Pre-print of full paper presented at CONVERSATIONS 2023 – the 7th International Workshop on Chatbot Research and Design, hosted by the University of Oslo, Norway, November 22-23, 2023. The final version of the paper will be published in the forthcoming Springer LNCS post-workshop proceedings.

In Search of Dark Patterns in Chatbots

Verena Traubinger[1][0000-0002-1786-672X], Sebastian Heil[1][0000-0003-2761-9009], Julián Grigera[2,3,4][0000-0002-7962-4312], Alejandra Garrido[2,3][0000-0002-5052-705X], and Martin Gaedke[1][0000-0002-6729-2912]

[1] Faculty of Computer Science, Chemnitz University of Technology, Germany
verena.traubinger@informatik.tu-chemnitz.de
sebastian.heil@informatik.tu-chemnitz.de
martin.gaedke@informatik.tu-chemnitz.de
[2] LIFIA, Fac. de Informática, Univ. Nac. de La Plata, Argentina
julian.grigera@lifia.info.unlp.edu.ar
garrido@lifia.info.unlp.edu.ar
[3] CONICET, Argentina
[4] CICPBA, Argentina

Abstract. While Dark Patterns are widely present in graphical user interfaces, in this research we set out to find out whether they are also starting to appear in Chatbots. Dark Patterns are intentionally deceptive designs that trick users into acting contrary to their intention - and in favor of the organization that implements them. Chatbots, as a kind of conversational user interface, can potentially also suffer from Dark Patterns or other poor interaction design, sometimes referred to as Usability Smells. This keeps users from easily achieving their goals and can lead to frustration or limitations for users. To find Dark Patterns and Usability Smells, we analyzed user reports of negative experiences. Since we found no well known dataset of reports, we created the *ChIPS* dataset with 69 complaints from different web sources, and then classified them as one of 16 established Dark Patterns, potential new Dark Patterns, Usability Smells, or neither. Results show that, even though there are instances of established Dark Patterns, negative experiences usually are caused by chatbot defects, high expectations from users, or non-intuitive interactions.

Keywords: Dark Patterns · Deceptive Design · Usability Smells · Conversational User Interfaces · Chatbots.

1 Introduction

Chatbots as conversational user interfaces (CUIs) are a technology that became a trend over the last few years [1] and even more so with the emergence of large language models (LLMs) like ChatGPT or Bard. Companies, organizations and government structures are using chatbots as an easily available alternative to give information to customers or citizens. As with any new technology, users encounter situations where the chatbot malfunctions, or it does not meet their (perhaps high) expectations. This can even make users think that chatbots purposely withhold information or offer the wrong one, as can be seen in subreddits like *r/assholedesign*[5]. Most often this has a negative impact on the user experience or even leads to a disadvantage, especially for customers who are affected if companies apply Dark Patterns against them.

In graphical user interfaces (GUIs), the term *Dark Pattern* (nowadays often termed *deceptive patterns* [19]) describes design choices implemented by companies with a malicious intent to gain profit at the expense of the customers [3]. This term has to be differentiated from *Usability Smells*, which are badly designed user interactions, but are lacking a malicious intent [10]. Due to the linear narrative in chatbot interactions, users are dependant on the chatbot to offer correct information, without having the option to easily verify the validity of the statements. While users have the option of searching the website for information that the chatbot may withhold, or to confirm information that it does provide, sometimes this information could not be available at all. This offers possibilities for companies to implement Dark Patterns not only in their GUIs, but also their CUIs, which are often used for customer service. Here, we want to address the research question of whether chatbots are designed with Dark Patterns or if negative experiences from customers are rather due to Usability Smells.

In this work we have evaluated a corpus of 69 examples of negative user experiences with chatbots, searching for well-known Dark Patterns, potentially new Dark Patterns, and Usability Smells. The findings show that previously established Dark Patterns from GUIs are adapted for chatbots, that no new Dark Patterns specifically for chatbots were identified and that most negative interactions with chatbots might currently be caused rather by poor design choices and implementation than from Dark Patterns or Usability Smells.

2 Related Work

In this section we first define the background concepts that provide the basis of our work. In the following subsection we highlight related works, from which we identify possibly relevant results for our own research.

Even before AI chatbots like ChatGPT became widely available, an increase in chatbot research is noticeable since 2016 [1]. From a user centered perspective, research focuses on the interaction with the chatbots and how the dialogue and user experience can be improved [20, 11, 5]. This also includes customer service chatbots, which are the main focus in this work, where studies were looking into the user satisfaction, communication journeys or how the introduction of chatbots is impacting the users [15, 6, 13].

[5] `https://www.reddit.com/r/assholedesign/`

2.1 Background

In this section we introduce the basic concepts of Dark Patterns and Usability Smells, before we give an overview over the sparse literature which exists in the context of conversational user interfaces.

