

From C-3PO to HAL: Opening The Discourse About The Dark Side of Multi-Modal Social Agents

Vino Avanesi
avanesi@uni-bremen.de
University of Bremen
Germany

Johanna Rockstroh
rockstro@uni-bremen.de
University of Bremen
Germany

Thomas Mildner
mildner@uni-bremen.de
University of Bremen
Germany

Nima Zargham
zargham@uni-bremen.de
University of Bremen
Germany

Leon Reicherts
l.reicherts.17@ucl.ac.uk
University College London
United Kingdom

Maximilian A. Friehs
m.a.friehs@utwente.nl
University of Twente
Netherlands

Dimosthenis Kontogiorgos
dimosthenis.kontogiorgos@hu-berlin.de
Humboldt University of Berlin
Germany

Nina Wenig
nwenig@uni-bremen.de
University of Bremen
Germany

Rainer Malaka
malaka@tzi.de
University of Bremen
Germany



Figure 1: AI generated image of a social robot that is both cute and creepy using the Image Creator from Microsoft Bing.

ABSTRACT

The increasing prevalence of communicative agents raises questions about human-agent communication and the impact of such interaction on people's behavior in society and human-human communication. This workshop aims to address three of those questions: (i) How can we identify malicious design strategies – known as dark patterns – in social agents?; (ii) What is the necessity for and the effects of present and future design features, across different modalities and social contexts, in social agents?; (iii) How can we incorporate the findings of the first two questions into the design of

social agents? This workshop seeks to conjoin ongoing discourses of the CUI and wider HCI communities, including recent trends focusing on ethical designs. Out of the collaborative discussion, the workshop will produce a document distilling possible research lines and topics encouraging future collaborations.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**; *Interaction design*.

KEYWORDS

social agents, dark patterns, human-agent communication, conversational user interfaces, multi-modal interaction

ACM Reference Format:

Vino Avanesi, Johanna Rockstroh, Thomas Mildner, Nima Zargham, Leon Reicherts, Maximilian A. Friehs, Dimosthenis Kontogiorgos, Nina Wenig, and Rainer Malaka. 2023. From C-3PO to HAL: Opening The Discourse

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CUI '23, July 19–21, 2023, Eindhoven, Netherlands

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0014-9/23/07.

<https://doi.org/10.1145/3571884.3597441>

About The Dark Side of Multi-Modal Social Agents. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3571884.3597441>

1 INTRODUCTION

Communicative bots such as voice assistants, social (ro-)bots, and taskbots are increasingly becoming part of everyday life [17, 34], with work showing how we integrate these agents in our daily routines [11] while creating emotional bonds with them [24, 25]. As these artificial agents grow in numbers and level of sophistication, so does the number of questions relating to human-agent communication. In particular, these agents raise the question: how do humans relate to artificial entities that are awarded the position of social agent [12, 16]? We aim to address this question in this workshop by focusing on three related questions. First, we ask: how can we identify malicious design strategies – often called dark patterns [15] – in social agents? Secondly: what is the necessity for and the effects of certain design features, across different modalities and social contexts, in social agents? And lastly: how can we incorporate the findings of the first two questions into the design of social agents?

1.1 From “Computers Are Social Actors” to Dark Patterns

The Computers Are Social Actors (CASA) framework is one of the well-known lines of inquiry within HCI research which has attempted to address this question of artificial entities as social agents [32, 33]. Central to the CASA framework is the proposition that humans, by defaulting to the cognitive bias of anthropomorphism, assign social actorship to computational artifacts if said entities display even the most minimal of social cues such as human-like visual or vocal qualities (e.g., [33] and [32]). However, the CASA framework has its limitations as users, technologies, and users’ interactions with these technologies have changed since the framework was first conceptualized in the early 1990s [14, 23].

While the CASA paradigm demonstrates the effectiveness of anthropomorphism, recent work shines a light on interrelated risks when exploited. Revolving around the concept of dark patterns, such work has mainly focused on screen-based strategies to misguide users into unfavorable or even harmful actions to the advantage of service providers. Although there is much variety among the different types [8, 15, 31], dark patterns can be defined – based on a meta-analysis by Mathur et al. [29] – as those features of an interface that “modify the underlying choice architecture for users” by shaping the “decision space” and “manipulating the information flow”. The point of convergence between dark patterns scholarship and the CASA framework is the notion that various cognitive biases common among humans are encountered through the design features of a computational artifact, be it a graphical user interface (GUI) [28] or a social robot [20, 40]. The point of divergence between these two bodies of work is that such engagement is understood in dark patterns scholarship as malicious in the sense that these cognitive biases are exploited to the detriment of users. Indeed, as very recent work by Alberts and Van Kleek suggests,

the CASA framework might benefit from an engagement with dark patterns research on GUIs [2].

