International Journal of Social Robotics
https://doi.org/10.1007/s12369-023-00982-6

From EU Robotics and AI Governance to HRI Research: Implementing the Ethics Narrative

Jesse de Pagter[1]

Accepted: 15 February 2023

Abstract
In recent years, the European Union has made considerable efforts to develop dedicated strategies and policies for the governance of robotics and AI. An important component of the EU's approach is its emphasis on the need to mitigate the potential societal impacts of the expected rise in the interactive capacities of autonomous systems. In the quest to define and implement new policies addressing this issue, ethical notions have taken an increasingly central position. This paper presents a concise overview of the integration of this ethics narrative in the EU's policy plans. It demonstrates how the ethics narrative aids the definition of policy issues and the establishment of new policy ideas. Crucially, in this context, robotics and AI are explicitly understood as emerging technologies. This implies many ambiguities about their actual future impact, which in turn results in uncertainty regarding effective implementation of policies that draw on the ethics narrative. In an effort to develop clearer pathways towards the further development of ethical notions in AI and robotics governance, this paper understands human-robot interaction (HRI) research as a field that can play an important role in the implementation of ethics. Four different complementary pathways towards ethics integration in (HRI) research are proposed, namely: providing insights for the improvement of ethical assessment, further research into the moral competence of artificial agents, engage in value-based design and implementation of robots, and participation in discussions on building ethical sociotechnical systems around robots.

Keywords European Union · Human-Robot Interaction · Robot and AI Ethics · Robotics and AI Governance · Trust in Robots

1 Introduction

Autonomous systems with interactive capabilities are currently encountering a wide range of expectations regarding their anticipated future impact. Examples can be found in expressions of both excitement and fear about these systems' autonomy, their potentially deceptive anthropomorphism, their impact on privacy and so on. As a result of these expectations, there are currently many concerns about the potential socially disruptive impacts of increasing interaction between humans and autonomous systems. In order to mitigate these concerns, there have been growing deliberations on the use of ethical approaches in the governance of robotics and AI. A prominent example in this context is the European Union (EU), where the anticipatory governance of robotics and AI has been on the agenda for almost a decade. In recent years, the European Commission has expressed its aspirations to maintain a strong emphasis on ethical approaches in its policy-making and strategy development around robotics and AI technologies. The EU's stance is currently characterized by an emphasis on the need to build trust in autonomous systems while promoting a human-centred technology development trajectory that is grounded in ethical reasoning. As such, the EU has generally occupied a leading role in the worldwide discussion on the governance of autonomous systems in general [1].

The aim of this paper is to provide critical but constructive insight into the current discursive frame within which EU robotics and AI policy-making plays out in terms of this ethics narrative. While scrutinizing how this narrative has emerged in EU robotics and AI policy-making, the paper looks critically at the ways in which the EU's policy plans

✉ Jesse de Pagter
jesse.de.pagter@tuwien.ac.at

[1] Institute for Management Science, TU Wien, Theresianumgasse 27, Vienna, Austria

Published online: 07 March 2023

can become manifested in research and development. Specifically, I focus on human-robot interaction (HRI) research, as a crucial field for studying interactions with embodied autonomous systems. In short, the main point of the paper is that on the one hand, policy-making narratives can be instrumental for achieving systemic changes towards building more inclusive, trustworthy technologies. However, on the other hand, in order to achieve this change there is a clear need to invest in further ethics implementation in HRI and social robotics research. In this way, the ethics narrative can both be deepened and broadened. By developing this argument, the paper focuses on the interface between robot ethics, robotics governance and HRI research. As such, the main underlying challenge of this paper is to draw connections between two different components: on the one hand, the portrayal and understanding of robotics and AI technologies as (future) objects of EU ethical governance, and on the other hand, the implementation of ethical deliberations in HRI research.

The outline of the paper is as follows: Sect. 2 provides a short overview of current developments regarding the role of the ethics narrative in the EU robotics and AI strategy. This is done by providing insight into the main themes from a range of different policy documents published by the EU in recent years. Section 3 provides a critical reflection by questioning the ethics narrative's uptake in policy-making discourse, while providing more context regarding its history in the anticipatory governance of emerging technologies. Based on that, the main argument of this section is developed, namely that it is important to see the ethics narrative as a shared effort between policy-makers and technical experts (e.g. HRI researchers). Departing from that argument, Sect. 4 provides four different pathways that can help to establish further ethics implementation in robotics and HRI research. To wrap up, the conclusion provides final remarks about the further development of ethical approaches and their role in shaping the future of a society with interactive autonomous systems.

2 The Ethics Narrative in EU Robotics & AI Policy-Making

As indicated above, this paper analyzes the rise of ethics in EU policy-making using a narrative approach. First of all, this type of analysis posits that these policy-making processes develop against a background of ongoing negotiations and disputes [2]. In this setting, narratives can be considered a driving factor behind the arguments and justifications used to establish newly developed policy plans. Tracing the emergence of a narrative framework is a common method of analysis in theories of policy change in fields such as political science, public administration and governance studies [3–5]. As such, it provides insight into the way something emerges on the agenda as a distinguishable policy problem that can be monitored and controlled via policy-making efforts [6]. Thus, the conceptualization of policy discourses as narratives and their subsequent analysis helps to describe how policy-making on a specific topic develops over time, which is also the goal of this section.

Narratives that drive the definition of new policy ideas can evolve considerably, especially when policy plans develop quickly. In terms of empirical access to these narratives, this can sometimes be a rather fuzzy process characterized by a state of continuous development. Particularly, important to consider in this regard is that the directions in which the definitions of policy problems and policy ideas evolve can be very much influenced by stakeholders in the policy-making process, such as lobbyists, (academic) experts and so on [7, 8]. In other words, EU robotics and AI governance in its current state is still in a process of configuration. Among other things, this means that many specificities of the policies are explicitly understood as not yet fully developed. Furthermore, it also means that during this process of configuration, certain (potential) characteristics and impacts of current and future autonomous systems are becoming defined as specific policy problems. Looking at policy-making in such a manner has consequences for this paper: rather than taking statements about ethics at face value and analyzing what their consequences are, the very notion and use of ethics itself is explicitly analyzed as a policy-making narrative that is under development.

The present section provides insight into the rise of ethical notions in EU robotics and AI policy through a narrative analysis of policy documents, including both foresight reports from expert bodies as well as documents that define official EU policy and legislation. Whereas the former are usually based on expert statements and do not represent an official policy position of the European Parliament or European Commission, they have nevertheless been included since they provide important insights into the underlying argumentation behind the EU's policy narratives. For the first two subsections below, these documents were selected by going through the EU document databases using robo* as the main search term, while also selectively going through results based on the search term automat* in order to make sure documents that have a close relation to the topic of robotization were not excluded. For the latter two subsections, it is important to note that policy-making itself shifted towards paying increased attention to the impact of large-scale autonomous systems. An important consequence of this was that the EU's governance plans were increasingly defined based on large-scale AI strategy plans. From this period, all documents from the European Commission and

European Parliament defining or altering this strategy have been included. Important to realize in this case is that even though the EU has increasingly started to use the term "AI" to refer to its robotics and AI strategy, the focus of the paper in general is still on robotic technologies, since the latter sections focus on the role of HRI research. The resulting insights describe how the EU represents the issues, goals and instruments of (future) robotics and AI governance. The method was designed to trace the different elements that make up the different potential issues, goals and instruments of governance.

2.1 Fostering Robotics Development and Managing its Economic Impact

In terms of robotics governance, policy-makers' interest in these technologies is certainly not new. In fact, there are several regulatory areas in which specific issues with respect to robotic technologies have been regulated for many years. Furthermore, predecessors to current robotic technologies have been around for more than half a century, and as such, they have been part of the EU's economic, industrial, and research policies in recent decades. In order to improve the competitive situation and foster consolidated approaches to technological and economic growth, the EU has made multiple efforts to develop and support specific strategic initiatives. For instance, in 2000, the European Robotics Research Network (EURON) was founded in order to encourage and promote research, education and technology transfer in the field of robotics [9]. Interestingly, attempts to include robot ethics in robotics policy were also made here, based on Isaac Asimov's Three Laws of Robotics and mainly aiming to develop a taxonomy of potential ethical issues. Even though the content is not yet defined in much detail, many issues are already present, such as reliability and predictability, traceability of robot actions as well as issues around safety and security [10]. Furthermore, in 2005, the European Robotics Platform (EUROP) was founded and functioned as an industry-driven initiative to strengthen the EU's competitive position in robotics research and development [11].

