# EU automation policy: Towards ethical, human-centered, trustworthy robots?

Jesse DE PAGTER[a,1]
[a] TU Wien

**Abstract:** Within the context of European Union policy-making, there has been a considerable rise in attention and resources being dedicated to policy plans concerned with (the future of) robotics. The goal of this short article is to provide commentary on such developments by providing a concise overview of the EU's policy-making activities around robotics, especially with regards to the anticipated societal effects of robotics. After that, the paper engages in a critical review of the policy-making efforts, thereby scrutinizing three of its central concepts, namely: the role of ethical approaches, the understanding of human-centered technology and finally the notion of trustworthy technology.

**Keywords:** human-centered technology, robot ethics, robotics governance, trustworthy technology

## 1. Introduction

During recent years, the governance of (future) robotics' has become an important part of the European Union's (EU) technology policy, not in the least because of robotics' increasingly widespread applications and far-reaching implications. Different voices, from politicians to academics to NGOs to journalists, have therefore been arguing that novel and revised approaches towards robotics governance are crucial in order to keep up with this transition (see e.g.: Koops et al., 2013; Nübler, 2016; Woo, 2014). In light of such challenges, different institutions and bodies of the EU have been making efforts to establish new European-wide robotics strategies. In this paper, the most prominent policy-making developments with respect to robotics are described. After that, a critical perspective is developed by providing insights into the EU's recent focus on ethics, trustworthiness and human-centered approaches in it's robotics strategy. The issue in that regard is to question the ways in which those higher-level concepts can be deployed and improved in order to develop post-automation practices based on democratic

1 Jesse de Pagter, Institute for Management Science, TU Wien, Theresianumgasse 27, 1040 Vienna, Austria. E-mail: jesse.de.pagter@tuwien.ac.at

deliberation, as they are described in the paper to which this piece is a complementary item (Ionescu & De Pagter, 2022).

## 2. Robotics in the EU: towards ethical, human-centered, trustworthy technology

To provide an overview of the main priorities in the current robotics policy of the EU, this paper covers a period of a bit less than a decade. The main focus is on the most recent years.

In 2012, the Directorate General for Information Society and Media (DG INFSO) changed its name and structure to become the Directorate-General for Communications Networks, Content and Technology (DG Connect). Since then, one of its units has been called 'Unit A.2 Robotics'. Later (in 2016), this unit became Unit 'A.1 Robotics & Artificial Intelligence' (DG Connect, 2017). Being part of the administrative body of the Commission, DG Connect is responsible for the implementation of the Commission's policy in its different areas of expertise. Furthermore it is involved in the development of long-term visions concerning the technologies that are part of its portfolio. With regards to robotics, the abovementioned unit handles the management of the robotics part of the EU's strategic research agenda, as well as the organization of so-called Public Private Partnerships (PPP's) in a robotics context. Recently the unit has become increasingly involved with the organization of discussions and research on legal and ethical issues with regards to robotics and AI (Bajart and DG Connect, 2017).

This focus on legal and ethical issues represents an important development in the thinking about the future of robotics, as they are a central instrument to curb the expected societal implications following the (anticipated) development of increasingly autonomous types of robots in different areas. An interesting and important moment in this regard was the legislative resolution of 17 February 2017 by the European Parliament called the 'European Civil Law Rules in Robotics'. Within this resolution, the Parliament proposed a set of civil law rules, addressing the following issues: civil use of robots; robotics research and innovation; ethical principles for robotics; the need for a dedicated European Agency for robotics and AI; intellectual property rights; standardization, safety and security; autonomous means of transport; issues with care robots; issues with medical robots; issues of human repair and enhancement; issues relating robotics to future education and employment; the environmental impact of robotics; liability issues concerning (future) robots; and finally the international (geopolitical) impact of robotics. As becomes visible through the content of this resolution, the Parliament's goal is to maintain an EU-agenda that is as comprehensive as possible, thereby aiming for an approach that is

“maximising benefit and minimising harm” (European Parliament, 2017).