Dark Patterns. The term Dark Pattern was first coined by Brignull in 2010 and describes the malicious use of interaction design which brings a disadvantage to the users, or in the words of Brignull: *"tricks used in websites and apps that make you do things that you didn't mean to"* [3]. Well known instances of Dark Patterns include Cookie Banners, in which the option "select all" is preselected or a subscription is forced to obtain access to a service. Since the term was introduced, Dark Patterns became a popular research topic. The first context to find these patterns was on e-commerce websites [3], then it was transferred generally to GUIs. In the last years, several researchers tried to consolidate this rather novel and explorative field with meta-studies and taxonomies [8, 9, 17]. Still, there is not yet one common list of Dark Patters, as there are constantly novel and rather unique patterns found for specific use cases like mobile applications [12], Internet of Things (IoT) devices [14], or video games [21] as the field in which Dark Patterns can be found is vast. As no taxonomy could yet consolidate all patterns, one of the best known lists of Dark Pattern types can be found on *deceptive.design*[6]. We have not found studies about Dark Patterns in chatbots, which motivated this work. Research has mostly focused Dark Patterns in the context of GUIs, but determining their existence in CUIs is not trivial, since CUIs tend to serve a more specific purpose and also have more limited interaction options. Particularly for chatbots, it is harder to detect intentionality for Dark Patterns, and in the case of Usability Smells, it is more difficult to tell them apart from simple bugs or limitations.

Usability Smells. Usability Smells are catalogued signs of poor design that often lead to usability problems [10]. In contrast with Dark Patterns, smells are not intentionally placed in a website in favor of the site owner, as a bad usability may even drive users away. An example of Usability Smells is "Unformatted Input", in which a plain text box is used when specifically formatted data must be filled (e.g. phone number) but no hint or restriction is provided to the users, making it hard to enter the data in the right way - even if the data is correct.

The advantage of cataloguing Usability Smells is to provide concise descriptions and help detect bad GUI designs that make it hard for users to fulfill their tasks. Similarly to code smells, Usability Smells point to potential problems that can be solved by Usability Refactorings, i.e., transformations to the user interaction that preserve the system functionality.

[6] `https://www.deceptive.design/types`, previously called darkpatterns.org

2.2 Dark Patterns and Conversational User Interfaces

As Dark Patterns in chatbots is a new research area, we searched for publications with similar technologies. To the best of our knowledge, only three publications fulfill our requirements. We set our focus on CUIs and other user interfaces, where the user has a more limited interaction range than in a pure graphical setting, in which undesired interactions can be ended with a mouse click or keyboard shortcut. By using the established Dark Pattern types on *deceptive.design* [3], we will relate the findings of these publications to the Dark Patterns research. Due to the scope of our own experiment, we did not include any other taxonomies for this comparison, even when the publications also include them.

A provocation paper on unethical Design in CUIs was published by Mildner et al. in 2022 [18], in which they propose to concentrate on five specific characteristics of Dark Patterns. As research concentrated until now mostly on GUIs, they argue that the findings and lessons from this research should be used as a basis for CUIs. By relating to Mathur et al.'s Dark Pattern characteristics [16], Mildner et al. want to open the discussion for CUIs by rather suggesting to match them to the following characteristics: asymmetric, covert, deceptive, hides information and restrictive [18]. These characteristics rely on the functionalities and not the manifestations of Dark Patterns, and though they were introduced for GUIs, the authors still see their potential to be adaptable for CUIs. They thus propose to shift the focus on the research from finding specific descriptions of Dark Patterns to their underlying cause.

In another study [19], Owens et al. analyze several characteristics of Voice User Interfaces (VUIs) which are prone to be exploited by Dark Patterns. From these, an expert panel built 12 scenarios which were either deceptive or non-deceptive. This was followed by a survey, in which participants had to rate a part of the scenarios and could offer their own previous experiences with deceptive VUI behaviour. In the survey, participants report incidents with VUIs, where they experienced "Nudging" to subscribe or buy something, a "Feeling of Lack of Control", where the were forced to subscribe to content before they were even able to leave the current conversation and "Unsatisfactory Responses", where the voice assistant did not understand the question or was giving a way longer answer than was wanted. The first two cases can be related to the *deceptive.design* Dark Pattern types of Nagging and Forced Action [3].

Another direction was chosen by Kowalczyk et al. who built a codebook of Dark Patterns found in lab recordings of the usage of IoT devices and their apps [14]. They also included smart speaker interfaces in their research, although the pool of these interactions is rather limited to 7 devices, from which one was tested on smart interactions, while the rest only included setup interactions. In their final codebook, the authors include among others the following Dark Pattern types from *deceptive.design* which were found in smart speakers: Forced action, Trick wording, Hidden subscription, Sneaking, and Hard to cancel [3]. In combination with the limitation on the registration process, it is unclear which and how many Dark Patterns can be found in regular interactions with smart speakers in the home.

