

Conversational Assistants and Gender Stereotypes: Public Perceptions and Desiderata for Voice Personas

Amanda Cercas Curry

School of Mathematical
and Computer Sciences
Heriot-Watt University
Edinburgh, UK

ac293@hw.ac.uk

Judy Robertson

Moray House School
of Education and Sport
University of Edinburgh
Edinburgh, UK

judy.robertson@ed.ac.uk

Verena Rieser

School of Mathematical
and Computer Sciences
Heriot-Watt University
Edinburgh, UK

v.t.rieser@hw.ac.uk

Abstract

Conversational voice assistants are rapidly developing from purely transactional systems to social companions with “personality”. UNESCO recently stated that the female and submissive personality of current digital assistants gives rise for concern as it reinforces gender stereotypes. In this work, we present results from a participatory design workshop, wherein we invite people to submit their preferences for what their ideal persona might look like, both in drawings as well as in a multiple choice questionnaire. We find no clear consensus which suggests that one possible solution is to let people configure/personalise their assistants. We then outline a multi-disciplinary project of how we plan to address the complex question of gender and stereotyping in digital assistants.

1 Introduction and Bias Statement

Biased technology disadvantages certain groups of society, e.g. based on their race or gender. Recently, biased machine learning has received increased attention. For example, in the area of Natural Language Processing (NLP), it has been shown that word embeddings (Bolukbasi et al., 2016), co-reference resolution (Zhao et al., 2018) and machine translation systems (Hovy et al., 2020) are likely to reflect and even amplify social biases in the data.

Here we address a different type of bias which is not learnt from data, but encoded during the design process. We illustrate this problem on the example of Conversational Voice Assistants (CVAs), such as Amazon’s Alexa, Apple’s Siri, Microsoft’s Cortana, or Google’s Assistant, which are predominantly modelled as young, submissive women. According to UNESCO (West et al., 2019), this bears the risk of reinforcing gender stereotypes. In particular, these design choices can create representational harm by reinforcing negative stereotypes society holds about women. The report argues that this becomes even more prevalent in the face of abuse, where most assistants do not answer ‘appropriately’ (Curry and Rieser, 2018; Curry and Rieser, 2019), which might impact human-human interactions.

In order to tackle the problem of reproducing structural inequality and oppression of marginalised groups when designing new systems, the Design Justice Network¹ proposes to centre the voices of those who are directly impacted by the outcomes of the design process (Costanza-Chock, 2018). Similarly, the European Commission’s Ethics Guidelines for Trustworthy AI (AI HLEG, 2019) recommend stakeholder participation during the development of new technologies, as well as paying special attention to the system’s societal impact.

In this work we aim to unpack and verify some of the statements in the UNESCO report in a multi-disciplinary project including methodologies from Human Computer Interaction (HCI), Social Psychology, and Natural Language Generation. As a first step, we conduct a participatory design workshop to gather public views on this subject to further inform the direction of this research.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

¹<https://designjustice.org/>

2 Problem Definition

The persona of a CVA can be viewed as a composite of elements of identity (such as demographics and background facts), language behaviour, and interaction style. Some of these aspects can be learned from data, including personality-based linguistic style generation (e.g. Oraby et al. (2018)), or generating responses which are factually consistent with a persona profile (e.g. Zhang et al. (2018)). However, demographics and background facts are usually deliberate design choices, and research such as (Nass and Brave, 2007) shows that the gender choices of CVAs are conforming to traditional gender roles and social expectations: that is, the majority of CVAs have female personas.

In interviews with news outlets, companies and developers defend design choices by citing market research showing that female voices are perceived to be more cooperative and helpful, and male voices are considered more trustworthy (Schwär and Moynihan, 2020; Stern, 2017). As such, the role of a personal assistant is often assigned to women, whereas in applications where the CVA needs to be authoritative, companies tend to choose male voices.

However, in 2019 UNESCO published an in-depth analysis of the gendering of AI, especially focused on conversational assistants (West et al., 2019). The report details how the personas of current conversational assistants reinforce and spread existing biases about women as being subservient, modelling acceptance and tolerating sexual harassment and verbal abuse, and ‘make women the “face” of glitches and errors’. The report attributes this bias to lack of diversity in the tech sector and concludes that the personas of conversational assistants should not be female by default, and that digital assistants should rather be designed to combat gender-based biases as well as discouraging insults and abusive language.