In fact, it can be argued that the conceptual framework of dark patterns scholarship is better suited than the CASA framework to answer the question of how humans relate to artificial entities, including conversational user interfaces (CUIs). Mildner et al. for instance have already argued for expanding dark patterns scholarship from its narrow focus on GUIs to also include CUIs [30].

1.2 Dark Patterns and Multi-Modal Interaction

Human-agent interaction often involves CUIs (e.g. Alexa, Google Home, Apple Siri), due to the intuitive nature of this form of communication. But this form of interaction is regularly extended through novel technologies adding embodiment to the agents, which changes how users perceive them [7, 22]. Considering this recent trend, a multi-modal approach becomes necessary in order to be able to identify dark patterns in social agents. This point is not just limited to physical embodiment, but can be also extended to virtually embodied agents. Two observations substantiate this argument.

First, although there is very little work on dark patterns in social agents, the existing literature points to the fact that such dark patterns differ from those found in both GUIs and CUIs, as they rely more – but not exclusively – on embodied or spatial qualities rather than textual or visual interface elements for their operation. This has been shown in work on, for example, a class of such agents: social robots (cf. [21, 22, 38, 40]). In this regard, Shamsudhin and Jotterand observe that the function of certain embodied design elements in social robots, like Sony’s Aibo, is to exploit anthropomorphism to mislead users into accepting the robot’s alleged autonomy whilst simultaneously obfuscating that users endure “affective blackmail” leveraged for corporate data-collection [40]. Similarly, Lacey and Caudwell, writing on home robots that function as social robots by virtue of their “cute aesthetic”, draw attention to the fact that dark patterns present in embodied interfaces function through strategies that differ significantly from those of their counterparts in GUIs and CUIs [22]. This aesthetic is a dark pattern in that it produces “an emotional response likely to generate misplaced trust with the user regarding how their data is collected and processed.” Referencing Lacey and Caudwell [22], Kubota and colleagues notably observe that the possibility of dark patterns in robot design and subsequent infringement on user autonomy might affect vulnerable groups first and foremost, and thus – to put it in terms of CUI 2023 – constitute non-inclusive design [21]. They observe that the personalization of robot behavior through individualized data collection runs the risk of being used to interact deceptively with users, leveraging an affective bond to obtain from users “sensitive information” and manipulate them into buying additional goods and services from the robot’s manufacturer.

The foregoing argument is extended through spatial data collection with mischievous intent, which requires us to take stock of the various modalities of interaction along which the data is collected. A possible starting point for work in this direction has been provided by Greenberg and colleagues, who have looked at dark patterns in proxemic interactions. Proxemics refer to how space is perceived and used in terms of communication through “the conscious or

unconscious setting of distances between various objects, agents, and oneself” [39]. Given that Greenberg et al. suggest that such proxemic interactions are the subject of data-collection and manipulation through dark patterns, their work intimates that research into dark patterns in social agents would do well to consider four other forms of non-verbal communication displayed by socially interactive agents: kinsesics [39], haptics [39], chronemics [39], and vocalics [1].

Following Saunderson and Nejat [39], it can be noted that to varying degrees, all of these dimensions of non-verbal communication are involved in the cognitive framing, recognition of and response to “robot emotions”, behavioral response, and task performance displayed by users in their interactions with social agents. Cognitive framing refers here to the “process observed in human psychology by which people develop a certain perspective or orientation on a topic”. In the sense that cognitive framing and the recognition of and response to agent emotions inform users’ actions by way of providing or influencing existing heuristics for automatic or non-reflexive behavior, there might be a clear connection with research on dark patterns given that heuristics render users “susceptible to cognitive biases – systematic deviations from rational judgment” [9].

1.3 Human Feedback Is Part of The Equation

In contrast to the aforementioned effects, certain design features of social agents might be necessary to create a degree of authenticity and build trust simultaneously. Features such as the adaption of the voice [26], an alignment with the user regarding lexical choices [41, 42] or augmenting an agent with embodiment [3, 7, 44] can have a positive effect on the user’s impression of the agent. Likewise, the previously stated effect of framing can, for example, be used to positively influence a user to demonstrate empathy towards a social agent by using anthropomorphic language [10].

