# Listening to the Voices: Describing Ethical Caveats of Conversational User Interfaces According to Experts and Frequent Users

Thomas Mildner University of Bremen Bremen, Germany

Marion Bartl University College Dublin Dublin, Ireland

Diego Garaialde University College Dublin Dublin, Ireland

Nina Wenig University of Bremen Bremen, Germany Orla Cooney University College Dublin Dublin, Ireland

Gian-Luca Savino University of St.Gallen St.Gallen, Switzerland

Leigh Clark Bold Insight UK London, Great Britain

Rainer Malaka University of Bremen Bremen, Germany Anna-Maria Meck LMU Munich Munich, Germany

Philip R. Doyle University College Dublin Dublin, Ireland

John Sloan Trinity College Dublin Dublin, Ireland

> Jasmin Niess University of Oslo Oslo, Norway

### **ABSTRACT**

Advances in natural language processing and understanding have led to a rapid growth in the popularity of conversational user interfaces (CUIs). While CUIs introduce novel benefits, they also yield risks that may exploit people's trust. Although research looking at unethical design deployed through graphical user interfaces (GUIs) established a thorough understanding of so-called dark patterns, there is a need to continue this discourse within the CUI community to understand potentially problematic interactions. Addressing this gap, we interviewed 27 participants from three cohorts: researchers, practitioners, and frequent users of CUIs. Applying thematic analysis, we construct five themes reflecting each cohort's insights about ethical design challenges and introduce the CUI Expectation Cycle, bridging system capabilities and user expectations while considering each theme's ethical caveats. This research aims to inform future development of CUIs to consider ethical constraints while adopting a human-centred approach.

### **CCS CONCEPTS**

• Human-centered computing  $\rightarrow$  Empirical studies in HCI; Interaction design theory, concepts and paradigms; Ubiquitous computing; • Social and professional topics  $\rightarrow$  User characteristics.

### **KEYWORDS**

CUI, conversational user interfaces, conversational agents, voice agents, chatbots, thematic analysis, ethical design, deceptive design patterns, dark patterns

CHI '24, May 11-16, 2024, Honolulu, HI, USA

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA,* https://doi.org/10.1145/3613904.3642542.

#### **ACM Reference Format:**

Thomas Mildner, Orla Cooney, Anna-Maria Meck, Marion Bartl, Gian-Luca Savino, Philip R. Doyle, Diego Garaialde, Leigh Clark, John Sloan, Nina Wenig, Rainer Malaka, and Jasmin Niess. 2024. Listening to the Voices: Describing Ethical Caveats of Conversational User Interfaces According to Experts and Frequent Users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3613904.3642542

### 1 INTRODUCTION

Conversational systems have been around since the mid-20th century [17, 99]. However, recent technological advances in natural language processing and understanding have led to conversational user interface (CUI) interactions becoming common in many people's daily lives [50]. Today, CUIs come in two key forms: text-based chatbots, which are often used in commercial settings to fulfil customer service roles (e.g., IBM Watson <sup>1</sup>, Amazon Lex <sup>2</sup>), and voice assistants, which are commonly integrated into smartphones and smart home devices (e.g., Apple's Siri <sup>3</sup>, Amazon Alexa <sup>4</sup>, Google Assistant <sup>5</sup>). Following the claims of CUI developers, the technology promises novel ways to interact with service providers efficiently, intuitively, and seamlessly [87, 92]. The rapid growth of these technologies is mirrored by increased interest among researchers within the human-computer interaction (HCI) community [27, 83, 91].

Yet, many users experience frustration when engaging with CUIs [23] that negatively impacts their overall experience. According to the literature, frustrations stem from exaggerated expectations among users regarding communicative competence that current CUIs cannot live up to. For instance, anthropomorphic CUI

<sup>© 2024</sup> Copyright held by the owner/author(s).

<sup>&</sup>lt;sup>1</sup>https://www.ibm.com/watson

<sup>2</sup>https://aws.amazon.com/lex/

<sup>3</sup>https://www.apple.com/siri/

<sup>\*</sup>https://alexa.amazon.com/ 5https://assistant.google.com/

design can lead to unrealistic expectations of near-human-level performance when interacting with such devices [26]. However, what users generally experience is a call-and-response type interaction rather than a free-flowing conversation. While recent efforts in generative artificial intelligence (AI) mitigate some of these shortcomings of traditional CUI interactions, recent work shows similar limitations remaining in large language models (LLM) [49]. In a similar vein, privacy and trust-related issues are said to limit the scope of tasks that users are willing to perform with a system [26, 62]. The lack of transparency around how personal data is recorded and used can also become an obstacle for users in taking full advantage of CUI capabilities, again limiting the contexts in which people are willing to use them [23, 84]. To mitigate these issues, recent work offers domain-specific design heuristics [56] or frameworks to guide user engagement [100]. Still, further guidelines are needed as the technology advances from its infancy. Suggesting a lack of sufficient guidelines, practitioners adopt graphical user interfaces (GUIs) best practices to account for the unique affordances that CUIs present [16, 46, 56]. However, we ought to be wary of the ethical implications as recent work in HCI voices serious concerns regarding unethical practices in contemporary GUIs - providing a taxonomy of deceptive design patterns often referred to as "dark patterns", which inhibit users' decision-making to benefit service providers [12]. Although malicious incentives prevail in CUIs [19, 79], we currently have a limited understanding of how unethical design manifests in the context of CUIs and which ethical caveats require consideration. We, therefore, see an opportunity to proactively address concerns akin to recently addressed deceptive design issues that required legislative and regulatory actions [18, 60, 81] to protect users in GUI contexts.

To that end, we aim to gain insights regarding ethical caveats for CUI design. Here, the dark pattern discourse supports this endeavour by providing a novel angle to understand unethical design in problematic CUI interactions. In the context of this work, we refer to ethical caveats as considerations practitioners need to make to avoid adopting design strategies or features that undermine users' autonomy or deceive them into making choices that are not necessarily in their best interests. To that end, this work explores ethical concerns around the design of current CUI systems, as well as potential issues that may arise in the future, among three specific cohorts: (1) researchers who focus their work on CUIs, (2) practitioners who develop CUIs, and (3) frequent users who engage with CUIs at least once a week. To our knowledge, this work presents the first to consider multiple perspectives to assess CUIs based on ethical caveats. In total, we interviewed 27 people to address the following research question:

**RQ:** Which ethical caveats should be considered when designing CUI interactions, and how should they be addressed?

In answering this research question, this work has two main contributions. Firstly, five themes outline ethical caveats in CUI design: Building Trust and Guarding Privacy, Guiding Through Interactions, Human-like Harmony, Inclusivity and Diversity, and Setting Expectations. Secondly, we introduce the CUI Expectation Cycle, a framework promoting ethical design considerations by incorporating our five themes. This framework responds to repeated demands

for design guidelines to mitigate problematic interactions and negative user experiences expressed by practitioners while addressing users' concerns and expectations.

### 2 RELATED WORK

The related work of this paper encompasses the current development of ethical awareness in HCI and CUI research. The section begins with recent work from the CUI community, addressing design and technology limitations and showcasing a need for explicit guidelines for practitioners. We will then briefly outline work addressing unethical design practices (a.k.a. dark patterns), currently most widely discussed in GUI research. The lack of such work in CUI research emphasises a need to consider similar issues when developing chatbots and voice assistants and motivates our research. Lastly, we will review recent advances in developing CUI-related guidelines and best practices.

### 2.1 Design and Technology-based Limitations

Due to the distinct affordances they present, CUIs have to overcome design challenges that cannot simply be borrowed from other fields [70]. In particular, the use of language, uttered or typed, as a primary input for CUIs poses problems regarding the ease with which users can explore their functionalities [9, 15, 96] or restricting the assessment of a system's capabilities and boundaries [30, 47, 62]. This is affirmed by users' inability to recall the exact command needed for a response from their device [21, 79], limiting interactions to memorised prompts.

Further difficulties are linked to expectations users have toward CUIs. These expectations are steep as it is proposed that users can talk to their devices intuitively via a principal way of communication: language. However, in daily use, CUIs often fall short of their users' expectations [23, 26, 62, 95]. The degree of dissatisfaction this encourages among users has even led some experts in the field to suggest practitioners should rethink the "Conversational" part of "Conversational User Interfaces" altogether [86]. High user expectations also stem from anthropomorphic design features that encourage users to see CUIs as "social actors", a term coined by Nass and Brave in their Computers are Social Actors (CASA) paradigm [71]. Anthropomorphic design has measurable effects through users aligning their speech patterns when talking to CUIs, as explored by Cowan et al. [22]. Although anthropomorphic design can serve beneficial purposes such as increased technology acceptance [22, 54, 90, 96] and can influence the user experience positively [96], there are several usability [54] and ethical problems [4, 32] associated with overusing anthropomorphic design features. Lacey and Cauwell [55], for instance, studied dark patterns in connection to social robots, which could exploit "cuteness" attributes to coerce users' decisions and are only possible through anthropomorphic design.

While Seymour and Van Kleek [90] found a link between relationship development and anthropomorphic CUI design, Kontogiorgos et al. [54] voice a note of caution: although an anthropomorphic social agent led to a higher degree of engagement and sociability compared to a smart speaker without human-like traits, they also found that task-completion time with an embodied anthropomorphic agent was 10% higher compared to a disembodied smart

speaker. They link this to the anthropomorphic agent being perceived as a "socially present partner in conversation" [54, p 138]. In a similar vein, Lee et al. [59] note a preference in users towards physically present CUIs over invisible ones based on a user study describing their mental model when interacting with CUIs. In recent work, Dubiel et al. [28] spotlight a connection between trust and anthropomorphic traits of CUIs. Realising ethical caveats, the authors propose four design strategies to calibrate trust while avoiding anthropomorphic features to enhance user engagement. Hence, whilst anthropomorphic agents often seem to be preferred, they can lead to increased expectations and inadequate attributions of capabilities. This makes the design and implementation of human-like cues a delicate balancing act. Including human traits in CUI design can foster engagement and acceptance. However, concurrently, potential negative consequences for usability if users overestimate their CUI's abilities need to be considered.

Problems also arise from more general technical difficulties associated with datasets used to train LLMs that underpin CUI capabilities. Specifically, LLMs are said to contain and reproduce highly biased worldviews, leading them to "overrepresent hegemonic viewpoints" [4, p 610]. A lack of data representing marginalised social groups increases their marginalisation even further. This is the case, for instance, with older adults, users with speech impairments or pronounced idiosyncrasies, or speakers with a strong colloquial dialect, vernacular, and/or accent for whom speech recognition performs poorer than for standard language speakers [3, 44, 53, 65, 85]. As a result, diverse cultural experiences and identities are undercatered, leading to "othering" referring to the alienation of certain (social) groups resulting in marginalisation or exclusion in social contexts [65]. Where contemporary and commercially available CUIs initially promise natural language understanding and natural interactions, this promise only holds for a certain group of users. For example, African-American Vernacular English (AAVE) speakers need to adapt their speaking style and/or apply less natural speech patterns, like code-switching, to cater for a CUI's limited capabilities [44]. These technological and design limitations hint at a lack of user-centred approaches at the end-user's expense. This is highlighted in work done by Blair and Abdullah [5], who identify the challenges of deaf and hard-of-hearing individuals using smart assistants in daily contexts. If marginalised groups are neglected in development processes, many will be unable to use the resulting systems [5, 44], posing unethical consequences [45]. This work considers different perspectives and, in part, investigates how marginalised groups can be better addressed in the design phases of CUIs.