Furthermore, in terms of economic and societal impact, the increasing uptake of robots has long been a topic of EU (macro-)economic policy-making. In a range of foresight documents from the last decade, different EU institutions and bodies have taken up various broader themes in relation to robots and their potential impact on society [12–15]. Such foresight studies generally maintain a strong focus on how new types of (mostly industrial) robots could be deployed in the near future, while also paying attention to the macroeconomic effects of this increase in automation. Foresight concerning these topics mostly focus on issues like the potential effects of robots on employment, while simultaneously emphasizing the positive potential of robotization for economic growth. In this way, a narrative is shaped which establishes (industrial) robots as potentially problematic in terms of socioeconomic impact, while simultaneously mitigating this impact because of the expected economic growth, which has the potential to foster a general increase in welfare and employment. Furthermore, such studies often emphasize topics like potential re-shoring and the general strengthening of the EU's industrial sector as an important effect of the increased implementation of robots.

2.2 The Ethics of Autonomous Systems as a Topic of Concern

Whereas the impact of robotics innovations has long been a theme in different types of governance, the explicit rise in attention paid to ethical issues surrounding robotics and AI in the EU policy-making context can be seen as more recent (even though it has been around for much longer in academic circles). Attention to ethical themes in this realm of EU policy-making was rather marginal before 2015, but has since evolved to become much more prominent. An important development in that regard are several reports by expert bodies of the European Parliament and European Commission. For instance, in 2016, the European Parliament's Panel for the Future of Science and Technology (*STOA*) issued a report titled 'Ethical Aspects of Cyber-Physical Systems' [16]. This report identifies in detail the impacts of and issues with cyber-physical systems in general, while also paying substantial attention to various types of robots in a range of areas of applications. In terms of what this study understands to be important ethical issues, the most prominent are concerns about privacy infringement, questions regarding liability and accountability with respect to autonomous machines, safety issues and health issues. Ethical issues are thus framed in terms of clearly definable problems that require specific legislative solutions.

A pivotal development for the further definition of robots and AI as a potential policy problem was when a new consensus started to take shape around potentially unprecedented forms of artificial agency. A general trigger moment for the discussion regarding the position of interactive robots (and other artificial agents) in EU policy-making was the European Parliament resolution of 16 February 2017 entitled 'Civil Law Rules on Robotics' [17]. The ideas expressed in this resolution had been informed by studies from the Parliament's Scientific Foresight Unit, the European Parliament Research Service and the Committee on Legal Affairs [16, 18]. Generally, in the EP resolution, there is a strong focus on the expected impacts of AI and AI-powered robots in society, while autonomous features of these systems play a central role in the way they are problematized. It is in

this context that the use of ethical approaches is mentioned abundantly and with respect to numerous applications. First of all, ethical principles are mentioned as constitutive components of guidelines that restrict both the development and use of (AI-powered) robots. This is seen as relevant for a range of different ethical issues, such as preventing potential harm to humans, protection of human liberty as well as privacy, and the risk of manipulation. Interestingly, the resolution also maintains an emphasis on issues related to accessibility and equity, such as ensuring right to care or general access to technological progress in the form of robotics. The resolution also envisions an important role for ethical committees that assess robotics and AI research. In the resolution itself, these ethical principles are listed, along with other principles and objectives, such as research and innovation goals, standardization, safety and security.

Furthermore, to provide more insight into the Parliament's line of argumentation, it is worth considering one short, but widely debated paragraph in this resolution. This paragraph caused a considerable stir, as it expressed the possibility of establishing forms of 'electronic personality' while arguing for 'creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently' [17]. This quote demonstrates how the Parliament anticipates that technological developments could lead to a situation in which new types of autonomous robots become engaged in (social) interactions. According to the proposal, this will impact the legal interpretation of autonomous robots' actions and liability and render the notion of electronic personhood necessary in order to deal with the legal ramifications. This paragraph caused quite a controversy: in 2018, an open letter was published by 156 AI and robotics experts from different European countries, rejecting the recommendation of a legal status for robots [19]. They mainly argued that the proposal overestimates actual technical developments in robotics, while also arguing that such a status for robots would be inappropriate from a legal and ethical point of view.

Apart from the discussion over the legal (and moral) status of autonomous systems, the resolution by the European Parliament marked an important moment for how such systems in general and interactive robots more specifically became defined in the EU's agenda: as a specific policy problem that requires substantial attention to ethical issues that stretch beyond existing boundaries between humans and machines, especially in terms of the ethical consequences of potential robot autonomy. Roughly around the same time, the European Commission (Juncker Commission) and its affiliated institutions also started to explicitly focus on robotics and AI as policy issues that required both dedicated policies as well as a consolidated strategy for technological research and development [20, 21]. Within a few years, the process of trust-building and the role of ethics in this process became important components of policy ideas with respect to an anticipatory governance and development strategy for robotics and AI. The Commission generally focused on those issues under the banner of AI technology, but robots were part and parcel of the considerations. This can partially be explained by the increased hype and buzzwords around AI technology that arose around that time.

An important development in the further establishment of a dedicated strategy on robotics and AI was in June 2018, when the European Commission established the independent High-Level Expert Group on Artificial Intelligence (AI HLEG). This group consisted of 52 experts from a range of different backgrounds. The group's mission was to advise the European Commission concerning the potentials and challenges of autonomous systems and support the implementation of a new EU strategy on robotics and AI. Societal issues were evidently considered a prominent topic not only in the group's mission but also in its composition, since quite a number of members were recognized experts in areas related to such issues.

2.3 Towards a Coordinated EU Robotics & AI Strategy

After its establishment, the AI HLEG published several deliverables, the most important of which was the 'Ethics Guidelines for Trustworthy AI' published in 2019. The document provides four ethical principles (respect for human autonomy, prevention of harm, fairness, explicability) that form the foundation of trustworthy robotics and AI. Furthermore, it lists seven key requirements (human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being, and accountability) for the realization of trustworthy AI [22]. An important key argument in this deliverable is that trust in robotics and AI can be built through regulation, legislation and standardization efforts, while also emphasizing a need to pay increased attention to topics like participation and inclusivity. Partly based on the key requirements from the first deliverable, a second deliverable was published, called 'Policy and Investment Recommendations for Trustworthy AI' [23]. The group also provided an 'Assessment List for Trustworthy AI', which included a web-based version of the assessment list [24]. Finally, a deliverable on 'Sectoral

Considerations on the Policy and Investment Recommendations' was provided [25].

In the spirit of developing a socially sustainable, EU-wide strategy for robotics and AI research and development, the 'Coordinated Plan on Artificial Intelligence' was put forward in April 2018 [26]. It contains the initial version of a coordinated plan on the EU level. Its content is quite closely related to the ideas the AI HLEG put forward, since the latter's deliverables were supposed to serve as a foundation for the further evolution of the EU's approach. Furthermore, the plan represents an important moment in terms of deeper EU integration on autonomous systems, since it implied that most EU member states would collaborate on a common strategy and approach. While several member states were already working on national approaches, this was a step that could have important consequences for the future of a harmonized policy (i.e. common policy across the EU) on autonomous systems in the EU. In this regard, another important communication document concerning such a strategy was 'Building Trust in Human-Centric Artificial Intelligence' in April 2019 [27]. As the title suggests, the document's main aim is to further elaborate the European Commission's focus on the development and implementation of trustworthy robotics and AI. Furthermore, the European Parliament drafted another important resolution with roughly the same theme, namely to establish a "comprehensive European industrial policy on artificial intelligence and robotics" [28]. This resolution emphasizes the importance of AI and robotics as technologies that can support society, while outlining how this can be achieved through research and development and industrial policy. Also in this resolution, substantial attention is paid to potential new legal and regulatory frameworks, as well as ethics as a type of soft governance - for instance, in the form of standards as well as ethics-by-design frameworks. This goes hand in hand with a stronger focus on a use of ethics that is less focused on defining ethical issues, but more on it becoming an inherent part of engineering, design and implementation processes.