When it comes to examining design features, it is worth taking a look at the interactions between humans and social agents as a practice embedded in a socio-cultural and communal context [27]. In this context, the aforementioned design features can facilitate successful communication and, at the same time, result in counter-productive social actions, such as showing offensive or disrespectful feelings [2] or cause users to develop expectations that the current state of the art cannot meet [19].

The challenge is to strike a balance between creating social agents that are authentic and facilitate successful interactions while also avoiding behavior that could harm the user or create false expectations. Correspondingly, the current workshop aims to address the intersection of dark patterns and design features features to develop strategies that focus on a human-centered design and reflect on design features across different modalities in a way that benefits the users.

2 ORGANIZERS

This workshop is organized by researchers with experience in the diverse fields connected to our aims. Together, the organizers have a history in publishing on a variety of topics including human-robot interaction [20], natural language processing [6, 18], conversational user interfaces [35, 37, 45–47], dark patterns [30, 31], multi-modal

communication [7, 36, 43], and pervasive and persuasive technologies targeting user agency [4, 5, 13]. The combination of these disciplines shall foster collaboration of our trans-disciplinary participants and bridge the gap between these fields of research.

Vino Avanesi University of Bremen, Germany. Vino Avanesi is a Ph.D. student in the Digital Media Lab at the University of Bremen, where he does research on malicious patterns in conversational agents from an ethical and political economic perspective.

Johanna Rockstroh University of Bremen, Germany. Johanna Rockstroh is a Ph.D. Student in the Digital Media Lab at the University of Bremen, where she does research on how to increase a chatbot’s empathy by using sentiment analysis and knowledge graphs.

Nima Zargham University of Bremen, Germany. Nima Zargham is a Ph.D. student in the Digital Media Lab at the University of Bremen. His research focuses on human-centred approaches for designing desirable speech-based systems.

Thomas Mildner University of Bremen, Germany. Thomas Mildner is a Ph.D. student at the Digital Media Lab at the University of Bremen. His research focuses on ethical design and developing technologies to support autonomous decision-making. His work thereby considers social media platforms but also ubiquitous technologies including CUIs.

Maximilian A. Friehs Assistant Professor at the University of Twente, Netherlands. His research focuses on manipulating human performance, cognition and experiences. This also includes investigating interactions between humans or human-virtual agent interactions under different constraints in order to elicit specific experiences or behaviours.

Leon Reicherts University College London, United Kingdom. Leon Reicherts is a Ph.D. student at the UCL Interaction Centre. His research is focused on how conversational interactions can augment users’ thought processes. The interfaces he designs aim to achieve this by prompting users to explore alternative possibilities, perspectives, and approaches for the cognitive task at hand.

Dimosthenis Kontogiorgos Humboldt University of Berlin. Dimosthenis Kontogiorgos is a PostDoctoral researcher at the Humboldt University of Berlin and is interested in how conversational agents’ embodiment and non-verbal behaviours affect the process of establishing, maintaining and repairing common ground.

Nina Wenig University of Bremen, Germany. Nina Wenig is a Post-doctoral researcher in the Digital Media Lab at the University of Bremen and works on applying artificial intelligence in new user interfaces. Her research focus is on automatic and manual image annotation, chatbots and natural language understanding.

Rainer Malaka University of Bremen, Germany. Rainer Malaka is professor for Digital Media at the University of Bremen. He is managing director of the Center for Computing Technologies (Technologiezentrum Informatik und Informationstechnik, TZI) and director of the PhD program Empowering Digital Media, funded by the Klaus Tschira Foundation. His research focus is on multi-modal interaction, language understanding, entertainment computing, and artificial intelligence.

3 PRE-WORKSHOP PLANS

This workshop aims to bring together people from academia and industry to engage in trans-disciplinary discussions on the topics of multi-modal anthropomorphism that may lead to problematic consequences for users. We invite people from various HCI-related communities to participate in this workshop. These include but are not limited to human-agent interaction, human-robot interaction, CUI, and dark pattern and persuasive technologies scholarships. To reach potential candidates, we will announce a *Call for Participation* on multiple threads, including popular social media (e.g., Twitter, Facebook, LinkedIn), as well as topic-related mailing lists. We will further directly invite people who publish articles in related fields in HCI and psychology conferences and journals.

