the box rather than having to go and create those things each time." This sentiment was echoed by participant P2 who appreciated the ability to "plug and play" to speed up their workflow.

Conversely, participant P5 spoke negatively about amateurs using these tools to create "substandard experiences" because new "clever platforms" allow them to "do it quickly." Participant P1 also spoke negatively of such tools stating that "in the hands of end-users" they simply lead to "chaos". Participant P2 also spoke of the potential problems with these tools when engineers use them. They had noticed that people often believe that knowing a language automatically makes you an apt conversation designer:

“They’re like ‘I’m well versed with the use of language. Why should I not do that?... I’m English or I’m American. I speak English. Why can I not do it?’, which of course I completely disagree with.”

Participant P9 also mentions this underestimation of the skills needed and equating knowing a language to being able to design conversations, relating it to the “*Dunning Kruger effect [where] you think you know what you don’t actually know*” and makes an insightful comparison: “*Yeah you can write but that doesn’t make you an author, right?*”

Participants also suggested that the lack of standardization brought on by this democratization is somewhat being addressed through the emergence of guidelines and best practices, created largely by platform providers. Participant P10 mentions this emergence and notices that “*more and more best practices [are] being talked about*”. As does participant P2, hinting, however, that it may be a double-edged sword because it leads people to believe “*they don’t really need a methodology because [they’ll] just do what Amazon recommends or Google recommends or Microsoft recommends.*”

4.1.2 Separation of Disciplinary Roles. Practitioners also spoke about the increasing separation of roles and the rising number of disciplines becoming involved in CA projects. Participant P3 recalls their experience on a previous project, stating that: “*it was a one man band and I had to do everything from design and the conversations to implementing and deploying the chatbot.*” However, as conversational design becomes more recognized as a “*true profession*” [P6], this is beginning to change.

As demand for CAs increases, participant P2 has seen larger companies want to hire CA practitioners as an in-house role as opposed working with consultancies or freelance practitioners. Participants P1 and P8 spoke of companies beginning to understand that more is needed than just engineering, but still with an uncertainty as to exactly what is required. Participant P5 also spoke of an uncertainty regarding knowing which skills and roles are needed given the still emerging nature of the space, stating that “*people don’t still know what to do, what’s involved in creating these sort of projects and the stakeholders [that] need to get involved.*” Finally, participant P9 shares the same views that many companies “*have limited understanding*” on how to push CA project forward beyond initial “*proof of concept[s]*”.

Despite this uncertainty, participant P5 also reports seeing the development of different roles and specialties among CA practitioners that “*didn’t exist 10 years, even 5 years ago.*” Participant P10 also mentions a “*culture of multiple roles being involved... and a separation of roles.*” They especially noticed that aside from those working on “*the nuts and bolts, the technical building of a conversational agent... people with more of an English background, instead of a software engineering background, are getting involved.*”

4.1.3 Improvement in CA Capabilities Leading to Better Conversational Partners. Participants P5 and P8 were quick to mention improvements in the underlying components of CAs, such as the accuracy of their models, the improvement in their natural language understanding capabilities, and so on. One effect of this improvement is that practitioners are also seeing a rise in conversational agents becoming more like “*human conversational partner[s]*” [P11]. As their capabilities improve, conversational agents are shifting away from having specific use-cases and tasks they can fulfill, and

more towards having general conversations. This shift can very prominently seen in the advent of ChatGPT as participant P11 describes:

“At the same time, there is renewed sort of engagement for perhaps one particular type of conversation layer, which is the companion AI and chatbots based on large language models. Which sort of again sparks an idea of conversational AI becoming something more resembling a human conversational partner than what we have seen so far.”

4.2 Practitioners’ Challenges

Narrowing the scope from broader trends in the space of CA design, the study’s next focus was on practitioners’ largest challenges when working in this space. Similarly to when talking about trends, practitioners mentioned several technical challenges with tools and workflows, as well as a number of socio-technical challenges regarding working with users, team members, and other stakeholders.

4.2.1 Lack of Conversational Archetypes. As CA capabilities continue to improve, practitioners are beginning to focus more on CA personality, persona and tone of voice alongside its dialogue and content. Practitioners spoke about the lack of conversational archetypes and personalities that are available to use out-of-the box, also explaining why they are so important.

Participant P1 describes this lack of archetypes and contrasts it to the multitude of user interface templates currently available:

“Another problem was there’s not a well defined set of conversational archetypes that people can adopt, for example if you’re building a mobile app, you have your login screen and you have your home screen and you have your hamburger menu.”