### 2.2 Summary of Dark Patterns

Research on dark patterns has illustrated a wide range of applications where problematic design occurs [37, 64, 69, 101] while demonstrating a limited focus in the scope of CUIs [19, 79]. Adjacent literature guides our work to understand design areas that require further attention from practitioners to avoid unethical design in CUIs. After Brignull coined the term "dark patterns" in 2010 [12] and initialised a first set of twelve dark patterns, numerous researchers have set out to describe unethical practices in GUI interfaces. In 2022, however, Brignull promoted the term "deceptive

designs" instead of dark patterns to address the risk of the term being racially misappropriated [12]. Recently, the ACM Diversity, Equity, and Inclusion Council [1] added the term to their list of controversial terminologies. However, critique against the term "deceptive design" has been voiced as it is deemed too vague to describe the scope and precision of its predecessor, lacking reference to pattern language [2] while linking "dark" not to malicious intent but something that is hidden [76]. Another argument is that such interfaces do not only deceive but also obstruct, coerce or manipulate users [64, 76]. We acknowledge that, at the time of writing this paper, no perfect term exists to convey all unethical and problematic issues. As the community seeks a better term [39], we opted to use the term "dark patterns" in this work to maintain consistency and continuity with previous research.

Today, the dark patterns terrain has widened throughout numerous domains, including, but not limited to, mobile applications [24], e-commerce [37, 63], and social media [43, 64, 68, 69, 88]. Collectively, dark pattern research has produced a taxonomy of individual design strategies (for an overview, we refer to Gray et al.'s ontology [38]). In an attempt to capture the diverse nature of dark patterns, Mathur et al. [63] describe five characteristics: asymmetry, covert, deceptive, hides information, and restrictive (see Appendix A, Table 4 for a more detailed overview of the characteristics). The authors ground their work on prior dark pattern research and their study based on a large-scale investigation of over 11,000 shopping websites. Each characteristic is introduced as a dimension where dark patterns manifest, while a single dark pattern can contain aspects of multiple characteristics. Moreover, characteristics are described through distinct mechanisms that restrict informed decisionmaking (e.g. user interfaces promote specific choices over others or hide relevant information from the user). Our work utilises these characteristics during the interviews to learn about interviewees' views on potential unethical designs in CUIs.

Despite this growing field of research, most work focuses on GUI artefacts, and only limited work considers dark patterns in the context of CUIs. Surveying CUI users on twelve scenarios involving voice-based systems, Owen et al. [79] establish a ground for future work by highlighting the importance of considering dark patterns in this environment, which finds further support by the provocation by Mildner et al. [66] from the same year. The need for further research is also mirrored in a recent review by De Conca [19], who showcases the presence of known dark patterns in speech-based CUI systems. Thereby, the work not only highlights potential differences of dark patterns in CUIs compared to their GUI counterparts. The work further concerns necessary regulatory actions, emphasising the need for more tailored considerations, as contemporary regulation mainly concerns GUI dark patterns.

Importantly, studies have repeatedly shown difficulty among users to recognise and identify dark patterns sufficiently [24], or with low accuracy [7], even if provided information about dark patterns [7, 67]. Users' inability to safeguard themselves places them in a vulnerable and exploitable situation that requires particular design considerations. In this context, it would be beneficial for practitioners to be guided by ethically aligned guidelines, which could aid in the development of user-centred systems that potentially empower users to make informed decisions.

Setting out to protect end-users of technologies, we aim to increase our understanding in this field by consulting practitioners, researchers, and frequent users. Led by the momentum of this discourse, we aim to draw attention to ethical caveats in CUI design as the technology becomes more ubiquitous. As practitioners often work under diverse ethical constraints, leading to the unconscious or unwilling deployment of unethical design [35], we aim to provide concepts to avoid implementing dark patterns to begin with.

### 2.3 Guidelines & Best-Practices

Spanning work including Nielsen's ten usability heuristics [73] and Friedman et al.'s Value Sensitive Design [31], HCI has produced a range of important frameworks, guidelines, and best practices to aid practitioner' efforts. Yet, often recorded frustration and negative experiences of CUI users indicate a lack of application or applicability of these aids in this domain. This issue has been addressed by various researchers [46, 56, 70]. Ghosh et al. [33] highlights a lack of accuracy for the system usability score [13, 14] — a prominent measure of subjective perceptions of usability of systems - when used to evaluate CUIs. To address this problem, recent work investigates the possibility of transferring concepts known to work for GUIs and other interfaces toward the context of CUIs. Langevin et al. [56], for instance, adapt Nielsen's heuristics within the context of CUI interaction. Similarly, Klein et al. [51] revise the widely used UEQ questionnaire [58] and extend it with scales for response behaviour, response quality, and comprehensibility to mirror aspects important for the user experience of CUIs. Indeed, various reviews of CUI research note the lack of validated measures of user perception as an ongoing problem that calls into question the reliability of this body of work due to a lack of continuity in how concepts are defined [16, 52, 89]. To date, there has been only one validated scale of this nature available to CUI researchers, known as the partner modelling questionnaire (PMQ) [25].

Still, guidelines tailored explicitly to CUIs are scarce. Additionally, they often exclude unique requirements of vulnerable groups who could especially benefit from interactions with hands-free and speech-based devices, as is the case for users with, for example, visual impairments [9]. Through our established themes and our framework, our work aims to provide some mindful guidance for practitioners and researchers to utilise a user-centred approach and reach a variety of different user groups.

### 3 METHOD

This study aimed to gain insights regarding CUI-specific ethical caveats and identify ways unethical practices manifest within CUIs. We, therefore, conducted a total of 27 semi-structured interviews split between three groups. The first group includes researchers focusing on CUI-related topics (N=9). The second cohort comprises practitioners who work in industries developing CUI technologies (N=8), whilst the third encompasses frequent users of CUI systems (N=10). The interviews were conducted online via video conference tools, which allowed for the recruitment of participants with an international scope.

### 3.1 Interview Protocol

The interview consisted of two parts. The first part focused on participants' general experience with CUIs. The second part is based on and inspired by Mathur et al.'s dark pattern characteristics, promising interesting insights into unethical design strategies by describing design choices and mechanisms that prohibit informed decision-making. As Mathur et al.'s definitions were constructed to convey similarities between certain groups of dark patterns, the original definitions were neither created for an interview context nor designed to account for CUI-based interaction. To address these limitations, we adapted the original questions to foster more relevant answers from our participants. While this enabled us to learn about participants' views on dark patterns in this context, we also added three questions targeting specific situations and demographics to gain a deeper understanding of circumstantial issues. The full interview protocol and the five dark pattern characteristics according to Mathur et al. [63] are included in the Appendix A. During the interview, participants were conditionally prompted to think about text-based and voice-based systems, both concerning their actual experiences and hypothetical future scenarios they might envisage.

### 3.2 Participants

In total, 27 participants were recruited for interviews. Researcher and practitioner cohorts were recruited from the authors' collective professional network and word of mouth, enlisting academic and commercial researchers, designers, and developers whose work focuses on CUIs. The third group, frequent users, were recruited via the online platform Prolific [61]. Recruitment criteria were used to ensure participants in this cohort were at least 18 years of age and used CUIs at least on a weekly basis, though no particular CUI system was stipulated. All participants were informed about the nature of the study, what participation involved, and their data rights before being asked to provide informed consent. Participants recruited via Prolific were rewarded with a £10 honorarium. This is in keeping with suggested hourly rates for ethical payments for research involving crowdworkers [57]. Researchers and practitioners participated voluntarily and were not furnished with an honorarium. Table 1 presents a full overview of all recruited participants. The interviews lasted an average of 42:30 minutes (SD = 11:29). Interviews with the researcher cohort took the longest (mean=48:00, SD = 08:48), while interviews with those held practitioners were slightly shorter on average (mean=44:47, SD = 13:50). The interviews conducted with frequent users were the shortest (mean=36:03, SD = 08:28).

Researchers. Eight academic researchers were recruited for our study (three female, four male, one preferred not to disclose their gender), four of whom were professors, one a postdoctoral researcher, and three who were PhD candidates. When conducting this study, the researchers' average age was 37.0 (SD = 13.0), and their average years of experience researching CUIs was 13.9 years (SD = 14.4).

ID	Age	Gender	Country of Residence	Occupation	Years of Experience
	1180	Gender	<del>-</del>	archer	Tears of Experience
 R1	38	male	USA	Professor	16
R2	27	female	USA	Professor	11
R3	35	male	USA	PhD Candidate	10
R4	33	female	South Korea	Assistant Professor	5
R5	69	male	United Kingdom	Professor	48
R6	32	male	Germany	Postdoctoral Researcher	12
R7	30	female	Switzerland	PhD Candidate	4
R8	30	not disclosed	Ireland	PhD Candidate	5
Mean	= 36.75			Mean	= 13.88
SD	= 13.46			SD	= 14.4
 P1	56	male	United Kingdom	Chief Science Officer	15
P2	49	female	USA	Manager	21
P3	40	male	USA	Educational Research	2
P4	32	male	UK	Tech. Consulting Manager	6
P5	40	female	Brazil	Research	9
P6	38	female	USA	Conversation Designer	10
P8	64	female	USA	Communication Consultant	35
P9	42	female	Ireland	Designer	15
P10	58	male	USA	Digital Business Executive	25
Mean	= 46.56			Mean	= 15.33
SD	= 10.74			SD	= 10.28
			Freque	nt User	
ID	Age	Gender	Country of Residence	Current Occupation	Highest Level of Education
 F1	26	female	Mexico	ESL Teacher	Undergraduate
F2	26	female	South Africa	Customer Service Rep	Postgraduate (or higher)
F3	30	male	South Africa	Media Analyst	Secondary/Vocational
F4	33	male	South Africa	Construction	Secondary/Vocational
F5	24	male	South Africa	Student	Undergraduate
F6	25	female	South Africa	Student	Postgraduate (or higher)
F7	57	male	United Kingdom	Retired	Postgraduate (or higher)
F8	45	female	Italy	Remote Freelancer	Secondary/Vocational
F9	22	male	Poland	Frontend Developer	Undergraduate
F10	29	female	Mexico	Healthcare Entrepreneur	Undergraduate
Mean SD	= 31.7 = 11.02				

Table 1: This table presents our interview participants in three cohorts: researcher, practitioner, and frequent users. Notably, we did not have a participant P7. The associated participant cancelled the interview after IDs were already given to all participants. To avoid confusion, we retained our initial structure.