While six different Dark Pattern types from *deceptive.design* [3] could be found in affiliated areas to chatbots, the publications are either set in an experimental setting or do not classify examples from everyday use. Similar results could thus be expected in other VUIs, conversational agents and chatbots, but a validation is needed. While the provocation paper calls to research on Dark Patterns in CUIs, it lays a focus on the underlying causes and not the basic existence of them. As no research yet indicates if Dark Patterns can be even found in real world chatbot interactions, we conducted an experiment by collecting and coding these interactions.

3 Methodology

For scoping we narrowed the research question about Dark Patterns in chatbots from the introduction to customer service chatbots, as established Dark Patterns in GUIs often occur in e-commerce [3]. We started our research by collecting self-reported negative experiences from publicly available web sources, built a corpus of complaints about chatbots, narrowed down the corpus with described inclusion and exclusion criteria and coded these complaints in our *ChIPS* dataset. The section concludes with a look at the results and the discussion of their implications for further research on Dark Patterns in chatbots.

3.1 ChIPS Dataset

For the search of occurrences of Dark Patterns in existing chatbots, a dataset is needed. To our knowledge there is currently neither a crawler for publicly accessible chatbots available, nor a definite dataset with independent chatbot interactions of real users. While possible Dark Patterns could also be found in a clinical explorative setting, some possible interactions are hidden behind login pages or an ID-number for a service/invoice/etc. We wanted to use real interactions by real persons to get access to possibly hidden Dark Patterns for which we built the *ChIPS*[7] dataset ourselves. The dataset includes 69 complaints describing negative user interactions with chatbot.

To collect them, we searched for reports in three different data sources: subreddits, the *deceptive.design* Hall of Shame[8], and the database of user complaints from the Consumer Financial Protection Bureau[9]. Similar to Systematic Literature Reviews, we split our research question to construct relevant search terms [2], i.e. the used technology being chatbots. Likewise, we considered all possible spellings of the concept ‘chatbot’: *“chatbot”* and *“chat bot”*.

Three strategies were employed to find 213 possible occurances of Dark Patterns, from which 69 were included in the final dataset after applying inclusion and exclusion criteria. The initial set of candidates, the final dataset and other

[7] Chatbot Interactions for a Dark Patterns Search
[8] `https://www.deceptive.design/hall-of-shame`
[9] `https://www.consumerfinance.gov/data-research/consumer-complaints/search`

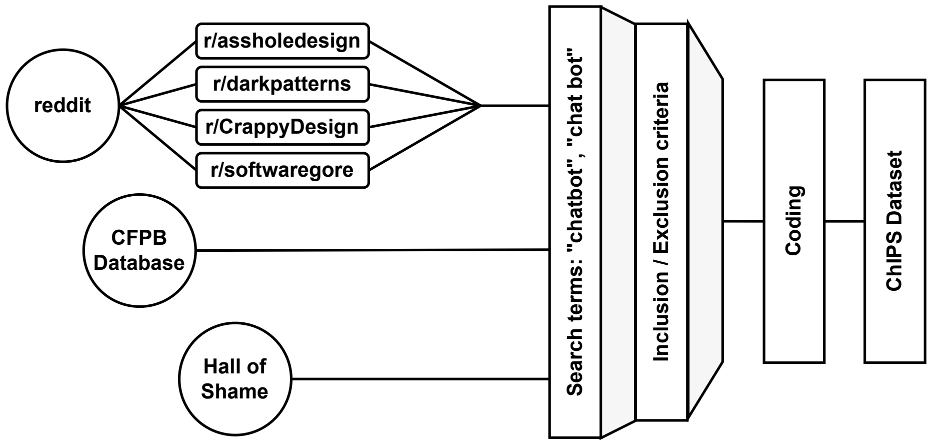

Fig. 1. The process of creating the *ChIPS* dataset.

data are publically accessible[10]. The collection process can be seen in Fig. 1. By adapting a previously used strategy by [7], we searched in several subreddits for negative experiences of users with chatbots. They showed that the subreddit r/assholedesign can be used as a potential source for Dark Patterns research. The subreddits' internal submission rules include a flow chart (Fig. 2) to decide if a post is relevant, which considers characteristics similar to the already established Dark Patterns definition. In addition to *r/assholedesign*, three other subreddits were used: *r/darkpatterns*[11], *r/CrappyDesign*[12], and *r/softwaregore*[13]. This decision was made, as users might be unreliable in posting experiences in the appropriate subreddit and because we formulated specific inclusion and exclusion criteria for the final *ChIPS* dataset which are not limited to one subreddit.