Participant P8 also goes on to speak of the importance of developing CA personalities “*because that’s how many people you can service.*” They describe how CA designers can either play it safe with a “*neutral*” personality where “*you don’t please anybody but you probably don’t offend anybody*”, or they can create a personality which might appeal to some people strongly, but risks “*alienating others.*”

4.2.2 Lack of Socio-Technical Tools. Practitioners also expressed frustration with current tools being too focused on technical aspects of CA design and offering no support for educating or explaining capabilities to non-technical stakeholders in order for them to participate meaningfully.

Participant P1 was frustrated by the lack of a “*metaphor that anybody uses, understands, or can apply, that sort of explains to them what they’re doing and what the possible outcomes are going to be*”, with most tools diving into technical terms such as intents, entities and utterances. They describe these tools as not giving users a “*humanized way of thinking about it.*” They mention that this is especially difficult for traditional software developers as it is “*made to look like other programming tools where you would expect to get the same result every time, but then all of a sudden you didn’t get the same result every time [because of] it’s non-deterministic nature.*”

Alternatively, participant P3 also discusses how tools are either user-friendly but lack important features or are comprehensive but too technical. On one hand they find that “*drag and drop systems like IBM Watson... or Google dialog flow don’t provide you with a*

lot of flexibility". On the other hand, platforms such as Rasa are much more flexible and powerful, but *"with Rasa you have to hire a team of both chatbot engineers and back end developers and software engineers specifically because Rasa uses those custom actions."*

4.2.3 Biases and the Need for Diversity. Practitioners discussed a lack of diversity within the space of CA design leading to potential biases. They mentioned a lack of diversity in terms of:

- **Machine Learning Models:**

Non-technical practitioners especially felt that they have to inevitably use models created by others that are inherently biased: *"Biases that are published by the model makers up front, and I think that people to go through like a click through tour on one of these things and be like 'see if you can get it to create a woman scientist, nope, it's impossible' and to understand that."* [P1]

- **Training Data:**

In turn, technical practitioners also felt that they themselves have to create models using training data that is inherently biased: *"I guess the problem is that you need to train the data, the language data... who gathers the data and curates the data and trains the systems. Is it just a young white male guy or could it also be more race and gender and also age inclusive? But yeah, as you suspect, it's maybe young male straight guys who think they know everything. But yeah, that's a rant for another day."* [P2]

"You [need to] have good training data - that is not biased - that you have an understanding of, you know, privacy or copyright or whatever in your ingestion of data - that you're not just going out and scraping the cloud arbitrarily." [P1]

- **Geopolitical Participation:**

Practitioners felt that the CA space was dominated by Western innovations and values: *"And that's about the same time that I heard about conversational AI and it was this new thing that seemed to be inspired by what was going on in Asia which, ironically, I don't see Asia as much in the conversation anymore."* [P10]

- **Team Members:**

Finally, practitioners felt that teams creating CAs lacked diversity among their members, whether technical or non-technical: *"I think things just sometimes fall through the cracks if the team isn't diverse enough... A typical example is that we were building a chatbot for people that were calling banks during COVID pandemic just to reduce the demand on the contact centers and something there that showed up is that if someone said 'I lost my wallet', it would recommend getting new card. But if it said 'I lost my purse', it wouldn't resolve anything. And by the time it was a bit too late. So I think having a diverse team is quite useful."* [P12]

"I think it just helps you with the kind of the unexpected responses. You know the kind of the things that you don't, or maybe the language that will be used that you don't expect to do." [P8]

4.2.4 Platform Saturation and Integration. One negative aspect of the increased popularity and demand for CAs is the resulting increase in the number of platforms available for creating them.

Practitioners report feeling overwhelmed with having to keep up with changes and developments in platform-specific knowledge.

Participant P3 spoke about this issue at length:

"There are many different trends coming up from research and making their way into commercial tools and conversation designers and chatbot developers are kind of struggling to keep up with all the innovation and existing tools... And conversation developers are always thinking, 'should I learn about things?' there's so much for them to learn and basically what I notice from this discourse is that they just gave up. Because they're already so busy like implementing a chatbot and a lot of this will be contract workers who are doing everything alone. So they have to design the conversations and deploy them. And they're just saying 'we don't have enough time as developers to learn all of these innovations...' But yeah, there's that kind of resistance in the development community and I do understand it because it's a lot to keep up with."

Participant P5 also complained that *"it's quite challenging trying to keep on top of all of the different technologies while still adhering to costing"* and participant P11 echoed the sentiment that *"keeping track of the different platforms and systems that are required to set-up a workable, working conversational agents is challenging... [they] make the job notoriously difficult."* They also mention the lack of connection and *"traceability across these different platforms."*

Participant P2 also spoke of this issue at length, especially in relation to integrating tools for documentation as their team had knowledge and artefacts spread out across *"word documents"* for *"conversation flows"* and *"Excel spreadsheets"* for *"training data"* and *"Confluence and JIRA"* to make *"tickets"* and they were worried that *"something may be lost in one channel."* They wished for either a single tool or tighter integration across tools to ensure *"no duplication but also no functionality aspect that gets lost."*

4.3 Design Processes

Finally, practitioners were asked to describe the typical design process they go through when creating CAs. The overall sentiment was that there is no standard process or best practice that every practitioner follows.