*Practitioners.* Nine practitioners participated in our interviews (five female, four male) with roles in conversation design (2), executive management (4), research (2), and consultation (2). Practitioners' average age was 47.0 years (SD=11.0), and they had an average of 15.3 years (SD=10.3) of experience working in this space.

*Frequent Users.* Finally, we recruited ten individuals for the frequent users' cohort (five female, five male; mean age = 31.7 years,

SD=11). To qualify as frequent users in this study, participants of this cohort had to use CUIs at least once per week basis. However, we did not require participants of this group to have long-term experience as the technology is still relatively young. We also asked them about the kinds of devices they used most often. Seven stated that they mostly accessed CUIs through their smartphones, whilst the other three mostly used smart speakers. CUIs used included

Amazon's Alexa, Apple's Siri, Google Assistant, Microsoft's Cortana, Samsung Bixby, and text-based chatbots from online services. Six said they used CUIs daily, while four stated using CUIs more than once a week. The highest level of education among this cohort was: undergraduate (4); secondary/vocational (3); postgraduate or higher (3). Participants' occupations span construction (1), customer service (1), front-end developer (1), media analyst (1), medical laboratory scientist (1), self-employed healthcare entrepreneur (1), students (2), and teacher (1).

### 4 THEMATIC ANALYSIS

After completing all 27 interviews, we transcribed and prepared the recorded material for analysis, de-identifying any traceable or personal information. Data transcription was carried out by a professional UK-based service provider. Based on these data, we conducted a reflexive thematic analysis [11], which was divided into four phases: familiarisation; code generation; construction of themes; and revising and refining themes.

### 4.1 Positionality

Authors of this work lived, gained education, and worked in Central Europe most of their lives, with WEIRD (Western, Educated, Industrialised, Rich, and Democratic) [45] backgrounds. Their research backgrounds contain expertise in design, linguistics, computer science, and psychology, with scholarly work oriented toward social justice and well-being topics. The interview transcripts were coded by four authors who have previously engaged in scholarly work in human-computer interaction, computer science, and psychology, each contributing more than three years of experience in their respective domains. At the time of conducting the study, no personal or professional conflicts with CUI systems or their developers occurred among the authors. Participants of the researcher and practitioner cohorts were recruited through professional networks with some personal relations. Where an author had a rather close relationship with a participant, we ensured that another author would conduct the interview instead. Frequent users were recruited through Prolific [61], where no previous relationships existed. Focused on ethical considerations in CUI design, this work is partly inspired by Mathur et al.'s [63] dark pattern characteristics to analyse problematic CUI interactions and propose counter-measures. To conclude, we acknowledge potential bias based on our cultural, academic, and personal backgrounds.

### 4.2 Coding of the Transcripts and Identifying Themes

For the first round of data analysis, we selected two interviews per participant group to generate an initial set of inductive codes. Following advice from Braun et al. [11], two researchers coded these six interviews, discussed their strategies and results, and drafted an initial codebook encompassing 65 codes. Through axial coding, the codebook was then reduced to 46 codes. All interview transcripts were then split among four authors. Each interview was thus coded by a single author following Braun et al.'s proposed methodology [10]. Although coding was carried out independently, authors did meet to discuss the procedure once before the coding, once after half of the interviews had been coded, and one last time

after coding had been completed. These discussions ensured coding strategies were aligned between authors, including resolving any issues authors may have had applying specific codes [11] and discussing any potential new discoveries in the data. Due to the collaborative nature of this process, indices of inter-rater reliability [11] are not provided.

Four authors then went on to identify themes for each cohort independently. Based on the codes, quotes, and annotations, each researcher applied affinity diagramming [6] to develop early iterations of these. During this process, these first themes were evaluated and refined by iteratively checking each for applicability to interview material. They were then discussed among researchers to establish agreement. Here, we followed Braun and Clarke's theme mapping [10] to collapse similar themes. If a theme was adapted, it was again revised against the whole dataset and other themes before being accepted. This last phase resulted in the construction of five themes and a corresponding ethical caveat for each.

### 5 FINDINGS

In this section, we describe the constructed themes based on reflexive thematic analysis. We highlight these themes across each of the interview cohorts individually. Notably, a degree of overlap is to be expected as themes tend to be intertwined, often in a supportive fashion. In total, we synthesised five high-level themes summarised in Table 2, where each theme is listed next to a statement for a particular ethical caveat. Additionally, we formulated guiding questions that refer to the ethical caveat to support the design of CUI interactions. Before each theme is addressed in detail, we present a high-level overview of our findings. The outline for each theme follows the same structure: a description followed by the perspectives of each of the three cohorts interviewed (researchers, practitioners, and users).

### 5.1 Building Trust and Guarding Privacy: Operating Extrinsic and Intrinsic Factors

This theme spotlights extrinsic and intrinsic challenges that result in trust and privacy deficits in CUI interactions. Across cohorts, a reoccurring theme echoed fears of untrustworthy handling of personal data, kindled by prior negative experiences and the reputation of companies and practitioners. A call from researchers to increase transparency to bridge users' concerns was underlined by users' mention of their own safeguarding strategies, which limit interactions to basic functionalities and do not require them to disclose private information. Practitioners acknowledged these problems but noted counterintuitive regulations and design limitations obscuring access to settings rooted in speech-based interactions.

Researcher. Researchers interviewed were aware of trust-related problems of CUIs, suggesting a need for greater transparency, "it's really difficult for users to know how data travels across different platforms" (R4). Connecting this problem to the "reputation of the companies behind each device" one participant said that "there's very little transparency [...] about how that data is being used and processed" (R3). Another link is drawn to deceptions based on humanlike features with which "we are undermining the trust when we are projecting something which isn't real" (R5). Meanwhile, there are "ethical issues with data sharing and things like that, but [users] care

Theme	Ethical Caveat	Guiding Question	
Building Trust and Guarding Privacy	Users feel vulnerable to use CUIs, posing a need for CUI developers to prioritise transparency and control over data handling.	Does the system/interaction provide accessible and transparent information about personal data with easy control thereof?	
Guiding Through Interactions	Guidelines and frameworks need to educate developers to develop accessible CUIs that empower users with diverse technological literacy to confidently interact with available features.	Does the system/interaction adequately inform users about its technical capabilities to enable full utilisation of its features?	
Human-like Harmony	Anthropomorphic features should be implemented with care and in line with a CUI's capabilities to support intuitive and authentic interactions, preventing unrealistic expectations.	Does the system/interaction clarify the presence of anthropomorphic features to avoid misconceptions and unrealistic expectations?	
Inclusivity and Diversity	The development and design of CUI interactions need to consider individual needs and characteristics of users, especially marginalised groups, ensuring equitable CUI interactions.	Does the system/interaction cater towards users with diverse needs, potentially through alternative interactions where otherwise inaccessible?	
Setting Expectations	CUI capabilities should avoid deceptive interactions and, instead, be transparent to users to prevent frustration and mistrust.	Does the system/interaction handle user prompts truthfully, clarifying the scope of its capabilities to provide realistic expectations?	

Table 2: This table summarises the five identified themes and design questions per ethical caveat.

more about: 'Can I get to what I need to get to fast enough?'" (R4). Looking at users' uncertainty from a technological lens, another researcher discussed how "people didn't know [the CUI] was listening to them all the time [...]. How does it know when I say 'Hey Alexa' or 'Hey Google' [...]? Is it already listening to me?" (R2).

Practitioner. Practitioners also reflected on privacy and trust issues. The current situation users are in when engaging with CUIs was described as "completely wild west, talking to a black hole", because "sometimes there is a lack of transparency, you just don't know what they're doing with your information and that's a problem for a lot of people" (P8). CUI's present some inherent difficulties when users try to access their privacy settings: "That info does seem to be kind of buried [...] you have to click in three or four different menus to get to the privacy stuff and I don't think it's explained" (P6). It is later reflected that users "don't get any kind of introduction to the app itself or the onboarding is very light" (P6). Regarding a lack of technological literacy among users to understand and trust their CUIs, one practitioner asked the question: "How do you prevent people from overhearing my voice and trying to imitate me, how does that thing know that it's not really me?". They went on to provide an answer: "There's a simple explanation. It's too smart, it knows the difference. The technical explanation is, it gets into how it analyses your voice" (P3).

Frequent User. Being the affected cohort, users mentioned various concerns regarding their privacy and why they hesitate to trust CUIs. One participant was "not fully comfortable disclosing [their] personal details even if it's a virtual assistant" (F5). Reflecting on potential trust connected to anthropomorphising CUIs, they later

stated that "it feels like you're speaking to a human and you're giving them your personal data" (F5). Talking about giving consent to things they never fully read, one participant said: "I don't think I've read them all, I just think that I know what I'm accepting. So I don't know any of the consequences that I might have with it" (F3). Users are "expected to know the risks to [their] privacy, to data, to enter into the real world. [Users] are expected to know it in advance." (F7). Some worry about exposing critical information publicly, for instance regarding their financial accounts: "It puts me in danger of whoever that's around that can hear how much money I have. Maybe I'm at risk of being robbed" (F2). Despite their concerns, the interviews outlined aspects of a privacy paradox. "Today, there is no alternative [to commercial CUIs]. [...] It is bad, but I don't see [an] alternative" (F8).

# 5.2 Guiding Through Interactions: Overcoming Knowledge Gaps

This theme illuminates the importance of providing users with informed guidance about possible CUI interactions, as a lack of technological literacy and limited experience results in difficulties in accessing available features. While researchers noted that research findings need to be better aligned with industry development, practitioners echoed a desire for further guidelines.

Researcher. Researchers reflected on a lack of communication between stakeholders, as well as peculiarities of CUI interaction that lead to unique restrictions. A level of trust and limited collaboration was emphasised by researchers, who suggested work in academia seems to "make very little impact on what big companies do. [...] If it is

profitable, they're going to do it" (R4). Another researcher speculated that there was a disconnect between the communities that design the components required to build CUIs, and the community of HCI researchers who explore and develop theories that explain how users interact with CUIs: "What I see is a bunch of different technology CUIs just being demonstrated, and then in our particular space [design research] we've got toolkits, and the toolkits define what's possible. [...] I'm struggling really to understand how come there's this disconnect between these communities?" (R5). It was noted that the communication between practitioners and users is also problematic "if users don't care and only developers do" (R4).

Practitioner. Practitioners again referred to a lack of best practices guidelines that would help them create better products. "[...] There's very little that's actually focused on conversational user interfaces... even to the point that we don't have heuristics really. [...] To create these heuristics for CUIs, that hasn't been done, so there's this massive gap out there." (P6). When contrasted against comments by researchers, this further highlights the disconnect between these communities. Practitioners also highlighted the lack of standardised guidelines for users when it comes to how best to interact with these kinds of systems. They suggest users are essentially left to rely on what they know from graphic counterparts: "For voice stuff, what's missing, there is no standard, like you know when you go to a website, and there's that little hamburger menu, those three lines." (P6). Instead, users have to compare their experiences with other, similar systems in the hopes that they operate the same way; "[...] it's still fairly new and [...] design best practice is still kind of evolving. [...] Maybe if there was some more standardisation in how the experience works between different companies or between different interfaces, people will get more used to it" (P4).