Drawing on past experiences, we anticipate a turnout of 10 to 15 attendees, accompanied by 4 to 6 position papers. Also, we will put online a workshop website containing the call and key dates, providing all required information for potential participants. Later, accepted submissions will be made available on this website before the workshop starts and kept there for archiving purposes.

The call will ask applicants to submit position papers of 3-4 pages in length, presenting personal work or discussing novel ideas related to the potential risks and harms of multi-modal features of social agents, how to address or counter them, and work concerning dark patterns. We encourage submissions describing research in progress, preliminary results to be discussed with the community, methodology proposals, and lessons learned in designing conversational agents for end users. We further encourage the submission of papers that address the critical issue of inclusive designs through conversational interfaces. All submissions will be independently reviewed by at least two organizers of this workshop before accepting applicants for this workshop.

4 OUTLINE OF THE WORKSHOP

The workshop comes at a time when the design of artificial social agents has to meet the users' expectations on different levels. There are various factors involved in designing social agents - many of these factors involve human-centred design whilst others do so only indirectly. With the help of the expertise from various research disciplines in the context of Human-Computer Interaction, the workshop aims to:

- (1) **Identify dark patterns in social agents:** Collect examples of social agents' characteristics and traits that relate to dark patterns in CUIs.
- (2) **Discuss the effects and necessity of present and future design features in social agents:** Evaluate design features in given social contexts and across different modalities for relevant applications, ranging from chatbots to embodied agents.
- (3) **Identify the impact on the design of social agents:** Develop ideas on incorporating the findings on dark patterns and human-like features in social agents into design features for future social agents.

These aims are implemented within this workshop's structure as seen in Table 1.

4.1 Format

The workshop will take place in a hybrid configuration to make it as accessible as possible. In the same regard, we scheduled the workshop to begin in the early afternoon to make it available in as many time-zones as possible for people to join online. We will use Zoom as our conferencing solution together with Miro for participants to engage during sessions. Depending on the turnout, we will try to have different groups for on-site and online participants to counter uneven effects that may lead to unfairness during discussions.

4.2 Session 1: Social actor or dark pattern

During the first session, participants will be split into groups of 4-6 people with different backgrounds. In these groups, they will look both at the present and future by reflecting on and envisioning emerging threats of (novel) technologies that may require attention to protect users from harmful interactions. To support their discourse, this session will incorporate a Miro board for grouping topics in the form of a card-sorting exercise. We will provide visual examples from the dark pattern literature to offer inspiration. After 45 minutes, each group will present key points of their agreement within 2-3 minute retrospectives.

4.3 Session 2: Fostering users' agency

The second session aims to bring together collected arguments and provoked thoughts. Again, participants will be divided into trans-disciplinary groups of 4-6 people (however, new groups will be formed), particularly mixing people from academia and industry. Within 30 minutes, focusing on users' agency, participants will be tasked with discussing design features social agents should include to support this aim. We will motivate participants to think about possibilities to align industry and user incentives while considering the necessity of guidelines and/or regulations in case features are otherwise exploited to trick users in the manner of dark patterns.

5 WEBSITE

This workshop will be supported by a website that will contain a call for participation including related work of our topic. Also, we will use the website to share accepted position papers of our participants¹.

6 POST-WORKSHOP PLANS

The main aim of this workshop is to bring people together from different fields and with different perspectives on the potential risks residing behind the design of social agents in a multi-modal context. Opening this workshop's results to the public, all accepted contributions of the participants will be collected and archived via our workshop's website. The participants will also be encouraged to submit their articles to *arXiv*. The ideas and discussion points throughout the session will be documented on an open online platform during and after the symposium. The workshop participants will be invited to collaborate on future projects arising from the workshop discussions and talks. Drawing from the potential need

¹If this workshop paper gets accepted, we will include a link to a subdomain of the Digital Media Bremen lab's website of the University of Bremen with the address <https://www.digital-media-lab.uni-bremen.de/multimodaldarkpatterns/>.

