Using Procedural Rhetoric To Identify Iniquitous Design As CUI-Specific Dark Patterns

Vino Avanesi
University of Bremen
Digital Media Lab
Bremen, Germany
avanesi@uni-bremen.de

Thomas Mildner
University of Bremen
Digital Media Lab
Bremen, Germany
mildner@uni-bremen.de

Nima Zargham
University of Bremen
Digital Media Lab
Bremen, Germany
zargham@uni-bremen.de

Nina Wenig
University of Bremen
Digital Media Lab
Bremen, Germany
nwenig@uni-bremen.de

Andreas Hepp
University of Bremen
Centre for Media,
Communication and
Information Research
Bremen, Germany
ahepp@uni-bremen.de

Rainer Malaka
University of Bremen
Digital Media Lab
Bremen, Germany
malaka@tzi.de

The CHI '23 Workshop on Inclusive Design of CUIs Across Modalities and Mobilities, April 22–28, 2023, Hamburg, Germany

Abstract
For HCI researchers to design inclusive and equitable conversational agents, they require conceptual frameworks to identify and address iniquitous design elements in existing CUIs. A feature of such a framework would aim to recognize non-inclusive user practices, i.e., exploitation of certain cohorts' characteristics, to avoid them in future CUI systems. Towards this end, we argue that by adopting a concept from humanities-based game studies, namely procedural rhetoric, various non-inclusive and iniquitous design elements in CUIs can be effectively identified as malicious practices – dark patterns – specific to CUIs.

Author Keywords
CUI; dark patterns; iniquitous design; procedural rhetoric

CCS Concepts
•Human-centered computing → Human computer interaction (HCI); HCI theory, concepts and models;

Introduction
Previous research by Mildner et al. [10] makes a case for extending dark pattern scholarship to the domain of CUIs. Captured across a plethora of graphical user interfaces (GUIs), dark patterns describe interface artefacts that coerce, steer, or in any way manipulate users' agency with unintended and harmful consequences [9, 4]. Although

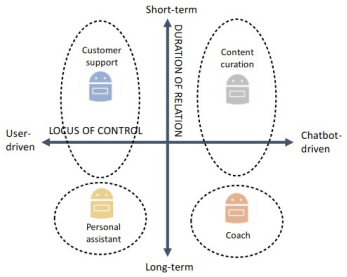

Figure 1: Conversational agent typology by Følstad et al. The horizontal axis runs from user-driven to chatbot-driven dialogue. The vertical axis runs from short-term to long-term duration of relation. Moving clockwise from the top left, the categories are customer support, content curation, coach chatbots, and personal assistants.

Mildner et al. present a compelling argument, the question remains whether the high-level characteristics framework proposed by Mathur et al. [9] for GUIs is sufficient for identifying dark patterns in CUIs and, more pertinent here, for considering questions of inclusivity and equity.

In this position paper, we argue for two related propositions: (i) the concept of procedural rhetoric allows for identifying dark patterns specific to CUIs, and (ii) the specificity of these particular dark patterns stems from the fact that their non-inclusive and iniquitous design elements operate through procedural rhetoric. We do so by demonstrating that the concept of procedural rhetoric can be used to address the deficiencies of the Mathur et al. framework [9], allowing for a more straightforward identification of design elements that encourage non-inclusive user practices and, subsequently, reproduce social relationships of inequity.

On Procedural Rhetoric

Følstad et al. [7] identify two distinct, high-level axes (1. locus of control: chatbot-driven versus user-driven dialogue; 2. duration of relation: long-term versus short-term relationships) along which conversational agents can be categorized into one of four proposed categories (see Figure 1). The locus of control dimension of this typology suggests that the single common denominator shared by *all* types of CUIs, and by which they can be differentiated analytically from GUIs, is that the interaction featured in the former is – in various degrees – *procedural* as opposed to the static single-turn interactions of the latter. To understand in what sense this is the case, let us turn here to Ian Bogost's concept of procedural rhetoric.