Frequent User. Frequent users echo the idea that the burden of figuring out the limitations of a speech agent falls largely on themselves. "Most of the time you have to understand Google Assistant and what its limitations are when you ask it to do specific tasks" (F5). Similarly, users "have to use the correct terms" and "learn to adapt to [their] language set" (F7). In reference to using a CUI on their smartphone, one participant reflects on the limited information available to them: "You'd have to actually sit on your phone and actually test out everything to see what works and what doesn't" (F3), whilst another suggests they draw expectations from smartphone experiences because "they are able [...] and that's what we want from our home devices" (F8).

# 5.3 Human-like Harmony: Providing Authentic Anthropomorphism

This theme describes the importance of implementing an appropriate amount of human-like features so that a system can support intuitive interaction without eliciting false beliefs that may leave users feeling deceived. Finding the right balance between humanness and technological limitations is challenged by the phenomena of anthropomorphism. That is the tendency to attribute human characteristics to non-human objects that most people engage into some degree [98] from early childhood [82]. Anthropomorphous behaviour appears significantly heightened in dialogue with technological

devices endowed with gendered voices and names [34]. Other influences that might encourage anthropomorphous behaviour in this context include: expectations for social affordances implied by representations of speech interfaces in media and advertising [70]; the fact that these systems conduct tasks typically carried out by humans using human language [72]; and that language use itself might be inherently social and agentic [29, 48]. When developing CUIs, practitioners should conscientiously navigate the amount of human-like features to avoid inadvertently manipulating our bias for anthropomorphic characteristics.

Researcher. Researcher interviewees provided thought-provoking ideas to enhance transparency around the ontological nature of speech agents and demonstrated awareness of potential ethical concerns associated with endowing CUIs with human traits. Interviewees suggested CUIs should be designed in such a way "that people are not fooled into thinking it's a social entity when it's just a machine" (R7), and that artificial voices should be designed to enable users to differentiate between humans and machines easily (R1, R5, and R8). Some even went as far as to state that "From an ethical standpoint [a CUI] should announce it's a machine" (R5). This would undoubtedly provide clarity for some users who experience confusion due to "not knowing what social situation [they are] in" (R1). Overall, researchers agreed there should be increased transparency for users when engaging with CUI technologies.

Practitioner. The need to increase transparency was also acknowledged by individuals from the practitioner cohort, who suggested CUIs "should flag the fact that it is a machine talking" (P9), arguing for the introduction of a "watermark [...] that indicate[s] that [the CUI] is fake" (P8). Although "some level of humanisation" was thought to cause no harm, this could be context-dependent (P9). Practitioners felt the responsibility lies in "how the bot identifies itself [...,] how human-like the conversation is" and "the visual representation" (P9). However, there is a risk for users who "still got [...] attached to [CUIs]" (P2), even though the interaction was not particularly "emotional or personal". In this regard, a counterargument mentioned that "people are lonely and [...] there is a benefit to having a virtual companion in some ways" (P2). Another practitioner did not see any issues endowing CUIs with human-like characteristics, believing people can still easily distinguish between humans and a device that "just speaks like a human" (P10).

Frequent User. Many of the frequent users' observations echoed facets of the other cohorts' observations. Demonstrating the kind of bond some users are said to have with their devices, one participant admits that "it really feels like you have your friend walking with you in your pocket" (F5) when referring to the virtual assistant built into their mobile phone. Similarly, another participant finds that "especially voice assistants [...] let users believe that they are human or, to an extent, friendly" (F6). It was felt that a lack of awareness may lead users to "interact [with CUIs] as [they] would with a human without knowing that it is not" (F1), with one participant recalling how elderly family members "treat the machine like a human being" and use it "with courtesy" (F10). However, some also highlighted limitations when interacting with certain CUIs, particularly those implemented in customer service. Expecting a human to help with a

problem, the limited responses and understanding of queries exhibited by some systems were said to create frustration and exacerbate a desire to "talk to a person, a real person" (F10).

### 5.4 Inclusivity and Diversity: CUIs in the Wild

This theme delves into the unexpected design and interaction challenges emerging when CUIs are introduced to diverse user groups with distinct characteristics and needs. Participants spotlighted several groups susceptible to poor design choices or technical limitations within CUI interactions. They emphasised a regression towards the mean, acknowledging how CUIs are designed with an "average" user in mind, leading to the "othering" [65] – the marginalisation or exclusion of certain people – of individuals and groups that do not fit this profile. This encompasses users with stronger accents, colloquial dialects, second language users, people with deficits in speech or cognition, or users with lower technical literacy.

Researcher. Researchers note that in designing CUIs, one has to be aware of varying user abilities and behaviours; "So certainly you have users with diminished capacity that are going to be particularly vulnerable." (R1). This was seen as particularly problematic around consent "if you know that certain groups of users can't give their informed consent, yet they're still using certain technologies, then you'd have to be more careful." (R4)

Practitioner. Practitioners also suggested, "[T]he use of these systems for vulnerable groups is problematic." (P1), whilst also identifying issues faced by minority groups: "in terms of demographic, in terms of like accent, people who have a strong accent because the language that they're trying to interact with isn't their native language. Basically, anyone who doesn't fall into the kind of the centre of the bell curve, I expect, is more susceptible [to problematic CUI design]" (P4).

Frequent User. Frequent users identified a number of groups they envision encountering issues with CUIs; "people who are not technology literate or digitally literate, they might be vulnerable." (F5), "there are certain people who are very gullible to [...] information in general, anything that is said online they absolutely believe." (F2). Frequent users also offered anecdotal examples of individuals that have been [other-ed] by CUIs, explaining that "a user would need to have a fairly good memory to be able to use [CUIs] effectively. I have a friend who is severely disabled, and she just cannot use it because she can't remember the commands." (F2). Another participant shared that "a friend of mine, her son is autistic [...] he has trouble with her Alexa, [...] sometimes he knows the answer but he'll ask the question for information. He'll expect it to conform exactly to what he knows, even using the correct words and he gets so frustrated when it doesn't" (F7).

## 5.5 Setting Expectations: Transparency to Mitigate Frustration

This theme emphasises the pivotal role of transparency when designing CUI interactions to set realistic expectations. Unintended device reactions can result in disappointment and frustration in the user when a device appears more capable than it actually is. To deliver users an optimal experience, commercial incentives should

be aligned with human-centred practices of HCI to avoid ethical concerns tied to deceptive and manipulative interactions. Instead, transparent and evident CUI capabilities should empower users to use the system autonomously and easily.

Researcher. The importance of setting the right expectations for what systems are capable of was noted among researchers for offering users a better experience that is more aligned with reality. Here it was stated that "functionality expectations are not set right" (R2) for contemporary CUIs, with introductions and manuals either not being readily available or ignored by users. Additionally, "people bring their previous experiences to an interaction when they start" (R7), suggesting users with prior knowledge have an advantage over those who are new to CUIs. If true, this highlights a lack of onboarding for novice users, with prior knowledge biasing experiences for frequent users. In this regard, "lack of transparency [...] really makes it easy to misunderstand what the system does, or is capable of doing" (R2).

Practitioner. Practitioners suggest a lack of best practices or guidelines for how and when to communicate a systems' features to users; "getting across to the user what the system can and cannot do is one of the most difficult problems" (P2). Current commercial CUIs are limited to presenting information synchronously on request. As the same metaphors working in GUI environments do not translate readily to CUIs, "discovery [of functionalities] is very difficult right now because there's no directory to help us find it" (P8). While multi-modal chatbots can mitigate certain problems, voice assistants without integrated displays cannot rely on the same design philosophies. "specifically for voice stuff, what's missing, there is no standard" (P6). It was further argued that "when [CUIs] can only answer a fraction of questions it is inherently putting the burden on the user to rephrase their question over and over" (P6). This difficulty is increased "if there are a lot of options, it's probably more of a challenge to make people aware of the choices that they have" (P4). In relation to human-like features, how CUIs "are designed can make them seem more intelligent than they are" (P9), creating unrealistic expectations that will eventually cause frustration in users.

Frequent User. Multiple users expressed frustrations they experience when using their CUIs; if a CUI "fails to give you what you want it to do, you end up having to do it yourself" (F5). An initial problem stems from raised expectations through advertisement of commercial products, where "everything [...] runs smoothly [...] creat[ing] that expectation in your head as that's how it works" (F1). Another participant explains that if "you don't know what [the CUI] has heard [...], you have to go to the app" (F10) as the device is incapable of providing insights into why errors occur. Limiting their expectations, one user adapted their behaviour to "only [use voice-based systems] for easy communication because [...] they are not set yet for more complicated things" (F8). Another user discussed the importance of good memory, which "users need to have [...] to be able to use [CUIs] effectively and that they "need to have read the literature for the [CUI] to know the limits" (F7).

### 6 ADDRESSING EXPECTATIONS: THE CUI EXPECTATION CYCLE

Based on our findings, we constructed the CUI Expectation Cycle (CEC), a framework that practitioners and researchers might use to understand and serve users' needs better. To this end, the CEC aligns user expectations with system capabilities. It aims to enable successful CUI interactions that meet the goals of its users, mitigating unethical design strategies and voiced frustrations of our participants in line with related work [44, 100]. The CEC incorporates the themes identified in our qualitative analysis, including garnered ethical caveats. Further, it partly builds on existing frameworks – Norman's *Action Cycle* [75] and Oliver's Expectation Confirmation Theory [77, 78] – to inform about the variables that shape user expectations.

### 6.1 Constructing the CEC

Drawing from dark pattern literature, our interviews confirm the findings of existing CUI research while providing a detailed overview of the interplay of design challenges. While identifying our themes, we assessed their interrelationships and noticed a strong gap between users' expectations and CUI capabilities, as described by our participants. Focusing on this gap, we distilled the themes into a framework while consulting related work. To this end, the CEC (see Figure 1) links core issues of CUI design with a particular focus on anthropomorphism [25, 26], deception [19, 79], inclusivity [44], and trust [62, 93]. Moreover, the CEC builds on two established frameworks. On the one hand, it utilises Norman's Action Cycle [75] by adopting users' execution and evaluation gulfs; in the case of our framework, we refer to them as bridges as they help to cross the Expectation Gap. We were also inspired by Oliver's Expectation Confirmation Theory [77, 78], a cognitive framework that describes customer expectations and subsequent (dis)satisfaction with purchased goods. Supported by these widely accepted frameworks and related literature, the CEC focuses on creating realistic expectations extended by carefully considering ethical caveats to avoid problematic design in the area of CUIs. Ideally, this framework enables the development of transparent and trustworthy CUI interactions, allowing satisfying and successful user engagement.