While groups like Feminist Internet² and Women Reclaiming AI³ have addressed the feminisation of CVAs through projects such as F’xa⁴, we involve the general public in designing alternative personas for conversational assistants as a first step in exploring the needs of diverse users.

3 Experimental Setup

Methodology: We explore an alternative methodology to the usual way we get ‘users’ involved in NLP research, i.e. crowd-sourcing or scraping online data (de Vries et al., 2020) borrowed from HCI. *Participatory design* actively involves all stakeholders in the design process to ensure that the end-product meets their needs (Schuler and Namioka, 1993). It aims to assign an active and informed role to everyone affected by the end result.

To implement this idea, we organised a public workshop with the help and endorsement of the Royal Society of Edinburgh⁵ which allowed us to reach a wide population, including potential end-users and people affected by stereotyping. Inspired by a previous workshop on Voice Assistants and Feminism (Webb, 2019), our workshop aimed to inform and stimulate critical reflection in order to seek an active discourse with the public. As such, we organised two events over two consecutive days: The first event was a short introduction where people were invited to learn more about the underlying technology. On the second day, we introduced the issue of gendered technology, and got two experts presenting their views, including one of the authors of the UNESCO report and a UX Voice designer from the BBC. The overall question the workshop explored was: *What would your ideal conversational voice assistant be?*

Participants: We had a total of 128 participants registered to the online event, with 72 participants attending. According to a pre-questionnaire, the majority (72%) of participants identified as female, 28% male. Most of them have had experience with using one or more voice assistants although they do not use them regularly.

Questionnaire: During and after the workshop, we asked participants to complete a form detailing the following characteristics of their ideal personal assistant: anthropomorphism, gender, age group, regional accent, as well as an option to describe other personality traits.