Time (in UTC+1)	Activity	Description
13:00	Welcome	The organizers will introduce themselves and present the agenda of the workshop.
13:15	Lightning introductions	Participants introduce themselves with their position papers in one presentation slide. If multiple participants authored the same position paper, one person should take lead but introduce their co-authors.
13:45	Session 1: Social actor or dark pattern	Based on their position papers, participants will be divided into small groups of 4-6 people. Within their groups, they will discuss emerging designs and technologies and their potential to be exploited as dark patterns.
14:15	<i>Coffee break</i>	
14:30	Sharing the highlights	One rapporteur of each group presents key highlights of their group's discussion including a Q/A per group.
15:00	Keynote with Q/A, dark patterns in a surveillance capitalistic world	Colin Gray, Associate Professor in UX pedagogy at Purdue University, has contributed essential research to dark pattern scholarship and will talk about the emerging risks of dark patterns when design regulations and guidelines fail to support users' agency.
15:45	<i>Coffee break</i>	
16:00	Session 2: Fostering users' agency	Based on the highlighted points of the previous session and keynote, groups of 4-6 participants will discuss how social agents can be designed and developed to set realistic expectations promoting users' agency.
16:30	Sharing the highlights	One rapporteur of each group presents key highlights of their group's discussion including a Q/A per group.
16:45	Preparing the continuation of the discourse	The organizers will conclude the workshop and promote online tools to carry the discourse onward.
17:00	End of the workshop	

Table 1: Workshop activity schedule

for guidelines and regulations, we will invite participants to collaborate and distill the results of this workshop into a refined piece, for example, in the form of an open letter to stakeholders or a refined set of guidelines that could be published in a future venue. Lastly, we will encourage all participants to submit extended versions of their submissions for a special journal issue which we will aim to realize.

7 CALL FOR PARTICIPATION

How much should communicative and social agents disclose that they are not human while their features mislead us to think otherwise? Recent technological advances, including generative AI that produces human-like language, speech syntheses, and realistic visuals, widen the scope of possibilities to create human-computer interactions that may be impossible to distinguish from human-human interaction in specific environments. Meanwhile, scholars of the HCI community have increasingly focused on malicious design strategies, producing a growing list of so-called dark patterns. However, this discourse has primarily been set on screen-based interactions. Considering the availability of technologies and the prevalence of social agents, which, in the wrong hands, can be used in similarly malicious manners, requires special awareness among stakeholders regarding the protection of users.

This half-day hybrid workshop will kick off this discourse by inviting participants from academia and industry to reflect on these technologies. We will carry over the ongoing discourse of dark patterns to social agents and conversational user interfaces and

consider the need for design guidelines and regulations to ensure users' safety.

Applicants are invited to submit position papers of 3-5 pages in length (ACM single-column format including references) to present their related research findings, novel ideas, and work-in-progress. We invite submissions covering, but not limited to, the following topics:

- Persuasive technologies for CUIs
- Embodied conversational agents
- Ethical caveats in CUIs and HRI
- Mechanisms of preventing deceptive design in CUIs and HRI
- Effects of specific dark patterns in CUIs/multimodal interfaces
- Tasks & activities in CUIs/multimodal interfaces at risk of dark patterns
- Dark patterns specific to CUIs/multimodal interfaces but not to GUIs

Submissions will be reviewed by at least two workshop organizers and admitted based on their quality, relevance to the topic, and diversity. For each submission, at least one author must attend the workshop. Key details can be found below:

- Submission deadline: 2nd of June, 2023
- Acceptance Notification: 14th of June, 2023
- Send position papers to: multimodaldarkpatterns@gmail.com
- Website: www.digital-media-lab.uni-bremen.de/multimodaldarkpatterns
- Date and time: 19th of July, 13:00 p.m. - 17:00 p.m. (UTC+2)

ACKNOWLEDGMENTS

This work was partially funded by the Klaus Tschira Foundation, by the FET-Open Project 951846 “MUHAI – Meaning and Understanding for Human-centric AI” funded by the EU program Horizon 2020, by the Leverhulme Trust (award DS-2017-026), the German Research Foundation (DFG) under Germany’s Excellence Strategy – EXC 2002/1 ‘Science of Intelligence’ (project number 390523135), as well as the DFG as part of Collaborative Research Center (Sonderforschungsbereich) 1320 “EASE – Everyday Activity Science and Engineering”, University of Bremen (<http://www.ease-crc.org/>).