The work of the Juncker Commission has been continued under the current Von der Leyen Commission (installed in November 2019). Ursula von der Leyen states that a main goal of her new Commission is to set the standard and thus 'put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence'. She writes this in the political guidelines for the Commission under her leadership, entitled 'A Union that Strives for More: My Agenda for Europe' [29]. Furthermore, a white paper entitled 'On Artificial Intelligence - A European approach to excellence and trust' was published by the European Commission in 2020 [30]. This white paper establishes a connection between a strong focus on building trust and the attachment to 'European values'. It states in this regard that '[g]iven the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection' [30]. Also important with regards to values and fundamental rights is a report by the EU Agency for Fundamental Rights (FRA) called 'Getting the future right – Artificial intelligence and fundamental rights' [31]. While this report focuses on AI technologies in general, it does provide important arguments and ideas in the search for ways to grapple with the human rights implications of new forms of automated decision-making.

2.4 Establishing a Common Strategy and Regulatory Framework

In April 2021, an updated version of the coordinated plan was communicated, entitled 'Fostering a European approach to Artificial Intelligence' [32]. In this longer version of the coordinated plan, the perceived need for trustworthy technology plays an important role in the way policy ideas are developed. Here as well, the ethical guidelines of the AI HLEG serve as an important inspiration. In general, the argument is that in order to achieve worldwide strategic leadership in AI and robotics, attention to trustworthiness and human-centeredness can be a key element that distinguishes the EU strategy from other global players, while further solidifying its status as a global regulatory power (for more context on this topic, see [33]).

Furthermore, also in April 2021, the European Commission came up with an important proposal for a legal framework by providing the draft for an 'AI Act', which is short for 'Regulation on Artificial Intelligence' [34]. This act lays down quite specific rules for the development, implementation and use of AI systems and can be seen as a major outcome of the activities and plans listed above. Furthermore, as is always important in an EU context, it aims for a harmonized European approach consisting of governance systems at the level of the Member States as well as cooperation mechanisms on the level of EU governance. Moreover, the terms 'trust' and 'trustworthy' are mentioned multiple times in this AI Act and the different explanatory documents attached to it. Interestingly, these ideas are also central to aspirations regarding the EU's future as a worldwide leader in the development of ethical robots.

Particularly interesting in relation to the process of building trust through ethical frameworks is the use of risk-based approaches in the AI Act. Risk plays a central role in these proposals, since the idea is to have different rules for different risk levels of AI-powered technologies. In this way, ethical risk assessment can become a way of enforcing ethical reasoning, with distinctions made between unacceptable risk, high risk and low or minimal risk. Overall, such

proposals can be seen as an important step in the notion of governing artificial agents, since the aim is to develop a harmonized set of rules that specifically target autonomous systems, while its explanations explicitly connect EU values and fundamental rights to risk assessments of those systems. While legal experts and relevant EU committees and bodies (e.g. the Parliament's IMCO and LIBE committees) are currently assessing the consequences of those developments, this already constitutes an important development in the worldwide discourse on the regulation of autonomous systems. At this moment, an important subject of discussion in the interpretation of the new proposals are statements about standard-setting [35]. In this regard, standard-setting bodies in the EU are expected to continue to play a central role in the coming years.

Finally, the emphasis on ethics has developed such that the design and engineering of interactive robots has become a topic of concern in terms of fundamental human rights. This development is currently an important subject, since, for instance, UNESCO adopted the 'Recommendation on the Ethics of Artificial Intelligence' on 24 November 2021, intended as a global standard-setting instrument [36]. With its relatively strong focus on human rights, the EU has profiled itself geopolitically as the place where robotics and AI technologies and their implementation are inherently ethical and developed in a human-centered manner. In this way, autonomous systems are on the way to becoming geopolitical objects of concern.

To summarize and conclude this overview, first of all, with regards to this section's approach, interpreting ethics reasoning as a narrative emphasizes its role as a central component of the policy process surrounding robotics and AI. As such, the ethics narrative provides specific discourses and solutions that allow the EU policy-making process to develop in a specific manner. Second, the ethics narrative can be seen as rooted in an emergent definition of autonomous systems as a topic of policy concern focused on potential ethical issues. The potential autonomy of these systems is playing an increasingly prominent role in this definition of these systems as a policy problem. Third and final, current interpretations of the AI Act include several ways in which policy ideas about ethical reasoning are becoming solidified into regulation, mostly based on identifying risk. Simultaneously, ethics is increasingly becoming part of a discourse around soft governance, which means that it is expected to have an effect on the development and implementation of robots and other types of autonomous systems.

3 From Policy Narratives to Ethical Robotics and AI

From the perspective of robot and AI ethics as an up-and-coming subfield of applied ethics, it is encouraging to see how ethical approaches have come to constitute a prominent component of policy plans for the governance of autonomous systems. This can first of all be seen as a sign that policy-makers take the mitigation of their potential impact seriously. Moreover, on a broader level, the attention to the ethical dimensions of these technologies can be constitutive of how the general culture in the EU develops in terms of its attitude towards social robots and other interactive autonomous systems. As such, the plans and strategies described above can be interpreted as a hopeful development that could potentially play an important role in initiating large-scale commitments to developing more pluralist, inclusive, equitable robotics and AI development. In short, such calls and commitments from a major player like the EU can be seen as an incentive for the further establishment of ethical governance in the research and development of robots and AI.

3.1 Problematizing the Ethics Narrative

In order to take this incentivizing function seriously, it is first and foremost important to maintain a critical stance - particularly towards the actual implications of the ethics narrative, but also with regards to the implementation of the policy plans that arise from it. AI and robot ethics might indeed have become a prominent component of the policy-making consensus. This indicates promising intentions, but these intentions do not necessarily guarantee effectiveness and applicability. Others have already argued that many of the ideas to include ethical perspectives in EU policy on robotics and AI contain arguments with a rather abstract, imprecise character [37]. In the same vein, AI and robot ethics have been described as potentially "toothless" [38]. It has also been argued that ethical approaches could easily turn into platitudes if their concepts and content are not continuously critically discussed [39].

Furthermore, there are voices that generally criticize the suitability of ethics as a tangible framework for a critical discourse on the societal impact of technological development. For example, it has been argued that the rise of ethics has hampered proper public debate about the respective technologies by crystallizing relevant issues into rigid areas of ethical expertise. In this way, the argument goes, ethical inquiries and discussions have become the property of the intellectual establishment, thereby neutralizing politicized, public discourse on the impact of emerging technologies [40]. Others have claimed that robot and AI ethics itself has

become "big business", and that this could in fact aid the establishment of a rhetoric of "ethics washing" large-scale technological developments that harm society [41–43]. In line with this point, ethicists have argued that a collaborative attitude within a system of technological innovation can be counterproductive and that sometimes, more politicized, disruptive strategies should be considered to generate change for the better in the long run [44].

Finally, an often-heard argument is that the integration of ethics perspectives can complicate the process of technological development. This point goes hand in hand with two more general problems, namely the long-standing impasse between ethical and economic imperatives when it comes to innovation and the general problematization of the idea that AI and robotics technologies can be governed at their current stage of development. For instance, when it comes to ethical considerations regarding AI and robotics, some argue that they establish unnecessary hurdles in the current race for robotics and AI dominance (e.g., with regards to military supremacy) [45]. The idea is that if ethical limitations in the EU are much more restrictive than in other jurisdictions, competitors worldwide could have an advantage in terms of development speed. Accordingly, it is useful not to disregard this narrative too quickly, but rather to consider which new frameworks can be employed to define and tackle this problem in a way that enables technological development. Luciano Floridi states, for instance, that the EU's normative, human-centered approach is unique on the world stage, but must be recognized as quite a difficult approach. Nevertheless, he also explicitly argues that "nobody ever said that doing the right thing was going to be cheap and easy" [46].