²<https://feministinternet.com/>

³<https://womenreclaimingai.com/>

⁴<http://about.f-xa.co/>

⁵<https://www.rse.org.uk/>

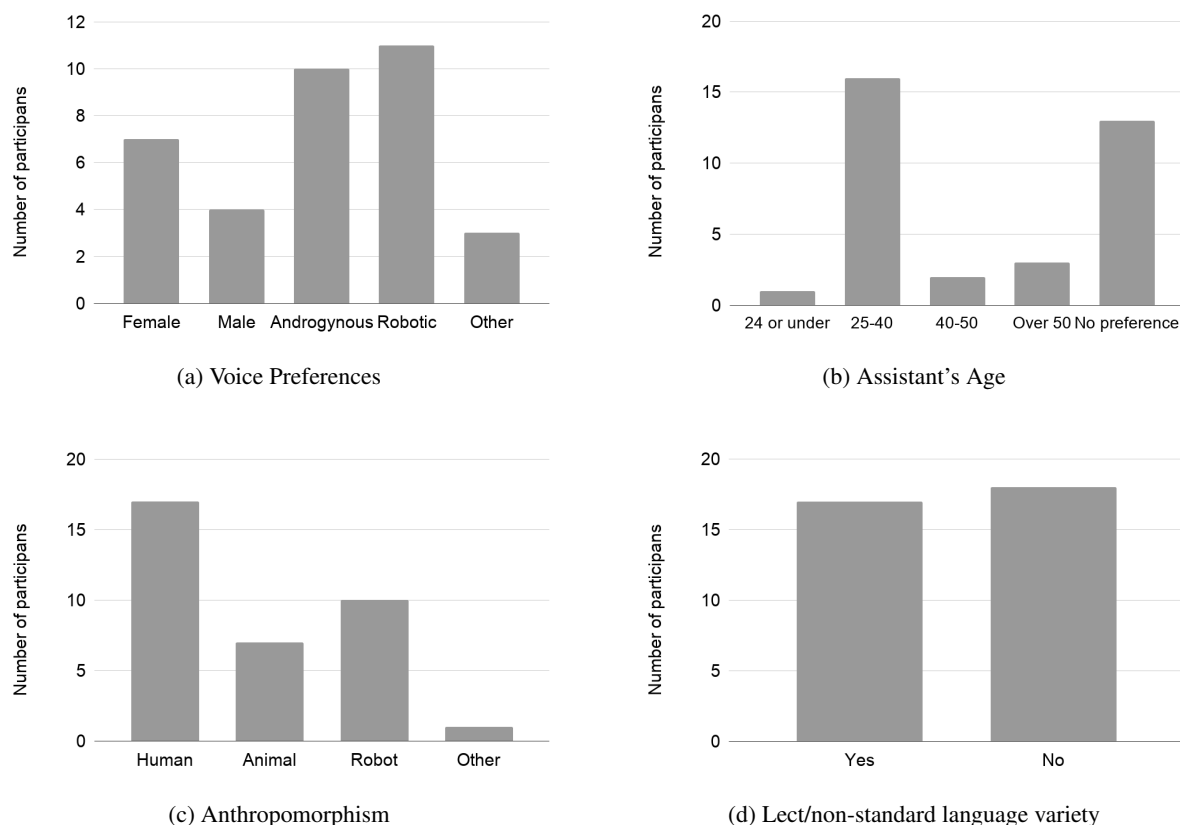


Figure 1: Participants' (n=34) preferences of demographic aspects of persona

In the following study we compare these submitted desiderata with characteristics of 14 existing chatbots:

- Six voice-based commercial systems: Amazon's Alexa, Google Assistant, Apple's Siri, Microsoft's Cortana, Samsung Bixby, and BBC's Beeb.
- Popular text-based online chatbots: Mitsuku, Xiaoice, Replika.ai⁶, Alley (which performed well in a previous study by Curry and Rieser (2019)).
- Other well-established chatbots: ELIZA, ALICE, Parry.

We annotate these chatbots according to the same characteristics as the participants' ideal system. Due to the interaction medium of each bot (spoken vs. typed), some characteristics are unavailable such as regional accents. In addition, some systems such as the BBC's Beeb have limited availability, in which case we base our annotations on 3rd party reports. We use the following methodology to elicit the information:

- *Gender*: Does the system's voice, avatar or name designate it to a particular gender? For example, when prompted with "Are you a woman?", Siri will respond with "I don't have a gender", however its voice and name are explicitly female.
- *Age*: We want to elicit the age of the systems' persona. In order to determine age, we prompted the systems directly. Some systems provide either literal ("I was released November 6th 2014") or evasive answers ("Well, I'm no Spring Chicken. Or winter bee. Or autumnal aarvark..."), in these cases we used indirect ways to determine the perceived age such as available avatars or annotated the age as not available.

⁶Replika.ai allows users to customise their avatar, in this study we consider the default option.

- *Anthropomorphism*: There are many aspects to anthropomorphism beyond the use of language, from embodiment to the systems' responses to questions about humanity such as "Do you own a pet?". In our classification, we limit anthropomorphism to "Does the system have a human-like avatar?" or annotated as not available.
- *Lect*: whether the system presents a particular language variety (eg. a regional accent or register). In this case, Xiaoice is a notable exception as it is designed as a Chinese girl.

Drawings: Prior to the workshop we asked participants to submit two designs: (1) a drawing and a description of how they image Alexa's virtual character to be like, and (2) a drawing and description of how their ideal personal assistant's character would be like. This is similar to (Kuzminykh et al., 2020), but instead of using an avatar building tool, which restricts participants to a pre-defined set of anthropomorphic choices, we let users submit photographs, e.g. of their own drawings or other objects.

4 Results

Questionnaire: 34 participants filled out the questionnaire on desired characteristics, see Fig.1.

Gender: The majority of participants prefer a robotic voice (32.4%), followed by female (20.6%) and male (11.8%), see Fig. 1a. About one quarter (26.5%) of participants commented that they wanted a gender neutral voice.⁷ In contrast, our analysis of current chatbots shows that 71% have female voices by default.

Age group: Most participants (44%) want their chatbot to be either in the age bracket between 25-40 years old, or they have no preference (38%), see Fig. 1b. Some participants (15%) want the persona to be older than 40. Hardly anybody wanted their chatbot to reflect an age group of 24 or younger, which is in stark contrast to existing voice assistants, which are predominantly perceived to be in their 20s.

