



Editorial: Responsible Robotics

Martim Brandão^{1*}, Masoumeh Mansouri² and Martin Magnusson³

¹Department of Informatics, King's College London, London, United Kingdom, ²School of Computer Science, University of Birmingham, Birmingham, United Kingdom, ³School of Science and Technology (AASS), Örebro University, Örebro, Sweden

Keywords: robotics, responsible innovation, responsible robotics, trustworthy robotics, critical robotics, AI and society, robot ethics

Editorial on the Research Topic

Responsible Robotics: Identifying and Addressing Issues of Ethics, Fairness, Accountability, Transparency, Privacy and Employment

1 RESPONSIBLE AI AND ROBOTICS

Recent work in both academia, industry, and journalism has brought widespread attention to various kinds of harmful impact that AI can have on society. These are very often concentrated on marginalized social groups. AI algorithms may unintentionally reinforce social prejudice Bolukbasi et al. (2016) and biased conceptions of gender Adams and Loideáin (2019); Hamidi et al. (2018), race Sweeney (2013), age Rosales and Fernández-Ardávol (2019) or disabilities Guo et al. (2020), they may lead to unfair access to opportunities Dastin (2018); Angwin et al. (2016), discriminatory pricing practices Bar-Gill (2019); Hannak et al. (2014), etc. Recent work has also shown that many seemingly technical issues in machine learning are actually socio-technical. For example: the over-fitting of machine learning models, the choice of dataset or learning objective, and other aspects of learning may lead to algorithms performing poorly on unrepresented or unmodeled groups of people Brandao (2019); Barocas et al. (2019); Buolamwini and Gebu (2018). A growing community of Fairness, Accountability, Transparency, and Ethics of AI¹ is now approaching these Research Topic from a socio-technical point-of-view, in order to identify, understand, and alleviate such issues.

Robotics, as a technology focused on automation and intelligent behavior, also abounds in similar ethical and social issues that need to be identified, characterized, and considered in design. While many of the same problems with AI will also be present in robotics, the physical nature of robotics raises new aspects of the social and ethical nature of these technologies. As one example: models that are considerably less accurate on certain groups of people can lead to physical safety differentials Brandao (2019), where robots or autonomous vehicles using those models are more likely to collide with those groups. Additionally, there are physical safety concerns with respect to surgical and other medical robots Yang et al. (2017); Ficuciello et al. (2019), as well as concerns of physical and political security—not least concerning autonomous weapon systems and the dual-use of robot technologies like autonomous cars and drones Brundage et al. (2018); Sparrow (2007).

The physical design and visual appearance of robots also introduce new aspects to responsible development. For example, people's moral evaluation of robot decisions can be affected by whether the robot is more or less human-like Malle and Scheutz (2016), the design of robots in a care setting

OPEN ACCESS

Edited and reviewed by:

Bertram F. Malle,
Brown University, United States

*Correspondence:

Martim Brandão
martim.brandao@kcl.ac.uk

Specialty section:

This article was submitted to
Ethics in Robotics and Artificial
Intelligence,
a section of the journal
Frontiers in Robotics and AI

Received: 06 May 2022

Accepted: 30 May 2022

Published: 21 June 2022

Citation:

Brandão M, Mansouri M and
Magnusson M (2022) Editorial:
Responsible Robotics.
Front. Robot. AI 9:937612.
doi: 10.3389/frobt.2022.937612

¹Example venues: ACM Conference on Fairness, Accountability, and Transparency (FAccT), AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES).

affects caregivers and caretakers van Wynsberghe (2021); Kubota et al. (2021), the choice of sensors, measurements and motion has an impact of privacy Calo (2011); Eick and Antón (2020); Luo et al. (2020), and the ethics of deception takes on new shape Danaher (2020).

The robotics community has been discussing ethics for long². Recent workshops have also started bringing attention to philosophical problems in robotics³ and issues such as bias⁴ and transparency⁵. These efforts share a common goal of developing robotics technologies responsibly—they are part of “Responsible Robotics” or “Trustworthy Robotics.”

A similar effort on “Critical Robotics” Serholt et al. (2021) has focused on questioning current practices in robotics research. These range from how older adults are represented in HRI Burema (2021) and ethical issues in education robots Serholt et al. (2017), to normative dimensions of speech used by researchers Brandao (2021), their technological optimism Šabanović (2010) and the influence of their social background in research directions Forsythe (2001); Šabanović (2010).

