Interdisciplinary Science Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/yisr20

Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies

Inga Ulnicane, Damian Okaibedi Eke, William Knight, George Ogoh & Bernd Carsten Stahl

To cite this article: Inga Ulnicane, Damian Okaibedi Eke, William Knight, George Ogoh & Bernd Carsten Stahl (2021) Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies, Interdisciplinary Science Reviews, 46:1-2, 71-93, DOI: 10.1080/03080188.2020.1840220

To link to this article: https://doi.org/10.1080/03080188.2020.1840220

Published online: 07 Mar 2021.

INTERDISCIPLINARY SCIENCE REVIEWS
2021, VOL. 46, NOS. 1–2, 71–93
https://doi.org/10.1080/03080188.2020.1840220

The Institute of Materials, Minerals and Mining

Routledge Taylor & Francis Group

Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies

Inga Ulnicane, Damian Okaibedi Eke, William Knight, George Ogoh and Bernd Carsten Stahl

Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

ABSTRACT

Recent advances in Artificial Intelligence (AI) have led to intense debates about benefits and concerns associated with this powerful technology. These concerns and debates have similarities with developments in other emerging technologies characterized by prominent impacts and uncertainties. Against this background, this paper asks, What can AI governance, policy and ethics learn from other emerging technologies to address concerns and ensure that AI develops in a socially beneficial way? From recent literature on governance, policy and ethics of emerging technologies, six lessons are derived focusing on inclusive governance with balanced and transparent involvement of government, civil society and private sector; diverse roles of the state including mitigating risks, enabling public participation and mediating diverse interests; objectives of technology development prioritizing societal benefits; international collaboration supported by science diplomacy, as well as learning from computing ethics and Responsible Innovation.

KEYWORDS

Artificial Intelligence; emerging technologies; governance; policy; ethics; regulation; societal challenges; Responsible Innovation

> Uncertainty about the impact of AI can be a concern but it is also an opportunity: the future is not yet written. We can, and should, shape it. (European Commission 2018a, 13)

Introduction

Is time travel among the numerous wonders promised by Artificial Intelligence (AI)? Could this transformative and revolutionary technology transport us back to 'good (or not so good) old times'? For researchers of governance, policy and ethics of emerging technologies, recent years of academic and public debates about AI have often presented an opportunity to travel several decades back

CONTACT Inga Ulnicane inga.ulnicane@dmu.ac.uk De Montfort University, The Gateway, Leicester LE1 9BH, UK

to the twentieth century, namely, to the times when emerging technologies were understood mostly in terms of their contributions to economic growth and national prestige. Actors involved in innovation were largely limited to industry and academia. The state mostly was seen as having a limited role of market correction and the field of computing ethics was less prominent. Those were times with little or no awareness of cross-cutting political, social and ethical issues of emerging technologies – from life sciences to information technologies – that shape our lives (e.g. Jasanoff 2016: Juma 2016) and specific ways to govern them (e.g. Kuhlmann, Stegmaier, and Konrad 2019) such as Responsible Innovation (e.g. Stilgoe, Owen, and Macnaghten 2013). The knowledge accumulated and lessons learned about governance, policy and ethics of emerging technologies over the past decades have not much featured in recent discussions about AI.

Is AI really such a *sui generis* phenomenon that its governance and ethics have little if anything to learn from other emerging technologies? A number of concerns and issues addressed in AI debates leave a déjà vu feeling because they are well-known from work on other emerging technologies. This suggests that when addressing concerns associated with AI, rather than 'reinventing the wheel', there are opportunities to learn from advances made and lessons derived from governance, policies and ethics of other emerging technologies. While each technology has some unique features, all emerging technologies have a number of common characteristics – such as radical novelty, relatively fast growth, coherence, prominent impact, and uncertainty and ambiguity (Rotolo, Hicks, and Martin 2015) – that pose some similar concerns and allow to learn across different technologies.

Against this background, this paper asks – what can AI governance, policy and ethics learn from other emerging technologies to address concerns and ensure that AI develops in a socially beneficial way? To answer this question, we draw on analysis of AI policy documents and on recent literature on governance, policy and ethics of emerging technologies. Discussions of concerns about AI and governance, policy and ethics of emerging technologies are diverse and extensive. It is beyond the scope of this article to cover them comprehensively and to study all of them in-depth. Therefore, the main focus here is on some of the key concerns and solutions identified in AI policy discussions and lessons from work on other emerging technologies that might be relevant for AI.