In short, the main issue with the ethics narrative in its current form is that the resulting plans and strategies mostly involve a top-down approach that is still quite strongly based on high-level principles. Even though it seems evident that the EU is poised to invest in human-centred, ethical robotics and AI, the pathways towards this goal are still rather unclear and are generally up for negotiation, especially in light of the need for economic and technological development. In the remainder of this paper, the main question is how the high-level ideas and concepts that have been used to establish the narrative of ethical robotics and AI governance in the EU can become a meaningful component of research and development processes in practice-related fields like HRI. That is to say, HRI is explicitly understood as a field that can play an important role in the quest for more tangible implementations of ethics.

3.2 The Ethics Narrative in a Context of Emerging Technology Governance

In order to answer that question, it is useful to first contextualize the use of ethical notions in the governance of technology. In that regard, the position of robotic and AI technologies as emerging technologies needs to be emphasized. The general benefit of doing so is that it provides better insights into the reasons for governmental interest in ethics as a solution in the quest for governance of autonomous systems. In the specific context of this paper, it is also instrumental for establishing the connection between the ethics narrative in policy-making and the integration of ethical notions in robotics and HRI research.

First of all, in a policy-making context, the term 'emerging technologies' generally encompasses notions of fast-paced technological development, with a high level of societal and economic impact that governments need to manage [47, 48]. When understood in such a context, the use of ethical approaches as part of public policy decision-making goes back further than one might expect. In fact, the implementation of ethics as a tool for the governance of emerging technologies can be traced back to the 1970s in the US and the 1980s in the EU [49]. In several areas of technological development, applied ethics approaches and committees have been advanced in response to societal and governmental concerns over technoscientific innovations. Good examples can be found in the establishment of applied ethics approaches in fields like nanotechnology (nanoethics), data technology (data ethics), biomedical technology (bioethics) and so on [50–52]. An important reason for the implementation of ethical perspectives in the context of emerging technologies is the emphasis on the potential high stakes, high expectations, and sizable amount of future unknowns surrounding these technologies [53, 54]. This is important for policy-makers because it captures aspirations for a more prosperous future on the one hand (e.g. in the form of economic growth), while also representing contradictory and contested understandings of what that future might entail [55].

As is the case with other emerging technologies, robotics and AI technologies are often presented in the context of a future in which new innovations are expected to have many unprecedented properties and abilities with revolutionary and transformative potential. Furthermore, the historical dimension of such technologies is important to consider in this regard. Robots and other automata have captured the imagination since ancient times [56]. Especially in the last century, robotics and AI technologies have become prominent, widely recognized technocultural icons that represent many different speculative expectations and imaginations [57, 58]. While robotics and AI technologies have become

established as a professional field of research and engineering during the last half a century, it is this mix of expectations that makes them important objects of policy discourse. More specifically, as Sect. 2 has demonstrated, the increasing interactive capabilities of autonomous systems have emerged as an important issue of concern in how current EU policy-making on robotics and AI is currently being defined. It is in such a context that ethics can function as an important tool for anticipating and mitigating potential issues that have not yet fully materialized [59].

In relation to that, alongside the function of ethical reasoning as a multi-purpose toolkit for governing a range of different potential impacts of emerging technologies, ethics has an important function for managing public trust. Emerging technologies typically enjoy relatively strong coverage in general public discourse drawing on the anticipated trajectories of their technological development [60]. As argued above, expectations about emerging technologies like robotics and AI often involve many questions and uncertainties about the specificities of their (future) impact, leading to increased general public attention to the impacts of such technologies [61]. Prominent examples in the case of robotics technologies are their general impact on economic growth, the impact of potential automation on certain areas of the job market, potential gaps in the legal system, effects on intimate relationships and so on. From a governance perspective, the attention paid to such technology-based changes and the generation of socially sustainable narratives around them can aid the process of building general public trust in the long run [62, 63].

3.3 The Discussion on Implementation

Having elaborated on criticisms of the robotics and AI ethics narrative, as well as its wider embedding in the context of emerging technology governance, the next step is to discuss how this narrative can become strengthened and broadened in order to enhance its actual effects. The growing applicability of concepts from the field of robot and AI ethics has already led to increasing reflections on the position and possibilities of ethics in general [64, 65]. Furthermore, in a more general sense, the implementation of ethics is a widely discussed topic in different applied ethics contexts [66–68]. Here, it should be taken into consideration that the role of the ethics narrative in AI and robotics development is far from fixed and must be further materialized in order to have a proper effect. Different kinds of stakeholders (from industry, academia, NGOs, etc.) can still use their resources to inform the debate over how this narrative develops and gets implemented. I focus on the idea that it is fruitful for the general maturity of the ethics narrative to connect the governance context with the HRI context, as a professional community with unique qualities and insights, to improve ethical perspectives regarding the interactions between humans and interactive autonomous systems. Furthermore, I argue that in order for the ethics narrative to have a positive, constructive effect, it is important that the term "ethics" does not degenerate into dogmatic debates concerning human control over technology, but rather becomes further established in critical, inter- and transdisciplinary inquiries over the roles ascribed to social robots and other autonomous systems.

The two different contexts that are connected with one another here - the governance context vs. the robotics and HRI research context - are characterized by very different epistemologies and practices. In short, autonomous systems such as robots as emerging objects of governance are vastly different from their understanding as objects of robotics research and design. Thus, from the perspective of HRI research, an easy way out would be to denounce policy-making efforts and maintain a strong distinction between these two epistemologies. However, this would be a rather simplistic way of approaching the problem that does not keep pace with promising trends in other fields. Whereas more traditional interpretations of policy-making are indeed based on a rather strong distinction between the policy-making context and research and engineering contexts, there are many new notions of governance that argue otherwise. In fact, one trend in policy-making is the notion that well-implemented policy can be achieved by developing a type of governance that identifies and avoids problems by design [69, 70].

In recent years, these kinds of notions have increasingly led to deliberations about types of governance that understand scientific and technical expertise as fields of practice that can contribute to the goals of newly developed governance frameworks [71]. For instance, in a paper on what they call "integrative expertise", Michael Poznic and Erik Fisher refer to this type of governance as "midstream", thereby distinguishing it from "upstream" research prioritization and "downstream" technology regulation [72]. In doing so, they focus on the normative aspects of technical experts' practices, thereby establishing the possibility to understand ethics as an integrated component of this midstream domain. They argue that this attention to the midstream can help with understanding and responding to socio-ethical considerations in the practice of developing novel technologies. In the same spirit, I argue here that robotics and HRI as fields of research can be seen as very well equipped to develop clearer pathways towards further defining the concepts and ideas that are currently shaping the narrative of ethical AI and robotics governance. Thus, the role of HRI would be one that focuses on its own capacities of altering its practices in such a way that is focused on the socio-technical integration

of the ethics narrative. By translating issues from the realm of governance to research areas like HRI, ethical notions can move beyond conceptual stages and provide new, more precise insights into how the ethical development and use of interactive robots can be achieved.

4 Manifesting the Ethics Narrative in HRI Research

Up until this point, the EU governance context was described as the main locus of change, whereas the social robotics and HRI research contexts were mostly mentioned as a crucial component in that development. In this section, however, the focus will shift to the latter. As has been made clear above, an important future prospect that has sparked the EU's focus on anticipatory AI and robot regulation is the emergence of new generations of AI-equipped robots with unprecedented interactive capabilities. The ramifications of these developments have also become an important point of discussion in robotics and HRI research, since they allow many more (unexpected) factors to affect the interactions between robots and humans. With this in mind, the subsections below explore the different ways in which ethics and ethical notions in the EU ethics narrative are currently (a) already being discussed and studied in HRI research, including prominent examples from the literature; and (b) how these efforts can be further developed and enriched in order to contribute to a more profound and effective ethics narrative on a midstream level.