### 6.2 The Elements of the CEC

Figure 1 illustrates the CEC and how the five themes relate to individual elements of the framework. Dividing user (at the top of the CEC) and CUI (at the bottom), the critical element of the CEC is the *Expectation Gap* (at the centre of Figure 1). However, the Expectation Gap can be overcome through two bridges that connect the CUI with the user (the Evaluation Bridge, left of the centre) and vice versa (the Execution Bridge, right of the centre). In jeopardising these bridges, two delimiters result from problematic or unethical design decisions (the Deception Delimiter on the outer left side and the Distrust Delimiter on the outer right, each pointing toward the bridges). Above the diagram are the five themes: Building Trust and Guarding Privacy, Guiding Through Interactions, Humanlike Harmony, Inclusivity and Diversity, and Setting Expectations. Each theme is connected to the relevant elements of the diagram through arrows with differing dotted lines and colours. Informed by our themes, the elements of the CEC encompass the ethical

caveats identified as considerations for the design of CUIs. Here, we describe each element more closely, beginning from the centre while highlighting elements and connected themes in *italic* font.

The Expectation Gap encompasses our Setting Expectations theme, which reflects the different perspectives across cohorts regarding technical capabilities and design decisions. As these decisions anchor user goals, it is pertinent to design accessible interactions that address users' needs depending on a CUI's purpose and its situational context. For instance, frequent users reported frustration with the complexity in which CUIs had to be prompted, while practitioners noted how many features were left unused. Consequently, reasonable bridges need to connect both sides. The CEC includes two such bridges (Execution Bridge and Evaluation Bridge) inspired by Norman's [75] gulfs of expectation and evaluation.

To the left of the CEC's centre, the *Evaluation Bridge* starts from the CUI and reaches toward the user. It thereby includes understandable and truthful communication of a CUI's capabilities to inform the user about possible interactions and set realistic expectations. This bridge is strongly connected to the *Guiding Through Interactions* theme, expressing a need to empower users to fully understand possible interactions and their consequences.

The *Execution Bridge*, right of the CEC's centre, starts at the user's side and links it to the CUI. The bridge describes how expectations are shaped into goals a user can realise. However, a variety of obstacles can hinder users from successfully executing interactions. In this regard, the *Inclusivity and Diversity* theme addresses the importance of accommodating users' individual characteristics in the design of CUIs.

In a well-designed system, both bridges support users in making realistic assumptions about a system and using it without frustration. In a sub-optimal system, users may encounter negative experiences and designs lacking ethical considerations, which could adversely affect their well-being, as dark pattern literature indicates [64, 69]. The CEC includes two delimiters fostering the negative influences of the two bridges. The Deception Delimiter (on the diagram's outer left side and pointing to the Evaluation Bridge) has a detrimental effect on the Evaluation Bridge, resulting in unrealistic expectations. Similarly, the Distrust Delimiter (on the diagram's outer right side and pointing toward the Execution Bridge) adversely affects the Execution Bridge by decreasing users' faith in the system and its responses. The two delimiters are informed by our Building Trust and Guarding Privacy and Human-like Harmony themes as they advocate transparent and authentic interactions. The different perspectives cast across our participants highlighted how each delimiter can be intrinsic and extrinsic in nature. To illustrate, previous individual experiences impact how a user engages with a CUI (i.e. intrinsic in nature), while design decisions influence how a CUI is perceived (i.e. extrinsic in nature). Importantly, either delimiter can be the source of unethical design practices commonly found in dark patterns.

### 6.3 Using the CUI Expectation Cycle

Although previous research has repeatedly addressed a gap between users' expectations and CUI devices [26, 56, 62], our participants, especially practitioners, requested design guidelines to overcome

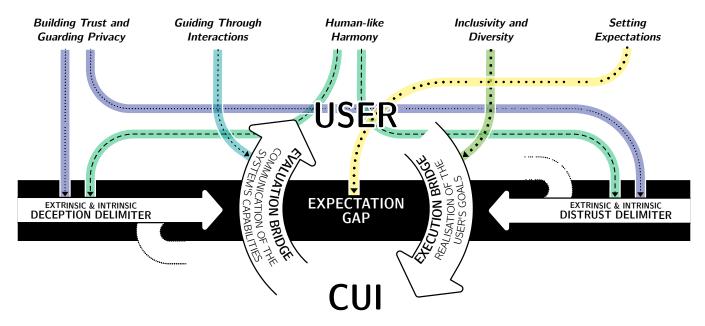


Figure 1: The CUI Expectation Cycle demonstrates how an expectation gap between user and system can be bridged by an evaluation and an execution bridge, inspired by Normans' two evaluation and execution gulfs [75]. However, deceptions can impact how users assess the system's capabilities, resulting in unrealistic expectations. Similarly, distrust limits users' faith in the system and its responses, influencing their execution of actions. Both delimiting factors can be of extrinsic as well as intrinsic nature. The diagram further features the five themes and connects them to respective sections of the CUI Expectation Cycle.

related issues. The CEC responds to these demands through a user-centred approach, embedding ethical caveats to avoid problematic designs. While the CEC could be used as a standalone framework, we believe it is best utilised next to contemporary efforts in CUI research to provide additional design considerations and constraints as alternatives to GUI-related best practices. To support practitioners' work, we created questions derived from the ethical caveats that allow tracing of related limitations (see Table 2). Following the CEC (Figure 1), practitioners can consult each theme in the corresponding design areas to ensure that ethical caveats are respected.

Generally, CUIs promise benefits for people with diverse abilities, like visual impairments [9]. However, studies demonstrating difficulties among marginalised demographics [44] suggest that these benefits are not exhausted, addressed in our *Inclusivity and Diversity* theme. CUIs, such as Amazon Alexa, often require additional GUI input to control device settings or personal data, restricting access. To cater to diverse user groups, practitioners could consider including alternative interactions that allow simple execution of planned goals. Asking our guiding question – "Does the system/interaction cater towards users with diverse needs, potentially through alternative interactions where otherwise inaccessible" – would remind practitioners about the importance of accessibility in their systems.

Taking the *Human-like Harmony* theme, for example, anthropomorphism can impact the deception and distrust delimiters. To mitigate deriving consequences, chatbot systems should clarify if no human is in the loop and state the scope of possible prompts. Design borrowing from other chat interfaces can cause additional confusion but could easily be avoided through a distinctive interface design. Similarly, anthropomorphic voice-based CUIs may confuse

some users, particularly those with little technical literacy, leveraging their expectations. Appropriate voice design could clarify a CUI's artificial nature, clarifying related misconceptions. Asking our guiding question – "Does the system/interaction clarify the presence of anthropomorphic features to avoid misconceptions and unrealistic expectations?" – helps to unravel otherwise deceiving design choices.

In line with prior work [56], our findings demonstrate that simple adoption of GUI best practices do not readily translate to CUI interactions as heuristics differ. While GUIs utilise graphical icons, buttons, or links [79], with some exceptions, CUIs' general input lie in uttered or typed-in commands. Though CUIs, including a GUI, can use these interactions, voice-based devices often require prompts to be spoken only. In most current CUIs, however, interactions require users to remember the correct prompts for the expected response, as recalled by some participants. Consequently, practitioners should account for human errors, support them better during interactions, and offer suitable alternatives. Adequate onboarding of the user is a crucial step to lower barriers and set expectations. While the CEC cannot account for malintent, we envision following its recommendations facilitates a better user experience and helps mitigate risks for users in consideration of our proposed ethical caveats.

### 6.4 Retrospective Application of the CEC

To assess the guiding utility of the CEC, we apply our framework retrospectively on a selection of four high-quality, contemporary papers relevant to CUI research while sharing topical overlaps with our contribution. Papers were chosen based on their research quality and rigour from esteemed conferences. Consequently, we apply the CEC to these papers' findings to illustrate its relevance while demonstrating how it can be used.

In their work, Yeh et al. [100] discuss common pitfalls of guiding users effectively in task-oriented chatbot interactions. A key finding of their study describes how lack of transparency segues into frustration. In addressing the complexity of making possible interactions evident to the user, the authors describe guidance strategies that inform users about a chatbot's capabilities adequately. In line with the CEC, the authors unravel the need to align expectations (Expectation Gap) between user and system through guidance as featured in our Guiding Through Interactions theme. Incorporated in the Evaluation Bridge, our Guiding Through Interactions theme helps overcome these difficulties.

Following the potential advantages of LLM-supported CUIs, Jo et al. [49] studied users' experience with a chatbot as an empathetic conversation partner in care environments. However, users reported a lack of emotional support, which they linked to limited personalisation and missing health history. Moreover, particular worries accompanied users' experience, for example, a fear that personalised and empathetic CUIs may reduce peoples' desire to engage in social activities. The authors identified problems and foreshadowed solutions akin to our *Inclusivity and Diversity* theme and reflected by our Deceptive and Trust Delimiters. As the *Setting Expectations* theme addresses requirements to bridge an *Expectation Gap*, our *Building Trust and Guarding Privacy* and *Human-like Harmony* themes could answer user's worries by making the system more transparent and trustworthy, increasing their experience.

Voice-based systems introduce additional layers to the interaction that require consideration. Language and speech barriers become obstacles for many marginalised groups that require special attention. Based on a study run in the U.S.A., Harrington et al. [44] showed how Black adults from lower-income households struggle when using voice-based CUIs. Participants mentioned falling back to code-switching, mentioning the fear of being misunderstood. Additionally, the researchers captured distrust toward Google Home in the health-related context of their study. These trust issues were amplified by a lack of knowledge and technological literacy, hindering users from experiencing the CUI's full potential. As a source for these problems, the authors identify a host in the ignorance of developers who neglect the cultural identity of their users. Providing practitioners with further insights, our Inclusivity and Diversity theme spotlights the importance of catering to diverse users, as ignorance and negligence can swiftly lead to unethical implications. As Harrington et al. describe, CUIs should foster trust and transparency (i.g. avoid Deception and Distrust Delimiters). Their study highlights the importance of careful and inclusive research design, given the WEIRD (western, educated, industrialized, rich, and democratic) [45] context in which many contemporary studies in HCI are conducted. It further demonstrates that some issues cannot be mitigated through LLMs alone.

Generally, LLMs open promising avenues for CUIs whether deployed through text or voice-based systems. Exploring these advantages to make mobile GUIs more accessible through an assistive CUI, Wang et al. [97] describe prompting techniques that allow users to engage with interfaces through speech, particularly helping

marginalised and vulnerable groups to access otherwise unattainable content. Their work is an excellent example of how the CEC's *Expectation Gap* can be bridged by showcasing how conversations could be customised around special needs, thereby increasing the inclusivity of this technology in the future in line with our *Inclusivity and Diversity* theme.

