
Gender Choices of Conversational Agent: How Today's Practice Can Shape Tomorrow's Values

Ji-Youn, Jung

Delft University of Technology
Delft, Netherlands
j.y.jung@student.tudelft.nl

Dave Murray-Rust

Delft University of Technology
Delft, Netherlands
d.s.murray-rust@tudelft.nl

Ujwal Gadiraju

Delft University of Technology
Delft, Netherlands
u.k.gadiraju@tudelft.nl

Alessandro Bozzon

Delft University of Technology
Delft, Netherlands
a.bozzon@tudelft.nl

Abstract

When creating conversational agents, designers have to make decisions about the way the agents present themselves. In this position paper, we identify and synthesize ethical dilemmas that conversational interface designers and researchers face around gender of conversational agents. First, we identify three layers that cause tension in designing conversational agents' gender: (i) interactional qualities; (ii) goal-orientation; and (iii) societal issues. We then argue that conversational agent designers and researchers can navigate this problem space by comparing two ethical frameworks: a utilitarian perspective and a dialogical ethics perspective. Finally, we argue that dialogical ethics can be a balanced, ethical lens that can help conversational agent designers and researchers make design decisions about the gender of an agent.

Author Keywords

Gender-inclusive, Dialogical ethics, Design ethics

CCS Concepts

•**Social and professional topics** → Gender; Please use the 2012 Classifiers and see this link to embed them in the text: https://dl.acm.org/ccs/ccs_flat.cfm

When designing the gender of a conversational agent(CA), several ethical dilemmas rise to the surface. Research find-

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).

Copyright held by the owner/author(s).

CHI'22, to be filled

ACM 978-1-4503-6819-3/20/04.

<https://doi.org/10.1145/3334480.XXXXXXX>

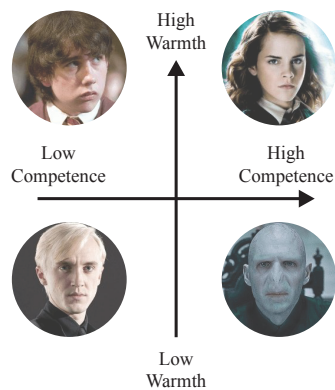


Figure 1: Stereotype Content Model [6], exemplified with famous novel/movie *Harry Potter*'s characters - (from top-left, clockwise) Neville Longbottom, Hermione Granger, Voldemort, and Draco Malfoy.

ings point us to evidence around how people perceive female agents as friendly and warm, male agents as competent and professional, and gender-neutral agents as creepy [14]. In addition, research has shown that male agents led to higher user trust in the context of airport security [16]. However, several researchers have raised concerns on how designing an agent-based on stereotyped gender roles could reinforce the existing power structures [9, 19, 14]. As technological artifacts and systems acquire embedded values [7] (p.49), it is an ethical design challenge that the conversational user interface (CUI) community holds.

Female gendered agents have a higher chance to experience verbal abuse and sexualization than male-gendered agents or an agent depicted as a nonhuman robot agent [17, 3]. Few studies in education technology have investigated the topic of abusive conversation and reinforcement of cultural stereotypes through gendered CAs [8, 21]. As the natural language processing technology evolves as they learn from the data that agent-user interaction generates, it is of great importance that the system can prevent agents from learning an abusive language. A good example of this danger is a Microsoft Tay, where the agent learned and generated racist, sexist, and anti-Semitic language to the public within 24 hours of its release.

Although studies around gender have a long history, the concept of gender has radically transformed in recent years, leading to a call for reflection. Several recent studies investigated this topic through a contemporary lens. For example, the CHI community has changed their survey system to ask participants' gender in CHI 2016 [18]. Similarly, recent studies have paid attention to practices in image description of web contents [2]. Stumpf et al. [19] presented a conceptual review paper on Gender-Inclusive HCI Research where

they asserted the need to produce gender-inclusive design, published in 2020. Gender is a complex subject, with a range of political, ethical and social issues attached, especially where devices interact directly with humans using natural language and human-like presentation. The interdisciplinary nature of the CUI community and the current engagement with gender in the design of CUIs makes it a compelling site for developing a robust and caring approach to designing gender.

