"Embedded Ethics and Social Sciences" in HRI Research: Scenarios and Subjectivities

1st Maximilian Braun
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
maximilian.braun@tum.de

2nd Svenja Breuer
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
svenja.breuer@tum.de

3rd Daniel Tigard
Institute for History and Ethics of Medicine
Technical University of Munich
Munich, Germany
daniel.tigard@tum.de

4th Ruth Mueller
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
ruth.mueller@tum.de

Abstract—In current robotics and AI research, much work is done to harness recent technological developments and make them fruitful for healthcare applications. Such research often works in a very technology-centered fashion and oftentimes struggles with considering healthcare-related aspects at early phases of research. Existing paradigms in HRI research, for instance, often reduce development and implementation environments to quantifiable material relationships that constitute such a delineation (i.e., between ideally-arranged experimentation sites and much messier implementation realities). But this alone does not capture the social, epistemic or symbolic dimensions that underlie them. Based on our experiences as embedded ethicists and social scientists, we offer two contributions to the HRI2022 Workshop on "Re-Configuring HRI". Our first contribution centers around creating an open exchange of perspectives between different stakeholders through scenario-based focus groups, whereas the second is a shift of the analytic focus from the technologies in-the-making to the subjectivities of those designing and developing them. Scenario-based focus groups can establish an early exchange of perspectives in the research process, while the researchers' subjectivities help us to gain a critical understanding about the wider normative dimensions and rationales that guide the technological research for domains like healthcare. Together, these two pillars help to foster a more open and informed exchange between scientists and developers on the one side and stakeholders and users on the other.

Index Terms—embedded ethics and social sciences, robotics, AI, healthcare, focus groups, subjectivities

I. INTRODUCTION

Political discourses all around the world expect an increased demand for healthcare and caregiving services due to a number of past and contemporary trends, such as a lack of qualified personnel in rural areas or healthcare professions in general and a constantly increasing life expectancy of elderly populations [8], [14]. Against this discursive backdrop, the idea that technology can help to mitigate these challenges is driving research efforts on robotic systems to be implemented in healthcare practices. A myriad of projects in fields such as Human Robot Interaction (HRI) [11], [31], Gerontology [19], [32] and Geriatronics [27] epitomize this trend, as they refer to these challenges as a motivation for research.

Especially with regard to robotics and AI, much work is done to harness recent technological developments and make them fruitful for healthcare applications. However, such research often works in a very technology-centered fashion and oftentimes struggles with considering healthcare-related aspects at early phases of research. To address this problem, it can help to create bridges between researchers and developers on the one hand and the real-world environments (including those affected by the research) on the other to allow for an open exchange of perspectives early on.

This is what we try to achieve in the Responsible Robotics (RR-AI) project at the Technical University of Munich. Our project follows an 'embedded ethics and social sciences' methodology [16] that draws from intensive empirical field work, comprising qualitative, semi-structured interviews, focus groups, and ethnography, to account for the imaginaries, subjectivities and values embedded in engineering, healthcare and caregiving practices [1]. Our team comprises ethicists and social scientists that are embedded in the development practices and processes of Geriatronics, an emerging field of research that tries to build robotic applications to support caregiving and healthcare tasks for elderly populations.

As an overall scope early in the RR-AI project, we tried to capture and understand the various perspectives of the many stakeholders involved, shedding light on the wider ecosystem around Geriatronics. With regard to the developers and users of Geriatronics systems, we tried to account for the mutual conceptions that these groups might hold of each others work environments in robotics laboratories and hospitals or elderly care homes. As these conceptions are usually present long before phases of implementation, they serve as a good entry point for exploring these stakeholders' perspectives at early phases of research and development, where often no direct interaction between them is foreseen.