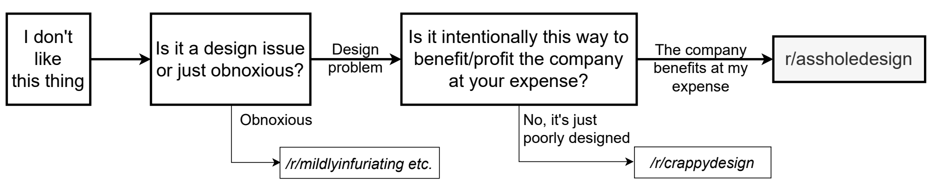

Fig. 2. Flow chart of the rules on posting in the subreddit /r/assholedesign[14].

[10] https://github.com/vertr/ChIPS-dataset
[11] https://www.reddit.com/r/darkpatterns/
[12] https://www.reddit.com/r/CrappyDesign/
[13] https://www.reddit.com/r/softwaregore/
[14] https://www.reddit.com/r/assholedesign/comments/lnymf2/meta_an_updated_flow_chart_to_help_cut_down_on/

Table 1. Inclusion and exclusion criteria for relevant complaints.

Inclusion criteria	**Exclusion criteria**
1. Posts made in the English language.	1. Negative experiences while accessing a chatbot or for the choice of a chatbot.
2. Post shall describe or show a negative experience while using a chatbot.	2. Post from developers who have problems while programming their own chatbot or for a company.
3. The chatbot has to be an official instance of a company which allows chat interactions.	3. Posts which include the same negative experience that was made not only with chatbots but also on other communication channels.
4. The experience shall be made personally by the reporting person with a chatbot of another party.	4. Posts which contain clearly recognizable software bugs, even for laypersons.
5. The described situation and problem have to be clearly identifiable.	
6. The interaction with the chatbot has to be based in the need for getting information or in claiming a service.	

As Dark Patterns in GUIs are already reliably found in the e-commerce or customer setting [3], the second strategy includes the official complaints database of the Consumer Financial Protection Bureau (CFPB). By not only relying on social media accounts, the diversity of the complaint authors could be widened. The last strategy consisted in searching the *deceptive.design* website, which offers a collection of Dark Patterns "in the Wild" in their Hall of Shame. All three web sources were searched with the previously introduced search terms. The results were not limited in the year they were submitted; however, as chatbots are a relatively new technology, the found interactions had a natural limit at ca. 2011.

3.2 Inclusion/Exclusion criteria

After determining the data sources, criteria for a filtering of complaints were built. These criteria should make sure to build a meaningful corpus for the coders. Both sets of criteria can be seen in Table 1.

The data gathering and preselection was made by one author, while two others functioned as coders for the preselected complaints. Generally, if an edge case was found, the example was added in the pool of included complaints, to let the coders decide whether it was a possible Dark Pattern, a Usability Smell or neither. Complaints in which an undesired behaviour was reported, but it was not clear which company was talked about, were excluded due to inclusion criteria 3 and 5. In the same line, reports in which the main focus was on an experienced bad customer service and chatbots were only negatively connotated as "bad", "unhelpful" or "abysmal" without a more detailed interaction descriptions, were excluded according to inclusion criterion 5.

The code labels were defined as follows: 16 labels that include all listed Dark Pattern types on *deceptive.design*, one label for Usability Smells, one label for new chatbot-specific Dark Patterns, one for no Dark Pattern detected, and one label named "cannot be determined" for cases where no other label fits. These 20 labels together with their explanations are included with the *ChiPS* dataset. The coders each received lists of all included complaints in randomized order and coded them. For complaints which were coded with different labels, we used a structured process of discussions between the coders and the author who built the *ChiPS* dataset to resolve this divergence. Relevant parts of these resolutions are also included in the result section. Findings from the coding were also initially discussed in this session and further specified in later ones.

3.3 Results

In this section we present the results from the coding process and some of the related resolutions between the coders, before the findings are discussed in the next section. Out of the 69 complaints, 11 included overall 6 established Dark Patterns in GUIs, 7 Usability Smells, 24 examples did not include enough context to decide on and 27 instances were coded as not containing any Dark Pattern. There were no new chatbot specific Dark Patterns found. Table 2 shows an overview over representative examples of the dataset. For the sake of brevity, only labels which were coded are included in the table. A detailed view on the coding is available in the shared data repository.