Collectively, these works illustrate relevant design challenges that the CEC emphasises. The addition of ethical caveats, elevating user experience, finds support as our *Inclusivity and Diversity* theme resonates in three of the four selected papers [44, 49, 97]. Moreover, the studies foreshadow benefits gained from aligning user and CUI expectations [97, 100] while informing about system capabilities [49] and making interactions transparent to increase trust [100]. We hope our framework can guide future research and design of CUIs in a human-centred manner.

### 7 DISCUSSION & FUTURE WORK

Our research aimed to gain insights into the ethical caveats and design considerations faced in current and future CUI research and development. Motivated by dark pattern scholarship, particularly Mathur et al. [63] dark pattern characteristics, we conducted 27 semi-structured interviews and identified five themes providing answers to our research question. Derived from these themes, we propose the CUI Expectation Cycle (CEC), a framework to guide future work in bridging CUI capabilities with users' expectations and goals. In this section, we return to our research question and discuss the different perspectives held by the interviewed cohorts. Lastly, we point toward some directions for future work.

### 7.1 Revisiting Our Research Question

With the main aim of this study to identify ethical caveats in CUI design, our findings answer our research question: Which ethical caveats should be considered when designing CUI interactions, and how should they be addressed? Our themes illustrate how the unique nature of CUIs cannot simply carry over expertise readily available for other domains. To fill this gap, the subordinate ethical caveats contain design considerations that, if disregarded, compromise user experience. These include the discoverability of features and possible interactions [9, 15, 30, 47, 96], often further limited to diverse and marginalised users [44], but also deceptive designs that lead to unrealistic expectation [62] and even unwanted interactions – as seen in dark pattern literature [37, 63]. Incorporating these themes, the CEC builds on established theory to provide practitioners and scholars with means to mitigate these effects.

### 7.2 Perceptions Across Researchers, Practitioners, and Users

Each developed theme considers the perspectives of each cohort, highlighting similarities and differences in the diverse code groups (the full codebook is included in the supplementary material of this paper). Although researchers and practitioners often shared similar expertise about the involved technologies, we noticed worries among users, illustrating a lack of the same technical understanding. Interestingly, this was further mirrored by similar views held among researchers and practitioners who – in sum – recognised

Themes							
Building Trust and Guarding Privacy	Guiding Through Interactions	Human-like Harmony	Inclusivity and Diversity	Setting Expectations			
Bad Defaults [8] Captive Audience [41] Disguised Ads [12] Disguised Data Collection [41] Hidden Legalese Stipulations [8] Hides Information [63, 64] Making Personal Information Public [41] Privacy Zuckering [12]	Confusion [20] Exploiting Errors [20] Manipulating Navigation [20] Misrepresenting [36] Obfuscation [20] Obstruction [37] Roach Motel [12] Trick [20]	Deceptive [63, 64] Hides Information [63, 64] Misrepresenting [36] Sneaking [37] Trick [20]	· Covert [63, 64] · Disparate Treatment [64] · Making Personal Information Public [41] · Social Network Of Proxemic Contacts Or Unintended Relationships [41]	· Asymmetry [64] · Automating The User [36] · Bait & Switch [8, 41] · Forced Actions [37] · Forced Work [20] · Hidden Legalese Stipulations [8] · Interface Interference [37] · Obfuscation [20] · Obstruction [37] · Restricting Functionalities [20]			

Table 3: This table summarises previously captured dark patterns within each theme.

a need to address users' concerns and develop CUIs that are more accessible and inclusive.

While all cohorts agree that CUIs should be more transparent about their capabilities and how data is handled, suggestions as to how this could be addressed differ. Researchers and frequent users pointed out that CUI design should reflect capabilities as anthropomorphic features overshadow limitations. Practitioners, noticing some relevance to announcing a CUI's artificial nature, argued for the benefits of CUIs offering companionship to lonely users. This is in line with previous work (e.g. [74, 94], and mirrored by users admitting emotional bonds with their devices. However, both frequent users and practitioners expressed frustration regarding user experience. Interestingly, though, they are divided as to which group should take responsibility. While frequent users are frustrated about the lack of discoverability in hard-to-navigate interfaces, practitioners argue that users are not exhausting their CUI's technical potential. This disparity is one example of the gap between users and practitioners. A potential solution was presented across cohorts when discussing the need for improved onboarding for setting realistic expectations. Here, researcher and practitioner cohorts were aware of a need to address diverse and marginalised users better, while practitioners, in particular, described current technical limitations. A general need for guidelines is voiced by both researchers and practitioners. While the latter echoed this sentiment but missed standardised guidelines for CUIs, the former also noticed a need for better communication as contemporary research aims to fill this gap.

### 7.3 Paving the Way for Future Work

Our decision to listen to different voices allowed us to garner insights from three perspectives with partially contrasting incentives. Although related work has outlined similar concerns raised by our participants [9, 19, 26, 80, 100], our approach to connecting the discourse to unethical design and dark pattern literature has led to novel findings, resulting in further design considerations and

bridging expectations between cohorts. Future work could aim to lift the tension between user and CUI and develop interfaces that meet their expectations. Thus, increasing user experience would be a success for the practitioners' work.

Our work can serve as a foundation for ethical design in CUI contexts. As the interview questions adopted Mathur et al.'s [63] dark pattern characteristics, we reviewed each theme as a potential host for dark patterns. For this, we follow Mildner et al. [69] and draw from contemporary typologies that collectively describe over 80 types of dark patterns [8, 12, 20, 36, 37, 41, 43, 63, 64, 69, 101]. Table 3 allocates identified dark patterns, suggesting that each theme features nefarious opportunities for manipulative interactions. Notably, most dark pattern types have been described in GUI contexts. Introducing this body of work to CUI research offers valuable insights. However, our themes foreshadow additional patterns unique to CUI interactions, extending previously described instances in CUI contexts [19, 79]. As the technology is still in its relative infancy, understanding unethical practices early can present an important head start in protecting users. Future work could build on our findings to gain a better understanding of the underlying technologies of CUIs that could be exploited to deploy novel types of dark patterns.

### 8 LIMITATIONS

Although we were careful when designing and conducting this study, our work has limitations. The interviews were conducted using an online format, which, in individual cases, led to connectivity problems, prolonging some interviews. Although we ensured that all interviewees were asked the same questions, technological obstacles may have influenced participants' responses.

Sample size limitations should also be acknowledged, with a relatively small sample representing each cohort. That said, steps were taken to ensure cohorts were relatively balanced in size and gender and to ensure practitioners and researchers represented a relatively broad range of backgrounds and experiences. The size of

each cohort sample is also common to qualitative research in the field, as similar qualitative studies estimate high saturation of codes within 7-12 interviews [42]. Nonetheless, this may pose limitations concerning the breadth of opinions expressed, which may not be representative across professions. Additionally, many practitioners have experience in research areas with multiple participants holding a Ph.D. For frequent users, we were mainly interested in people who use CUI devices weekly. However, we acknowledge that we did not further inquire about the duration of participants' interactions with CUIs. Moreover, the fact that participants could afford devices may suggest a selection bias in this sample. While we aimed to sample frequent users from across the world, the size of this study hinders a representative sample. We used Prolific [61] to recruit matching participants, where we had to trust the platform and participants' self-evaluation to meet our criteria.

With regard to our themes, we recognise two limitations. First, participants drew heavily from their experience with voice-based interfaces. While we prompted each interviewee to consider both voice-based and text-based CUIs, replies from frequent users indicate insufficient experience with the latter. We were careful to frame our themes around all kinds of CUIs, but the lack of chatbot experience among the frequent user group may have introduced a bias toward voice-based systems. Second, not every design issue in our themes necessarily implies malicious intent. The current discourse accompanying dark patterns and malicious designs discusses whether the malevolent intents of practitioners are relevant or if any potential harm suffices to fulfil its criteria [40, 43, 76]. Future work could address this gap to study potential links between intents and dark patterns to aid regulators and create guidelines for ethically aligned user interfaces.

### 9 CONCLUSION

Recent interest in HCI has raised awareness of unethical design in technologies. In this paper, we identified ethical caveats related to CUI technologies by conducting 27 interviews between researchers, practitioners, and frequent users of CUI systems. Based on our analysis, we identified five themes covering each group's perspectives and informing about exploitative designs' unethical consequences. In line with prior research, we noticed broken expectations between users and their devices' capabilities and learned about the underlying problems. We hope that the five ethical caveats, derived from our themes, can be used to support user-centred designs and create systems that are better aligned with service providers' intentions and users' expectations. To mitigate users' frustration and ensure more transparency and ethically aligned interactions, this work contributes a framework, the CUI Expectation Cycle, to connect users' expectations with CUIs' capabilities.

### **ACKNOWLEDGMENTS**

Special thanks go to Justin Edwards, whose great support has aided this work significantly throughout the project's course. The research of this work was partially supported by the Klaus Tschira Stiftung gGmbH. This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 12/RC/2289\_P2. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

### REFERENCES

- ACM. 2023. Words matter: Alternatives for charged terminology in the computing profession. https://www.acm.org/diversity-inclusion/words-matter
- [2] C. Alexander, S. Ishikawa, and M. Silverstein. 1977. A Pattern Language: Towns, Buildings, Construction. OUP USA. https://books.google.ch/books? id=hwAHmktpk5IC
- [3] Saminda S. Balasuriya, Laurianne Sitbon, Andrew A. Bayor, Maria Hoogstrate, and Margot Brereton. 2018. Use of voice activated interfaces by people with intellectual disability. In OzCHI '18: Proceedings of the 30th Australian Conference on Computer-Human Interaction. ACM, Melbourne Australia, 102–112. https://doi.org/10.1145/3292147.3292161
- [4] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmitchell Shmargaret. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, Canada, 610–623. https://doi.org/10. 1145/3442188.3445922
- [5] Johnna Blair and Saeed Abdullah. 2020. It Didn't Sound Good with My Cochlear Implants: Understanding the Challenges of Using Smart Assistants for Deaf and Hard of Hearing Users. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 4, Article 118 (dec 2020), 27 pages. https://doi.org/10.1145/3432194
- [6] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. Synthesis Lectures on Human-Centered Informatics 9, 1 (April 2016), 1–115. https://doi.org/10.2200/ S00706ED1V01Y201602HCI034 Publisher: Morgan & Claypool Publishers.
- [7] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I Am Definitely Manipulated, Even When I Am Aware of It. It's Ridiculous!" Dark Patterns from the End-User Perspective. In Proceedings of the 2021 ACM Designing Interactive Systems Conference (Virtual Event, USA) (DIS '21). Association for Computing Machinery, New York, NY, USA, 763–776. https://doi.org/10.1145/3461778.3462086
- [8] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. Proc. Priv. Enhancing Technol. 2016, 4 (2016), 237–254.
- [9] Stacy M. Branham and Antony R. Mukkath Roy. 2019. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. In ASSETS '19: Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility. ACM, Pittsburgh USA, 446—458. https://doi.org/10.1145/3308561.3353797
- [10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa
- [11] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. The-matic Analysis. In Handbook of Research Methods in Health Social Sciences, Pranee Liamputtong (Ed.). Springer Singapore, Singapore. https://doi.org/10.1007/978-981-10-5251-4\_103
- [12] Harry Brignull. 2010. Deceptive Design formerly darkpatterns.org. https://www.deceptive.design/ Visited on 2023-01-24.
- [13] John Brooke. 2013. SUS: A Retrospective. Journal of Usability Studies 8, 2 (2013), 29–40. http://delivery.acm.org/10.1145/2820000/2817913/p29-brooke.pdf?ip=193.1.166.20&id=2817913&acc=ACTIVESERVICE&key=846C3111CE4A4710. FFC467975145E027.4D4702B0C3E38B35.4D4702B0C3E38B35&\_acm\_=1553781942\_85763446e3705c872a65d0c0431bbc79
- [14] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
- [15] Mei-Ling Chen and Hao-Chuan Wang. 2018. How Personal Experience and Technical Knowledge Affect Using Conversational Agents. In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (Tokyo, Japan) (IUI '18 Companion). Association for Computing Machinery, New York, NY, USA, Article 53, 2 pages. https://doi.org/10.1145/3180308.3180362
- [16] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, and Benjamin R Cowan. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (09 2019), 349–371. https://doi.org/10.1093/iwc/iwz016 arXiv:https://academic.oup.com/iwc/articlepdf/31/4/349/33525046/iwz016.pdf
- [17] Kenneth Mark Colby. 1981. Modeling a paranoid mind. Behavioral and Brain Sciences 4, 4 (1981), 515–534. https://doi.org/10.1017/S0140525X00000030