Existing conceptions in HRI research and robotics research more generally (e.g., the structured/unstructured divide) often reduce development and implementation environments to quantifiable material relationships that constitute such a delineation (i.e., between ideally-arranged experimentation sites and much messier implementation realities). But this alone does not capture the social, epistemic or symbolic dimensions that underlie them. The problem is that, by missing these other dimensions, HRI research will not become attuned to differences in practices, values and working cultures that underlie the materiality of development and work environments, which are important to understand and care for the concerns that users, developers and stakeholders of the technologies might have. Our ethnographic lens can help to see beyond a material understanding of application contexts or environments as material arrangements and extend our view to other normative dimensions that shape and are shaped by the research efforts in Geriatronics [3], [5].

II. Contributions to Forming New Paradigms in HRI Research

With this agenda in mind, we offer two contributions to the HRI 2022 Workshop on 'Re-Configuring HRI'. Our first contribution centers around creating an open exchange of perspectives between different stakeholders, whereas the second is a shift of the analytic focus from the technologies in-the-making to the subjectivities of those designing and developing them.

A. *1st Contribution: Conducting Scenario-based Focus Groups with Users*

New technoscientific developments are always accompanied by particular visions, use cases and application scenarios that are picked up in the public discourse to debate the societal dimensions of these developments. The problem is that these visions can be quite disconnected from the actual research and often do not mirror actual laboratory work, but instead they reflect imagined applications of a more or less distant future [7], [18], [21]. This future-orientation runs the risk of blurring the lines between fiction and reality, driving stakeholders and the wider public's attention away from the actual research practices, where their perspectives and concerns would actually deserve consideration to allow for more responsible ways of conducting research [2], [22], [23], [25].

While imaginations of robots as companions and assistants in almost every aspect of our everyday lives have proliferated in the past decades in mainstream media and popular culture, the everyday reality created by robotics researchers and developers still looks quite different. This is due in part to the fact that, while most popular imaginaries already depict them as widely applicable (if not general-purpose) technologies, robots have so far mainly been used in so-called structured environments and industrial applications, where it was often not only the robot, but also the contained environment that got implemented into work settings.

Visions around the use of robots in elderly care and healthcare are a particularly illuminating example. In the past years, more and more actors started to depict robots as venturing into the unstructured, wild environment of elderly care, where they would solve tasks, such as delivering food or suggesting medical diagnosis [26], [29]. Such depictions again pose a risk of blurring the lines between fiction and reality, since such depictions rather refer to ongoing, open-ended research efforts and not accomplished implementation examples. As a remedy, we provide stakeholders with a more realistic sense of what the researchers are currently working on, based on our insights gained through ethnography and early interviews with Geriatronics researchers.

We conducted focus groups with nursing students which relied on the current robotics research rather than more general conceptions about robotics in elderly care. We confronted the students with concrete application examples currently worked on in the laboratory and asked them what they think about them, how these applications could interfere with existing caregiving practices and what might speak in favor of or against their implementation. One example was the use of a telemedical system that allows caregivers to conduct blood pressure measurements remotely. This confronting is another mode of (re)configuring existing depictions, ideas and understandings of the laboratory work, since healthcare practitioners get access to scenarios that reflect existing development work that open up the logics and imaginaries behind them.

As a next step, we plan to conduct workshops in which we bring engineers and healthcare workers together. This enables us not only to give healthcare practitioners even more direct access to the development perspective, but also engages developers with the hopes, ideas and concerns of caregiving and healthcare professionals. In this way, we facilitate a dialogue between often separately conceived environments (i.e., the laboratory and the context of implementation) that can go beyond narratives of an alleged inevitability, an undisputed beneficence and the radical newness of robotic applications in caregiving and healthcare.

Among the key concerns in our recent focus groups and future workshops have been potential power asymmetries [9], [23], [28]. Special attention has to be paid to allowing caregivers and healthcare practitioners to voice their concerns and hopes freely and openly, but also to consider and hold them up in case they should not align with dominant understandings of how research and innovation should happen. Otherwise, such an approach runs the risk of turning them into the passive recipient of the technologies whose views and concerns are framed as obstacles to overcome, not as serious concerns able to re-configure the trajectory of technological development [4], [15].