We present two research gaps that are controversial but are ethical in nature. First, while research has been carried out on the way that users apply gender stereotypes to CAs based on a range of markers [14, 5, 8], little work has been done to calibrate the undesirable outcomes when the anthropomorphized agent does not align with expected stereotypes such as 'female as warm' and 'male as competent'. While this surfaces the implicit assumptions and values embedded in the technology, it only recognizes the danger of reinforcing the stereotype, and does not suggest solutions to overcome the challenge of gender stereotypes.

Second, the consequences of having a non-binary gendered agent have not been investigated deeply. While the concept of fluid gender as a person's identity has recently become an important part of socio-technical debate, most of the HCI community findings that show the effect of a gendered agent only explore the binary male/female concept of gender [19](p.3). The current state of the real-world debate around gender is moving on from this framing, so it is critical that the CA community studies the effect of non-binary gendered agents.

In this position paper, we present our arguments on why these topics matter in the CUI community through an ethical lens, based on the previous research findings. We argue that the CUI community should consider adopting dia-

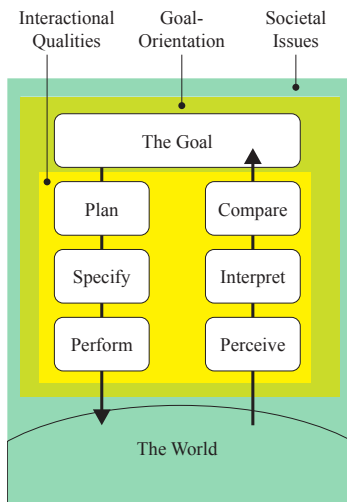


Figure 2: Layers of CA design consideration, augmented on Norman's action cycle [15]

logical ethics in their research practice, which will help develop more robust ethical guidelines for designing gendered agents.

Agents with Gender Stereotypes

Classic works in social perception have shown that people tend to perceive others in category-based information processing automatically. Furthermore, the most salient social categories that help people assign stereotypes are reported to be age, race, and sex [6]. This sensitive categorization develops from an early age, where they also develop their identity through a sense of membership in one gender or others [4]. While such automatic process of categorization make information processing cost-effective, it comes with a cost of stereotyping and discrimination.

As users anthropomorphize CAs [10], previous works have reported how people apply gender stereotypes to conversational agents that shows gender markers [5]. Stereotype content model(SCM), a renowned model in social psychology, explains group stereotypes form along two dimensions of warmth and competency (Figure 1). Considerable amount of findings of user perception towards CAs has been based on this model, where they used warmth and competency as a measure to mediate stereotypical categorization, user expectation, and evaluations [14].

Layers of CUI Design Consideration

Here, several layers can be revealed in considerations for designing gendered conversational agents (Figure 2). First, *interactional qualities* where the point of interaction sparks factors such as enjoyment or engagement. For example, female voices tend to be rated as more likeable [14]. Leading consumer CAs such as Amazon Alexa or Google Assistant is often designed with default feminine voice, and also some widespread consumer reports support user

preference towards female-voiced CAs. Second, the *goal-orientation* level where the characteristics of CAs help the user achieve the final goal of their interaction. For example, users are more likely to choose a male agent to help them solve stereotypically male tasks, such as tasks that require mathematical ability [5]. Finally, a *societal* layer where the danger of reinforcing a gender stereotype exist by deploying gendered agents that conforms to user's — often unintentional— stereotypical expectation [5, 19, 14].

Based on this inspection, CA designers encounter challenging tension between each layer, as illustrated in Figure 2. Existing and widely used axes of warmth and competence help designers provide CA design considerations in 'interactional qualities' and 'goal-oriented' layers. However, few works have explored the consequences and implications of the 'societal' layer, when gendered agents might foster reinforcing gender stereotypes.

Gender neutral agents are unlikeable?