Three participants in particular described frustration with the lack of a *"rigorous and agile process in place"* [P3] as opposed to some people resorting to *"a blind process where they like typing a bunch of stuff, they train their model and then they hope that it didn't break"* [P1]. Participant P6 especially speaks of their experience educating practitioners on how to use a standardised process:

"We just saw so many bad experiences. That they're like 'ohh wow, this is really poor', nobody had any idea what to do... We always say everybody's selling hammers and we're the only ones training carpenters."

Each practitioner followed their own variation of a CA design process, with no two processes being identical. A summary of these processes as described by the practitioners is provided diagrammatically in Figure 1 below. All the processes tended to follow these major phases: an initial planning phase where requirements are collected and the project is understood, a prototyping phase where conversation and dialogue flow design takes place, an often iterative phase where the prototype is tested and feedback is collected, a development phase where the technical implementation of the

CA takes place, and a final phase where deployment, maintenance, testing and optimization occurs. Practitioners used similar language to describe their activities across these phases, such as mentioning requirements, prototyping, conversation design, development and deployment.

Individual steps and activities varied between these phases and their order and repetitions also varied. Some practitioners' processes focused more on the conversations themselves, while others took a more technical approach such as looking at intent structures and platform selection. None of the practitioners mentioned any scholarly design processes.

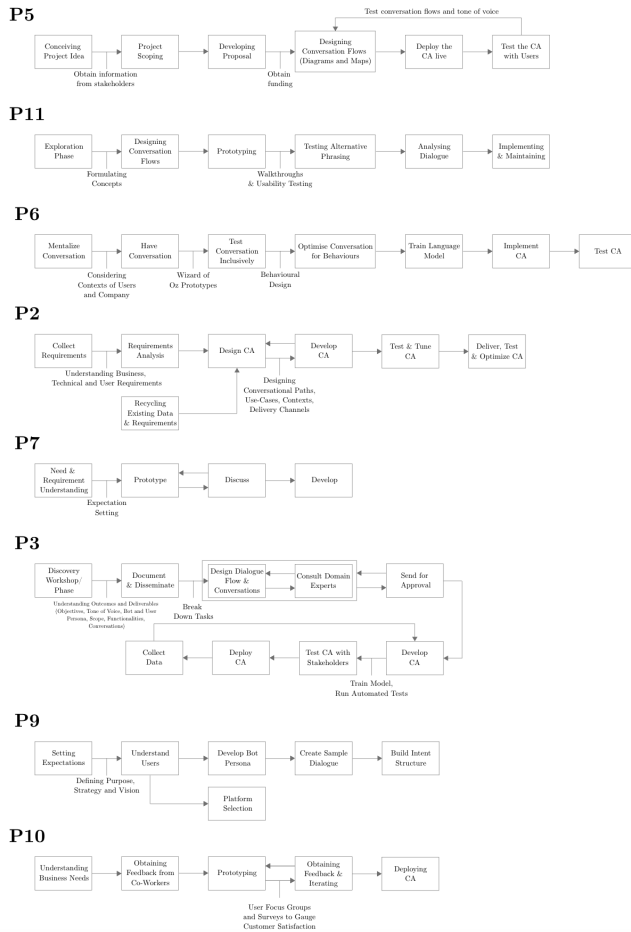


Figure 1: A diagrammatic representation of the design processes followed by 8 CA practitioners based on their descriptions.

5 DISCUSSION

During the interview study, practitioners touched upon several topics which have been of great interest to the CA research community. Accordingly, this section synthesises these findings into five unmet practitioner needs. For each need, relevant literature is consulted to provide context regarding current research directions related to this need. Finally, open problems are highlighted.

5.1 The Need to Mitigate Misuse of No-Code Solutions

Several practitioners expressed concern with how other stakeholders such as end users or engineers use no-code solutions and the problematic mindsets that they facilitate. They also mention that one solution to combat this misuse is the emergence of best practices and guidelines when creating CAs.

This emergence points to the developing maturity of the field and is especially needed by the stakeholders and amateur CA builders that practitioners have described as ‘not knowing what they’re doing’. The recommendations followed by practitioners have largely come from platform providers, such as IBM’s Alma Design System¹ and Microsoft’s Bot Framework².