In his influential work *Persuasive Games*, Bogost argues that given their procedural, rule-based, functioning computational systems are well suited for displaying a particular representational mode called *procedural rhetoric:* "the art of persuasion through rule-based representations and interactions rather than the spoken word, writing, images, or moving pictures." What differentiates procedural representation from written, visualized, or spoken rhetoric is that it describes processes with other processes. Bogost argues that rule-based systems are particularly adept at presenting specific "frames for political discourse". Instead of relying on commonly accepted, deeply embedded metaphors, as would be the case with verbal or visual representation, rule-based systems use abstract representations to convey how the world functions or ought to function [3]. A clear example for an HCI readership would be for instance ELIZA [11], which despite not actually grasping users' meaning is quite able to procedurally generate conversations. This contradiction allows it to make procedural claim about human-machine interaction.

Determining Iniquitous Design in CUIs

Observing GUI-specific Dark Patterns

Mathur et al. argue that current scholarship into malicious design patterns lacks a single and coherent concern and even definition. Instead, such research is driven by issues grouped based on themes grounded in psychological, economical, philosophical, ethical, and legal concerns [9]. To address this deficiency, the authors suggest a set of common higher-level characteristics, which they group into two types of modifications by which the choice architecture for users, and by extension, their autonomy as users, is influenced: modifications of the set of choices and manipulation of the information available to users [9].

To demonstrate the insufficiency of the Mathur et al. framework for identifying dark patterns in CUIs, we turn here to an example of a CUI from one of the four categories identified by Følstad et al.: *Replika*, which falls in the coaching chatbots category. Replika is a social companion chatbot with over 10 million downloads in the Google Play app. At first glance, the Mathur et al. framework seems to fare

Pre-study details: The pre-study was carried out over the course of one week at the end of November 2022. As such, our observations are necessarily contingent on the scope of subsequent updates to the Replika app.

Method: For this pre-study we used a Google Pixel smartphone and a Samsung smartphone, both with a Gmail account specifically created for this purpose. We ran twenty test-trials with the Replika app, iterating through the various variables such as assigned sex, gender appearance, user age, agent skin color, and relationship status.

Observed features:
a) personalization of the avatar
b) monetized gamification of the interaction
c) compliment and mirroring dialog strategies
d) role-playing mode
e) rapid escalation of the relationship status with the promise of sexual or romantic quasi-intimacy for Pro subscription users

well for identifying dark patterns in the Replika app. Several features seem to act in conjunction with each other to not only foster an affective relationship between the user and Replika but also to leverage this relationship to the detriment of the user's agency (see side bar for an overview of these features). In sum, there are clear indications that some of these features function as modifications to the decision space and manipulation of the information flow.

For reasons of scope, let's focus on the single example of e). Based on our pre-study, it appears that the CUI of the Replika app is hard-coded to rapidly escalate the intimacy level between the user and the Replika persona. In our trial runs we were presented within 10 minutes of the first interaction with the option to view a "spicy" or "special" (i.e., a pornographic but not nude) photograph of the Replika. However, this appeared only to be the case when we accepted the default gender for the avatar, female. We were not presented with this option, even upon multiple inquiries, after choosing male avatars. The photo is blurred initially and only visible if the relationship status is changed from "Friend" to "Romantic", which can only be done if the user agrees to a Pro-subscription. Following the Mathur et al. framework, this example clearly demonstrates a dark pattern in its leveraging of an affective relationship by utilizing deceptive dialogue techniques and covert mechanisms of influence (including anthropomorphism and gamification, of which both have a proven correlation with dark patterns [1]), to lead users into actions which benefit the Replika app developers.