Recently, radical changes in scientific understanding of the endocrine system insisted on biological plasticity. In addition, a traditional understanding of dichotomous gender has changed where a concept of non-binary gender has risen in society. A recent study conducted by Lopatovska et al. [13] reported that nearly half of the participants indicated a preference for a non-binary voice before the experiment. In contrast, gender-ambiguous voices still received the lowest acceptance during the interaction compared to the binary voices. This aligns with previous findings [14] where authors mentioned gender-ambiguous voice creates categorical tension, which leads to unpleasant feelings ("strange, dislikeable, dishonest and unintelligent") towards human users as a possible explanation.

The industry has been developing gender-neutral voices in

recent years. One of the examples is Project Q, a genderless voice assistant to reflect the diversity and reduce the gender bias that got introduced in 2019.¹ Another example is Sam, which got introduced in 2020 and developed by Accenture Labs in collaboration with CereProc.² However, it is early to judge the implication or the consequences of adopting this technology in the real-world.

It is important to notice the tension between people's conscious preference toward non-binary gender agents and unconscious disapproval. Such phenomenon can be explained as a dissonance between people's awareness towards a socially ethical behavior (*societal layer*) and an interactional quality. It will be an assignment for agent designers and researchers to clarify the trade-off between these colliding needs.

Ethics in support of gender choices for conversational agents

Like any ethical dilemmas, different ethical standards can provide moral frameworks to approach the gender choices of CAs. However, we pick two contrasting positions, utilitarian ethics and dialogical ethics.

Utilitarian ethics

Utilitarian ethics will say that the choice which produces the greatest good for the greatest number is the most ethical one. From a Utilitarian point of view, designing an agent that conforms to the gender stereotype seems like a preferable solution to achieve the goal of making agents capable of interacting with users in an efficient and pleasant manner. For example, in the contexts where the credibility of the agent is critical, such as an E-Commerce software where

agents are designed to explain or sell a service or a product, designing an agent with a gender that conforms to the "gender" of its product has shown to improve perceived credibility of the agents (e.g., cosmetics for female agents, hammers for male agents). In addition, perhaps more critically, if there were to be a situation where agents have to direct people for fire evacuation, deploying a male agent that was shown to be more authoritative [20] makes the most sense.

However, if agent designers decide to apply gender stereotypes blindly, they risk strengthening the perception that can result in unfair understanding towards the CA. The game design field has been experiencing a similar dilemma, where gendered game character design can potentially impact youth's understanding of desired gender roles. Moreover, it can shape an individual's body image, self-esteem, self-perception, and expectations from other genders. Some research findings reported over-representation of gender stereotypes in video games [12].

CUI field also holds a similar dilemma to the game design domain, where an overwhelming number of agents today shows feminine markers (e.g., Microsoft's Xiaoice, Amazon's Alexa, Pandorobot's Kuki AI (previously Mitsuku)). As agent abuse and other related challenges call for participation in academic research, we argue that CUI researchers should critically reflect on our research practice that is not solely based on utilitarian values.

Dialogical Ethics

Dialogical ethics is a stance that locates ethics in the communicative ground between people rather than in philosophical thought. Unlike some traditional ethics, dialogical ethics does not censor an opinion or impose any advance restrictions. Instead, dialogue is seen as an epistemological vehicle for learning, where we are doomed to be interdependent

¹<https://www.genderlessvoice.com/>

²<https://youtu.be/mL1n5AEFLI4>

in our aim of co-creating moral answers.

Previously, feminist ethicist Koehn [11] suggested dialogical ethics in response to some problems she identified in the traditional feminist ethics, such as care ethics. Adam [1] also recognizes dialogical ethics to provide a more balanced approach to be looked into in her book, where she links feminist ethics with computer ethics.

Instead of giving a single answer, dialogical ethics introduce an ethical way of making a conversation to avoid any wrongdoing in the conversation itself. Surely a conversation could drive one further away from the other if one is to have a self-righteous or a rigid position. Therefore, it is not the verbal exchange of the conversation that matters, but *how* the conversation is made. In conclusion, dialogical ethics focus on fostering mutually acceptable consensus and, therefore, developing a community [11].

According to dialogical ethics, the ethical way of dealing with the dilemma when designing conversational agents would be to openly approach the people who assert the need to develop CAs that go against gender stereotypes. Here, the focus is to have "*the right*" conversation and to prevent any attempt to abstract away from the relevance of the related party's point of view in arriving at a description of the problem at hand [11]. Therefore, an ethical way of solving our dilemma, in the lens of dialogical ethics, could start with developing a guideline to have a respectable conversation around gender and how every party can get an open attitude to acknowledge an insight from each one's point of view.