We used questions in the focus groups that motivate the participants to reflect on both the opportunities and the challenges, as well as the wider implications, of the presented application scenarios. We suggest that focus groups and workshops include a clear description of a relevant application scenario, be backed up by empirical material (rather than exaggerated

demos or advertisements) and pose open-ended questions with sufficient time to discuss them. Furthermore, participants should be encouraged to reflect on their own professions in the light of a potential uptake of robotic applications. This includes structural aspects that might change through the implementation of the respective technologies, questions of what role robotics could and should play in their work in general and what a desirable research trajectory would be from their point of view.

B. 2nd Contribution: Analyzing Researchers' Subjectivities

The other contribution we offer is a focus on the researchers' subjectivities. Scientific and technological research in the laboratory often strives to realize certain notions of objectiveness and generalizability, creating a dichotomy between the scientific knowledge and its subjective aspects [6], [13]. This view runs the risk of neglecting the individual researchers as epistemic subjects in scientific and technological research, each of them holding her own set of values, motivations and beliefs. In opposition to that, we acknowledge the researchers' subjectivities and their agency in the research process in order to understand the trajectory of emerging fields and trending topics of research. From this view, we consider the process in which researchers become experts and advocates of a new or trending domain of research and how they accommodate, comply or resist the normative expectations and structures directed towards their work.

Looking at the researchers as epistemic agents helps to trace the normative dimensions that co-determine the ways in which research can be conducted. It can also serve as an entry point to shed light on the "epistemic cultures" [12], "epistemic living spaces" [3] or "scientific identities" or "communities" [10] potentially at play in trending domains of research. These domains always follow a political Zeitgeist in terms of the rules and protocols of how research can be arranged and conducted, which in turn influences what researchers consider as valuable and feasible research. Project-based research settings, for instance, have an influence on how researchers sample study participants [20] and often result in epistemic endeavors that value research output that is translatable into commercial products [30]. Researchers now can, for instance, ignore, comply or actively resist these tendencies, finding themselves in a "mangle" of how they understand themselves in relation to them and that they can account for in conversations with us [24].

To shed light on this mangle, we conducted in the RR-AI project semi-structured interviews with the junior and senior researchers of Geriatronics that focused on their current work biographical information and social aspects of their work life. We paid attention to how the researchers articulated their own agency, with a special focus on how they a) relate it to the many expectations that are directed towards their work and the institution they work for in the wider discourse, and b) account for the relationships with internal and external stakeholders, colleagues or their supervisors. As an example, some researchers distinguished between a "human" and a "robot"

side in Geriatronics and situated themselves either on the one or the other side. Also, a strong commercial framing was present in our conversations with them, where many would refer to their own research as developing products whose societal value they need to demonstrate and promote. These accounts point to ongoing negotiations about the epistemic nature of the project and about the researchers grappling with being cast (and mangled) as a neoliberal agent who represents the developed technologies and argues for their support and uptake. By analyzing these negotiations, we can trace the ways in which individual subjectivities might be collectivized and taken up (or not) in the institutionalization of emerging research domains.

Methodologically, this contribution requires a certain amount of trust, adequate deliberation practices and spaces. Utilizing the "peer-to-peer" interviewing methodology [17], we established such spaces by approaching the researchers individually and inviting them to reflect on and articulate their personal and professional situation in conversations with us. This methodology also resonates well with the relationship we hold with the researchers, as we collaborate with them also beyond the interviews (e.g., by conducting user studies) and since we are peers to them in the sense that we are also in PhD and postdoc phases of our academic careers. The interviews lasted for around 1-2 hours, which provided the researchers with enough time to elaborate on aspects of their biography, their career and the currently envisioned research trajectories for a longer amount of time, with several spontaneous *in situ* occurrences of reflection that go beyond established narratives that are present in the wider discourse.