With this article, we aim to contribute to the social studies of AI and emerging technologies more broadly with a particular focus on their governance and policy. Our examination of recent policies for AI is complementary to other articles in this special issue on AI discontents (Garvey 2021, this issue) engaging, for example, with critical analysis of AI for good initiatives (Holzmeyer 2021, this issue), emerging AI applications in healthcare (Datta Burton et al. 2021, this issue) and social criticism of computing (Loeb 2021, this issue). While our article takes a 'bird's eye' view on common trends in recent policies

for AI, similar to other contributions in this special issue (Adams 2021; Blackwell 2021, this issue) we recognize the importance of diverse national, regional and local contexts and cultures.

This article proceeds as follows: first, insights from AI policy documents on AI framing, a mix of hopes and concerns and suggested solutions in terms of ethics and regulation are discussed; second, we reflect on the lessons that recent literature on governance, policy and ethics of emerging technologies can offer to AI; and finally, we conclude summarizing the suggestions from this literature that could be relevant for AI governance.

AI policy debates: framing of hopes, concerns and solutions

Recent advances in AI, driven by developments in hardware and big data (see, e.g. Marcus and Davis 2019), have triggered active public debates around the world about the benefits and concerns related to AI and appropriate public policies (Ulnicane et al. forthcoming). According to the OECD, in early 2020 at least 50 countries[1] have developed, or are in the process of developing, national AI strategies.[2] Additionally, international organizations, consultancies and stakeholders have also launched their AI policy documents. We have analysed 49 AI policy documents launched from 2016 to 2018 by national governments, international organizations, consultancies and stakeholder organizations in Europe and the United States (for the list of documents, see the Appendix).[3] In our analysis, we focused on how these policy documents frame AI, associated benefits and concerns, and what mechanisms they suggest for addressing them.

In these policy documents, we find various definitions and understanding of AI as well as concerns about difficulties to define AI (The 2015 panel 2016; Villani 2018; European Commission 2018a; EESC 2017). While it is recognized that the concept of AI has existed for more than 60 years, these documents highlight that the real-world applications have only accelerated during the past decade (Campolo et al. 2017; Crawford and Whittaker 2016; European Commission 2018a Executive Office of the President 2016a; UNI Global Union 2017). Wide-ranging and long-lasting effects of AI across many areas of life and numerous sectors of economy are often discussed. Sometimes they are presented as a reason why AI is different from other technologies, as can be seen in a French document which states that

[1]Most of these AI strategies have been launched in Europe, North America and major Asian economies. Similar geographical concentration can be seen with regard to AI ethics guidelines (Jobin et al 2019).

[2]AI strategies and public sector components https://oecd-opsi.org/projects/ai/strategies/ Last accessed 15 February 2020.

[3]We selected these policy documents according to a set of criteria: they have strong focus on AI; they focus on overarching AI policy rather than AI policy in a specific domain such as education and health; and they address policy questions rather than just ethical principles. For more information on methodology, see Ulnicane et al 2020.

> The key factor setting AI apart from other scientific disciplines is its all-encompassing impact society wide. This is not just some passing trend or media phenomenon, far from it: its implications are posed to be long-lasting and game-changing worldwide. AI is seeping into all sectors – economic, social, political and cultural alike […] The key question now is nothing less than what kind of society we wish to live tomorrow. (Villani 2018, 63)

Attitudes towards these numerous long-term effects of AI represent a complex mix of optimism on how they should improve 'our lives' (Villani 2018) and caution or even concern pointing out that 'it is necessary to look carefully at the ways in which these technologies are being applied now, whom they're benefitting, and how they are structuring our social, economic, and interpersonal lives' (Crawford and Whittaker 2016). Some documents more emphasize expected benefits of AI, others focus more on concerns and yet other ones attempt to balance opportunities and challenges associated with AI. A common feature is that, typically, the impact of AI is seen to be significant across many areas, from jobs, health and education to transport and security. While discussing opportunities, some documents predominantly focus on economic impact in terms of increases in growth and productivity, while others take a broader view considering societal benefits. Latter ones include hope that AI 'will be central to the achievement of the Sustainable Development Goals (SDGs) and could help solve humanity's grand challenges' (ITU 2017).

