Panel Proposal

Exploring the Possibility and Ethics of AI Paternalism in Health Apps

Organizers and Participants

- Michael Kühler, University of Twente, m.c.kuhler@utwente.nl
- Lena Busch, University of Münster, lena.busch@uni-muenster.de
- Katja Stoppenbrink, University of Münster, katja.stoppenbrink@uni-muenster.de
- Cristina Zaga, University of Twente, c.zaga@utwente.nl

General Description

Health apps are supposed to promote their users' health by tracking health related data and influencing their users to act in a healthier way. Hence, they may be considered persuasive technologies, even if they usually cannot be considered actors themselves.
Now, consider health apps which include AI technology and are capable of

1. analyzing their users' behavior in light of the individual user's and more general health related data,
2. drawing conclusions as to which behavior would benefit the individual user, and
3. influence the individual user's behavior accordingly, for example by way of making "nudging" suggestions on what to do or by means of gamification.

Arguably, such AI based health apps may now be considered a sufficiently autonomous actor when it comes to influencing the users' behavior for their own good. If so, it seems that AI based health apps do not only raise common ethical questions about their influence on the users' autonomy but also gain a paternalistic potential, which needs to be analyzed and critically addressed in more detail. The panel is intended to do just that and aims at an interdisciplinary discussion about the possibility and ethics of AI paternalism in health apps.
Keywords: AI, autonomy, paternalism, self-determination theory, responsible design

Individual Abstracts

Michael Kühler: Exploring the concept of AI paternalism

Consider the following traditional philosophical notion of paternalism:
"X acts paternalistically towards Y by doing (omitting) Z:

1. Z (or its omission) interferes with the liberty or autonomy of Y.
2. X does so without the consent of Y.
3. X does so only because X believes Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y." (Dworkin, https://plato.stanford.edu/entries/paternalism/)

Based on the assumption that the phenomenon of AI paternalism in health apps is not implausible to begin with, I discuss how well the above definition can capture this novel phenomenon. Firstly, the AI technology involved does not have beliefs or intentions (condition 3) based on which it acts. Yet, its programming is certainly goal-oriented, namely improving the user's good health, and includes calculations concerning which of the user's behavior would contribute best to reaching this goal. Secondly, while it might be argued that condition 2 is not satisfied, given that users voluntarily have to install and use even an AI based health app, I argue that particular instances of the app's persuasive functioning nevertheless may be considered as influencing behavior without the user's consent in this particular instance. This is so mainly because the influence on the user's autonomy (condition 1) is best analyzed in terms of so-called soft paternalism or

nudging. In conclusion, the traditional notion of paternalism is in need for some specific revision in order to be applicable and useful for addressing the phenomenon of AI paternalism.
Keywords: AI, autonomy, paternalism

Lena Busch: Autonomy support, motivation and physical activity in health apps: a perspective from self-determination theory and empirical evidence

Fitness apps are promising digital tools to support self-tracking and physical activity. Specific app functions such as normalized step targets represent controlling conditions that can affect controlled vs. autonomous motivation and thus motivated physical activity, all of which are crucial aspects for preparing a well-considered discussion of a corresponding paternalist potential.
In my talk, I provide an empirical approach on this from sport and exercise psychology. First, the concept of autonomy, autonomy need satisfaction, and autonomous motivation are introduced from the perspective of self-determination theory (SDT). SDT is a well-established theory describing and explaining motivated behavior in psychology. Second, the results and implications of an empirical study are presented.
The specific aim of the experimental study was to examine the effects of a normalized step target on autonomy need satisfaction, autonomous vs. controlled motivation, and physical activity by using SDT as a theoretical framework. In a six-week RCT, participants in two groups used fitness app devices to track their physical activity. Participants in one group had a normalized step target of 10,000 whereas the other group had not any step target. Participants in a third control group tracked their physical activity without fitness app support. In sum, self-tracking via fitness apps can support physical activity, and normalized step targets can undermine autonomous motivation. Lack of normalized targets can support autonomy need satisfaction and physical activity but can also foster amotivation. Thus, it is advised to support autonomous goal setting in fitness app users.
Keywords: autonomy, fitness app, self-determination theory

Katja Stoppenbrink: Implications of transtemporal considerations and different conceptions of autonomy for the ethics of AI paternalism

Self-tracking or, more generally, health apps are intended to support their users in sustaining healthy lifestyle practices. Ideally, they are employed following free and informed choices. However, initial decision-making and ongoing interaction with one's app need to be distinguished. While supposedly there is no threat to users' autonomous decision-making in first-time employment of an app, over time, recommendations made by an app may interfere with users' autonomy in that they are meant to function as 'gentle reminders' and 'book-keepers' at the same time. A fortiori, this applies to 'learning apps' relying on artificial intelligence (AI). Proposals to conceptualize this function as 'nudging' either face an objection from semantic vagueness or – given a certain understanding of nudges – an objection from (soft) paternalism. At any rate, the qualification of an app as 'nudge provider' does not do any normative or evaluative work itself. Against the background of the standard definition of 'paternalism' (Dworkin, *Stanford Enc Philos* 2020) with its three conditions (see the citation in Kühler's abstract above), there may be 'app-paternalism' in that an action-token suggested by an app is not included in the initial general consent to employ the app (condition 2) and does not correspond to the user's overall welfare (condition 3). Cost-benefit assessments may even change over time through app use itself, e.g. when attitudes towards app employment change. I discuss the implications of different conceptions of autonomy (condition 1) and transtemporal considerations for how 'app paternalism' and, more specifically, 'AI paternalism' should be judged from an ethical point of view.