First, we want to highlight the findings for already established Dark Patterns. From the distribution of found labels in Figure 3 we can see that the following patterns were found in chatbots: *Nagging*, *Confirmshaming*, *Hard to cancel*, *Forced action*, and *Obstruction*. These examples show that Dark Patterns that are already used in GUIs can be transferred in conversational interactions. The following list presents first the results for already established Dark Patterns before we continue with the other coded labels.

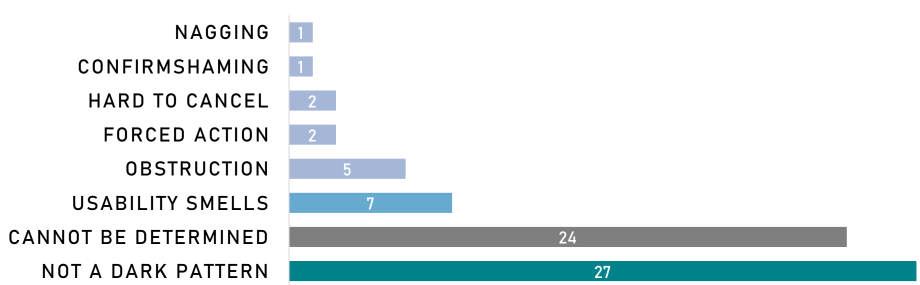

Fig. 3. A bar chart with the distribution of all coded labels in the *ChIPS* dataset.

Table 2. Descriptions of representative examples for coded labels.

Code labels	Representative examples
Nagging	The user is browsing a homepage and gets disturbed by popups of a “Sales Special” chatbot which is hard to close.
Confirmshaming	The user wants to cancel a subscription, in response to which the chatbot uses guilt-tripping and persuading language to convince them to not cancel.
Hard to cancel	The user wants to cancel an internet subscription and is then sent by the chatbot to other communication channels to finalize the cancellation.
Forced action	The chatbot automatically asks the user to accept “alerts and updates” to be able to start chatting.
Obstruction	The user wants to dispute a fraudulent charge and even though the chatbot can not help, it also denies contact to live agents who could help.
Usability Smells	The user wants to use the chatbot and when entering their 17-digit account number, the chatbot does not recognize it because it expects 11 digits.
Cannot be determined	The user chatted with a chatbot before they were referred to a live agent, at which point the connection seemed to break and they were repeatedly thrown out of the chat.
Not a Dark Pattern	The user searches for an explanation of a credit score drop and the chatbot could only answer basic Q&A pairs.

Nagging (1 occurrence). The only coded instances of Nagging is concerning a chatbot which apparently pops up to offer information about a “Sales Speciale” while the user is visiting the homepage. Most chatbots either have to be navigated to as a separate page element or they are part of the homepage and included in the lower right corner of the interface. Most often, chatbots are only visible as an inconspicuous icon than can be extended if the user wishes to enable the chat function. To turn them into popups, which could also be accompanied by alert notifications[15], will negatively impact the user experience.

Confirmshaming (1 occurrence). The complaint coded as Confirmshaming is a chatbot of Modern Milkman, a company that delivers milk in glass bottles to avoid pollution. According to the report, the chatbot sends several messages with statistics, trying to convince the user of the company benefits, guilting them into continuing the subscription and even asks them to “*keep fighting the good fight against nasty single-use plastics*”, if they insist on cancelling. The difference from this Dark Pattern occuring in graphical user interfaces is that the chatbot also uses emojis and is able to flood the chat with seemingly pre-programmed messages, possibly in an attempt to dissuade the user from cancelling the subscription.

[15] An example can be found here: `https://asana.com/de`

Hard to cancel (2 occurrences). This pattern was found in two instances, from which we will highlight one here. There, the user wants to cancel an internet service, but after he chose the option to cancel the corresponding service, the chatbot refers them to other contact options and even tells the user that they will try to "*keep [them] as a customer*". This procedure of leading the customer from one contact to another without them being able to directly cancel is a typical example of this Dark Pattern[16] and has the goal to make the cancellation as complex and time-consuming as possible.

Forced action (2 occurrences). In one of these complaints, the user is forced in a WhatsApp chat to opt in "to receive alerts and updates" before any services can be used. This example sparked some debate between the coders as this wording could simply mean a confirmation to receive messages at all, but the formulation could also include ads or other unwanted messages. It is also unclear if this consent can be revoked later on. Additionally to this unclear information, the user is only able to choose between Yes and No which prevents them from getting more details regarding the terminology. By assuming that these terms include messages beyond any necessary messages for the conversation, the coders reached a consent to label this example as Forced action.