- [18] European Commission. 2016. GDPR-16 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://ec.europa.eu/commission/ sites/beta-political/files/data-protection-factsheet-changes\_en.pdf
- [19] Silvia De Conca. 2023. The present looks nothing like the Jetsons: Deceptive design in virtual assistants and the protection of the rights of users. Computer Law & Security Review 51 (Nov. 2023), 105866. https://doi.org/10.1016/j.clsr. 2023 105866
- [20] Gregory Conti and Edward Sobiesk. 2010. Malicious interface design: exploiting the user. In Proceedings of the 19th international conference on World wide web -WWW '10. ACM Press, Raleigh, North Carolina, USA, 271. https://doi.org/10. 1145/1772690.1772719
- [21] Eric Corbett and Astrid Weber. 2016. What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (Florence, Italy) (MobileHCI '16). Association for Computing Machinery, New York, NY, USA, 72-82. https://doi.org/10.1145/2935334.2935386
- [22] Benjamin R. Cowan, Holly P. Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human-computer dialogue. *International Journal* of Human-Computer Studies 83 (2015), 27–42 pages. https://doi.org/10.1016/j. ijhcs.2015.05.008
- [23] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can i help you with?": infrequent users' experiences of intelligent personal assistants. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, Vienna Austria, 1–12. https://doi.org/ 10.1145/3098279.3098539
- [24] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/ 3313831.3376600
- [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 (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 244, 14 pages. https: //doi.org/10.1145/3411764.3445206
- [26] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, Taipei Taiwan, 1–12. https://doi.org/10.1145/3338286.3340116
- [27] Philip R Doyle, Daniel John Rough, Justin Edwards, Benjamin R. Cowan, Leigh Clark, Martin Porcheron, Stephan Schlögl, María Inés Torres, Cosmin Munteanu, Christine Murad, Jaisie Sin, Minha Lee, Matthew Peter Aylett, and Heloisa Candello. 2021. CUl@IUI: Theoretical and Methodological Challenges in Intelligent Conversational User Interface Interactions. In 26th International Conference on Intelligent User Interfaces Companion (College Station, TX, USA) (IUI '21 Companion). Association for Computing Machinery, New York, NY, USA, 12–14. https://doi.org/10.1145/3397482.3450706
- [28] Mateusz Dubiel, Sylvain Daronnat, and Luis A. Leiva. 2022. Conversational Agents Trust Calibration: A User-Centred Perspective to Design. In Proceedings of the 4th Conference on Conversational User Interfaces. ACM, Glasgow United Kingdom, 1–6. https://doi.org/10.1145/3543829.3544518
- [29] Caitlin Fausey, Bria Long, Aya Inamori, and Lera Boroditsky. 2010. Constructing Agency: The Role of Language. Frontiers in Psychology 1 (2010). https://doi. org/10.3389/fpsyg.2010.00162
- [30] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161. https://doi.org/10.1016/j.ijhcs. 2019.07.009
- [31] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. In Early engagement and new technologies: Opening up the laboratory, Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman (Eds.). Vol. 16. Springer Netherlands, Dordrecht, 55–95. https://doi.org/10.1007/978-94-007-7844-3\_4 Series Title: Philosophy of Engineering and Technology.
- [32] Radhika Garg and Subhasree Sengupta. 2020. He Is Just Like Me: A Study of the Long-Term Use of Smart Speakers by Parents and Children. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 1, Article 11 (mar 2020), 24 pages. https://doi.org/10.1145/3381002

- [33] Debjyoti Ghosh, Pin Sym Foong, Shan Zhang, and Shengdong Zhao. 2018. Assessing the Utility of the System Usability Scale for Evaluating Voice-Based User Interfaces. In Proceedings of the Sixth International Symposium of Chinese CHI (Montreal, QC, Canada) (Chinese CHI '18). Association for Computing Machinery, New York, NY, USA, 11–15. https://doi.org/10.1145/3202667.3204844
- [34] Li Gong and Clifford Nass. 2007. When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. Human communication research 33, 2 (2007), 163–193.
- [35] Colin M. Gray and Shruthi Sai Chivukula. 2019. Ethical Mediation in UX Practice. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow Scotland Uk, 1–11. https://doi.org/10.1145/3290605.3300408
- [36] Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. 2020. What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. Association for Computing Machinery, New York, NY, USA, 61–73. https://doi.org/10.1145/3357236.3395486
- [37] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108
- [38] Colin M. Gray, Cristiana Santos, and Nataliia Bielova. 2023. Towards a Preliminary Ontology of Dark Patterns Knowledge. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 286, 9 pages. https://doi.org/10.1145/3544549.3585676
- [39] Colin M. Gray, Cristiana Teixeira Santos, Nicole Tong, Thomas Mildner, Arianna Rossi, Johanna T. Gunawan, and Caroline Sinders. 2023. Dark Patterns and the Emerging Threats of Deceptive Design Practices. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 510, 4 pages. https://doi.org/10.1145/3544549.3583173
- [40] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (Feb. 2021), 1–38. https://doi.org/10.33621/jdsr.v3i1.54 Number: 1.
- [41] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark Patterns in Proxemic Interactions: A Critical Perspective. Association for Computing Machinery, New York, NY, USA, 523–532. https://doi.org/10.1145/2598510. 2598541
- [42] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods 18, 1 (Feb. 2006), 59–82. https://doi.org/10.1177/1525822X05279903
- [43] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. Proceedings of the ACM on Human-Computer Interaction 5, CSCW2 (Oct. 2021), 1–29. https://doi.org/10.1145/3479521
- [44] Christina N. Harrington, Radhika Garg, Amanda Woodward, and Dimitri Williams. 2022. "It's Kind of Like Code-Switching": Black Older Adults' Experiences with a Voice Assistant for Health Information Seeking. In CHI Conference on Human Factors in Computing Systems. ACM, New Orleans LA USA, 1–15. https://doi.org/10.1145/3491102.3501995
- [45] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. Nature 466, 7302 (2010), 29–29.
- [46] Samuel Holmes, Anne Moorhead, Raymond Bond, Huiru Zheng, Vivien Coates, and Michael Mctear. 2019. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. In Proceedings of the 31st European Conference on Cognitive Ergonomics. ACM, BELFAST United Kingdom, 207–214. https://doi.org/10.1145/3335082.3335094
- [47] Mohit Jain, Ramachandra Kota, Pratyush Kumar, and Shwetak N. Patel. 2018. Convey: Exploring the Use of a Context View for Chatbots. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1-6. https://doi.org/10.1145/3173574.3174042
- [48] Haiyan Jia, Mu Wu, Eunhwa Jung, Alice Shapiro, and S. Shyam Sundar. 2013. When the Tissue Box Says "Bless You": Using Speech to Build Socially Interactive Objects. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (Paris, France) (CHI EA '13). Association for Computing Machinery, New York, NY, USA, 1635–1640. https://doi.org/10.1145/2468356.2468649
- [49] Eunkyung Jo, Daniel A. Épstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3581503
- [50] Hyunhoon Jung, Hyeji Kim, and Jung-Woo Ha. 2020. Understanding Differences between Heavy Users and Light Users in Difficulties with Voice User Interfaces. In Proceedings of the 2nd Conference on Conversational User Interfaces (Bilbao, Spain) (CUI '20). Association for Computing Machinery, New York, NY, USA, Article 51, 4 pages. https://doi.org/10.1145/3405755.3406170
- [51] Andreas M. Klein, Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. 2020. Construction of UEQ+ scales for voice quality: measuring user experience

- quality of voice interaction. In Mensch und Computer 2020 Tagungsband, Florian Alt, Stefan Schneegass, and Eva Hornecker (Eds.). ACM, New York, 1—-5. https://doi.org/10.1145/3404983.3410003
- [52] A Baki Kocabalil, Liliana Laranjo, and Enrico Coiera. 2018. Measuring user experience in conversational interfaces: a comparison of six questionnaires. In Proceedings of the 32nd International BCS Human Computer Interaction Conference 32. 1–12.
- [53] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, D Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. PNAS 117, 14 (2020), 7684–7689
- [54] Dimosthenis Kontogiorgos, Andre Pereira, Olle Andersson, Marco Koivisto, Elena Gonzalez Rabal, Ville Vartiainen, and Joakim Gustafson. 2019. The Effects of Anthropomorphism and Non-Verbal Social Behaviour in Virtual Assistants. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (Paris, France) (IVA '19). Association for Computing Machinery, New York, NY, USA, 133--140. https://doi.org/10.1145/3308532.3329466
- [55] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'Dark Pattern' in Home Robots. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, Daegu, Korea (South), 374–381. https://doi.org/10.1109/ HRI.2019.8673274
- [56] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445312
- [57] Laura Lascau, Sandy J. J. Gould, Anna L. Cox, Elizaveta Karmannaya, and Duncan P. Brumby. 2019. Monotasking or Multitasking: Designing for Crowdworkers' Preferences. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow Scotland Uk, 1–14. https: //doi.org/10.1145/3290605.3300649
- [58] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society (Graz, Austria) (USAB 2008). 63–76. https://doi.org/10.1007/978-3-540-89350-9 6
- [59] Sunok Lee, Minji Cho, and Sangsu Lee. 2020. What If Conversational Agents Became Invisible? Comparing Users' Mental Models According to Physical Entity of AI Speaker. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 3, Article 88 (sep 2020), 24 pages. https://doi.org/10.1145/3411840
- [60] California State Legislature. 2018. CCPA-18 2018. California Consumer Privacy Act of 2018 [1798.100 1798.199] (CCPA). https://leginfo.legislature.ca.gov/faces/codes\_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.
- [61] Prolific Academic Ltd. 2023. Prolific | Online participant recruitment for surveys
- and market research. https://prolific.co/ (visited on 2023-01-24).