This focus on industry-based guidelines comes despite several scholarly sources also providing such recommendations [34]. One possible explanation can be found in criticisms for research-based guidelines and recommendations for AI-based systems in general. Overall, these guidelines are found to be too abstract, vague, or difficult to implement on-the-ground [56]. They are also found to be difficult to measure in order to evaluate their success or effectiveness [26]. In fact, studies examining their impact have found that they have no effect on practice [37].

Future work should examine the reasons behind this void between research-based recommendations and industry-based practices and the sluggish conversion rates from the prior to the latter. A potentially viable solution is to focus on and incorporate practitioners’ experiences and viewpoints into the creation of these principles, which was the main inspiration for the approach followed by this study.

5.2 The Need for Operationalising CA Capabilities and Personalities

Several practitioners spoke with excitement about CA’s improving capabilities allowing them to take on more ‘general’ use-cases and conversations with no specific task. This industry-based interest goes together with an academic interest as significant research is being conducted in these directions.

These types of CAs are becoming known as “social conversational user interfaces” [32] and are indeed taking on social roles such as partners [43], team members [5], and friends³.

All conversations sit somewhere on a spectrum between social and transactional conversations, depending on the context and objective of the conversation [10]. Social conversations aim to build a relationship between participants through establishing trust and reaching a common ground. Transactional conversations have a goal which typically all participants are aware of and where each one plays a specific role, culminating in the achievement of that goal. Conversations with CAs are typically viewed as transactional [14], with only few applications providing social conversations.

In particular, there is much interest in CAs’ ‘humanness’. From a user perspective, how human-like a CA is, whether in terms of its conversation style or through being embodied or anthropomorphized

¹IBM’s Alma Design System can be viewed at: https://researcher.watson.ibm.com/researcher/view_group.php?id=8426

²Microsoft’s Bot Framework can be viewed at: <https://dev.botframework.com/>

³Such as the Replika chatbot: <https://replika.com/>

using an avatar, has been found to influence CA adoption [13] and knowledge transfer between CA's and humans [51]. As such, the 'humanness' of CAs is sometimes used as an evaluation metric in studies [19].

This increased interest in CA humanness goes hand-in-hand with an interest in CAs' personalities. Moving beyond the content of their dialogues, it is becoming crucial that CAs adopt a persona or tone of voice in order to bring it to life more effectively [38]. Work has already been done on creating design patterns for different CA personalities or archetypes [3] and using CA personalities as a starting point when collaboratively designing them with other stakeholders [12]. There has also been work on creating Personality-Adaptive Conversational Agents (PACA) that assume different personalities either when discussing different topics [2], to mimic the user's personality or to complement their personality (e.g. by being direct with impatient users).

The study's participants expressed an interest and increasing focus on these aspects in the CAs they design and build but felt frustrated at the lack of ready-made and practical tools, patterns, and archetypes that they could incorporate into their CAs. Since research is already being conducted regarding CA archetypes and PACAs, future work can help bridge the research-practice gap by distilling research-based innovations into actionable tools and resources for practitioners.

5.3 The Need for Socio-Technical Tools

Most of the tools used by CA practitioners were to aid with technical aspects of designing and building CAs. [4] make the distinction between commercial tools for creating dialogue-based conversational agents, scientific tools for creating embodied conversational agents, and open-source tools for creating applications such as interactive stories or narrative games that include a conversational agent. Commercial tools can be further divided into "dialogue management systems" which introduce a layer of abstraction through visual elements such as flowcharts and node-link diagrams to allow non-technical users to design conversation flows with a low barrier to entry (such as one participant's mention of Google's DialogFlow), and CA creation tools that allow for the technical specification of a ready-to-deploy CA (such as the same participant's mention of Rasa).

In all cases, these tools provide no support for the socio-technical aspects involved in designing CAs such as talking to stakeholders and understanding their requirements, educating stakeholders and setting their expectations, working with ethics and values, and so on. AI practitioners from other studies have also reported a similar situation for different types of AI-based systems [61]. Educating other non-technical team members, such as designers, is especially important [59] and is also not explicitly supported in existing tools.

Academics have also mentioned using metaphors and personifying AI systems to explain more effectively to non-technical stakeholders [40, 41] and make them more relatable. Despite these potential benefits, one participant complained that the tools they used offered no means of using these metaphors and used technical terms instead.

As such, practitioners have turned to design tools and techniques to collaboratively design CAs with non-technical stakeholders and

team members [9]. There's also been an increase in tools aiming to lower the barrier-to-entry for non-technical individuals to create CAs [7], but as discussed with participants, these come with their own disadvantages.