4.1 Ethical Assessment

One of the most common, straightforward forms of implementing ethics in research and development processes across the board is to discern and define potential ethical issues. This also plays a major role in the EU governance plans, since a central aim of these plans is to stimulate the development of autonomous systems that serve societal needs, while simultaneously avoiding or mitigating ethical issues. Furthermore, the ethical risk assessment procedures that are part of the AI Act will likely be strongly grounded in this type of ethics implementation. An important way to achieve this kind of ethics consideration is through the institutionalization of ethical assessments in the form of committees that review certain decisions and establish a consensus on specific technologies based on expert review procedures. Apart from that, this type of ethics integration is generally concerned with the ways in which ethical principles, guidelines and standards can be employed by researchers and other practitioners. Ethical assessment procedures in such contexts often include elements of prediction and foresight through the anticipation of consequences of development and implementation.

In general, this way of applying ethics fits quite well with most of the existing HRI research and the specific ramifications of this approach are discussed quite regularly in the field. First of all, important attempts have been made to discuss what the main ethical issues in HRI are, while defining ethical codes and guidelines based on these discussions [73, 74]. Furthermore, even though ethical assessments and ethical standards are often portrayed as an endeavor that restricts robotics development and innovation, several practitioners have argued that increasing attention to ethics now will likely lead to more socially sustainable innovations in the long run [75–77]. A crucial argument in this regard is that when it comes to interactions between robotic systems and their users, the identification of ethical issues and implementation of ethical principles, guidelines and/or standards is often argued to aid in configuring human-robot interactions themselves. For instance, persistent attention to principles like non-maleficence or social justice can help to increase the acceptance of social robots and other autonomous systems, especially in the longer run [78]. Furthermore, this has led to important discussions in HRI on the assessment of norms and values that require recognition of the needs of minorities in autonomous systems [79, 80].

Further development and enrichment in order to contribute to a more profound and effective ethics narrative can be considerably advanced through research that explicitly engages with existing high-level principles that are currently discussed in governance. In line with the arguments above regarding technological innovation and ethics, it can be fruitful to see the ethical governance of interactive robots itself as an instance of sociotechnical innovation. Because of its focus on empirical analyses of interactions, HRI can add substantial insights to the ways in which high-level notions can become part of ethics assessments in HRI practice. Translating such considerations and values into implementable rules, standards and guidelines allows them to evolve and be tested as components of robotics research. Crucial in this context are current advancements in the field of ethical standards. Robotics and HRI expertise plays an important role in the further development of these ethical standards [81].

4.2 Moral Competence for Artificial Ethical Agency

Thinking about normative reasoning in artificial agents has a rather long history in fields such as logical reasoning and other realms of AI and robotics research [82, 83]. As has also been argued in the EU context, the very prospect of more sophisticated automated decision-making systems renders it desirable to formalize and implement ethical reasoning - the

main rationale being that human dependence on such automated systems requires their decisions to remain within the borders of what is determined to be ethical. When robot ethics as a field began to take shape, this was an important part of considerations on how to make ethics a relevant notion for robotics development, and it remains an important point of consideration in current discussions on the implementation of ethics in autonomous systems. One of the most common examples illustrating the importance of ethics for autonomous systems can be found in the trolley problem, which is often used to show how autonomous systems (e.g., autonomous vehicles) need to be prepared to make decisions of an ethical nature.

In its most strict sense, this kind of research concerns autonomous agents that can be considered morally competent. Thus, with regards to HRI research, the integration of ethics and ethical approaches can be achieved by formalizing ethical norms so that they can become embedded in the automated reasoning of robotic artifacts. It is particularly relevant when the artifacts that have emerged from such research are tested in interaction research. In this case, social robots as autonomous agents could exhibit behavior that can be considered "ethical" [84]. There are several case studies that discuss and study the direct implementation of ethics in this sense as a part of autonomous agents' behavior [85–88]. Other useful studies in this realm analyze human norms themselves and the way autonomous agents can engage in norm-conforming ways of interacting [89]. Furthermore, the notion of trust in technology has been evolving together with initiatives that implement ethical rules to guide the behavior of artificial agents and ensure trustworthy interactions [90].

Apart from the research that is needed to further develop moral reasoning in robots, this is very much a project where moral HRI can help us understand what it means to have moral machines and how this affects robotics' socio-technical systems [91]. In terms of the high-level EU plans, this type of ethics involves the ambitious and important endeavor to create autonomous machines that can be inherently trustworthy. Because of its direct research focus on interactions, as well as its high level of technical expertise, HRI has a lot to offer here to help make the very notion of the moral, trustworthy robot a success. For instance, an important task of HRI research in this context is to understand how humans apply moral norms to (social) robots and judge their behavior [92]. Findings from such research can help to determine how social robots should behave, thus contributing to the discussion on meaningful ethical governance of more sophisticated forms of artificial agency. Thus, in this interaction context, it is important to develop elaborate definitions of what constitutes normative behavior and to embed ethics itself within such definitions in a formalized manner.

4.3 Value-Based Design and Implementation

Another promising path to developing a more mature ethics narrative that is gaining traction at the moment concerns the proliferation of value-based design and implementation processes. This kind of ethics implementation has also gained traction in the EU context, for instance through the notion of ethics-by-design. In short, the idea here is that deliberation on and implementation of values can help ensure a strategy in which societal and ethical issues become part and parcel of robotics research and development processes, as well as a central consideration when implementing robots in existing contexts [93, 94]. As such, central to this approach are the analysis of values and the establishment of new norms in the research, design, and development process for social robotics applications [95]. From this point of view, responsible research and innovation practices are crucial to achieving this type of ethics integration.

By making the design, engineering and implementation of robotics more focused on values that are important in open, democratic societies, innovation processes can likewise become more open, responsible, and inclusive [96]. Examples of such values can be found in more general longstanding democratic notions like transparency and accountability, but often require detailed deliberation in order to be properly implemented [90, 97, 98]. Many of these topics are currently the subject of discussions and research in HRI. This includes discussions about the norms and values of roboticists and HRI practitioners themselves, as well as the implementation of new norms and values [99, 100]. Research approaches like Value-Sensitive Design (VSD) and Participatory Design (PD) have proven useful here, as they generally place a strong emphasis on integrating ethical principles and values [96, 101–104]. Furthermore, standardization can also play an important role in this context, since it can establish design requirements which are useful for establishing new types of societal engagement with robotics engineering and design [105].

From a governance perspective, these kinds of ethics implementations draw attention to the question of how research and innovation systems can be modified to establish responsible, ethical innovation [106]. As such, design itself becomes acknowledged as a specific mode of ethical inquiry. One that is strongly grounded in immediate practical engagement with ethical issues. In this regard, it is also important to note that ethical notions from within HRI have become connected to broader value-based innovation frameworks in the governance of science and technology. Useful examples in this context are the responsible research

and innovation (RRI) framework or the Sustainable Development Goals (SDGs) [107, 108]. In addition to their goal of improving the innovation process, these frameworks can also actively increase public trust in technologies [109].

4.4 Ethical Sociotechnical Systems

Finally, ethics can play a role in the establishment of autonomous systems as components and drivers of large-scale sociotechnical systems. From a philosophical point of view, this notion of ethics tends to see ethical approaches mostly as a form of critical engagement with wider sociocultural developments around robots. Mark Coeckelbergh, for instance, proposes a social-relational approach to robot ethics which aims to pay attention "to the moral significance of how we humans talk about robots, do things with robots, and live with robots" [110]. Other examples can be found in the integrative approach developed by Seibt et al. They argue that in HRI and social robotics, the aim has generally been to "investigat[e] what social robots *can* do, while robo-ethicists and policy-makers deliberate *afterwards* what social robots (*may* or) *should* (not) do, relative to the professional discussion in applied ethics". Instead, the real question should be "what social robotics applications *can* and (*may* or) *should* do" [95].