In context of this paper's focus on emerging technologies, it is interesting that one document explicitly indicates that benefits are related to AI interaction with other technologies, stating that 'in combination with other emerging and converging technologies, AI has the potential to transform our society through better decision-making and improvements to the human condition' (Bowser et al. 2017). This idea sees AI as a critical component of the so-called fourth industrial revolution that includes the fusion of physical, digital and biological technologies (Schwab 2017). Some documents simplify or selectively use examples from previous emerging technologies, for example, by suggesting that 'the history of technology development tends to show that a foundational technology, such as the Internet, can serve everyone' (Accenture 2017) and comparing 'the advent of AI technologies to the development of the commercial internet in the 1990s to provide insight into how policy-makers may champion pro-growth policies while maintaining an appropriate level of oversight and accountability for consumers' (Thierer, Castillo O'Sullivan, and Russell 2017). These two examples of internet history ignore more problematic issues such as digital divide or surveillance.

Concerns and challenges

The policy documents address a wide variety of concerns and challenges associated with AI including ethics, safety, privacy, transparency and accountability,

work, education and skills, inequality and inclusiveness, law and regulations, human rights and fundamental norms and values, governance and democracy, and warfare (e.g. EESC 2017). Some concerns are phrased more as individual level ethical questions about personal integrity, autonomy, dignity and freedom of choice (e.g. EESC 2017), while others address macro-level issues of geopolitics, power and populist political movements (Campolo et al. 2017). Some of the key concerns are summarized in this quote from a European Union document:

> Workers fear they will lose their job because of automation, consumers wonder who is responsible in case a wrong decision is taken by an AI-based system, small companies do not know how to apply AI to their business, AI startups do not find the resources and talent they need in Europe, and international competition is fiercer than ever with massive investments in the US and China. (European Commission 2018b, 1)

Some policy documents state that challenges posed by AI are similar to those raised by other emerging technologies but others argue that such comparisons are not relevant for AI due to its major differences from earlier technologies. A US document states that 'as with most transformative technologies, AI presents some risks in several areas, from jobs and the economy to safety, ethical, and legal questions' (Executive Office of the President 2016a). Similar ideas can be found in the European stakeholder document which tells that 'as with every disruptive technology, AI also entails risks and complex policy challenges in areas such as safety and monitoring, socio-economic aspects, ethics and privacy, reliability, etc.' (EESC 2017) However, a document from the European Parliament suggests that AI is very different from previous technologies highlighting that 'the split between past and future societal models will be such that we cannot expect to take the emergence of information technology, the internet or mobile phones as a starting point for reflection' (European Parliament 2016). This quote from a report from the UK Parliament summarizes some old and new concerns posed by AI:

> Many of the challenges now linked to AI are far from new. For instance, concerns about increasing inequality gaps, stereotypes and biases, shortages of skills, and abuse of power have existed in our society for centuries now. AI is not the creator of these problems. Rather, in many ways, AI is simply resurfacing prevailing problems and urging society to acknowledge their existence and provide solutions. On the other hand, AI technologies are of such high impact and progress at such rapid speeds that some issues developing are authentically new. Some of these include increasingly automated decision-making, potentially catastrophic security threats, technological unemployment, and transformations in current notions of privacy, agency, consent, and accountability. (BIC/APPGAI 2017a, 6)

Proposed solutions: ethics and regulation

How can these old and new concerns raised or reinforced by AI be addressed? Almost every AI policy document calls for an appropriate ethical and legal framework and related activities such as ethics education and research on ethical,

legal and social implications of AI. A US document points out that this call has similarities and differences with other technologies: 'as with any technology, the acceptable uses of AI will be informed by the tenets of law and ethics; the challenge is how to apply those tenets to this new technology, particularly those involving autonomy, agency and control' (Executive Office of the President 2016a). Due to the global reach of AI, several documents suggest that ethics guidelines and sometimes even regulation should be coordinated or adopted at international level (e.g. EGE 2018; Rathenau Institute 2017).