III. CONCLUSION

In this position paper, we suggested two contributions to add to the common repertoire of HRI research and development practices at early phases of research and development, namely conducting scenario-based focus groups and analyzing researchers' subjectivities using interviews and ethnography. Overarching to both approaches is a high sensitivity towards the openness and context-dependency of technological applications and the respect for alternative interpretations of how technologies might or might not contribute to existing practices in the context of implementation. From this perspective, scenario-based focus groups can establish an early exchange of perspectives in the research process, while the researchers' subjectivities help us to gain a critical understanding about the wider normative dimensions and rationales that guide the technological research for domains like healthcare. Together, these two pillars help to foster a more open and informed exchange between scientists and developers on the one side and stakeholders and users on the other.

ACKNOWLEDGMENT

REFERENCES

[1] Breuer, S., Braun, M., Tigard, D., Buyx, A., Mueller, R. (unpublished). Embedded ethics and social science in HCI: Tracing engineers imaginaries of healthcare. ACM Transactions on Computer-Human Interaction (TOCHI), 2021.

[2] Felt, U. (2016). Response-able practices or new bureaucracies of virtue: The challenges of making RRI work in academic environments, Pre-print; Published by the Department of Science and Technology Studies, University of Vienna, November 2016. Available at http://sts.univie.ac.at/publications

[3] Felt, U., Fochler, M. (2012). Re-ordering epistemic living spaces: On the tacit governance effects of the public communication of science. In S. Roedder, M. Franzen, P. Weingart (Eds.), The sciences media connection public communication and its repercussions: Springer Science+Business Media B.V.

[4] Felt, U., Wynne, B., Callon, M., Goncalves, M. E., Jasanoff, S., Jepsen, M., . . . Tallacchini, M. (2007). Taking European knowledge society seriously: Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission ([1st ed.] ed.). Luxembourg: Office for Official Publications of the European Communities.

[5] Fischer B., Oestlund B., Peine A. (2020). Of robots and humans: Creating user representations in practice. Social Studies of Science, 50(2), 221-244. doi:10.1177/0306312720905116

[6] Fleck, L. (1980). Entstehung und Entwicklung einer wissenschaftlichen Tatsache - Einfuehrung in die Lehre vom Denkstil und Denkkollektiv Frankfurt am Main: Suhrkamp.

[7] Fortun, M. (2005). For an ethics of promising, or: A few kind words about James Watson. New genetics and society, 24(2), 157173. doi:10.1080/14636770500184792

[8] Hergesell, J., Maibaum, A., Meister, M. (2020). Genese und Folgen der Pflegerobotik. Die Konstitution eines interdisziplinaeren Forschungsfeldes. Weinheim: Beltz Juventa.

[9] Joly, Pierre-Benoit (2015). Governing emerging technologies? The need to think outside the (black) box. In science and democracy: Making knowledge and making power in the biosciences and beyond, edited by Stephen Hilgartner, Clark Miller, and Rob Hagendijk, 133-55. London, UK: Routledge.

[10] Kastenhofer, K., Molyneux-Hodgson, S. (2021). Community and identity in contemporary technosciences (Vol. 31): Springer Nature.

[11] Kim, K. J., Park, E., Shyam Sundar, S. (2013). Caregiving role in humanrobot interaction: A study of the mediating effects of perceived benefit and social presence. Computers in Human Behavior, 29(4), 1799-1806. doi:10.1016/j.chb.2013.02.009

[12] Knorr-Cetina, K. (1999). Epistemic cultures: How the sciences make knowledge. Cambridge, Mass.: Harvard Univ. Press.

[13] Latour, B., Woolgar, S. (2013). Laboratory life. Princeton University Press.