When looking at the current landscape of HRI research, the empirical and/or applied nature of the field generally does not facilitate the development of concepts referring to such macro-level ethics. Nevertheless, there are several research trends that can help to further this perspective. First of all, one of the most important developments is the realization that the rising feasibility of robots taking part in real-world scenarios is currently leading to discussions over the methodologies used to conduct HRI research. The limitations of current experimental methodologies are being increasingly addressed, and several researchers have already argued for changes and additions to these mainstream methodologies. For instance, Kerstin Dautenhahn writes that HRI methodologies sometimes fail to address "how real people, in real-world environments, would interact face to face with a real robot" [111]. As a part of this development, more HRI studies are now incorporating qualitative methods [112–114]. These methods can help to study interactions in which complex, autonomous robots interact with people in unrestricted (or less restricted) contexts. Others have also argued that robot ethics and HRI need to jointly develop models of analysis that can capture the complexity of interactions by looking at the entire system in which robots are to play a facilitating role (e.g. the healthcare system) [104]. Finally, the entanglements that constitute the (mundane) relationships between humans and technological artifacts are increasingly addressed in research, which could in turn be integrated with critical HRI perspectives [115, 116].

The ethics narrative in governance, in turn, largely takes a fundamentally critical approach. Even though this might not directly benefit governance or innovation strategies, it is crucial to maintain an open critical culture at a time of rapid technological development. In this regard, the practices of robot and AI ethicists on the one hand and HRI researchers on the other need to become ever more firmly integrated as part of interdisciplinary methodological frameworks. When established within the context of HRI research, these perspectives can facilitate research on how interaction in real-world contexts develops. In this way, more sophisticated concepts can help to create improved policy ideas for the EU's robotics and AI governance by helping us better understand the political economy of robotics and AI and the different roles ethics can play in changing it for the better.

5 Conclusion

The main point of this paper has been to engage in deliberations on the connection between the ethics narrative in EU governance and the implementation of ethical notions in HRI research. I have done so by providing an overview of the integration of this ethics narrative in the EU's policy plans and demonstrating how this narrative provided new definitions of policy issues. In the face of emerging implementation of autonomous social robots, it is likely that ethical approaches continue to be relevant for discussions of issues pivotal to democratic societies, such as human rights and democratic values. This paper has argued for an interpretation of EU policies and strategies not just as a way of developing new regulations inspired by ethics, but very much also as a trigger for and incentive towards developing more profound ethical approaches in a context of rapid and uncertain technological developments in social robotics. As autonomous systems are projected to have a strong impact in our society, it is important to see the emergence of this narrative as an opportunity for substantial change while simultaneously understanding that new ideas will have to materialize quickly in order to ensure that our society is able to deal with the impact. As argued above, HRI can have an important role in solidifying and implementing this narrative and the paper has provided a call to adopt an understanding of policy-making and technical expertise that in conjunction with one another.

These new ideas can serve as points of departure for deliberation on the ethical futures we consider desirable and which ones we would prefer to avoid. As such, they can be very informative while contributing to the emergence of a technological culture that has the ability to simultaneously

engage with the sociotechnical potential of autonomous systems as well as with the ethical consequences of these futures. In that regard, HRI can be seen as an epistemic community with its own unique position when it comes to the responsibility of implementing and materializing ethical notions. Furthermore, in order to draw ethics out of the buzzword context and elicit these profound effects, there is a clear necessity for developing new insights about ethics implementation. This paper has attempted to do exactly that in the hope to provide new pathways for a new generation of ethical inquiry. Especially now that ethics is such a prominent topic in general, it is likely that new pathways and approaches will emerge. In that regard it is important to remark that this should not be confined to EU policy-making. Whereas this paper maintained a strong emphasis on EU governance of autonomous systems, other governance context such as EU member-states, other major geopolitical players, intergovernmental organizations) should be very much seen as pivotal to this process. The same counts for other, non-governmental contexts of ethics integration (e.g. in business contexts) [66, 117].

Funding This work was supported by the TrustRobots Doctoral College, TU Wien.
Open access funding provided by TU Wien (TUW).

Data Availability The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

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

References

1. Smuha NA (2021) From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. Law Innov Technol 13:57–84. https://doi.org/10.1080/17579961.2021.1898300
2. McBeth MK, Lybecker DL (2018) The Narrative Policy Framework, Agendas, and Sanctuary Cities: the construction of a public Problem. Policy Stud J 46:868–893. https://doi.org/10.1111/psj.12274
3. Allan BB (2017) Producing the climate: States, scientists, and the Constitution of Global Governance Objects. Int Org 71:131–162. https://doi.org/10.1017/S0020818316000321
4. Hampton G (2009) Narrative policy analysis and the integration of public involvement in decision making. Policy Sci 42:227–242. https://doi.org/10.1007/s11077-009-9087-1
5. Jones MD, McBeth MK (2010) A Narrative Policy Framework: clear enough to be wrong? Policy Stud J 38:329–353. https://doi.org/10.1111/j.1541-0072.2010.00364.x
6. McBeth MK, Shanahan EA, Arnell RJ, Hathaway PL (2007) The intersection of Narrative Policy Analysis and Policy Change Theory. Policy Stud J 35:87–108. https://doi.org/10.1111/j.1541-0072.2007.00208.x
7. Legro JW (2000) The Transformation of Policy Ideas. Am J Polit Sci 44:419–432. https://doi.org/10.2307/2669256
8. Sabel C, Herrigel G, Kristensen PH (2018) Regulation under uncertainty: the coevolution of industry and regulation. Regul Gov 12:371–394. https://doi.org/10.1111/rego.12146
9. euRobotics (2020) AbouteuRobotics. https://www.eu-robotics.net/eurobotics/about/about-eurobotics/about-eurobotics.html. Accessed11Apr2022
10. Veruggio G (2006) The EURON Roboethics Roadmap. In: 2006 6th IEEE-RAS International Conferenceon Humanoid Robots. pp 612–617
11. DG INFSO (2006) EUROP--theEuropeanRoboticsPlatform
12. Connect DG, Jäger A, Moll C (2016) et al Analysis of the impact of robotic systems on employment in the European Union
13. DGRI (2020) Unlocking the potential of industrial human–robot collaboration: a visionon industrial collaborative robots for economy and society
14. Eurofound (2018) Game changing technologies: exploring the impact on production processes and work. Research Report
15. Joint Research Centre (2020) Global race for robotisation Looking at the entire robotisation chain
16. EPRS,STOA (2016) STUDY Ethical aspects of cyber-physical systems: scientific foresight study: in-depth analysis
17. European Parliament (2017) RESOLUTION Civil Law Rules on Robotics
18. DG IPOL, Nevejans N, European P (2016) etal STUDY European civil law rules in robotics
19. Nevejans N (2018) Open letter to the European Commission. Artificial Intelligence and Robotics
20. DGRI,European Group on Ethics in Science and New Technologies to the European Commission,European Commission (2018) Statement on artificial intelligence, robotics and "autonomous" systems: Brussels, 9 March 2018
21. EC (2018) Artificial Intelligence for Europe
22. AIHLEG (2019) DELIVERABLE Ethics Guidelines for Trustworthy AI
23. AIHLEG (2019) DELIVERABLE Policy and Investment Recommendations for Trustworthy Artificial Intelligence
24. AIHLEG (2020) Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment
25. AIHLEG (2020) Sectoral considerations on the policy and investment. Recommendations for Trustworthy Artificial Intelligence
26. European Commission (2018) COMMUNICATION Coordinated Plan on Artificial Intelligence
27. European Commission (2019) COMMUNICATION Building Trust in Human-Centric Artificial Intelligence
28. European Parliament (2019) RESOLUTION on a comprehensive European industrial policy on artificial intelligence and robotics