Interestingly, while policy documents often mention AI ethics and law next to each other, suggesting that they are closely linked, on a closer reading and observation of practical developments, the major differences in attitudes to ethics and law emerge. The documents reveal a lot of enthusiasm for ethics codes and guidelines based on human rights and values that would guide the development and use of AI. Only occasionally more cautious notes are mentioned reminding that AI ethics codes 'should be accompanied by strong oversight and accountability mechanisms', that 'more work is needed on how to substantively connect high level ethical principles and guidelines for best practices to everyday development processes' and the 'need to move beyond individual responsibility to hold powerful industrial, governmental and military interests accountable' (Campolo et al. 2017).

If the overall attitude towards ethics guidelines is enthusiastic, then views on regulation are typically more cautious with caveats added. Often it is suggested that while regulation is needed to avoid AI related risks, it is important to 'avoid the risk of over-regulation, as this would critically hamper' innovation (European Commission 2017) and that 'regulation that stifles innovation, or relocates it to other jurisdictions' would be counterproductive (The 2015 panel 2016). Some caveats added to regulation discussion include pointing out that there is a lot of uncertainty about this new technology and 'taking the right approach to laws and regulations on AI will also require a good understanding of what AI can, cannot and will be able to do in the short, medium and long term' (EESC 2017) and that 'attempts to regulate "AI" in general would be misguided, since there is no clear definition of AI (it isn't any one thing), and the risks and considerations are very different in different domains' (The 2015 panel 2016). This cautious approach includes statements that the officials are monitoring the developments, and reviewing and adapting existing legal frameworks (European Commission 2018c). Interestingly, these caveats and cautious statements are hardly ever mentioned when discussing codes of ethics, to which they might also be relevant.

While ethics guidelines and regulations are interconnected as regulations can be a way to protect values and implement voluntary ethical guidelines via binding legislation, there are important political, technical and other types of differences between the two, which is part of the explanation why more has been done in launching AI ethics guidelines than adopting regulation. A recent review of 84 AI ethics guidelines adopted by governments, international

organizations, think tanks, companies and professional organizations (Jobin et al 2019) confirm vibrant developments in this field. Much less activity has taken place in adopting regulation. For example, the European Union has been much faster in launching its AI ethics guidelines, while discussions on appropriate regulation and legislation take much longer (Ulnicane 2021; Vesnic-Alujevic, Nascimento, and Polvora 2020).

Experts have suggested that active work on ethics and little progress on regulation are connected because:

> AI ethics initiatives have thus far largely produced vague, high-level principles and value statements that promise to be action-guiding, but in practice provide few specific recommendations and fail to address fundamental normative and political tensions embedded in key concepts (for example, fairness, privacy). Declarations by AI companies and developers committing themselves to high-level ethical principles and self-regulatory codes nonetheless provide policymakers with a reason not to pursue new regulation. (Mittelstadt 2019, 501)

This idea that companies strategically promote ethics as part of their public relations to delay or avoid binding regulation is echoed by others. For example, Thilo Hagendorff (2020) points out that the focus of companies on ethics, which lacks mechanisms to enforce its normative claims, discourages efforts to create a binding legal framework. A number of stories from Europe and the US (Coeckelbergh and Metzinger 2020; Metzinger 2019; Ochigame 2019) provide empirical insights that support the above claims.

If regulation tends to be avoided or delayed and the role of ethics is contested, then what remains to address the concerns posed by AI? Below we discuss a number of insights from other emerging technologies on their governance, policy and ethics that can provide suggestions for addressing AI related concerns.

Towards a 'good governance' of AI? Lessons from governance, policy and ethics of emerging technologies

Recent years and decades have seen the emergence and development of ideas, concepts and practices for shaping the development and use of emerging technologies towards societal benefits. In particular, work on the inclusion of diverse stakeholders, consideration of various roles of the state, broad societal goals of technology, international cooperation and science diplomacy, computing ethics and Responsible Innovation are discussed here. While each of these ideas, concepts and practices also have some limitations, they can provide useful starting points to broaden and enrich approaches to AI, which faces a number of challenges similar to those for other technologies. These are some opportunities to shape AI development and use nationally and internationally as well as at the level of research projects and laboratories, which need to be further contextualized to specific locations, cultures, temporalities and applications.