REFERENCES

- [1] 2016. *Communication in the Real World: An Introduction to Communication Studies* (01 ed.). University of Minnesota Libraries Publishing. <https://doi.org/10.24926/8668.0401>
- [2] Lize Alberts and Max Van Kleek. 2023. Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially. *ArXiv abs/2302.04720* (2023).
- [3] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 2571–2582. <https://doi.org/10.1145/3025453.3026033>
- [4] Vito Avanesi. 2020. Second-screen subsumption: Aesthetics of control and the second-screen facilitation of hyper-attentive watching-labour. *Convergence* 26, 2 (2020), 369–385. <https://doi.org/10.1177/1354856518781650> arXiv:<https://doi.org/10.1177/1354856518781650>
- [5] Vito Avanesi and Jan Teurlings. 2022. “I’m Not a Robot,” or am I?: Micro-Labor and the Immanent Subsumption of the Social in the Human Computation of ReCAPTCHAs. *International Journal of Communication* 16, 0 (2022). <https://ijoc.org/index.php/ijoc/article/view/17251>
- [6] Daniel Beßler, Robert Porzel, Mihai Pomarlan, Abhijit Vyas, Sebastian Höffner, Michael Beetz, Rainer Malaka, and John A. Bateman. 2020. Foundations of the Socio-physical Model of Activities (SOMA) for Autonomous Robotic Agents. *CoRR abs/2011.11972* (2020). arXiv:2011.11972 <https://arxiv.org/abs/2011.11972>
- [7] Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays. In *CUI 2021-3rd Conference on Conversational User Interfaces*. ACM, New York, NY, USA, 1–11.
- [8] Harry Brignull, Marc Miquel, Jeremy Rosenberg, and James Offer. 2010. Dark Patterns-User Interfaces Designed to Trick People. <http://darkpatterns.org/visited-on-2021-08-25>.
- [9] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. <https://doi.org/10.1145/3290605.3300733>
- [10] Kate Darling. 2017. “Who’s Johnny?” Anthropomorphic Framing in Human-Robot Interaction. In *ROBOT ETHICS 2.0: From Autonomous Cars to Artificial Intelligence*, Patrick Lin, Keith Abney, and Ryan Jenkins (Eds.). Oxford University Press, New York, 89–110. <https://doi.org/10.1093/oso/9780190652951.001.0001>
- [11] 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>
- [12] 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>
- [13] Maximilian A Friehs, Martin Dechant, Sarah Schäfer, and Regan L Mandryk. 2022. More than skin deep: about the influence of self-relevant avatars on inhibitory control. *Cognitive Research: Principles and Implications* 7, 1 (2022), 31.
- [14] Andrew Gambino, Jesse Fox, and Rabindra Ratan. 2020. Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication* 1 (Feb. 2020), 71–86. <https://doi.org/10.30658/hmc.1.5>
- [15] 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>
- [16] David J. Gunkel. 2020. *An introduction to communication and Artificial Intelligence*. Polity Press, Medford, MA.
- [17] Andreas Hepp. 2020. Artificial companions, social bots and work bots: communicative robots as research objects of media and communication studies. *Media, Culture & Society* 42, 7-8 (Oct. 2020), 1410–1426. <https://doi.org/10.1177/0163443720916412>
- [18] Sebastian Höffner, Robert Porzel, Maria M. Hedblom, Mihai Pomarlan, Vanja Sophie Cangalovic, Johannes Pfau, John A. Bateman, and Rainer Malaka. 2022. Deep understanding of everyday activity commands for household robots. *Semantic Web* 13, 5 (Aug. 2022), 895–909. <https://doi.org/10.3233/SW-222973>
- [19] Bing Cai Kok and Harold Soh. 2020. Trust in Robots: Challenges and Opportunities. *Current Robotics Reports* 1 (2020), 297–309. <https://doi.org/10.1007/s43154-020-00029-y>
- [20] 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*. 133–140.
- [21] Alyssa Kubota, Maryam Pourebadi, Sharon Banh, Soyon Kim, and Laurel D Riek. 2021. Somebody That I Used to Know: The Risks of Personalizing Robots for Dementia Care. *Proceedings of We Robot, 2021* (Jan. 2021), 24. <https://ssrn.com/abstract=3910089>
- [22] 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>
- [23] Eun-Ju Lee. 2010. The more humanlike, the better? How speech type and users’ cognitive style affect social responses to computers. *Computers in Human Behavior* 26, 4 (2010), 665–672. <https://doi.org/10.1016/j.chb.2010.01.003>
- [24] Minha Lee, Lily Frank, Yvonne De Kort, and Wijnand IJsselstein. 2022. Where is Vincent? Expanding Our Emotional Selves with AI. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. <https://doi.org/10.1145/3543829.3543835>
- [25] Minha Lee, Lily Frank, Yvonne De Kort, and Wijnand IJsselstein. 2022. Where is Vincent? Expanding Our Emotional Selves with AI. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. <https://doi.org/10.1145/3543829.3543835>
- [26] Gesa Alena Linnemann and Regina Jucks. 2018. ‘Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?’—Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction. *Interacting with Computers* 30, 3 (03 2018), 173–186. <https://doi.org/10.1093/iwc/iwy005> arXiv:<https://academic.oup.com/iwc/article-pdf/30/3/173/24805335/iwy005.pdf>
- [27] Alasdair MacIntyre. 2007. *After virtue: a study in moral theory* (3 ed.). University of Notre Dame Press, Notre Dame.
- [28] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–32. <https://doi.org/10.1145/3359183> arXiv:1907.07032 [cs].
- [29] Arunesh Mathur, Jonathan Mayer, and Mihir Kshirsagar. 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*. 1–18. <https://doi.org/10.1145/3411764.3445610> arXiv:2101.04843 [cs].
- [30] Thomas Mildner, Philip Doyle, Gian-Luca Savino, and Rainer Malaka. 2022. Rules Of Engagement: Levelling Up To Combat Unethical CUI Design. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 26, 5 pages. <https://doi.org/10.1145/3543829.3544528>
- [31] Thomas Mildner, Gian-Luca Savino, Philip R Doyle, Benjamin R Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. *arXiv preprint arXiv:2303.00476* (2023).
- [32] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (Jan. 2000), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- [33] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Conference companion on Human factors in computing systems - CHI '94*. ACM Press, Boston, Massachusetts, United States, 204. <https://doi.org/10.1145/259963.260288>
- [34] 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* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3174214>
- [35] Leon Reicherts, Gun Woo Park, and Yvonne Rogers. 2022. Extending Chatbots to probe users: Enhancing complex decision-making through probing conversations. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–10.