Obstruction (5 occurrences). Of the five complaints that were coded as Obstruction, four include instances where the chatbot seemingly prevents the users to get in contact with live agents. In the chosen report to highlight this pattern, the user wants to dispute a fraudulent charge, but is unable to do so via the online portal and is also not able to reach a live agent as a contact. In the fifth complaint, the chatbot was not able to disclose the terms and services that apply to using it. In the case of the Obstruction Dark Pattern, there was debate among the coders about whether it could be a specific Dark Pattern for chatbots. Generally, an Obstruction like this could also occur within a GUI which does not offer any contact option at all. While there can be made an argument that an available communication channel offers another premise than only a graphical user interface, the coders decided to not regard this as chatbot specific, as the same situation can occur with automated telephone services.

Usability Smells (7 occurrences). Here we describe some examples of Usability Smells found in negative interactions with CUIs. To the best of our knowledge, this is the first study to analyze Usability Smells in chatbots.

From the 7 instances of usability smells that we found among the complaints, there were some known Usability Smells for GUIs which include for example missing information on the waiting time, a sudden disconnection after being put in a queue as no live agents were available, or a time counter until a live agent is available which increased over time instead of decreasing. Two of the reports

[16] sometimes also called 'Roach Motel'

indicate that, when trying to connect to a real agent, the chatbot does not provide a time estimation, typically shown as a queue. This could be compared to a known Usability Smell in GUIs named "No Loading Indicator", which describes a situation in which a time-consuming process does not provide clear indication of the remaining time, or at least a "Loading..." page / spinner. Waiting queues are such a typical interaction with chatbots that not having one could be considered a chatbot-specific Usability Smell. These examples show that some negative interactions with chatbots can be traced back to not understandable and insufficient feedback for the user on processes that are running in the background. Even though a situation where customers are not able to talk to a person might seem malicious, we did not label it as a Dark Pattern, as the companies have no gain from this.

Cannot be determined (24 occurrences). Some of the complaints that were labelled as "cannot be determined" have a potential for specific chatbot Dark Patterns, although this would depend on particular circumstances. One user was for example removed from the chat after the chatbot transferred the conversation to a live agent. This forced disconnect was seemingly due to a technical error, especially as the chatbot previously mentions possible connection problems on the side of the user. For the customer it still seemed as being ignored and forcefully removed from a chat, where from their viewpoint, the other side did not react on any messages. While it is possible to intentionally end a chat-conversation at some point, it is unclear how a company might benefit from this, to make this a Dark Pattern. The communicated technical problems were also the reason why this example was not labelled as a Dark Pattern.

One other complaint that could not be determined due to missing information concerned the complaint of a user who received a refund on tickets, but it was unclear to them that this refund would be split between all tickets. As this was a customer complaint that did not include a screenshot of the conversation, the coders decided to label this instance as "cannot be determined", although it might be a "Trick wording" Dark Pattern. This pattern is also prone to be potentially used in a conversational user interface. Not only because a conversation has generally the possibility to include misunderstandings, but even more so as most currently used chatbots are trained with information from FAQs, homepages or other available documents. Depending on this, trick wordings that are already in use on the graphical user interface might thus be transferred to a conversational one.

Not a Dark Pattern (27 occurrences). Besides undeterminable occurances, most other chatbot interactions did not include a Dark Pattern.

Even though users seem to want to often interact with live agents, chatbots still are expected to be available 24/7, which is often not the case. Cases in which the access was restricted per-se were excluded from the dataset in accordance to exclusion criteria 1. In the dataset two complaints were included for similar situations with the Playstation chatbot, where access to the chatbot was generally possible, but only granted once per day, which was apparently only clear to the users after they were already able to open the chat. Three reports were labelled as not a Dark Pattern, because users were unsatisfied as the chatbot was technically not able to fulfill their high expectations. One user locked themselves out of their 2-factor-authentication and tried several ways to talk with someone from the company. As the chatbot could not help with restoring the 2-factor-authentication, it apparently did "not respond to anymore messages" from the users. Another user had questions on their credit score, which the chatbot could not answer as it was apparently only programmed to answer questions that are part of the FAQ section. A similar example was also found with the Venmo chatbot, which was unable to help with payment problems and could seemingly also only answer questions from the FAQ. Most interactions that were found are dated before LLMs became widely available. To interact with these chatbots, the prompts have to be formulated very precisely and in an understandable way for the chatbot. Some users reverted to short or single word prompts like "hold payment" or "agent" or used colloquial language like "Y'all". This led to the chatbot either not understanding the prompts or giving false information. This in turn was interpreted by the users as the chatbot being unwilling to give them information or to consciously decide to give them false information. While this was an unsatisfactory and possibly disturbing experience to the users, it is missing the intentionality to be classified as a Dark Pattern.

3.4 Discussion

By building the dataset, both in the gathering and the coding process, we could gain some insights about the nature of interaction with chatbots, and users dispositions and expectations towards them. Telling Dark Patterns apart from Usability Smells, and even bugs, was not a trivial task - as was shown during the post-coding discussions.