Beyond government: governance

AI policy documents (e.g. BIC/APPGAI 2017a) tend to mention governance as a way to facilitate benefits and mitigate risks associated with AI (Ulnicane et al 2020). However, AI documents (and the academic literature) hardly ever define or explain what do they mean by governance. Either governance is mentioned close to government implying that governance is something that government does or it is mentioned next to ethics suggesting that governance is similar to and reinforces ethics. Looking at the social science literature on governance and its role in emerging technologies can help not only to clarify definitions but also to reveal the potential of systematically addressing governance issues for AI.

In the social science literature, a move from government to governance (Pierre and Peters 2000) stands for the involvement of more diverse non-governmental actors, groups and networks from civil society and private sector in decision-making and coordination. According to Susana Borras and Jakob Edler (2014, 13–14), governance is understood 'as the mechanisms whereby societal actors and state actors interact and coordinate to regulate issues of societal concern'. Coordination including a broad range of stakeholders is of high relevance for emerging technologies. Specific governance approaches have been suggested to address the uncertainty surrounding future developments, societal benefits and risks associated with emerging technologies. For example, the notion of 'tentative governance' (Kuhlmann, Stegmaier, and Konrad 2019) emphasizes the role of flexibility, learning and reflexivity in governing emerging technologies.

Elements of governance which emphasize the importance of interaction between government and a broad range of societal actors in decision-making can be found in AI policy documents which suggest multi-stakeholder approaches, inclusion and dialogue to ensure that AI is developed and used according to interests and needs of the society (Ulnicane et al 2020). How these important ideas are implemented is crucial. Some early experiences with multi-stakeholder AI forums (e.g. Metzinger 2019) suggest that they face the risks well-known in social science literature on the collective action, namely the problem that arises when business interests are better organized and resourced, and forums for public participation can be captured by vested interests (Olson 1974). It is important to be aware of such risks and take actions to ensure balanced participation of diverse interests.

Beyond market correction: diverse roles of the state

What is the role of the state in the governance of emerging technologies? This is a highly relevant question for AI because of concerns about the dominant role of big technology companies in this field in which 'the vast majority of the development of AI and all its associated elements (development platforms, data, knowledge and expertise) is in the hands of the "big five" technology

companies' (EESC 2017). If power and resources in this field are concentrated in a small number of large companies, then the question arises what can the state do to influence its development in societal interests? Policy documents suggest a number of activities expected from the state including regulation and supporting research and retraining. For example, a French document assigns the state the role of a key driver for 'laying the foundation for innovation and providing stakeholders with the means and the resources for breaking new ground' (Villani 2018). The AI Now 2017 report with a particular focus on the US situation highlights that the role has shifted during the history of AI:

> In the mid-twentieth century, advanced computing projects tended to be closely associated with the state, and especially the military agencies who funded their fundamental research and development. Although AI emerged from this context, its present is characterized by a more collaborative approach between state agencies and private corporations engaged in AI research and development. (Campolo et al. 2017, 23)

There are important questions to be asked about this collaborative approach between the state and private companies not only in the context of the US administration but more generally: Whose interests do they serve? Do they consider broader societal interests or are they predominantly geared towards economic interests of companies? Are other groups of civil society, consumers and users included in these collaborative relationships? Are these relationships transparent and accountable to the public?

One way to approach these questions is to reflect on the role of the state in the field of emerging technologies. If traditionally the role of the state with respect to new technologies was associated with market correction, then recent work in the innovation policy studies has undertaken a more systematic approach to identify diverse roles that the state plays. Borras and Edler (2020) have identified 13 roles of the state and many of them are highly relevant for addressing some of the AI related concerns. This includes such roles as, for example, a *mitigator* trying actively to reduce the negative effects that arise as a consequence of socio-technical change; an *enabler* of societal engagement encouraging the involvement of stakeholders in participatory processes to define direction of change; and a *moderator* acting as arbitrator or negotiator between different social and political positions among agents regarding the direction of transformation of a socio-technical system (Borras and Edler 2020). Considering the variety of roles that the state can play allows to think more broadly about the opportunities for the state to shape the development and use of AI.