While the literature calls for adopting a socio-technical perspective, including different non-technical stakeholders, and using techniques such as metaphors and personification, these calls are not echoed in the software and tools being used by practitioners. Similarly to the previous section, it becomes clear that while research sought after by practitioners does exist, it is not distilled into forms usable by them at the moment. Participant P3 mentioned a "resistance in the development community" to fast-paced innovations that overwhelm practitioners who must understand and adapt to them alongside their usual work. This struggle might be one of the causes for a lack of wider permeation from research to industry and should be investigated in more detail.

5.4 The Need for Increasing Diversity in the CA Space

There have been a number of academic works on the importance of democratizing AI [21]. AI practitioners are thought of as homogeneous, leading to the creation of AI-based systems that do not reflect the diversity of their users [55]. This lack of accounting for diversity and the resulting emergence of biases is further exacerbated by the opaqueness of workflows and pipelines for AI-based systems, especially with relation to data [21].

As such, there have been countless recent calls for an increase in the diversity of practitioners and of the disciplines involved beyond engineering or science-based ones [18, 33, 50].

It was thus incredibly significant to discover practitioners' complaints of biases and a lack of diversity beyond male engineers in the context of CAs, which are largely AI-based. These complaints point to the diffusion of these limitations into even relatively emerging branches of AI-based systems and adds weight to the urgency with which they must be addressed.

Academics have already begun calling for more interdisciplinary work when it comes to CA design [33, 34] and a number of studies have focused on collaboratively designing CAs to mitigate the harms of homogeneous design teams [24]. Practitioners' accounts and real-world examples of the harms resulting from a lack of diversity and interdisciplinarity adds a multi-faceted perspective to these calls that solidify these harms as real and present.

5.5 The Need for a Unified, Practical Design Process

Formal design processes have emerged from industrial sources. For example, Rasa, an open-source conversational agent development platform that supports building, deploying, training, and testing agents, conceived the methodology of conversational driven development (CDD) [42]. CDD is meant to evolve the development of the agent's conversations in a very user-centric approach. The idea is to listen to users early on and then continuously learn from real-world conversations instead of using specifically-made training data. Such an approach improves diversity in language, usage scenarios, style of speaking, and so on.

A number of scholarly design processes for CAs have also emerged recently. For example, [58] has introduced an incredibly thorough procedure model for projects involving chatbots. [4] focuses on creating dialogues and conversations, providing a process for this specific aspect within CA design.

Despite their rigor and sophistication, research-based design processes have seen lower and slower adoption rates in industry among practitioners. There has also been little work done on identifying which processes practitioners use or whether they use their own processes, or a hybrid of both.

As discussed earlier, there was a significant variation in the design processes followed by different practitioners. This variation points to a need for the creation of a standardized design process for use by industry practitioners. In fact, [23] speaks of the “shortage of integrative perspectives on CA development and design” hinting at the silo-ed nature of work in the space of CA design at the moment with CAs being built in isolation and with little documentation [39, 45]. In [23]’s review of design principles for conversational agents produced from existing studies, all the papers reviewed had low or moderate re-usability. In the same review, several CA design principles were also found to not directly address real-world problems and were varying in accessibility and actability [23]. This is far from the first instance where scholarly works were unsupportive of practice [44].

The Natural Conversation Framework (NCF) offers an extensive set of simple conversational user interaction patterns and is made use of by IBM’s Watson Assistant [38]. Leveraging interdisciplinary insights [38] to provide more detailed guidance in the form of a design system which is a concept several designers will already be familiar with [39]. The NCF’s interaction model consists of a set of 15 patterns or sequences which each have several possible expansions, reaching a total of over 50 sub-patterns. This greatly addresses issues of re-usability and standardisation.

Even though more prescriptive design processes for CA have begun emerging [58], they are still not being adopted at a large scale within industry. This lack of adoption could be due to any number of “translational barriers” described by [16], and even though these barriers are addressed by frameworks such as that of [38, 39], the lack of wide-spread adoption persists. At this stage, it is difficult to determine which, if any, process is ‘ideal’, however, consolidating these processes and harnessing their different strengths can bring the field closer to a unified and rigorous practitioner-based design process for CAs, instead of having to reinvent the wheel every time.

5.6 Future Work Needed to Address Research-Practice Gaps

This interview study has taught us much in terms of the experiences and challenges that practitioners in the CA space face during their work. Conversational UX is an emerging discipline and yet it borrows both beneficial and problematic aspects from related fields such as AI-based systems and human-computer interaction. Research-practice gaps and socio-technical gaps exist in all these fields and addressing them is pertinent to ensuring that works within the research community reach their full potential in terms of impacting concrete industrial practices and the outcomes they produce.