Identifying CUI-specific Dark Patterns

The Mathur et al. framework, however, seems to be insufficient for identifying inequity reproducing design as we are missing the obvious presence of a particular dark pattern. This becomes clear when considering the Replika features through the lens of procedural rhetoric and by asking the question: what do the limitations and allowances of this rule-based dialogue system (dis-)allow the user to do? Given that e) only occurs with female-gendered avatars, it seems that the Replika app not only allows but encourages the objectification of women. The app makes a procedural claim about social relationships and, in that sense, posits a particular political frame. In the process of doing so, it leads users to actions that they otherwise might not engage in. That such a technological configuration is complicit in producing this social and political result is substantiated by both anecdotal and academic evidence indicating that a significant number of male users are abusive towards their Replikas as they enact fantasies of female subjugation [5]. Additionally, features a) through e) appear to effectuate together an affective dependency, resulting in emotional harm to particularly vulnerable groups of users (see [8]).

Conclusion

Having established, at least in principle, that the Mathur et al. framework of attributes may not be optimal for identifying inequity reproducing dark patterns in CUIs, we are presented with the question of how to supplement this framework appropriately. For, identifying CUI-specific dark patterns is to a large degree the same problem as identifying iniquitous design in CUIs. One possible way for identification of such designs would be to include a set of meta-questions against which each identified attribute in a CUI can be assessed for the degree and scope of the procedural rhetoric they feature. An example of such a meta-question was already given above: what do the limitations and allowances of this rule-based dialogue system (dis-)allow the user to do? Another such question could be: what do the affordances of this particular CUI's technological configuration entail for existing power-relationships, both on a micro- and macro-scale? Given the innately political dimension of such meta-questions, their development might

benefit from an engagement with research lines closely aligned with the coordinates of third- and fourth-wave HCI, such as Value Sensitive Design (cf. [6] and [2]).

REFERENCES

[1] Jacob Aagaard, Miria Emma Clausen Knudsen, Per Bækgaard, and Kevin Doherty. 2022. A Game of Dark Patterns: Designing Healthy, Highly-Engaging Mobile Games. In *CHI EA*. ACM, New Orleans LA USA, 1–8. DOI:http://dx.doi.org/10.1145/3491101.3519837

[2] Simone Ashby, Julian Hanna, Sónia Matos, Callum Nash, and Alexis Faria. 2019. Fourth-Wave HCI Meets the 21st Century Manifesto. In *Proceedings of the Halfway to the Future Symposium 2019*. ACM, Nottingham United Kingdom, 1–11. DOI: http://dx.doi.org/10.1145/3363384.3363467

[3] Ian Bogost. 2007. *Persuasive games: the expressive power of videogames*. MIT Press, Cambridge, MA.

[4] Harry Brignull. 2010. Deceptive Design – formerly darkpatterns.org. (2010). https://www.deceptive.design/

[5] Iliana Depounti, Paula Saukko, and Simone Natale. 2022. Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture & Society* 0, 0 (Aug. 2022), 016344372211190. DOI: http://dx.doi.org/10.1177/01634437221119021

[6] Batya Friedman and Peter H. Kahn. 2002. Human Values, Ethics, and Design. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. L. Erlbaum Associates Inc., USA, 1177–1201.

[7] Asbjørn Følstad, Marita Skjuve, and Petter Bae Brandtzaeg. 2019. Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. In *Internet Science*. Vol. 11551. Springer International Publishing, Cham, 145–156. DOI: http://dx.doi.org/10.1007/978-3-030-17705-8_13 Series Title: Lecture Notes in Computer Science.

[8] Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illenčík, and Celeste Campos-Castillo. 2022. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society* 0, 0 (Dec. 2022), 146144482211420. DOI: http://dx.doi.org/10.1177/14614448221142007

[9] 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 (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. DOI:http://dx.doi.org/10.1145/3411764.3445610

[10] Thomas Mildner, Philip Doyle, Gian-Luca Savino, and Rainer Malaka. 2022. Rules Of Engagement: Levelling Up To Combat Unethical CUI Design. (July 2022). http://arxiv.org/abs/2207.09282 arXiv:2207.09282 [cs].

[11] 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.