Roughly simultaneous to those developments, a High-Level Expert Group on Artificial Intelligence (AI HLEG) was called into existence by the Commission (European Commission, 2018a). Even though its name indicates a focus on Artificial Intelligence, the development of robotics is central to the group's considerations. This connection between AI and robotics is understandable from a perspective of policy-making, since it is the fusion of robotics- and AI-technology that is an important factor behind the emergence of embodied, autonomous, intelligent systems. As such, this development is expected to bring about drastic technological and socioeconomic changes in the near future. The AI HLEG has published two important deliverables, namely (1) ‘Ethics Guidelines on Artificial Intelligence’ in which 7 key requirements are listed that should define the human-centric approach (AIHLEG, 2019a). Partly based on those requirements, a second document has been developed: (2) ‘Policy and Investment Recommendations’ in which the group emphasizes the importance of "trustworthy AI" (AIHLEG, 2019b). Next to the setting up of this body of expertise, a ‘European approach to Artificial Intelligence and Robotics’ was put forward in April 2018, as well as a communication concerning an AI strategy towards ‘Building Trust in Human-Centric Artificial Intelligence’ in 2019 (European Commission, 2019, 2018b). Its development is quite closely related to the AI HLEG, since the latter's deliverables should stand at the basis of this approach.

Finally, under the new Von der Leyen Commission (in place since 1 December 2019), several statements and plans have been published emphasizing the priority of those issues. The general direction is one which should ensure AI and robotics that is “human-centric and trustworthy" (European Commission, 2021, p. 5). An important goal is to set up a "regulatory framework to ensure trust in AI systems while promoting the EU’s value-based approach'' (idem, p. 9). Important thereby is to note that this trend of emphasizing the need for ethical, human-centered, trustworthy robotics in EU policy-making is not solely about adapting society to the challenges of robotics. Interestingly a narrative has been developing in recent years, connecting it to geoeconomic leadership. The notion being that the EU can be a worldwide leader in the development of ethically informed approaches that foster human centric, trustworthy robotic technologies, as is for instance explained in a ‘White Paper On Artificial Intelligence’ which describes the Commission's approach to robotics and AI (European Commission, 2020). The following section engages with this strategy critically.

## 3. Society & robots in the EU

The section above has mentioned how the focus on ethics, human-centered robotics and trustworthy robots is already initiating activities and commitments in the direction of building a more pluralist and inclusive narrative on the future of robotics. An example being for instance the composition of the above-mentioned expert group (AI HLEG) in which different approaches to the future of robotics are represented (from roboticists, to philosophers, to start-up CEOs, to trade unionists). Arguing from the point of view that such activities and commitments are enabling new discourses on the future of robotics, the goal of this paper's remainder is on the one hand to develop a critical perspective on the EU's efforts towards human-centered, trustworthy, ethical robotics, while simultaneously employing them as openings towards discussions in new directions.

*The use of ethical approaches*

In the EU's current policy-discourse and beyond, ethics are employed to play a crucial role as a safeguard against (potentially) harmful robotic technologies. Roughly two different types of application can be differentiated. First of all, ethical frameworks are (further) developed in the form of useful collections of rules and guidelines. As such, they are envisioned to stimulate and enforce a situation in which roboticists and robotics corporations develop robotic products that are kept from crossing certain ethical boundaries. A second area of application when it comes to ethics is based on the idea that ethical rules themselves can be employed as codes of conduct. As such they are to be programmed into artificial agents themselves which then should have the ability to respond to situations in an ethical manner.