29. Von der Leyen U (2019) A Union that strives for more. My agenda for Europe: political guidelines for the next European Commission 2019–2024. Publications Office of the European Union, Luxembourg
30. European Commission (2020) WHITEPAPER On Artificial Intelligence - A European approach to excellence and trust
31. European Union Agency for Fundamental Rights (2020) REPORT getting the future right artificial intelligence and fundamental rights. Publications Office of the European Union, Luxembourg
32. European Commission (2021) COMMUNICATION Fostering a European approach to Artificial Intelligence
33. Bradford A (2020) The Brussels Effect: how the European Union Rules the World. Oxford University Press
34. European Commission (2021) PROPOSAL FOR a REGULATION of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence. Artificial Intelligence Act) and amending certain Union legislative acts
35. Veale M, Borgesius FZ (2021) Demystifying the Draft EU Artificial Intelligence Act
36. UNESCO (2021) DRAFT Recommendationon the Ethics of Artificial Intelligence
37. Fosch Villaronga E, Golia AJ (2019) Robots, standards and the law: rivalries between private standards and public policymaking for robot governance. Comput Law Secur Rev 35:129–144. https://doi.org/10.1016/j.clsr.2018.12.009
38. Rességuier A, Rodrigues R (2020) AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society 7:2053951720942541. https://doi.org/10.1177/2053951720942541
39. Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nat Mach Intell 1:501–507. https://doi.org/10.1038/s42256-019-0114-4
40. Felt U, Wynne B (2009) Taking European knowledge society seriously. In: Science et devenir del' homme, 2009, N°59, fasciculethé matique "Science in Society: Dialogues and Scientific Responsibility"-Sciencein Society: Dialogues and Scientific Responsibility. European Conference, Paris, FRA, 2008-11-24. MURS, Paris(FRA)
41. Bietti E (2020) From ethics washing to ethics bashing: a view on tech ethics from with in moralphilosophy. In: Proceedings of the 2020 Conferenceon Fairness, Accountability, and Transparency. As sociation for Computing Machinery, New York, NY, USA, pp 210–219
42. Phan T, Goldenfein J, Mann M, Kuch D (2021) Economies of Virtue: the circulation of 'Ethics' in Big Tech. Sci as Cult 0:1–15. https://doi.org/10.1080/09505431.2021.1990875
43. Richardson K (2019) The business of Ethics, Robotics, and Artificial Intelligence. In: Heffernan T (ed) Cyborg Futures. Springer, Berlin, pp 113–126
44. Sætra HS, Coeckelbergh M, Danaher J (2021) The AI ethicist's dilemma: fighting Big Tech by supporting Big Tech. AI Ethics. https://doi.org/10.1007/s43681-021-00123-7
45. Manson K (2021) US has already lost AI fight to China, says expentagon software chief. Financial Times
46. Floridi L (2021) The european legislation on AI: a brief analysis of its Philosophical Approach. Philos Technol 34:215–222. https://doi.org/10.1007/s13347-021-00460-9
47. Rotolo D, Hicks D, Martin BR (2015) What is an emerging technology? Res Policy 44:1827–1843. https://doi.org/10.1016/j.respol.2015.06.006
48. Taeihagh A, Ramesh M, Howlett M (2021) Assessing the regulatory challenges of emerging disruptive technologies. Regul Gov 15:1009–1019. https://doi.org/10.1111/rego.12392
49. Tallacchini M (2009) Governing by values. EU Ethics: Soft Tool, Hard Effects. Minerva 47:281. https://doi.org/10.1007/s11024-009-9127-1
50. Kearnes M, Grove-White R, Macnaghten P et al (2006) From Bio to Nano: Learning Lessons from the UK Agricultural Biotechnology controversy. Sci as Cult 15:291–307. https://doi.org/10.1080/09505430601022619
51. Schaper-Rinkel P (2013) The role of future-oriented technology analysis in the governance of emerging technologies: the example of nanotechnology. Technol Forecast Soc Chang 80:444–452. https://doi.org/10.1016/j.techfore.2012.10.007
52. Zwitter A (2014) Big Data ethics. Big Data & Society 1:205395171455925. https://doi.org/10.1177/2053951714559253
53. Kuhlmann S, Stegmaier P, Konrad K (2019) The tentative governance of emerging science and technology—A conceptual introduction. Res Policy 48:1091–1097. https://doi.org/10.1016/j.respol.2019.01.006
54. Sandler RL (2014) Ethics and Emerging Technologies. Palgrave Macmillan UK, London
55. Szekely I, Szabo MD, Vissy B (2011) Regulating the future? Law, ethics, and emerging technologies. J of Inf Com & Eth in Society 9:180–194. https://doi.org/10.1108/14779961111167658
56. Rossi C, Russo F, Russo F (2009) Automata (towards automation and Robots). Ancient Engineers& inventions. Springer Netherlands, Dordrecht, pp 269–301
57. Geraci RM (2010) Apocalyptic AI: visions of heaven in robotics, artificial intelligence, and virtual reality. Oxford University Press, New York
58. Romic B (2018) Robotic Art and Cultural Imagination. In: Proceedings of EVA Copenhagen 2018. BCS Learning& Development, Copenhagen, pp 1–10
59. Brey P (2012) Anticipatory Ethics for Emerging Technologies. Nanoethics 6:1–13. https://doi.org/10.1007/s11569-012-0141-7
60. Goyal N, Howlett M, Taeihagh A (2021) Why and how does the regulation of emerging technologies occur? Explaining the adoption of the EU General Data Protection Regulation using the multiple streams framework. Regul Gov 15:1020–1034. https://doi.org/10.1111/rego.12387
61. Frewer L (1999) Risk perception, Social Trust, and public participation in strategic decision making: implications for Emerging Technologies. Ambio 28:569–574
62. Stebbing M (2009) Avoiding the Trust Deficit: Public Engagement, values, the Precautionary Principle and the future of Nanotechnology. Bioethical Inq 6:37–48. https://doi.org/10.1007/s11673-009-9142-9
63. Wynne B (2006) Public Engagement as a Means of restoring Public Trust in Science - Hitting the Notes, but missing the music? Community Genetics. Basel 9:211–220
64. Danks D (2022) Digital Ethics as Translational Ethics: In: Vasiliu-Feltes I, Thomason J (eds) Advances in Human and Social Aspects of Technology. IGI Global, pp 1–15
65. Stahl BC (2021) Concepts of Ethics and their application to AI. In: Stahl BC (ed) Artificial Intelligence for a better future: an ecosystem perspective on the Ethics of AI and emerging Digital Technologies. Springer International Publishing, Cham, pp 19–33
66. Eitel-Porter R (2021) Beyond the promise: implementing ethical AI. AI Ethics 1:73–80. https://doi.org/10.1007/s43681-020-00011-6
67. Grunwald A (2011) Responsible innovation: bringing together technology assessment, applied ethics, and STS research. Enterp Work Innov Stud 31:9–31
68. McLennan S, Fiske A, Tigard D et al (2022) Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med Ethics 23:6. https://doi.org/10.1186/s12910-022-00746-3
69. Greer SL, Vasev N, Jarman H et al (2020) It's the governance, stupid! TAPIC: a governance framework to strengthen decision making and implementation. World Health Organization, Copenhagen