Beyond growth: policy to address societal challenges

What is the objective of AI policy? Many AI policy documents still follow traditional paradigm of technology policy focusing on economic contribution of AI to increase growth and productivity. However, some documents also

include more recent shift in technology policy towards societal objectives mentioning potential contribution of AI to addressing important social challenges, achieving the United Nations Sustainable Development Goals (ITU 2017; Vinnova 2018) and the European Green Deal tackling climate and environmental-related challenges (European Commission 2020). In policy documents, traditional and novel ideas about the objectives of technology policy occasionally are mixed, as it can be seen in this quote from this French document which talks about the paradigm shift towards energy-efficiency but at the same time sticks to traditional discourse of growth and economics:

> A truly ambitious vision for AI should therefore go beyond mere rhetoric concerning the efficient use of resources; it needs to incorporate a paradigm shift toward a more energy-efficient collective growth which requires an understanding of the dynamics of the ecosystems for which this will be a key tool. We should take the opportunity to think of new uses for AI in terms of sharing and collaboration that will allow us to come up with more frugal models for technology and economics. (Villani 2018, 102)

The traditional technology policy paradigm focusing on economic growth and productivity is increasingly challenged in the context of climate change and resource consumption (see e.g. De Saille et al. 2020). An emerging policy paradigm focuses on societal objectives, namely, how new technologies can help to address societal challenges such as climate change, growing inequality, demographic change or resource scarcity (Diercks, Larsen, and Steward 2019). It takes a broader approach to the innovation process including not only industry, academia and government but also civil society and considering not only supply-side of technology but also demand-side of its uses (Diercks, Larsen, and Steward 2019). Of course, many of these ideas are not completely new and some of them can be traced back to the long-standing work on social function of science (Bernal 1939), the Moon and the Ghetto debate on differences in technological progress in different domains (Nelson 1977, 2011), and efforts to prioritize social justice issues in research and development policy-making agendas (Woodhouse and Sarewitz 2007). While societal challenges such as climate change are global in their nature, it is also important to contextualize them and recognize their plurality (Wanzenbock et al. 2020). This new transformative innovation policy paradigm, which brings these ideas together in a systematic framework, can be relevant for AI policy.

Beyond global competition: international research collaboration and science diplomacy

If many countries and international organizations have developed their national AI strategies, then how do they interact with each other? Policy documents and media promote seemingly contradictory discourses of global competition vs. global cooperation on AI (Ulnicane et al. forthcoming). The global competition discourse presents AI development as 'a new space race' where

countries compete to dominate the AI development. This discourse focuses on which countries are making major investments and advances in AI. An example of this global competitiveness narrative can be seen in this quote from the UK document stating that

> A challenging race to make most of the opportunities posed by AI has begun. China, the US, Russia, Canada, Japan, and many more countries have passed ambitious strategies in which they have put AI as a priority in their political agendas. (BIC/APPGAI 2017b, 5)

A number of countries and organizations have expressed their ambitions to be global leaders and defined their approaches to global competition, for example, the European Union has announced that it wants to lead based on its values (Ulnicane 2021). This global competitiveness discourse demonstrates the priority countries and organizations assign to AI but it can also have negative effects such as hampering global collaboration. Focus on global competitiveness among countries has been called 'a dangerous obsession' (Krugman 1994) because it suggests that relationships among countries are 'a zero-sum game' where one country wins and other looses rather than 'a positive-sum game' where the overall size of the pie increases and everyone gains.

At the same time, there are many calls for global cooperation and coordination on AI development and internationally recognized ethical and legal frameworks to address common challenges (e.g. EGE 2018; Rathenau Institute 2017). This quote from the EU High-Level Expert Group on AI summarizes some of the main arguments for global cooperation:

> Just as the use of AI systems does not stop at national borders, neither does their impact. Global solutions are therefore required for the global opportunities and challenges that AI systems bring forth. We therefore encourage all stakeholders to work towards a global framework for Trustworthy AI, building international consensus while promoting and upholding our fundamental rights-based approach. (European Commission 2019, 5)

International cooperation on AI is already taking place in the European Union, OECD, G20 and other organizations (e.g. European Commission 2020). An important part of AI policy development around the world is that countries learn from each other's AI initiatives and strategies (e.g. BIC/APPGAI 2017b) and such cross-country learning is facilitated by international forums such as the World Economic Forum. Suggestions have been made to launch international cooperation initiatives for AI similar to success stories in other fields such as the Intergovernmental Panel on Climate Change (European Commission 2018b) and the European Organization for Nuclear Research CERN for AI experts to discuss and develop technology (European Commission 2017).

