This article has been published and should be cited as:

Šabanović, S., Charisi, V., Belpaeme, T., Bethel, C. L., Matarić, M., Murphy, R., & Levy-Tzedek, S. (2023). "Robots for good": Ten defining questions. *Science Robotics*, *8*(84), eadl4238.

This is the authors version; the published version can be accessed at: https://www.science.org/doi/10.1126/scirobotics.adl4238

Title "Robots for Good": Ten Defining Questions

Authors: Selma Šabanović* (Indiana University Blooomington), Vicky Charisi* (European Commission Joint Research Center), Tony Belpaeme (Universiteit Gent), Cindy L. Bethel (Mississippi State University), Maja Matarić (University of Southern California), Robin Murphy (Texas A&M University), Shelly Levy-Tzedek* (Ben-Gurion University of the Negev)

* contributed equally to the work

1-line abstract: We suggest 10 questions to guide reflection and assessment of the "good" in robotics projects

Paper:
In July 2023, the UN's AI for Good Global Summit displayed over 50 robots to members of the public, UN staff and policy makers, and the media. Along with invited speakers from academia, industry, and government, these robots were meant to demonstrate how emerging technologies can benefit society. As robotics researchers participating in this meeting, we were struck by tensions in how social "good" was defined in relation to robotic technology [1]. In particular, we noted several humanoid robots presented as imminent Artificial General Intelligence solutions to the rising need for eldercare, supporting creativity, companionship, service, and assistance. Robots that were being partially remote controlled were described as autonomous; some developers shifted responsibility for answering questions about the ethical impacts of their projects to the robots themselves. The resulting discourse around these robots did not provide an honest and clear vision of what robots can really do or how they could, realistically and responsibly, lead toward societal "good".

This led us to question what "good" means in the context of robotics. Some robotic applications might seem inherently good for society, such as disaster relief, educational, healthcare, and assistive robots. All sorts of robotics research and development can, however, be "laundered" and cast as "good" by tenuously relating it to these applications. We need to be much more specific and explicit about how robotics will lead to measurable benefits and for whom before we call it “good” for society. We first need to operationally define what “good” is, what values we embed in such technology, and how we can be transparent in the choices we make. We need to consider how we are communicating the values, choices, and expected “good” outcomes related to robots to different audiences, taking into account their technology literacy, lived experiences, and expectations. We acknowledge that when we mention “good”, we are speaking about several “goods” – to the user, to the public, to science, to economics, to national security, among others – that might be complementary or in conflict with one another.

We present a (non-exhaustive) series of questions that every “robotics-for-good” project should return to repeatedly throughout its development, and that the broader community (you too, Reviewer 2) can use to critically consider projects which claim they are for the societal “good.” These are also questions that robotics projects should explicitly address in the communication of their work. We caution that these questions are meant to help robotics developers be more reflective of the strengths and limitations of their work, and are not meant as a checklist for certifying “good”.

How did you come up with a definition for “good” in your project, and what specific societal “goods” will it lead to? Your project might be motivated by an evidenced-based demand pull from the target community, by an “innovation push” from foundational scientific advances, or by other motives.

Who is this actually good for? Identify the explicit end-users that your robot can assist, such as affluent older adults who can afford the technology, as well as others who stand to benefit from the project implicitly, such as developers or employers. Also explain **in what specific ways is the project good for these populations?** For example, the robot might aim to enable functional improvement of motor/social skills in users [2], or lead to improved work efficiency in an organization.

Who is this project not good for, and in what ways? Workers may experience a change in their workload when a robot is introduced, or bias toward women may be encouraged by the deployment of gendered technologies [3]. The robot may also not be accessible to certain populations due to cost, maintenance needs, or other design features [4].

How do you measure whether the outcome of the robot is good or not? Be specific about whether you need to track behavioral, functional, psychological, and social change in users, and how you will measure these changes via clinical scales, self-report questionnaires, interviews, or other methods [5].

Which stakeholders were involved in the design? State whether end-users, their caregivers, clinicians, educators [6], managers, policy makers, or other groups were involved in the design process, how they were involved, and what stages of design they were involved in, including developing initial design ideas, iterative user testing, and final evaluations.

What are the limitations of the robot – what can it not do? Be transparent about the robot's sensing and reasoning capabilities, particularly those that might be misinterpreted by users (such as displaying emotions) [7]. This will ensure that everyone has a clearer view of what the robot can achieve and how its functions need to be supported by others, including users, informal caregivers, and staff who handle and maintain the robot.