  [62] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, San Jose California USA, 5286-5297. https://doi.org/10.1145/2858036.
- [63] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale. Proceedings of the ACM on Human-Computer Interaction 3, CSCW (nov 2019), 1–32. https://doi.org/10.1145/3359183
- [64] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. https: //doi.org/10.1145/3411764.3445610
- [65] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. Frontiers in Artificial Intelligence 4 (2021), 169.
- [66] Thomas Mildner, Philip Doyle, Gian-Luca Savino, and Rainer Malaka. 2022. Rules Of Engagement: Levelling Up To Combat Unethical CUI Design. In 4th Conference on Conversational User Interfaces. ACM, Glasgow United Kingdom, 1–5. https://doi.org/10.1145/3543829.3544528
- [67] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In Proceedings of the 2023 ACM Designing Interactive Systems Conference. ACM, Pittsburgh PA USA, 2362–2374. https://doi.org/10.1145/3563657.3595964
- [68] Thomas Mildner and Gian-Luca Savino. 2021. Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451659

- [69] Thomas Mildner, Gian-Luca Savino, Philip Doyle, Benjamin Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (April 23-28, 2023) (CHI '23). Association for Computing Machinery, New York, NY, USA, 1-15. https://doi.org/10.1145/3544548.3580695
- [70] Christine Murad and Cosmin Munteanu. 2020. Designing Voice Interfaces: Back to the (Curriculum) Basics. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/ 3313831.3376522
- [71] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (2000), 81-103 pages. https://doi.org/10.1111/0022-4537.00153
- [72] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, Massachusetts, USA) (CHI '94). Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703
- [73] Jakob Nielsen and Rolf Molich. 1990. Heuristic Evaluation Of User Interfaces. In Proceedings of the 1990 CHI Conference on Human Factors in Computing Systems (April). ACM, Seattle, Washington, USA, 249–256.
- [74] Jasmin Niess and Paweł W. Woźniak. 2020. Embracing Companion Technologies. In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (Tallinn, Estonia) (NordiCHI '20). Association for Computing Machinery, New York, NY, USA, Article 31, 11 pages. https://doi.org/10.1145/3419249.3420134
- [75] Donald A. Norman. 2013. The design of everyday things (revised and expanded edition ed.). Basic Books. New York. New York.
- [76] Ikechukwu Obi, Colin M. Gray, Shruthi Sai Chivukula, Ja-Nae Duane, Janna Johns, Matthew Will, Ziqing Li, and Thomas Carlock. 2022. Let's Talk About Socio-Technical Angst: Tracing the History and Evolution of Dark Patterns on Twitter from 2010-2021. http://arxiv.org/abs/2207.10563 arXiv:2207.10563 [cs].
- [77] Richard L. Oliver. 1977. Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology* 62, 4 (1977), 480–486. https://doi.org/doi.org/10.1037/0021-9010.62.4.480
- [78] Richard L. Oliver. 1980. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research* 17, 4 (1980), 460–469. https://doi.org/10.1177/002224378001700405
- [79] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In Proceedings of the 2022 European Symposium on Usable Security (Karlsruhe, Germany) (EuroUSEC '22). Association for Computing Machinery, New York, NY, USA, 64—78. https://doi.org/10.1145/3549015. 3554213
- [80] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Ta-dayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In Proceedings of the 2022 European Symposium on Usable Security (Karlsruhe, Germany) (EuroUSEC '22). Association for Computing Machinery. New York. NY. USA. 64-78. https://doi.org/10.1145/3549015.3554213
- Machinery, New York, NY, USA, 64-78. https://doi.org/10.1145/3549015.3554213
  [81] European Parliament. 2022. Digital Servicess Act\*\*\*I. European Parliament
  [A9-0356/2021]. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0014 EN.html
- [82] J. Piaget. 1997. The Child's Conception of the World. Routledge, London, UK. https://books.google.de/books?id=wxWd6bY2FAkC
- [83] Martin Porcheron, Leigh Clark, Matt Jones, Heloisa Candello, Benjamin R. Cowan, Christine Murad, Jaisie Sin, Matthew P. Aylett, Minha Lee, Cosmin Munteanu, Joel E. Fischer, Philip R. Doyle, and Jofish Kaye. 2020. CUI@CSCW: Collaborating through Conversational User Interfaces. In Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (Virtual Event, USA) (CSCW '20 Companion). Association for Computing Machinery, New York, NY, USA, 483–492. https://doi.org/10.1145/3406865.341887
- [84] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3174214
- [85] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, Montréal Canada, 102–112. https://doi. org/10.1145/3173574.3174033
- [86] Stuart Reeves. 2019. Conversation Considered Harmful?. In Proceedings of the 1st International Conference on Conversational User Interfaces (Dublin, Ireland) (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 10, 3 pages. https://doi.org/10.1145/3342775.3342796

- [87] Leon Reicherts, Yvonne Rogers, Licia Capra, Ethan Wood, Tu Dinh Duong, and Neil Sebire. 2022. It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. ACM Transactions on Computer-Human Interaction 29, 3 (June 2022), 1–41. https://doi.org/10.1145/3484221
- [88] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. 2022. Understanding Account Deletion and Relevant Dark Patterns on Social Media. Proceedings of the ACM on Human-Computer Interaction 6, CSCW2 (Nov. 2022), 1–43. https://doi.org/10.1145/3555142
- [89] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in Human–Agent Interaction: A Survey. ACM Comput. Surv. 54, 4, Article 81 (may 2021), 43 pages. https://doi.org/10.1145/3386867
- [90] William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. Proc. ACM Hum.-Comput. Interact. 5, CSCW2, Article 371 (oct 2021), 16 pages. https://doi.org/10.1145/3479515
- [91] Jaisie Sin, Heloisa Candello, Leigh Clark, Benjamin R. Cowan, Minha Lee, Cosmin Munteanu, Martin Porcheron, Sarah Theres Völkel, Stacy Branham, Robin N. Brewer, Ana Paula Chaves, Razan Jaber, and Amanda Lazar. 2023. CUI@CHI: Inclusive Design of CUIs Across Modalities and Mobilities. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 341, 5 pages. https://doi.org/10.1145/3544549.3573820
- [92] Alain D. Starke and Minha Lee. 2022. Unifying Recommender Systems and Conversational User Interfaces. In Proceedings of the 4th Conference on Conversational User Interfaces (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 7, 7 pages. https: //doi.org/10.1145/3543829.3544524
- [93] Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanatto. 2018. Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In *Proceedings of the Technology, Mind, and Society*. ACM, Washington DC USA, 1–6. https://doi.org/10.1145/3183654.3183691
- [94] Alexandra Voit, Jasmin Niess, Caroline Eckerth, Maike Ernst, Henrike Weingärtner, and Paweł W. Woźniak. 2020. 'It's Not a Romantic Relationship': Stories of Adoption and Abandonment of Smart Speakers at Home. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (Essen, Germany) (MUM '20). Association for Computing Machinery, New York, NY, USA, 71–82. https://doi.org/10.1145/3428361.3428469
- [95] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 254, 15 pages. https: //doi.org/10.1145/3411764.3445536
- [96] Katja Wagner and Hanna Schramm-Klein. 2019. Alexa, Are You Human? Investigating Anthropomorphism of Digital Voice Assistants A Qualitative Approach. In Proceedings of the 40th International Conference on Information Systems, ICIS 2019, Munich, Germany, December 15-18, 2019, Helmut Krcmar, Jane Fedorowicz, Wai Fong Boh, Jan Marco Leimeister, and Sunil Wattal (Eds.). Association for Information Systems, Munich, Germany, 1-17. https://aisel.aisnet.org/icis2019/human\_computer\_interact/human\_computer\_interact/7
- [97] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling Conversational Interaction with Mobile UI using Large Language Models. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, 1–17. https://doi.org/10.1145/3544548.3580895
- [98] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. Perspectives on Psychological Science 5, 3 (2010), 219–232. https://doi.org/10. 1177/1745691610369336 PMID: 24839457.
- [99] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1 (1966), 36–45
- [100] 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 CHI Conference on Human Factors in Computing Systems. ACM, New Orleans LA USA, 1–16. https://doi.org/10.1145/3491102.3501941
- [101] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In Proceedings of the 8th International Conference on the Foundations of Digital Games (FDG 2013) (May 14-17). Society for the Advancement of the Science of Digital Games, Chania, Crete, Greece, 39–46. http://www.fdg2013. org/program/papers.html

### A INTERVIEW QUESTIONNAIRE

Here, we provide a Table 4 including Mathur et al.'s [63] original characteristics as well as adapted questions. Below the table is the full script of all nine interview questions. We conditionally prompted interviewees to talk about graphical-based systems and voice-based systems as well as contemporary and future designs to get multi-faceted answers.

Original and adapted questions from the dark pattern characteristics by Mathur et al. [63]				
Characteristic	Question (Originals are <i>italic</i> whereas interview questions are not)			
Asymmetry	Does the user interface design impose unequal weights or burdens on the available choices presented to the user in the interface?  Can you tell me about how the design of conversational user interfaces may place specific burdens on people in terms of understanding what choices are available to them during interactions?			
Covert	Is the effect of the user interface design choice hidden from the user? How clear are the consequences of making particular choices to people when using conversational user interfaces?			
Deceptive	Does the user interface design induce false beliefs either through affirmative misstatements, misleading statements, or omissions?  Can you tell me about how the design of conversational user interface may induce false beliefs?			
Hides Info.	Does the user interface obscure or delay the presentation of necessary information to the user? How could aspects of the design of conversational user interface obscure or delay important information from the user?			
Restrictive	Does the user interface restrict the set of choices available to users?  Do you feel that all available choices are clear to people when using conversational user interfaces?			

Table 4: This table lists the descriptive questions by Mathur et al. (2019) [63] for their dark pattern characteristics. It further contains the adapted interview questions.

- (1) Could you please describe your own experience with CUIs?
- (2) Can you tell me about how the design of conversational user interfaces may place specific burdens on people in terms of understanding what choices are available to them during interactions?
- (3) How clear are the consequences of making particular choices to people when using conversational user interfaces?

- (4) Do you feel that all available choices are clear to people when using conversational user interfaces?
- (5) Can you tell me about how the design of conversational user interface may induce false beliefs? (these might include false beliefs about the system or in terms of the information it provides.)
- (6) How could aspects of the design of conversational user interface obscure or delay important information from the
- (7) Can you tell me about some inherent limitations of conversational user interfaces that are not always apparent to people who use them?
- (8) Can you tell me about situational contexts that might make people more vulnerable to problematic design in CUI interactions?
- (9) Can you tell me about specific groups of people who might be particularly vulnerable to problematic design in CUI interactions?