"Do I have a say?": Using conversational agents to re-imagine human-machine autonomy

Supraja Sankaran
Department of Industrial Design,
Eindhoven University of Technology
s.sankaran@tue.nl

Chao Zhang
Department of Psychology,
Utrecht University
c.zhang@uu.nl

Mathias Funk
Department of Industrial Design,
Eindhoven University of Technology
m.funk@tue.nl

Henk Aarts
Department of Psychology,
Utrecht University
h.aarts@uu.nl

Panos Markopoulos
Department of Industrial Design,
Eindhoven University of Technology
p.markopoulos@tue.nl

ABSTRACT

With human-centered AI gaining traction, needs for algorithmic transparency, explainability, empathy and ethical considerations in artificial agents are forming core research issues. However, with intelligent agents getting increasingly independent and human-like, there is an increase in perceived threat to human autonomy. Will this *perceived* threat eventually become an *actual* threat where humans lose control over their own goals, decisions and actions? With this provocation paper, we want to urge researchers working on human-agent interactions and conversational agents (CAs) to explicitly consider the impact of intelligent agents on human autonomy. We present arguments and highlight the critical need to focus on human autonomy when interacting with CAs by presenting some core research considerations.

CCS CONCEPTS

• **Computing methodologies** → *Philosophical/theoretical foundations of artificial intelligence.*

KEYWORDS

human-agent interactions, human autonomy, conversational agents

ACM Reference Format:
Supraja Sankaran, Chao Zhang, Mathias Funk, Henk Aarts, and Panos Markopoulos. 2020. "Do I have a say?": Using conversational agents to re-imagine human-machine autonomy. In *2nd Conference on Conversational User Interfaces (CUI '20), July 22–24, 2020, Bilbao, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3405755.3406135

1 INTRODUCTION

Conversational agents (CAs) have permeated into everyday applications such as e-commerce, reservation systems, tech support and personal devices such as smartphones and smart speakers. Yet, as with many state-of-the-art AI systems, CAs are still a long way from having natural everyday conversations with humans [15]. Increasingly, there are attempts to artificially introduce "humanness" when interacting with CAs using sentiment or emotion detection, contextual modeling, commonsense reasoning, response filtering (to prevent profanity, abuse, or hate speech), and personalization [18]. As CAs become more complex an embody greater intelligence, there could be both positive and negative repercussions of the agents' actions and decisions.

For designing meaningful conversations researchers are exploring when and how users can have conversations with agents [7, 26] and distinguishing how different populations of users such as children or elderly interact with CAs [10, 20]. Abdolrahmani et al. have demonstrated how CAs could empower blind people and evoking a sense of independence [2]. Conversely, there are also examples of robotic pets and anthropomorphic CAs designed to provide "companionship" but are instead perceived as devaluing social interactions [8]. Researchers have warned of AI devaluing humanity [23] and the potential threat it poses towards human autonomy [3]. Studies have shown that users perceived a greater threat to their autonomy when an agent used a more controlling language [14]. Therefore, to study the impact of CAs on human autonomy, it is imperative to understand when and how to convey user needs, goals and intentions to the agent and formulate appropriate responses of the agent.

From a social and ethical perspective, human autonomy is a key component of democratic constitutions [27], and shapes modern society as we know it [4]. Social scientists and policy makers are exploring ethical and moral aspects that are associated with intelligent agents which can take over human decision making and action [6], pointing to a more systematic research agenda to address the perceived threat to human autonomy when interacting with intelligent agents [17]. In this paper, we discuss the impact of CAs on human autonomy and present key considerations needed to respect human autonomy.

2 ENDANGERED EVERYDAY AUTONOMY

Autonomy is crucial for human well-being and development [16], although the term *autonomy* has different meanings across different disciplines. In social sciences and philosophy, autonomy is considered as a human need and fundamental right to have *freedom of*

CUI '20, July 22–24, 2020, Bilbao, Spain

ACM ISBN 978-1-4503-7544-3/20/07…$15.00
https://doi.org/10.1145/3405755.3406135

choice, to *determine one's own goals* and *course of actions* [19, 22]. In human-agent interaction and intelligent systems, autonomy refers to the ability of an agent (either man or machine) to *have control* and to *independently make decisions* [1, 5]. When interacting with CAs, we consider human autonomy as the ability to *have a 'say'* in decision making and goal pursuit.