What are the potential negative consequences of the robot? Consider whether the robot's use may result in perpetuating biases [8], socially isolating users, causing psychological, emotional, or social dependence, or hidden costs for maintenance, wireless infrastructure support, training, and outages [9].

What structures need to be put in place to guard against the potential negative consequences, and by whom? It may be necessary to develop policy guidelines and regulatory frameworks that can created by the UN, IEEE, national and local governments, or other institutions, to ensure safe, inclusive, and equitable use of robots [10].

How have you protected the users against the potential harms of the robot? Consider how to address concerns regarding legal liability, or include fail-safes for the robot's operation with redundancy.

How and to what degree do you make the capabilities, limitations, potential consequences, biases of the robot transparent to specific audiences? The design and presentation of the robot may be constructed to guard against misperception. On the other hand, you may choose not to be fully transparent if you expect that to produce greater benefit to users; in that case, you should make clear why the degree of transparency you choose in your design is "good" for users and society more broadly.

We recognize robots are not inherently good or bad, which is why these questions focus not on the technology alone but on the people that develop it, their values and goals, the ways and contexts in which robots will be used, and the societal norms and regulations that need to be in place to ensure "good" (or at least better) outcomes. We recognize these can be difficult questions to answer, but it is important to start asking them regularly and openly before robots come into broader use. We want to inspire developers to engage in ethical reflection and consideration of "good" very specifically, not just as a general idea that is difficult to assess. Developers should be ready to assume responsibility through answering these questions regarding the later uses of the robots they design, and their consequences. Equally importantly, societal stakeholders should be more regularly involved in discussions about robot development to provide their knowledgeable perspectives; regulatory bodies and international institutions like the UN should insist on in-depth consideration of these tensions and how we can address them in the design and application of future robots, alongside environmental consideration of technology development. Just because we can build new robots, doesn't mean we should, if we are not clear on exactly how they will be "good."

References [up to 10]:

[1] https://www.theguardian.com/technology/2023/jul/06/humanoid-robots-gather-ai-summit-geneva

[2] Raz, D., Barkan-Slater, S., Baum-Cohen, I., Vissel, G., Lahav-Raz, Y., Shapiro, A., Levy-Tzedek, S., A novel socially assistive robotic platform for cognitive-motor exercises for individuals with Parkinson's Disease: A participatory-design study from conception to feasibility testing with end users. *Frontiers in Robotics and AI*, *10* (203).

[3] K. Winkle, D. McMillan, M. Arnelid, K. Harrison, M. Balaam, E. Johnson, E., I. Leite, Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI. *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 72-82 (2023, March).

[4] A. Demofonti, G. Carpino, L. Zollo, M. J. Johnson, Affordable robotics for upper limb stroke rehabilitation in developing countries: A systematic review. *IEEE Transactions on Medical Robotics and Bionics* **3**(1), 11-20 (2021).

[5] Y. Koren, R. Feingold Polak, S. Levy-Tzedek, Extended interviews with stroke patients over a long-term rehabilitation using human–robot or human–computer interaction, *International Journal of Social Robotics* **14**(8), 1893-1911 (2022).

[6] H. R. Lee, S. Šabanović, W. L. Chang, S. Nagata, J. Piatt, C. Bennett, D. Hakken, Steps toward participatory design of social robots: Mutual learning with older adults with depression, *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, 244-253 (2017, March).

[7] S. Turkle, *Alone together: Why we expect more from technology and less from each other*, Basic Books (2012).

[8] A. Howard, J. Borenstein, Hacking the human bias in robotics, *ACM Transactions on Human-Robot Interaction* **7**(1), 1-3 (2018).

[9] R. R. Murphy, V. B. M. Gandudi, J. Adams, A. Clendenin, J. B. Moats, Adoption of Robots for Disasters: Lessons from the Response to COVID-19, *Foundations and Trends in Robotics* **9**(2), 130-200 (2021).

[10] V. Charisi, S. Chaudron, R. Di Gioia, R. Vuorikari, M. Escobar Planas, M. J. I. Sanchez, I., E/ Gomez Gutierrez, Artificial intelligence and the rights of the child: Towards an integrated agenda for research and policy (No. JRC127564). Joint Research Centre, European Commission (2022).