One of the most relevant findings was that, as chatbots only offer a linear and restricted interface, users often face situations in which they are either dissatisfied or confused. These situations can often be linked to Usability Smells, which can be solved by applying Usability Refactorings that may lead to a more intuitive user interface design. Examples of these refactorings include clear indications of time stamps or avoiding to send the same automated standard message in one conversation to the user several times. This is in line with early guidelines for good chatbot user interfaces which also include transparency on the waiting times for messages and to avoid automated messages in situations where they might be unwanted or could be confusing [4]. Many of our collected complaints could have been avoided by following these or similar guidelines in the implementation, both for Usability Smells and general better user interface designs. Most chatbots do not have publicly available information on their programming, training or the rules on which the conversations operate. Especially laymen might find it hard to comprehend the technical limitations in their interactions, and could interpret this as a malicious intent from the company or even an anthropomorphized version of the chatbot itself.

From the 5 Dark Patterns which we found in chatbot interactions, 3 were also found in similar studies [14, 19], which together with this paper provides a further validation of Dark Patterns in conversational user interfaces. The commonly identified Dark Patterns both in the literature and our *ChIPS* dataset are Forced action, Nagging and Hard to Cancel. There are not yet many chatbots which are selling products or services to users, which might be a reason that we did not find many Dark Patterns overall in our research. The majority of Dark Patterns in the e-commerce sector are related to selling/subscriptions and not customer service for which many chatbos in e-commerce are used today.

Our *ChIPS* dataset indicates that users have difficulties in two main aspects of the user interactions. First, they expect the chatbot to be able to perform more functionalities than is technically possible. To avoid this, the available functionalities (like answering FAQs or providing basic information about accounts or bills) should be made clear before the users interact with the chatbot. Secondly, and also connected to the first aspect, users are often not able to formulate prompts which are understandable for the chatbot. Some chatbots already give examples on how the users should interact with them, but especially through emerging AI trained chatbots and media representation, users might continue to have difficulties to communicate their needs to a simple rule based chatbot which is only trained on limited data. These two implications were also found in a previous study about communication journeys of users with customer service chatbots [6]. Established interaction designs, commonplace in GUIs, are still missing for chatbots, and users might falsely expect to be able to communicate with one chatbot in a certain way due to prior experience with others.

Rather than Dark Patterns, it is possible that users of chatbots currently mostly suffer from a relatively new technology that is still evolving and for which there are not yet general default designs or best practices established.

3.5 Limitations and Threats to Validity

For the data gathering of the examples that were used, we want to address some limitations which might be threats to the validity. Generating a dataset of chatbot interactions was a challenging task, since is is based on reports. This was affected by internal, external and construct validity threats.

Internal validity threats are mainly linked to the subjective nature of the reports, and also the coders' subjective criteria for labeling them. For one, we are relying on self reported situations of negative experiences, where the understanding of a situation is heavily influenced by the emotions of the customers who are likely stressed at the time of the report, and thus might skew them. This sometimes leads to coders not being able to determine the code label as information was missing or vaguely described. Stressed or angry users are also prone to misrepresent or exaggerate their negative interaction. For instance, in the particular case of the Obstruction Dark Pattern, many reports claim that the organizations were withholding the contact to live agents, which could be only the users' perception. Especially the customer complaints from the CFPB include often long winded tales about several failed interactions, where chatbots were only a small part of the story. This could also bias the coders, since they are far from being a neutral description of the facts. We mitigated this threat by consolidating the labels in a discussion session after the individual coding. However, we can not exclude a possible influence on the coders from other situations that were part of the narratives.

An external bias on the validity could be caused by the size of the dataset. Our intent to label real life interactions lead to a limited number of complaints that could be collected, because the experiences can not be reproduced in a lab setting. We attempted to mitigate this threat by searching in different sources, so as to get a set that is as diverse as possible, but this could be improved by collecting a higher number of reports from three different web sources.

Construct validity could also have been compromised in the labeling process, since the assessment of the presence of only Dark Patterns could have led the coders to attribute any problems in the interaction with a chatbot to a Dark Pattern, perhaps forcing the labels. We prevented this by adding the special label "Usability Smell" that describes an involuntary interaction problem.

4 Conclusion and Future Work

In this work we showed that Dark Patterns that are already established in GUIs can be transferred from graphical to conversational user interfaces and are already used in chatbots. In particular we also found the following Dark Pattern types: Nagging, Confirmshaming, Hard to cancel, Forced action and Obstruction. Our results underline the importance to create new experiments for finding traces of Dark Patterns that are specific for interactions with conversational user interfaces. The occurrence of Usability Smells and complaints in which the situation was not clearly determinable implies that users have both sometimes too high expectations on chatbots, and are often not able to formulate prompts in a way that is understandable for the chatbot.