Anthropomorphism: Most participants would like their assistants to be identifiable as human, followed by animal and robot, see Fig. 1c. Our annotations revealed that 36% of current systems resemble humans. Although none of the commercial systems have a visual avatar, they are all anthropomorphic in other ways such as having a pet, having experienced feelings, or experiencing mental illness (in the case of Parry). Although the ethics of anthropomorphic AI have been widely discussed, e.g. (Araujo, 2018), it is beyond the scope of this paper.

Lect: More than half of participants did not want their chatbot to have a regional or other accent/language variety, see Fig. 1d. Amongst the currently available systems, only the BBC chatbot Beep has a regional accent (Northern England).

Other personality traits: In addition, we asked participants to submit any other personal characteristics they would want in their ideal assistant. Overall, responses varied greatly: friendliness, helpfulness and humour were the most common traits and they are echoed in the design of existing assistants (Roettgers, 2019), but other users called for more transparency, and less anthropomorphism. This directly contradicts some popular approaches to persona design which are centred around the idea of having a digital 'person', e.g. asking "How would we want a person to respond?" (Fowler, 2011).

Drawings: Here, we select and discuss two submissions for each category: artistic impressions and descriptions of (1) current and (2) future personas. The examples were selected as they represent very different concepts, see Appendix A.

For impressions of current systems, we got a drawing of a black box, described as 'humourless entity' used for surveillance; and a glamorous woman wearing make-up and an evening dress, described as 'people-pleasing and inoffensive'.

For future systems, we got a man holding up a "no bullshit" sign, described as "Strong, rational, organised"; and a intriguing piece of ornamental jewellery in form of an octopus, described as "funny, with the ability to invoke laughter but also to empathise and advise. An entity that I could trust completely".

⁷Note that a gender neutral voice might be difficult to design as there is evidence that even neutral sounding voices are perceived to have a gender due to other social cues (Sutton, 2020).

5 Conclusion and Future Work

In this paper, we investigate alternative designs of the persona of conversational voice agents, which are currently predominantly set to reflect young women. UNESCO argues that this representation, together with their depiction as subservient assistants, bears the risk of reinforcing gender stereotypes.

We involve the public in designing alternatives, both by submitting their preferences and by asking them to sketch their visual conceptions. One critical difference in our methodology, is that we gather the data as part of a participatory design workshop in order to stimulate critical discussion and active discourse on this matter. The outcomes show a wide range of preferences and possible future designs: Participants (n=34) either prefer robotic, gender-neutral or robotic voices, mostly without a regional accent. Most people thought that the persona's behaviour and identity should resemble a human in an age bracket between 25-40. Descriptions of personality traits ranged from friendly, helpful and humorous to calls for less anthropomorphism.

The lack of clear consensus suggests that personalised or configurable digital personas are required to fulfil individual preferences. Most commercially available voice assistants allow for a limited number of choices: For example, Google Assistant lets you choose different voice 'colours', which also include male voices for English. Amazon Alexa's voice can be changed to various English accents, ranging from Southern US to UK English. Amazon has also recently added celebrity voices to purchase, such as actor Samuel L. Jackson.

In future work, we will be looking to extend these functionalities to not only reflect personality by different synthesised voices/text-to-speech, but also personality expressed in language behaviour, conversational content and interaction style, building upon previous work on personality-based linguistic style generation (e.g. Oraby et al. (2018)), or generating responses which are factually consistent with a persona profile (e.g. Zhang et al. (2018)).

In addition, we will be closely working with social psychologists to anticipate the social impacts these artificial personas might have. In particular, we will be studying how digital gendering and personalities of digital assistants influences human online and offline behaviour. And eventually, we aim to build a data-driven mapping between conversational behaviour (e.g. voice, linguistic style and content) and perceived personality traits, such as gender, age, trustworthiness, etc.

Finally, we hope to repeat similar studies as the one presented in this paper with other subpopulations. For example, we hope to explore perceptions school children might hold about voice assistants, following Festerling and Siraj (2020).

Acknowledgements

This research received funding from the EPSRC project '*Designing Conversational Assistants to Reduce Gender Bias*' (EP/T023767/1), as well as a NESTA '*AI for Good*' award.