In the race of developing more 'human-like' CAs and to artificially recreate naturalness in conversations, the impact of these advancements on human autonomy has been overlooked [3]. Moreover, evaluations of CAs have also been limited to completeness and accuracy of conversations, and basic user experience [7]. Although important, we believe that Turing tests and task-oriented user experience assessments as a sole benchmark are insufficient [18, 25]. It is important to go beyond functional tasks to assess and balance the power relation between agents and human users based on users' needs and goals.

2.1 Autonomy beyond Emotional Capacity

In 2019, George [12] presented a provocation on the need, use and potential misuse of building emotional capacity in AI at CUI. Adding to that, we would like to raise the issue of perceived threat to human autonomy when building in more "humanness" into CAs. Specifically, it is important to explore how much control CAs should have in decision making when users' needs and goals are not fully understood by an agent. Will a situation arise where humans do not have a 'say' in decisions and most tasks are performed by agents?

2.2 From Science Fiction to the Everyday

Let us consider the scenario in the movie *'Her'* [24]. The protagonist (Theodore) falls in love with his personal CA. Initially, the agent helps him overcome his loneliness but later the agent autonomously arranges blind dates and personal encounters without considering Theodore's needs or consulting with him. This represents a situation where a user's autonomy is seriously jeopardized by an intelligent agent. Similarly, psychologist Robert Epstein was allegedly fooled into thinking that a chatbot was a real person [11]. This also raises the moral question–"is endowing human traits in agents an ethical violation?".

Delving into current applications of CAs, we can assume that if an agent considers our request to order a meal at a fast food restaurant and steers us towards a healthier choice (based on our goals), it would not be *perceived* as a threat to our autonomy (given a generally positive stance towards healthy food). However, it might be perceived as a substantial threat if the agent becomes more controlling by autonomously deciding (without asking/providing choices) about when, where or how much we could eat. In an extreme case, our autonomy could be more seriously threatened if agents start taking critical decisions and actions such as consenting for surgery or changing financial investment plans without consulting our needs or informing us about why the action was taken. While CAs can support human autonomy, they certainly undermine autonomy when interaction and communication are constrained by the agent–either in content and quality, or in what is left out and which options are not presented. In consequence, an individual might no longer be able to play their human role in the interaction, as they are implicitly coerced into following the lead of the agent.

2.3 Perspective-taking CAs

Therefore, CAs need to be capable of *'perspective taking'* and applying a *'theory of mind'* [21], to be able to consider users' desires and beliefs, to provide them with choices, and to answer users' questions about decisions they make (e.g., "why did you suggest this instead of that?" and "is there something else that can be done apart from the suggested options?"). To this end, we would like to put forward some core considerations to be reflected upon when developing CAs.

- **Goals:** How to determine and balance users' beliefs, desires and goals to the knowledge and goals of agents? How will agents represent users' goals and align themselves? How will agents deal with conflicting and dynamically changing goals of one or more users?
- **Information/Choices:** How should agents communicate known and withheld choices and information? When, how and in which part of conversations should users be provided with choices? How will agents inquire about missing information for decision-making? How will agents represent users' past actions, decisions and information?
- **Action acceptance:** When do agents transcend beyond acceptable functionality and task engagement to perform tasks that surpass or contradict the desires of the human user? How can agents justify an action that is not defined directly by user goals? How can agents justify doing nothing in certain situations? When do users perceive a threat to their autonomy?

3 CONTINUING THE CONVERSATION

We put emphasis on the need to pay attention to human autonomy and the (perceived) threat to autonomy when designing and evaluating CAs as an important user experience component. This could potentially be done by assessing users' psychological reactance and social agency [9, 13]. In this provocation paper, we presented scenarios and critical considerations around the impact of CAs on human autonomy. The discussion could be extrapolated to other domains such as policy making, law and defence wherein the negative impact is not restricted to an individual's autonomy but ripples out to impact the autonomy of their family, friends or even the society at large. While we present the arguments and discussions here in relation to CAs, the points are fundamental provocations that could steer discussions around other applications where humans are interacting, *conversing* and negotiating with intelligent agents. As HCI researchers, designers and developers of intelligent systems aimed at enhancing human capabilities and augmenting human performance, it is essential to continuously reflect upon and ensure that fundamental human-aspects such as autonomy are not supplanted in the drive towards making systems more human-like.