This study serves as a first foray in the existence of Dark Patterns in chatbots. We hope that this work can inspire the CUI community to look into the gray area between Dark Patterns and bad usability, which needs to be further differentiated. Future work includes searching for Dark Patterns in other kinds of chatbots, for example AI companions like Replika[17] or chatbots in the gaming industry. Another research direction also includes the automated detection of Dark Patterns and Usability Smells in chatbots.

References

1. Adamopoulou, E., Moussiades, L.: Chatbots: History, technology, and applications. Machine Learning with Applications **2**, 100006 (2020)

[17] https://replika.com/

2. Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. Journal of Systems and Software **80**(4), 571–583 (apr 2007). https://doi.org/10.1016/j.jss.2006.07.009, https://linkinghub.elsevier.com/retrieve/pii/S016412120600197X
3. Brignull, H.: Deceptive patterns. https://www.deceptive.design/ (2023), last accessed 6 Sept 2023
4. Budui, R.: The User Experience of Customer-Service Chat: 20 Guidelines. https://www.nngroup.com/articles/chat-ux/ (2019), last accessed 26 Oct 2023
5. Følstad, A., Brandtzaeg, P.B.: Users' experiences with chatbots: findings from a questionnaire study. Quality and User Experience **5**(1), 3 (2020)
6. van der Goot, M.J., Hafkamp, L., Dankfort, Z.: Customer service chatbots: A qualitative interview study into the communication journey of customers. In: International Workshop on Chatbot Research and Design. pp. 190–204. Springer (2020)
7. Gray, C.M., Chivukula, S.S., Lee, A.: What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. In: Proceedings of the 2020 acm designing interactive systems conference. pp. 61–73 (2020)
8. Gray, C.M., Sanchez Chamorro, L., Obi, I., Duane, J.N.: Mapping the Landscape of Dark Patterns Scholarship: A Systematic Literature Review. In: Companion Publication of the 2023 ACM Designing Interactive Systems Conference. pp. 188–193 (2023)
9. Gray, C.M., Santos, C., Bielova, N.: Towards a preliminary ontology of dark patterns knowledge. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–9 (2023)
10. Grigera, J., Garrido, A., Rivero, J.M., Rossi, G.: Automatic detection of usability smells in web applications. International Journal of Human-Computer Studies **97**, 129–148 (2017)
11. Hernandez-Bocanegra, D.C., Ziegler, J.: Conversational review-based explanations for recommender systems: Exploring users' query behavior. In: Proceedings of the 3rd Conference on Conversational User Interfaces. pp. 1–11 (2021)
12. Hidaka, S., Kobuki, S., Watanabe, M., Seaborn, K.: Linguistic Dead-Ends and Alphabet Soup: Finding Dark Patterns in Japanese Apps. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2023)
13. van Hooijdonk, C., Martijn, G., Liebrecht, C.: A framework and content analysis of social cues in the introductions of customer service chatbots. In: International Workshop on Chatbot Research and Design. pp. 118–133. Springer (2022)
14. Kowalczyk, M., Gunawan, J.T., Choffnes, D., Dubois, D.J., Hartzog, W., Wilson, C.: Understanding Dark Patterns in Home IoT Devices. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–27 (2023)
15. Kvale, K., Freddi, E., Hodnebrog, S., Sell, O.A., Følstad, A.: Understanding the user experience of customer service chatbots: what can we learn from customer satisfaction surveys? In: International Workshop on Chatbot Research and Design. pp. 205–218. Springer (2020)
16. Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M., Narayanan, A.: Dark patterns at scale: Findings from a crawl of 11K shopping websites. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–32 (2019)
17. Mathur, A., Kshirsagar, M., Mayer, J.: 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. pp. 1–18 (2021)
18. Mildner, T., Doyle, P., Savino, G.L., Malaka, R.: Rules Of Engagement: Levelling Up To Combat Unethical CUI Design. In: Proceedings of the 4th Conference on Conversational User Interfaces. pp. 1–5 (2022)
19. Owens, K., Gunawan, J., Choffnes, D., Emami-Naeini, P., Kohno, T., Roesner, F.: Exploring deceptive design patterns in voice interfaces. In: Proceedings of the 2022 European Symposium on Usable Security. pp. 64–78 (2022)
20. Yildiz, E., Bensch, S., Dignum, F.: Incorporating social practices in dialogue systems. In: International Workshop on Chatbot Research and Design. pp. 108–123. Springer (2021)
21. Zagal, J.P., Björk, S., Lewis, C.: Dark patterns in the design of games. In: Foundations of Digital Games 2013 (2013)