This study has two main limitations: having a limited sample size and being conducted before the proliferation of generative AI and large language models. Over the coming years, the experiences of CA practitioners are expected to change drastically and this study might need to be repeated in the future to compare answers and analyse changes. Nevertheless, the recurrence of certain themes and complaints sheds light on the existence of unmet needs that practitioners have which the research community has been already addressing and working on. This tells us that future work needs to be directed towards ensuring that needed research makes it across the research-practice gaps and is understandable, desirable, and usable by practitioners. While each need may require a different form of ‘future work’, a common point is that the operationalisation of research into practical tools and resources that practitioners can use is a crucial aspect to address several unmet needs. [16] explain that addressing the “applicability” gap from applied research to design practice requires the “translation and synthesis of knowledge into usable resources” [p.5] referring to the domain of Human-Computer Interaction (HCI). The findings of this study support such a claim within the domain of CA design as well. It appears to us that on one side, there is practitioner demand for support of their practices, needs, and interests, and on the other side, there is a supply of relevant research-based insights and innovations; what remains is to cross the gulf between both through the creation of practical tools and resources.

6 CONCLUSION

This study has explored CA practitioners’ experiences and perspectives regarding current trends in the space of CA design, the challenges they face in their work, and the design processes they follow. The aim of the study is to foreground CA practitioners’ challenges and experiences to bring an industry-based perspective within the scholarly sphere, and to identify research-practice gaps where certain practical needs were not currently being addressed.

In terms of trends, practitioners spoke of improvements in CA capabilities and how this has led to more general and social use-cases for CAs. They also spoke of the advantages and drawbacks of the increase in no-code solutions, and the separation of roles among CA practitioners leading to increasing interdisciplinarity in the field. Practitioners also spoke of the challenges they face in terms of integrating between platforms in an overly saturated market, a lack of socio-technical tools and conversational archetypes despite the increasing interest in both these factors, and rampant biases due to a lack of diversity across several levels. Finally, they described the design processes they follow, showing a lack of standardization and integration between the work of different practitioners and between academia and industry.

While several of these themes and issues are also emerging in academic literature, it was crucial to understand their concurrent development in industrial and commercial contexts. By highlighting links between themes discussed in the interviews and relevant literature, this work helps highlight existing research-practice gaps where practitioners’ needs are not being addressed. These needs included the need for (i) mitigating misuse of no-code solutions,

(ii) operationalising CA capabilities and personalities, (iii) socio-technical tools, (iv) increasing diversity, and (v) a unified and practical design process. Future work needs to dive deeper into the reasons behind these gaps (i.e. whether they stem from a lack of literature or a lack of permeation of the literature into industrial settings and the reasons behind this deficiency), and how they can be addressed. Perhaps one final aim of this paper is also to reinvigorate interest in tackling the challenges and issues presented, whose relevance can be seen to lie well beyond the scope of academic research.

ACKNOWLEDGMENTS

This work has received no external funding or any resources from third parties. The authors all report no conflict of interest in this work.