REFERENCES

[1] Hussein A Abbass. 2019. Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation* 11, 2 (2019), 159–171.

[2] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 249–258. https://doi.org/10.1145/3234695.3236344

[3] Janna Anderson, Lee Rainie, and Alex Luchsinger. 2018. Artificial intelligence and the future of humans. *Pew Research Center* (2018).

[4] Albert Bandura. 2006. Toward a psychology of human agency. *Perspectives on psychological science* 1, 2 (2006), 164–180.

[5] K Suzanne Barber and Cheryl E Martin. 2000. Autonomy as decision-making control. In *International Workshop on Agent Theories, Architectures, and Languages*. Springer, 343–345.

[6] Rafael Calvo, Dorian Peters, Karina Vergobbi Vold, and Richard Ryan. 2019. Supporting human autonomy in AI systems: A framework for ethical enquiry. Springer.

[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, and et al. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. 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 475, 12 pages. https://doi.org/10.1145/3290605.3300705

[8] Simon Coghlan, Jenny Waycott, Barbara Barbosa Neves, and Frank Vetere. 2018. Using Robot Pets Instead of Companion Animals for Older People: A Case of "Reinventing the Wheel"?. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 172–183. https://doi.org/10.1145/3292147.3292176

[9] Maartje MA de Graaf, S Ben Allouch, and JAGM Van Dijk. 2015. What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *International Conference on Social Robotics*. Springer, 184–193.

[10] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is It OK If I Eat You?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 595–600. https://doi.org/10.1145/3078072.3084330

[11] Robert Epstein. 2007. From Russia, with love. *Scientific American Mind* 18, 5 (2007), 16–17.

[12] Stella George. 2019. From sex and therapy bots to virtual assistants and tutors: how emotional should artificially intelligent agents be? *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–3. https://doi.org/10.1145/3342775.3342807

[13] Aimi Shazwani Ghazali, Jaap Ham, Emilia Barakova, and Panos Markopoulos. 2019. Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. *Advanced Robotics* 33, 7-8 (2019), 325–337.

[14] Aimi S Ghazali, Jaap Ham, Emilia I Barakova, and Panos Markopoulos. 2017. Pardon the rude robot: Social cues diminish reactance to high controlling language. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 411–417.

[15] Hector J Levesque. 2017. *Common sense, the Turing test, and the quest for real AI*. MIT Press.

[16] Mirtha Rosaura Muniz Castillo. 2009. Autonomy as a foundation for human development: A conceptual model to study individual autonomy. (2009).

[17] High-Level Expert Group on Artificial Intelligence. 2019. Ethics guidelines for trustworthy AI. *European Commission* (2019).

[18] Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, et al. 2018. Conversational ai: The science behind the alexa prize. *arXiv preprint arXiv:1801.03604* (2018).

[19] Richard M Ryan and Edward L Deci. 2006. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of personality* 74, 6 (2006), 1557–1586.

[20] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R Cowan. 2019. Voice assistants and older people: some open issues. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.

[21] Brian Scassellati. 2002. Theory of mind for a humanoid robot. *Autonomous Robots* 12, 1 (2002), 13–24.

[22] Anastasia Vugts, MARIËTTE VAN DEN HOVEN, EMELY DE VET, and Marcel Verweij. 2018. How autonomy is understood in discussions on the ethics of nudging. *Behavioural Public Policy* (2018), 1–16.

[23] Joseph Weizenbaum. 1976. Computer power and human reason: From judgment to calculation. (1976).

[24] Wikipedia contributors. 2020. Her (film) — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Her_(film)&oldid=943862688. [Online; accessed 12-March-2020].

[25] Wikipedia contributors. 2020. Loebner Prize — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Loebner_Prize&oldid=936700092. [Online; accessed 12-March-2020].

[26] Nicole Yankelovich. 1996. How do users know what to say? *interactions* 3, 6 (1996), 32–43.

[27] Fareed Zakaria. 1997. The rise of illiberal democracy. *Foreign Aff.* 76 (1997), 22.