Final Remarks

As society's gender identity and discourse change, it is natural that the CUI community also faces ethical dilemmas as people anthropomorphize CAs. While many researchers

expressed concerns about adopting gender stereotypes to CA design, there have been no specific guidelines on dealing with such dilemmas with an ethical lens. This position paper shows how ethics can guide the CA designers and researchers to navigate this space. "Just as people bring gender expectations *to* technology, they can draw gender expectations *from* technology [14]". We believe that the CUI community is one of the important contributors to investigate what could best benefit society.

REFERENCES

- [1] Alison Adam. 2005. *Gender, ethics and information technology*. Springer.
- [2] Cynthia L Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis) Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [3] Sheryl Brahn and Antonella De Angeli. 2012. Gender affordances of conversational agents. *Interacting with Computers* 24, 3 (2012), 139–153.
- [4] Alice H Eagly, Anne E Beall, and Robert J Sternberg. 2005. *The psychology of gender*. Guilford Press.
- [5] Friederike Eyssel and Frank Hegel. 2012. (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230.
- [6] Susan T Fiske and Steven L Neuberg. 1990. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology*. Vol. 23. Elsevier, 1–74.

- [7] Luciano Floridi. 2010. *The Cambridge handbook of information and computer ethics*. Cambridge University Press.
- [8] Magnus Haake and Agneta Gulz. 2008. Visual stereotypes and virtual pedagogical agents. *Journal of Educational Technology & Society* 11, 4 (2008), 1–15.
- [9] Charles Hannon. 2016. Gender and status in voice user interfaces. *Interactions* 23, 3 (2016), 34–37.
- [10] Ji-Youn Jung, Sihang Qiu, Alessandro Bozzon, and Ujwal Gadiraju. 2022. Great Chain of Agents: The Role of Metaphorical Representation of Agents in Conversational Crowd-sourcing. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA*. ACM New York, NY, USA, 32 pages.
- [11] Daryl Koehn. 2012. *Rethinking feminist ethics: Care, trust and empathy*. Routledge.
- [12] Xeniya Kondrat and others. 2015. Gender and video games: How is female gender generally represented in various genres of video games? *Journal of comparative research in anthropology and sociology* 6, 01 (2015), 171–193.
- [13] Irene Lopatovska, Diedre Brown, and Elena Korshakova. 2022. Contextual Perceptions of Feminine-, Masculine- and Gender-Ambiguous-Sounding Conversational Agents. In *Paper to be published in Proceedings of the 2022 iConference*.
- [14] Clifford Ivar Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge.
- [15] Donald A Norman. 1986. Cognitive engineering. *User centered system design* 31 (1986), 61.
- [16] Jay F Nunamaker, Douglas C Derrick, Aaron C Elkins, Judee K Burgoon, and Mark W Patton. 2011. Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems* 28, 1 (2011), 17–48.
- [17] Annika Silvervarg, Kristin Raukola, Magnus Haake, and Agneta Gulz. 2012. The effect of visual gender on abuse in conversation with ECAs. In *International conference on intelligent virtual agents*. Springer, 153–160.
- [18] Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (2019), 62–65.
- [19] Simone Stumpf, Anicia Peters, Shaowen Bardzell, Margaret Burnett, Daniela Busse, Jessica Cauchard, and Elizabeth Churchill. 2020. Gender-inclusive HCI research and design: A conceptual review. *Foundations and Trends in Human-Computer Interaction* 13, 1 (2020), 1–69.
- [20] Silke ter Stal, Monique Tabak, Harm op den Akker, Tessa Beinema, and Hermie Hermens. 2020. Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *International Journal of Human-Computer Interaction* 36, 9 (2020), 881–892.
- [21] George Veletsianos, Cassandra Scharber, and Aaron Doering. 2008. When sex, drugs, and violence enter the classroom: Conversations between adolescents and a female pedagogical agent. *Interacting with computers* 20, 3 (2008), 292–301.