Neither of those applications are without issues. In both cases, ethically defined values are as it were ''added'' to either the practice of designing robots or the design of the robots themselves, which can easily lead to checklist approaches to ethics (Kiran et al., 2015). Rather, ethical discussions would become more useful if they would function as central starting points in debates on robots which can serve as an important inspiration to the (continuous) deliberations towards technodemocratic commons within societies. In that regard, deliberations on ethical approaches are still very much desirable: if the ideas about their function changes and they can start serving as points of departure rather than outcomes of debates. In other words, they play an important role in the politics that are engaged with the way in which ethical futures are imagined. As such they can be very informative and lead to more sophisticated visions that can contribute to the emergence of a technological culture which has a stronger focus on accountability and integrity.

*The understanding of ''human-centered'' technology*

From the (academic) perspective of the social sciences and humanities in general as well as from the perspective of peer production, it is both convenient and encouraging to see an explicit pledge for human-centric technology in the EU's robotic policy. Especially because it invites for elaborations on new discourses and practices around the societal role of robotic technologies. Looking at the notion of human-centredness in the EU documents, this focus seems to be translating specifically into the credo of ''putting people at the centre'' by emphasizing general human values, such as dignity, autonomy and fairness (AIHLEG, 2019a; DG IPOL et al., 2016; von der Leyen, 2020).

Nevertheless, even though the EU's efforts are encouraging, they can easily turn into platitudes. It is in fact quite difficult to argue why current providers of intelligent digital products (e.g. big tech companies) are not human-centric. Comparable to the greenwashing phenomenon in environmental marketing, also here it is easier to make a good impression, rather than inducing actual change (Delmas and Burbano, 2011; Hao, 2020). Furthermore, a more thorough problem is that ''human-centric'' robotics and AI pre-assume most of all that clear definitions of ''the human'' are possible. Even though this human-centredness may sound like an attractive combination of maintaining control over technology while simultaneously investing into technological progress and economic growth, it actively maintains an approach that is inherently rooted in a deep appreciation and longing for a humanistic nostalgia. It is therefore important that such human-centred approaches do not turn into dogmatic debates concerning human control over technology, but are rather used to develop speculative inquiries into the role of intelligent technology towards a new future (Bratton, 2016). Especially in times of severe challenges, the humanist legacy is likely to be insufficient for the future: more imaginative approaches are needed.

*The process of defining and building trustworthiness*

Finally, trustworthiness means different things in the EU's documents. The main focus is on creating trust through regulation, legislation and standardization of robotics, while also emphasizing participation and inclusivity (AIHLEG, 2019a, pp. 21–23). Whereas this is a hopeful development, it is first of all important to note that the process of building trustworthy technology should be approached with care.

It is first of all important to analyze the definitions and concepts behind trustworthiness critically. Moreover, an increase in emphasis on narratives around societal trust in robotics would be desirable. This means that not only the trust of individuals towards robotic

objects is taken into account, but also that society-wide narratives are included in the understanding of robots' (as artifacts) and robotics' (as research area and industry) trustworthiness. Furthermore, an important strategy to scrutinize the EU approach would in this regard be to develop a strong focus on the different sociotechnical imaginaries of robotic futures in the EU, thereby opening up space for new commons to evolve around robotics. In this regard, engagement in those imaginaries can help to understand the roles of emerging technologies in the societies of the future, as well as help to inquire the ways their different futures are incoherent with each other. Thus, it is important that speculative futures concerning emerging technologies are to be taken seriously and engaged with, thereby recognizing the sociotechnical potential of robotics.

## 4. Conclusion

While the EU's current approach towards robotics is defined by a rather strong focus on the development of high-level principles, an important element is crucial in making its approach a success: the fostering of a new technological culture in which the visions regarding robotics are based on plural understandings of this future. As the main paper by Ionescu & De Pagter (2022) shows, abstract values such as ''democratization'' of technology are often complex to be applied in technological practices. Therefore, in order to make sure that ''ethical'', ''human-centred'' and ''trustworthy'' technologies are not only functioning as buzzwords, an important task of the peer production community is to push the notion of a pluralist technological culture (Bensaude Vincent, 2014). A culture in which speculative attempts towards democratic technologies and practices will be grounded firmly. In this way robotics could potentially become enablers of new forms of equality and equity in the future, rather than further endangering that future.