REFERENCES

- [1] A. Abaza. 2021. MLOps: Why is it the Most Important Technology in the Age of AI? <https://www.synapse-analytics.io/post/mlops-why-is-it-the-most-important-technology-in-the-age-of-ai>
- [2] R. Ahmad, D. Siemon, and U. Gnewuch. 2022. Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Information Systems Frontier* (2022).
- [3] M. Azmandian, J. Arroyo-Palacios, and S. Osman. 2019. Guiding the Behavior Design of Virtual Assistants. Proceedings of the ACM International Conference on Intelligent Virtual Agents.
- [4] T. Beinema, H. op den Akker, D. Hofs, and B. Schooten. 2022. The WOOL Dialogue Platform: Enabling Interdisciplinary User-Friendly Development of Dialogue for Conversational Agents. *Open Research Europe* 2, 7 (2022).
- [5] E. Bittner, S. Oeste-Reib, and J. Leimeister. 2019. Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. Proceedings of the Hawaii International Conference on System Sciences (HICSS).
- [6] V. Braun and V. Clarke. 2012. *Thematic analysis*. American Psychological Association, 57–71. <https://doi.org/10.1037/13620-004>
- [7] J. Brummelen, K. Weng, P. Lin, and C. Yeo. 2020. Convo: What does conversational programming need? An exploration of machine learning interface design. *Computing Research Repository (CoRR)* (2020). <https://arxiv.org/abs/2003.01318>
- [8] L. Byrd. 2020. Tech Inclusion: How culture may enhance conversational AI technology. <https://uxmag.com/articles/tech-inclusion-how-culture-may-enhance-conversational-ai-technology>
- [9] H. Candello, M. Pichiliani, C. Pinhanez, S. Vidon, and M. Wessel. 2020. Co-Designing a Conversational Interactive Exhibit for Children. Proceedings of the ACM Interaction Design and Children Conference: Extended Abstracts.
- [10] C. Cheepen. 1988. *The predictability of informal conversation*. Pinter.
- [11] V. Chen, J. Li, J. Kim, G. Plumb, and A. Talwalkar. 2022. Interpretable Machine Learning: Moving from myths to diagnostics. *ACMQueue* 19, 6 (2022).
- [12] Z. Chen, Y. Lu, and A. Nieminen, M. and Lucero. 2020. Creating a Chatbot for and with Migrants: Chatbot Personality Drives Co-Design Activities. In *Proceedings of the ACM Designing Interactive Systems Conference*.
- [13] S. Choudhary, N. Kaushik, and B. Sivathanu. 2022. Modelling the Key Enablers and Barriers of AI-Based Conversational Agents' Adoption: an Ism and Micmac Approach. *Journal of Content, Community & Communication* 6 (2022).
- [14] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. Proceedings of the CHI Conference on Human Factors in Computing Systems.
- [15] L. Colusso, C. Bennet, G. Hsieh, and S. Munson. 2017. Translational Resources: Reducing the Gap Between Academic Research and HCI Practice. Proceedings of the Designing Interactive Systems (DIS) Conference.
- [16] L. Colusso, R. Jones, S. Munson, and G. Hsieh. 2019. A Translational Science Model for HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300231>
- [17] C. Custis. 2021. Operationalizing AI Ethics through Documentation: ABOUT ML in 2020 and Beyond. <https://partnershiponai.org/about-ml-2021/>
- [18] F. Delgado, S. Barocas, and K. Levy. 2022. An Uncommon Task: Participatory Design in Legal AI. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022).
- [19] S. Diederich, A. Brendel, S. Morana, and L. Kolbe. 2022. On the Design of and Interaction with Conversational Agents: An Organising and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems* (2022).
- [20] DigitalCatapult. 2020. Lessons in Practical AI Ethics.
- [21] C. Dignazio and L. Klein. 2020. *Data Feminism*. MIT Press.
- [22] U. Ehsan, K. Saha, M. Choudhury, and M. Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proceedings of the ACM on Human-Computer Interaction* 34 (2023).
- [23] E. Elshan, C. Engel, P. Ebel, and D. Siemon. 2022. Assessing the Reusability of Design Principles in the Realm of Conversational Agents. Proceedings of the International Conference on Design Science Research in Information Systems and Technology (DESRIST).
- [24] Daniel Fitton, Janet C Read, Gavin Sim, and Brendan Cassidy. 2018. Co-Designing Voice User Interfaces with Teenagers in the Context of Smart Homes. Proceedings of the ACM Conference on Interaction Design and Children.
- [25] A. Folstad, T. Araujo, E. Law, P. Brandtzaeg, S. Papadopoulos, L. Reis, M. Baez, G. Laban, P. McAllister, C. Ischen, R. Wald, F. Catania, R. von Wolff, S. Hobert, and E. Luger. 2021. Future directions for chatbot research: an interdisciplinary research agenda. *Computing* (2021), 2915–2942.
- [26] T. Hagendorff. 2020. The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines* 30 (2020), 99–120.
- [27] M. Harbers and A. Overdiek. 2022. Towards a living lab for responsible applied AI. In *Proceedings of the Design Research Society Conference (DRS)*.
- [28] A. Jobin, M. Ienca, and E. Vayena. 2019. The global landscape of AI ethics guidelines. *Natural Machine Intelligence* 1 (2019), 389–399.
- [29] Krishika Haresh Khemani and Stuart Reeves. 2022. Unpacking Practitioners' Attitudes Towards Codifications of Design Knowledge for Voice User Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 55, 10 pages. <https://doi.org/10.1145/3491102.3517623>
- [30] A. Kocaballi, L. Laranjo, and E. Coiera. 2019. Understanding and measuring user experience in conversational interfaces. *Interacting with Computers* 31, 2 (2019), 192–207.
- [31] S. Laumer, F. Gubler, A. Racheva, and C. Maier. 2019. Use Cases for Conversational Agents: An Interview-Based Study. Proceedings of the Americas Conference on Information Systems (AMCIS).
- [32] Minha Lee, Lily Frank, Yvonne De Kort, and Wijnand IJsselstein. 2022. Where is Vincent? Expanding Our Emotional Selves with AI. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. <https://doi.org/10.1145/3543829.3543835>
- [33] T. Lewandowski, J. Delling, C. Grotherr, and T. Böhmman. 2021. State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review. Proceedings of Pacific Asia Conference on Information Systems (PACIS).
- [34] T. Lewandowski, M. Heuer, P. Vogel, and T. Böhmman. 2022. Design Knowledge for the Lifecycle Management of Conversational Agents. Proceedings of Wirtschaftsinformatik (WI).
- [35] Michal Luria and Stuart Candy. 2022. Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 419, 13 pages. <https://doi.org/10.1145/3491102.3517536>
- [36] A. Marinho, A. Poulsen, M. Kroesen, and C. Chorus. 2021. Perspectives about Artificial Moral Agents. *AI and Ethics* 1 (2021), 477–490.
- [37] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Lake Buena Vista, FL, USA) (ESEC/FSE 2018). Association for Computing Machinery, New York, NY, USA, 729–733. <https://doi.org/10.1145/3236024.3264833>
- [38] R. Moore and R. Arar. 2019. *Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework*. ACM Books.
- [39] R. Moore, E. Liu, S. Mishra, and G. Ren. 2020. Design Systems for Conversational UX. UI@CHI Workshop Paper, CUI'20.
- [40] D. Murray-Rust, I. Nicenboim, and D. Lockton. 2022. Metaphors for designers working with AI. In *Proceedings of the Design Research Society Conference (DRS)*.
- [41] I. Nicenboim, E. Giaccardi, and J. Redström. 2022. From explanations to shared understandings of AI. In *Proceedings of the Design Research Society Conference (DRS)*.
- [42] A. Nichol. 2020. Conversation-Driven Development. <https://rasa.com/blog/conversation-driven-development-a-better-approach-to-building-ai-assistants/>
- [43] J. Rezwana and M-L. Maher. 2022. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interactions* (2022).
- [44] David J. Roedl and Erik Stolterman. 2013. Design Research at CHI and Its Applicability to Design Practice. In *Proceedings of the SIGCHI Conference on*

- Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 1951–1954. <https://doi.org/10.1145/2470654.2466257>
- [45] A. Sameera and J. Woods. 2015. Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications (IJACSA)* 6, 7 (2015), 72–80.
- [46] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 857–868. <https://doi.org/10.1145/3196709.3196772>
- [47] A. Selbst. 2019. Accountable Algorithmic Futures. <https://points.datasociety.net/building-empirical-research-into-the-future-of-algorithmic-accountability-act-d230183bb826>
- [48] Dilruba Showkat and Eric P. S. Baumer. 2022. "It's Like the Value System in the Loop": Domain Experts' Values Expectations for NLP Automation. Proceedings of the Designing Interactive Systems (DIS) Conference.
- [49] B. Sihneiderman. 2020. Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction (THCI)* 12, 3 (2020).
- [50] J. Stray. 2020. Aligning AI Optimization to Community Well-Being. *International Journal of Community Well-Being* 3 (2020), 463.
- [51] P. Sukhwal, W. Cui, and A. Kankanhalli. 2023. Knowledge Transfer between Humans and Conversational Agents: A Review, Organizing Framework, and Future Directions. Proceedings of the Hawaii International Conference on System Sciences.
- [52] L. Tudor, D. Dhinakaran, B. Kyaw, T. Kowatsch, S. Joty, Y. Theng, and R. Atun. 2020. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *Journal of Medical Internet Research (JMIR)* 22, 8 (2020).
- [53] K. Vasileiou, J. Barnett, and S. Thorpe. 2018. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology* 18, 148 (2018).
- [54] M. Wahde and M. Virgolin. 2022. *Conversational Agents: Theory and Applications*. World Scientific Publishing Company.
- [55] S. West, M. Whittaker, and K. Crawford. 2019. Discriminating Systems: Gender, Race, and Power.
- [56] J. Whittlestone, R. Nyrupe, A. Alexandrova, and S. Cave. 2019. The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.
- [57] B. Whitworth and A. Ahmad. 2014. *Socio-Technical System Design*. Interaction Design Foundation.
- [58] R. Wolff, S. Hobert, and M. Schumann. 2022. Chatbot Introduction and Operation in Enterprises – A Design Science Research-based Structured Procedure Model for Chatbot Projects. In *Proceedings of the Hawaii International Conference on System Sciences*.
- [59] X. Yang, M. Aurisicchio, and W. Baxter. 2019. Understanding Affective Experiences with Conversational Agents. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
- [60] S. Yuan, B. Brüggemeier, S. Hillmann, and T. Michael. 2020. User Preference and Categories for Error Responses in Conversational User Interfaces. Proceedings of the Conference on Conversational User Interfaces (CUI).
- [61] Sabah Zdanowska and Alex S Taylor. 2022. A Study of UX Practitioners Roles in Designing Real-World, Enterprise ML Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 531, 15 pages. <https://doi.org/10.1145/3491102.3517607>