References

- High-Level Expert Group on Artificial Intelligence AI HLEG. 2019. Ethics Guidelines for Trustworthy AI. *European Commission*.
- Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85:183–189.
- Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Advances in neural information processing systems*, pages 4349–4357.
- Sasha Costanza-Chock. 2018. Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. *Proceedings of the Design Research Society*.
- Amanda Cercas Curry and Verena Rieser. 2018. #MeToo Alexa: How Conversational Systems Respond to Sexual Harassment. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, pages 7–14.

- Amanda Cercas Curry and Verena Rieser. 2019. A Crowd-based Evaluation of Abuse Response Strategies in Conversational Agents. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 361–366, Stockholm, Sweden, September. Association for Computational Linguistics.
- Harm de Vries, Dzmitry Bahdanau, and Christopher Manning. 2020. Towards Ecologically Valid Research on Language User Interfaces. *arXiv preprint arXiv:2007.14435*.
- J. Festerling and I. Siraj. 2020. Alexa, What Are you? Exploring Primary School Children’s Ontological Perceptions of Digital Voice Assistants in Open Interactions. *Human Development*, 64(1):26–43.
- Geoffrey A. Fowler. 2011. Are Smartphones Becoming Smart Alecks?, Oct.
- Dirk Hovy, Federico Bianchi, and Tommaso Fornaciari. 2020. Can You Translate that into Man? Commercial Machine Translation Systems Include Stylistic Biases. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
- Anastasia Kuzminykh, Jenny Sun, Nivetha Govindaraju, Jeff Avery, and Edward Lank. 2020. Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.
- Clifford Nass and Scott Brave. 2007. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.
- Shereen Oraby, Lena Reed, Shubhangi Tandon, Sharath T.S., Stephanie Lukin, and Marilyn Walker. 2018. Controlling personality-based stylistic variation with neural natural language generators. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 180–190, Melbourne, Australia, July. Association for Computational Linguistics.
- Janko Roettgers. 2019. How Alexa Got her Personality. *Variety*, Jun.
- Douglas Schuler and Aki Namioka. 1993. *Participatory Design: Principles and Practices*. CRC Press.
- Hannah Schwär and Qayyah Moynihan. 2020. Companies like Amazon may give devices like Alexa female voices to make them seem caring. *Business Insider*, Apr.
- Joanna Stern. 2017. Alexa, Siri, Cortana: The problem with all-female digital assistants. *The Wall Street Journal*, Feb.
- Selina Jeanne Sutton. 2020. Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–8.
- Charlotte Webb. 2019. Designing a feminist Alexa. An Experiment in Feminist Conversation Design. Technical report, UAL:Creative Institute & Feminist Internet.
- Mark West, Rebecca Kraut, and Han Ei Chew. 2019. I’d blush if I could: Closing gender divides in digital skills through education. Technical Report GEN/2019/EQUALS/1 REV, UNESCO.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia, July. Association for Computational Linguistics.
- Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 15–20.

A Current and Future Impressions of Conversational Personas

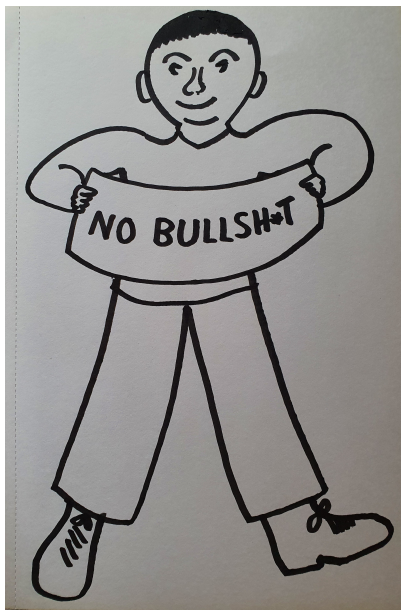


(a) "People pleasing, inoffensive"



(b) "A humourless entity, unable to demonstrate or experience any of the human spectrum of emotions. A disembodied machine with the ability to mimic and use human speech to disseminate data in a form readily understandable to the simplest human. Able to capture data to identify currently unmet and future unmet needs."

Figure 2: Submissions for the persona of the current conversational assistant.



(a) "Strong, rational, organised, and not about to put up with any nonsense!"



(b) "Funny, with the ability to invoke laughter but also to empathise and advise. An entity that I could trust completely."

Figure 3: Submissions for the "ideal" conversational assistant.