Keywords: AI, autonomy, paternalism

Cristina Zaga: Towards Responsible AI Design: methodological gaps in the design of machine learning driven health apps

Mobile applications are often intended to support healthy lifestyles. These applications use machine learning algorithms to analyze and parse users' actions with the ultimate goal of influencing their decisions and behaviors towards a pre-determined "healthy behavior". Even though most of these apps have, in practice, little autonomy and "intelligence", their automated nudging behaviors informed by people's data, communicate as a paternalist agent. The paternalistic agency creates tension between the autonomy of the apps and human control.
However, the majority of the health applications available on the market have not been developed taking into account and anticipating the human-technology relations and their impact on society. The design of machine-learning driven technology for health suffers from a methodological gap. Human-computer interaction, psychology, and engineering methods with their focus on utility, productivity, and – in certain instances – normative ontologies fall short in accounting for the emerging paternalism of technology. Moreover, designing algorithms and apps behaviors is not yet a democratic endeavor; therefore, it is hard to find a common ground between the various disciplines and stakeholders involved in the design process. Fruitful co-creation and the establishment of a shared understanding among stakeholders of machine learning-driven technology is difficult without dedicated methods. In this panel, I argue that mediation theory and transdisciplinarity could represent productive theoretical perspectives to take a moral and democratizing turn in designing health applications. I discuss how these theories could be applied to the development of health apps within a transdisciplinary and responsible design methodological framework informed by citizen science practices.
Keywords: paternalistic agency; transdisciplinarity; responsible design

CVs

Michael Kühler

Michael Kühler is Assistant Professor at the Department of Philosophy at the University of Twente and "Privatdozent" at the University of Münster. His research interests include ethics, metaethics and applied ethics, esp. ethics of technology. His recent publications include "Romantic Love Between Humans and AIs: a Feminist Critique," together with Andrea Klonschinski, in: Cushing, Simon (ed.): *Philosophy of Love and Loving*, Houndmills: Palgrave Macmillan, forthcoming; "Technological Moral Luck," in: Beck, Birgit/Kühler, Michael (eds.): *Technology, Anthropology, and Dimensions of Responsibility*, Techno:Phil 1, Stuttgart: Metzler, 2020, 115-132, URL: https://doi.org/10.1007/978-3-476-04896-7_9.

Lena Busch

Lena Busch is a Post-doc at the University of Münster in Germany. She has a background in psychology, specifically sport and exercise psychology and clinical psychology. She worked at the institute of sport and exercise psychology in Münster, participating in an interdisciplinary DFG-funded research group on trust and communication in a digitized world. Her research interests involve motivation, trust, digitalization, and perfectionism. Her recent publication is "The Influence of Fitness-App Usage on Psychological Well-Being and Body Awareness—A Daily Diary Randomized Trial" together with Till Utesch, Paul-Christian Bürkner, & Bernd Strauss, in: *Journal of Sport and Exercise Psychology*, 2020. Advance online publication. http://dx.doi.org/10.1123/jsep.2019-0315

Katja Stoppenbrink

Katja Stoppenbrink is a Stand-in Professor at the Department of Philosophy of the University of Münster in Germany. She has a background in philosophy, law and history. Her research interests range from ethical theory and applied ethics (esp. medical, technological, business and legal ethics) to the philosophy of human rights and questions of social and political philosophy. She is currently preparing a book on disability and social inclusion of persons with disabilities. Her most recent publication: "Person und Inklusion [Person and Inclusion]," in: *Der Begriff der Person in systematischer wie historischer Perspektive*, ed. by Michael Quante, Hiroshi Goto, Tim Rojek & Shingo Segawa, Paderborn: Mentis, 2020, 177–197, DOI: https://doi.org/10.30965/9783957437167_013.

Cristina Zaga

Cristina Zaga is Assistant Professor at the Human-Centred Design Group (Design and Production Management department) and a researcher at The DesignLab at the University of Twente. Cristina's research focuses on design methods for embodied AI and interactive agency. At the DesignLab, she investigates methodologies, methods, tools, and techniques to connect science, technology, and society through responsible design and citizen science. Her most recent publication is: Zaga C. Designing the future of education: From tutor robots to intelligent playthings. Tijdschrift voor human factors. 2019 Nov 2; 44(3).