- [36] 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 (2022), 1–41.
- [37] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions On Proactive Smart Speakers. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. <https://doi.org/10.1145/3469595.3469629>
- [38] Laurel D. Riek. 2017. Healthcare robotics. *Commun. ACM* 60, 11 (Oct. 2017), 68–78. <https://doi.org/10.1145/3127874>
- [39] Shane Saunderson and Goldie Nejat. 2019. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. *International Journal of Social Robotics* 11, 4 (Aug. 2019), 575–608. <https://doi.org/10.1007/s12369-019-00523-0>
- [40] Naveen Shamsudhin and Fabrice Jotterand. 2021. Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin? In *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*, Fabrice Jotterand and Marcello Ienca (Eds.). Springer International Publishing, Cham, 89–110. https://doi.org/10.1007/978-3-030-74188-4_7 Series Title: Advances in Neuroethics.
- [41] Laura Spillner and Nina Wenig. 2021. *Talk to Me on My Level – Linguistic Alignment for Chatbots*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3447526.3472050>
- [42] Ilaria Torre, Jeremy Goslin, and Laurence White. 2020. If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior* 105 (2020), 106215. <https://doi.org/10.1016/j.chb.2019.106215>
- [43] Tony Veale, Philipp Wicke, and Thomas Mildner. 2018. Duets Ex Machina: On The Performative Aspects of "Double Acts" in Computational Creativity.
- [44] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, Article 281, 12 pages. <https://doi.org/10.1145/3290605.3300511>
- [45] Nima Zargham, Dmitry Alexandrovsky, Jan Erich, Nina Wenig, and Rainer Malaka. 2022. "I Want It That Way": Exploring Users' Customization and Personalization Preferences for Home Assistants. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 270, 8 pages. <https://doi.org/10.1145/3491101.3519843>
- [46] Nima Zargham, Michael Bonfert, Robert Porzel, Tanja Doring, and Rainer Malaka. 2021. Multi-Agent Voice Assistants: An Investigation Of User Experience. In *20th International Conference on Mobile and Ubiquitous Multimedia* (Leuven, Belgium) (MUM 2021). Association for Computing Machinery, New York, NY, USA, 98–107. <https://doi.org/10.1145/3490632.3490662>
- [47] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. <https://doi.org/10.1145/3543829.3543834>