### References


AIHLEG, 2019a. Ethics Guidelines for Trustworthy AI. Accessed 11.4.20. Available at: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

AIHLEG, 2019b. Policy and Investment Recommendations for Trustworthy Artificial Intelligence. Accessed 11.4.20. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60343

Bajart, A., DG Connect, 2017. Artificial Intelligence activities. Accessed 4.12.20. Available at: https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/6%20Overview%20of%20current%20action

%20Connect.pdf
Bensaude Vincent, B., 2014. The politics of buzzwords at the interface of technoscience, market and society: The case of 'public engagement in science.' Public Underst Sci 23, 238–253. https://doi.org/10.1177/0963662513515371
Bratton, B., 2016. On Speculative Design Accessed 3.1.18. Available at: http://dismagazine.com/discussion/81971/on-speculative-design-benjamin-h-bratton/
Delmas, M.A., Burbano, V.C., 2011. The Drivers of Greenwashing. California Management Review 54, 64–87. https://doi.org/10.1525/cmr.2011.54.1.64
DG Connect, 2017. Robotics and Artificial Intelligence (Unit A.1) Accessed 12.4.20. Available at: https://ec.europa.eu/digital-single-market/en/content/robotics-and-artificial-intelligence-unit-a1
DG IPOL, Nevejans, N., European Parliament, Policy Department C: Citizens' Rights and Constitutional Affairs, European Parliament, Committee on Legal Affairs, 2016. European civil law rules in robotics: study. Accessed 13.1.20. Available at: https://op.europa.eu/en/publication-detail/-/publication/19ea0f1c-9ab0-11e6-868c-01aa75ed71a1
European Commission, 2021. Communication: Fostering a European approach to Artificial Intelligence. Accessed 21.4.21. Available at: https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2021)205&lang=en
European Commission, 2020. White paper on artificial intelligence – a European approach to excellence and trust. Accessed 11.10.20. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:0065:FIN
European Commission, 2019. Communication: Building Trust in Human Centric Artificial Intelligence. Accessed 11.10.20. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0168&qid=1630917561643
European Commission, 2018a. High-Level Expert Group on Artificial Intelligence. Accessed 11.4.20. Available at: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
European Commission, 2018b. A European approach to artificial intelligence. Accessed 11.4.20. Available at: https://ec.europa.eu/digital-single-market/en/artificial-intelligence
European Parliament, 2017. Resolution: Civil Law Rules on Robotics. Accessed 5.1.20. Available at: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html
Hao, K., 2020. We read the paper that forced Timnit Gebru out of Google. Here's what it says. Accessed 12.5.20. Available at:

https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/
Ionescu, T.B., De Pagter, J., 2022. “Meet Your Personal Cobot! Will it Change You? Can You Resist It?” Framing Participatory Research in Makerspaces as a Trading Zone. Journal of Peer Production 15.
Kiran, A.H., Oudshoorn, N., Verbeek, P.-P., 2015. Beyond checklists: toward an ethical-constructive technology assessment. Journal of Responsible Innovation 2, 5–19. https://doi.org/10.1080/23299460.2014.992769
Koops, B.-J., Di Carlo, A., Nocco, L., Casamassima, V., Stradella, E., 2013. Robotic Technologies and Fundamental Rights: Robotics Challenging the European Constitutional Framework. International Journal of Technoethics 4, 15–35. https://doi.org/10.4018/jte.2013070102
Nübler, I., 2016. New technologies: A jobless future or golden age of job creation. International Labour Office Research Department Working Paper 13, i–25.
von der Leyen, U., 2020. State of the Union Address by President von der Leyen. Accessed 4.12.20. Available at: https://ec.europa.eu/commission/presscorner/detail/en/speech_20_1655
Woo, M., 2014. Robots: Can we trust them with our privacy? Accessed 13.7.20. Available at: https://www.bbc.com/future/article/20140605-the-greatest-threat-of-robots