70. Guston DH (2014) Understanding 'anticipatory governance'. Soc Stud Sci 44:218–242. https://doi.org/10.1177/0306312713508669
71. Pohl C (2008) From science to policy through transdisciplinary research. Environ Sci Policy 11:46–53. https://doi.org/10.1016/j.envsci.2007.06.001
72. Poznic M, Fisher E (2021) The Integrative Expert: Moral, Epistemic, and Poietic Virtues in Transformation Research. Sustainability 13:10416. https://doi.org/10.3390/su131810416
73. Riek LD, Howard D (2014) A Code of Ethics for the Human-Robot Interaction Profession. In: Proceedings of We Robot. Miami, p 10
74. Wullenkord R, Eyssel F (2020) Societal and ethical issues in HRI. Curr Robot Rep 1:85–96. https://doi.org/10.1007/s43154-020-00010-9
75. Demir KA (2017) Research questions in Robot Ethics. Mugla J Sci Technol 3:160–165. https://doi.org/10.22531/muglajsci.359648
76. Zawieska K (2018) Is roboethics really optional? In: An Alternative HRI methodology: The use of ethnography to identify and address Ethical, Legal, & Societal (ELS) issues Workshop at the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI2018). ACM/IEEE, Chicago, IL
77. Torresen J (2018) A review of future and ethical perspectives of Robotics and AI. Front Robot AI 4:75. https://doi.org/10.3389/frobt.2017.00075
78. Bartneck C, Lütge C, Wagner A, Welsh S (2021) Trust and Fairness in AI Systems. In: Bartneck C, Lütge C, Wagner A, Welsh S (eds) An introduction to Ethics in Robotics and AI. Springer International Publishing, Cham, pp 27–38
79. Howard A, Borenstein J (2018) The Ugly Truth about ourselves and our Robot Creations: the Problem of Bias and Social Inequity. Sci Eng Ethics 24:1521–1536. https://doi.org/10.1007/s11948-017-9975-2
80. Howard A, Kennedy IIIM (2020) Robots are not immune to bias and injustice. https://doi.org/10.1126/scirobotics.abf1364
81. Winfield A (2019) Ethical standards in robotics and AI. Nat Electron 2:46–48. https://doi.org/10.1038/s41928-019-0213-6
82. Anderson SL (2016) Asimov's "Three Laws of Robotics" and machine metaethics. Science Fiction and Philosophy. John Wiley & Sons, Ltd, pp 290–307
83. Bonnemains V, Saurel C, Tessier C (2018) Embedded ethics: some technical and ethical challenges. Ethics Inf Technol 20:41–58. https://doi.org/10.1007/s10676-018-9444-x
84. Gips J (1994) Toward the ethical Robot. In: Ford KM, Glymour C, Hayes P (eds) Android Epistemology. MIT Press, pp 243–252
85. Malle BF (2016) Integrating robot ethics and machine morality: the study and design of moral competence in robots. Ethics Inf Technol 18:243–256. https://doi.org/10.1007/s10676-015-9367-8
86. Malle BF, Scheutz M (2017) Moral competence in Social Robots. Machine Ethics and Robot Ethics. Routledge
87. Bringsjord S, Arkoudas K, Bello P (2020) Toward a General Logicist Methodology for Engineering Ethically Correct Roborts. pp 291–297
88. McBride N, Hoffman RR (2016) Bridging the ethical gap: from Human Principles to Robot instructions. IEEE Intell Syst 31:76–82. https://doi.org/10.1109/MIS.2016.87
89. Wallach W, Allen C (2009) Moral machines: teaching robots right from wrong. Oxford University Press, Oxford; New York
90. Wortham RH, Theodorou A (2017) Robot transparency, trust and utility. Connection Sci 29:242–248. https://doi.org/10.1080/09540091.2017.1313816
91. Doyle-Burke D, Haring KS (2020) Robots Are Moral Actors: Unpacking Current Moral HRI Research Through a Moral Foundations Lens. In: Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings. Springer-Verlag, Berlin, Heidelberg, pp 170–181
92. Malle B, Scheutz M, Arnold T, et al (2015) Sacrifice One For the Good of Many?: People Apply Different Moral Norms to Human and Robot Agents. ACM/IEEE International Conference on Human-Robot Interaction 2015:117–124.https://doi.org/10.1145/2696454.2696458
93. Coeckelbergh M (2013) Human being @ risk: enhancement, technology, and the evaluation of vulnerability transformations. Springer, Dodrecht
94. Coeckelbergh M (2015) The tragedy of the master: automation, vulnerability, and distance. Ethics Inf Technol 17:219–229. https://doi.org/10.1007/s10676-015-9377-6
95. Seibt J, Damholdt MF, Vestergaard C (2020) Integrative social robotics, value-driven design, and transdisciplinarity. Interact Stud 21:111–144. https://doi.org/10.1075/is.18061.sei
96. Cheon E, Su NM (2016) Integrating Roboticist Values in to a Value Sensitive Design Framework for Humanoid Robots. In: The Eleventh ACM/IEEE International Conferenceon Human Robot Interaction. IEEE Press, Christchurch, New Zealand, pp 375–382
97. Dignum V (2017) Responsible Artificial Intelligence: Designing Ai for Human values. ICT, ITU Journal. Discoveries 9
98. Wortham RH, Theodorou A, Bryson JJ (2017) Robot transparency: improving understanding of Intelligent Behaviour for designers and users. In: Gao Y, Fallah S, Jin Y, Lekakou C (eds) Towards Autonomous Robotic Systems. Springer International Publishing, Cham, pp 274–289
99. Liu H-Y, Zawieska K (2020) From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. Ethics Inf Technol 22:321–333. https://doi.org/10.1007/s10676-017-9443-3
100. Winfield AFT, Winkle K, Webb H et al (2021) Robot Accident Investigation: a case study in responsible Robotics. In: Cavalcanti A, Dongol B, Hierons R,etal(eds) (eds) Software Engineering for Robotics. Springer International Publishing, Cham, pp 165–187
101. Darriba Frederiks A, Octavia JR, Vandevelde C, Saldien J (2019) Towards participatory design of social robots. In: IFIP Conference on Human-Computer Interaction. Springer, pp 527–535
102. Winkle K, Senft E, Lemaignan S (2021) LEADOR: a method for End-To-End Participatory Design of Autonomous Social Robots. Frontiers in Robotics and AI8
103. Azenkot S, Feng C, Cakmak M (2016) Enabling building service robots to guide blind people a participatory design approach. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp 3–10.
104. van Wynsberghe A, Li S (2019) A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI). MB 9:11–21. https://doi.org/10.2147/MB.S160348
105. Siau K, Wang W (2018) Building Trust in Artificial Intelligence, Machine Learning, and Robotics. Cut Bus Technol J 31:8
106. Geels FW (2004) From sectoral systems of innovation to socio-technical systems: insights about dynamics and change from sociology and institutional theory. Res Policy 33:897–920. https://doi.org/10.1016/j.respol.2004.01.015
107. Asveld L, van Dam-Mieras R, Swierstra T et al (2017) Responsible innovation 3: a european agenda? Springer Berlin Heidelberg, New York, NY
108. van den Hoven J, Vermaas PE, van de Poel I (2015) Handbook of Ethics, values, and Technological Design. Springer Netherlands, Dordrecht
109. Rask M, Mačiukaitė-Žvinienė S, Tauginienė L, et al (2016) Innovative public engagement: A conceptual model of public engagement in dynamic and responsible governance of research and innovation
110. Coeckelbergh M (2011) Is Ethics of Robotics about Robots? Philosophy of Robotics Beyond realism and

individualism. Law Innov Technol 3:241–250. https://doi.org/10.5235/175799611798204950
111. Dautenhahn K (2018) Some brief thoughts on the past and future of Human-Robot Interaction. J Hum-Robot Interact 7:4:1–4. https://doi.org/10.1145/3209769
112. Hasse C, Trentemøller S, Sorenson J (2019) Special issue on Ethnography in Human-Robot Interaction Research. Paladyn. J Behav Rob 10:180–181. https://doi.org/10.1515/pjbr-2019-0015
113. Weiss A, Spiel K (2021) Robots beyond Science Fiction: mutual learning in human–robot interaction on the way to participatory approaches. AI & Soc. https://doi.org/10.1007/s00146-021-01209-w
114. Fischer K, Seibt J, Rodogno R et al (2020) Integrative Social Robotics Hands-on. Interact Stud 21:145–185. https://doi.org/10.1075/is.18058.fis
115. Maibaum A, Bischof A, Hergesell J, Lipp B (2022) A critique of robotics in health care. AI & Soc 37:467–477. https://doi.org/10.1007/s00146-021-01206-z
116. Rommetveit K, van Dijk N, Gunnarsdóttir K (2020) Make way for the Robots! Human- and machine-centricity in constituting a european public–private Partnership. Minerva 58:47–69. https://doi.org/10.1007/s11024-019-09386-1
117. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399. https://doi.org/10.1038/s42256-019-0088-2