2 THIS RESEARCH TOPIC

This Research Topic gathers a diverse set of articles on Responsible Robotics. They range from user studies and philosophical inquiry, to modeling, algorithmic, and governance methods. Our goal when organizing this Research Topic was exactly to join various approaches in a single edition—to allow for greater multidisciplinary exchange under the common mission of Responsible Robotics. We believe that Responsible Robotics should focus both on *identifying* social and ethical issues, and on *designing* methods to account for (and alleviate) such issues—thus the focus of this edition on both understanding and *acting* on social and ethical issues.

Two articles in the Research Topic are focused on eliciting social and ethical issues *from users and stakeholders*. Lutz and Tamò-Larrioux investigate privacy concerns of lay users and their impact on technology use intentions, when using social robots that are either privacy-friendly or privacy-invasive (e.g., listen to conversations, share data with third parties). Colombino et al. use ethnographic studies, interviews and futuristic autobiographies to identify organizational principles, potential roles, and ethical design considerations for a robot that collaborates with disabled employees.

Three articles are more focused on methods, or socio-technical solutions to ethical problems in robotics. Webb et al., for example, focus on methods for conducting investigations of accidents involving humans and robots. In particular, they propose and

preliminarily evaluate a role-play-based methodology for investigating accidents, and to evaluate the testimonies that humans can give in forensic investigations of such accidents. Hurtado et al. focus on issues of harmful social bias in robot learning and how they could be detected and alleviated. Namely, they show through various examples how social robot navigation techniques that mimic human behavior may lead to harmful behavior, such as higher intrusion of personal space or longer waiting times for some groups compared to others. Winfield et al. focus on issues of transparency from a governance perspective. They describe a new draft standard on transparency for autonomous systems, with several contributions such as transparency levels, measurability, stakeholders, and example-based guidance on using the draft standard.

We then dive into philosophical inquiry and frameworks for robot ethics. Rhim et al. combine work in moral philosophy and psychology to propose a model that explains human decision-making in moral dilemmas involving autonomous vehicles. Pirni et al. consider aspects of autonomy and vulnerability in the ethics of designing care robots. And Kuipers argues that AI and robotics technologies rely heavily on over-simplified models, and that the widespread use of such models can lead to the erosion of trust and cooperation effectiveness. The article can serve as an argument for why more attention should be given to the *modeling* of complex socio-technical factors in AI/robotics.

Finally, two articles in the Research Topic dive into issues of jobs and economics in robotics and automation. Studley argues that we should consider how robotics impacts global supply chains, international development, and global economic disparities. Kyvik Nordås and Klügl then use modeling to understand the uptake of automation technologies and its relationship with unemployment and engineering, consultancy, and manufacturing jobs. The authors use this analysis to suggest an automation policy focus on user costs and education.

We believe that the contributions collected in this Research Topic can be relevant to roboticists, AI practitioners, policy makers and any other stakeholders concerned with the societal impacts of AI and robotics. We hope this Research Topic will stimulate future work on responsible robotics.

We end with an important remark. While the abundance of social and ethical issues raised in this editorial and this Research Topic might feel overwhelming or hopeless, we believe the opposite is the case. Responsible Robotics is about clearly identifying potential issues, because by doing so it is also possible to work towards responsible methods that mitigate them. This ultimately facilitates the application of robotics and AI in ways that increase safety, efficiency, and wellbeing in many areas of life: transportation, healthcare, work life, just to name a few.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

²ICRA 2007/2009/2011 workshops on Roboethics, ICRA 2014 workshop on “Robotics and Military Applications”.

³Robophilosophy Conference.

⁴ICRA 2019 workshops on “Bias-sensitizing robot behaviours” and “Unlearning biases in robot design”.

⁵HRI 2022 workshop on “Fairness and Transparency in HRI,” ICRA 2020 workshop “Against robot dystopias”.