[14] Maibaum, A., Bischof, A., Hergesell, J., Lipp, B. (2021). A critique of robotics in health care. Ai and Society. doi:10.1007/s00146-021-01206-z

[15] Marris, C., Calvert, J. (2020). Science and Technology Studies in policy: The UK Synthetic Biology Roadmap. Science, Technology, and Human Values, 45(1), 3461. doi:10.1177/0162243919828107

[16] McLennan, S., Fiske, A., Celi, L. A., Mueller, R., Harder, J., Ritt, K., . . . Buyx, A. (2020). An embedded ethics approach for AI development. Nature Machine Intelligence, 1, 389. doi:10.1038/s42256-020-0214-1

[17] Mueller, R., Kenney, M. (2014). Agential conversations: Interviewing postdoctoral life scientists and the politics of mundane research practices. Science as Culture, 23(4), 537-559. doi:10.1080/09505431.2014.916670

[18] Nordmann, A. (2007). If and then: A critique of speculative NanoEthics. NanoEthics, 1(1), 3146. doi:10.1007/s11569-007-0007-6

[19] Peine, A., Marshall, B. L., Martin, W., Neven, L. (2021). Sociogerontechnology (1st Edition ed.). London: Routledge.

[20] Penkler, M., Felder, K., Felt, U. (2019). Challenging diversity: Steering effects of buzzwords in projectified health care. Science, Technology, and Human Values, 45(1), 138-163. doi:10.1177/0162243919841694

[21] Sand, M. (2016). Responsibility and visioneering: Opening pandoras box. NanoEthics, 10(1), 7586. doi:10.1007/s11569-016-0252-7

[22] de Saille, S., Medvecky, F. (2016). Innovation for a steady state: a case for responsible stagnation. Economy and Society, 45(1), 1-23. doi:10.1080/03085147.2016.1143727

[23] Scott, D. (2021): Diversifying the deliberative turn: Toward an agonistic RRI. In: Science, Technology, and Human Values, December 2021. doi:10.1177/01622439211067268.

[24] Sigl, L. (2019). Subjectivity, governance, and changing conditions of knowledge production in the life sciences. Subjectivity, 12(2), 117-136. doi:10.1057/s41286-019-00069-6

[25] Stilgoe, J., Owen, R., Macnaghten, P. (2013). Developing a framework for responsible innovation. Research Policy, 42(9), 15681580. doi:10.1016/j.respol.2013.05.008

[26] Taylor, A., Chan, D. M., Riek, L. D. (2020). Robot-centric perception of human groups. ACM Transactions on Human-Robot Interaction, 9(3), 1-21. doi:10.1145/3375798

[27] Troebinger, M., Jaehne, C., Qu, Z., Elsner, J., Reindl, A., Getz, S., . . . Haddadin, S. (2021). Introducing GARMI - a service robotics platform to support the elderly at home: Design philosophy, system overview and first results. IEEE Robotics and Automation Letters, 6(3), 5857-5864. doi:10.1109/LRA.2021.3082012

[28] Valkenburg, G. (2020). Consensus or contestation: Reflections on governance of innovation in a context of heterogeneous knowledges. In: Science, Technology, and Society, 25(2), 341356. doi:10.1177/0971721820903005.

[29] Valls-Peris, N., Domnech, M. (2020). Roboticists imaginaries of robots for care: The radical imaginary as a tool for an ethical discussion. Engineering Studies, 12(3), 157-176. doi: 10.1080/19378629.2020.1821695.

[30] Vermeulen, N. (2015). From virus to vaccine: Projectification of science in the VIRGO consortium. In B. Penders, N. Vermeulen, J. N. Parker (Eds.), Collaboration across health research and medical care. London: Routledge.

[31] Weber, J. (2005), Helpless machines and true loving care givers: a feminist critique of recent trends in humanrobot interaction. Journal of Information, Communication and Ethics in Society, 3(4), 209-218. doi:10.1108/14779960580000274

[32] Wouters, E. (2022). Digital transformations in care for older people: critical perspectives. Journal of Gerontological Social Work, 63(6-7), 611-624.