REFERENCES

- Adams, R., and Loideáin, N. N. (2019). Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: the Role of International Human Rights Law. *Camb. Int. Law J.* 8, 241–257. doi:10.4337/cilj.2019.02.04
- Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). *Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks*. Wilmette, IL: Benton Institute for Broadband & Society.
- Bar-Gill, O. (2019). *Algorithmic Price Discrimination when Demand Is a Function of Both Preferences and (Mis) Perceptions*. Chicago, Illinois: University of Chicago Law Review, 86.
- Barocas, S., Hardt, M., and Narayanan, A. (2019). Fairness and Machine Learning (fairmlbook.org). Available at: <http://www.fairmlbook.org>.
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings. *Adv. neural Inf. Process. Syst.* 29.
- Brandao, M. (2019). "Age and Gender Bias in Pedestrian Detection Algorithms," in Workshop on Fairness Accountability Transparency and Ethics in Computer Vision, CVPR.
- Brandao, M. (2021). "Normative Roboticians: the Visions and Values of Technical Robotics Papers," in IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 671–677. doi:10.1109/RO-MAN50785.2021.9515504
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *arXiv Prepr. arXiv:1802.07228*.
- Buolamwini, J., and Gebru, T. (2018). "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Vol. 81 of *Proceedings of Machine Learning Research*. Editors S. A. Friedler and C. Wilson (New York, NY, USA: PMLR), 77–91.
- Burema, D. (2021). A Critical Analysis of the Representations of Older Adults in the Field of Human–Robot Interaction. *AI Soc.* 2021, 1–11.
- Calo, R. (2011). "Robots and Privacy," in *Robot Ethics: The Ethical and Social Implications of Robotics*.
- Danaher, J. (2020). Robot Betrayal: a Guide to the Ethics of Robotic Deception. *Ethics Inf. Technol.* 22, 117–128. doi:10.1007/s10676-019-09520-3
- Dastin, J. (2018). "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," in *Ethics of Data and Analytics* (Boca Raton, FL: Auerbach Publications), 296–299.
- Eick, S., and Anton, A. I. (2020). "Enhancing Privacy in Robotics via Judicious Sensor Selection," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 7156–7165. doi:10.1109/ICRA40945.2020.9196983
- Ficuciello, F., Tamburrini, G., Arezzo, A., Villani, L., and Siciliano, B. (2019). Autonomy in Surgical Robots and its Meaningful Human Control. *Paladyn, J. Behav. Robotics* 10, 30–43. doi:10.1515/pjbr-2019-0002
- Forsythe, D. (2001). *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Redwood City, California: Stanford University Press.
- Guo, A., Kamar, E., Vaughan, J. W., Wallach, H., and Morris, M. R. (2020). Toward Fairness in AI for People with Disabilities SBG@a Research Roadmap. *SIGACCESS Access. Comput.* 2020, 1. doi:10.1145/3386296.3386298
- Hamidi, F., Scheuerman, M. K., and Branham, S. M. (2018). "Gender Recognition or Gender Reductionism? the Social Implications of Embedded Gender Recognition Systems," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–13.
- Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. (2014). "Measuring Price Discrimination and Steering on E-Commerce Web Sites," in *Proceedings of the 2014 conference on internet measurement conference*, 305–318. doi:10.1145/2663716.2663744
- Kubota, A., Pourebadi, M., Banh, S., Kim, S., and Riek, L. (2021). Somebody that I Used to Know: The Risks of Personalizing Robots for Dementia Care. *Proc. We Robot*.
- Luo, Y., Yu, Y., Jin, Z., Li, Y., Ding, Z., Zhou, Y., et al. (2020). "Privacy-aware UAV Flights through Self-Configuring Motion Planning," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 1169–1175. doi:10.1109/ICRA40945.2020.9197564
- Malle, B. F., and Scheutz, M. (2016). "Inevitable Psychological Mechanisms Triggered by Robot Appearance: Morality Included?," in 2016 AAAI Spring Symposium Series.
- Rosales, A., and Fernández-Ardèvol, M. (2019). Structural Ageism in Big Data Approaches. *Nord. Rev.* 40, 51–64. doi:10.2478/nor-2019-0013
- Šabanović, S. (2010). Robots in Society, Society in Robots. *Int. J. Soc. Robotics* 2, 439–450.
- Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., et al. (2017). The Case of Classroom Robots: Teachers' Deliberations on the Ethical Tensions. *AI Soc.* 32, 613–631. doi:10.1007/s00146-016-0667-2
- Serholt, S., Ljungblad, S., and Ni Bhroin, N. (2021). Introduction: Special Issue—Critical Robotics Research. *AI Soc.* 2021, 1–7.
- Sparrow, R. (2007). Killer Robots. *J. Appl. Philos.* 24, 62–77. doi:10.1111/j.1468-5930.2007.00346.x
- Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Commun. ACM* 56, 44–54. doi:10.1145/2447976.2447990
- van Wynsberghe, A. (2021). Social Robots and the Risks to Reciprocity. *AI Soc.* 2021, 1–7. doi:10.1007/s00146-021-01207-y
- Yang, G.-Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., et al. (2017). Medical Robotics-Regulatory, Ethical, and Legal Considerations for Increasing Levels of Autonomy. *Sci. Robot.* 2, eaam8638. doi:10.1126/scirobotics.aam8638

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Brandão, Mansouri and Magnusson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.