The development of international cooperation in AI can benefit from lessons learned on global cooperation among researchers and science diplomacy among policy-makers in other fields. Extensive social science studies of

international scientific collaboration demonstrate the increase in scientists collaborating across national borders due to various reasons such as bringing together diverse types of expertise, seeking reputation, pooling resources for large-scale infrastructure and addressing cross-border problems (e.g. Wagner, Whetsell, and Mukherjee 2019). While such collaborations can bring many benefits, they also encounter difficulties such as high transaction costs due to diverse cultures and long geographical distances (Wagner, Whetsell, and Mukherjee 2019). To facilitate international collaborations among scientists, policy-makers have launched science diplomacy activities at the intersection of science policy and foreign affairs to support joint efforts to address global challenges (e.g. Flink and Rüffin 2019). When designing international cooperation initiatives for AI, which have to address not only complex scientific and technological issues but also often sensitive topics of diverse political and economic systems, regulatory environments and cultural traditions, there are many opportunities to learn what works in other fields and how. Does AI really need a large-scale, single-site physical research infrastructure such as CERN, or would a networked and distributed collaboration be more appropriate in this area?

Beyond biomedical ethics: ethics of computing

The dominant approach to ethics as applied to AI is based on biomedical ethics (see e.g. Mittelstadt 2019). Biomedical ethics arose following the second World War and the Nazi atrocities committed in the concentration camps. The principles were formed during the Nuremberg trials of the war criminals (Freyhofer 2004). These were developed in the World Medical Association's (2008) Helsinki Declaration and formalized through the Belmont report (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). A cornerstone of biomedical ethics is its reliance on mid-level ethical principles (beneficence, non-maleficence, justice and autonomy) (Beauchamp and Childress 2009). This approach of basing applied ethics on mid-level principles is dominant in AI ethics, as Anna Jobin and colleagues' (2019) review shows.

While the adoption of biomedical ethics offers advantages, notably the immediate recognition of approach and the possibility of utilizing existing structures of biomedical ethics (institutional review boards, research ethics committees, established review processes), it is important to point out that the 'principilsm' (Clouser and Gert 1990) of biomedical ethics has always been subject to controversy. Even from within the biomedical field it has been regarded as overly rigid (Klitzman 2015) and there have been questions of its consistency and applicability beyond the biomedical field (Schrag 2010; Stark 2011). A key concern with high relevance to AI is that biomedical ethics is based on the assumption that the underlying research is fundamentally ethically desirable

(understanding disease, finding cures) and the main point of ethics is the protection of patients. While this assumption holds for much of biomedical research, it is arguably much weaker elsewhere, including in AI, where it is not *a priori* obvious that a new technology or innovation *per se* is desirable.

In addition, the focus on biomedical ethics renders invisible a large body of work undertaken on ethical questions of computing. This work started in the early days of digital computing (Wiener 1954) and has become a more formal sub-discipline with dedicated journals and conferences since the 1980s (Moor 1985). There are bodies of work around concepts such as computer ethics (Gotterbarn 1995; Johnson 2001), information ethics (Floridi 1999), digital ethics and others, which are dedicated to ethical aspects of information and communication technologies. Importantly, a dedicated sub-field of Intercultural Digital Ethics brings in diverse cultural and social perspectives to examine ethical issues of digital technologies (Aggarwal 2020). AI, traditionally classified as one field of computer science, has been subject of these studies well before the current AI publicity. However, by relying on biomedical ethics, much of this work has been rendered less visible than it arguably deserves to be, thus limiting its ability to contribute to making AI beneficial.

Many issues at the core of current AI debates such as privacy, autonomy, agency, trust and inclusion have been extensively addressed in the field of computing ethics (Stahl, Timmermans, and Mittelstadt 2016), which have demonstrated the breadth of influence that technology can have on all aspects of life affecting power and politics, economics, education, the environment and other areas. As there is a lot of continuity and path-dependence from computing to AI technologically as well as in terms of its wide-ranging impacts, AI can benefit by building on the work done and knowledge accumulated on ethical and societal issues of computing.

Beyond ethical principles: Responsible Innovation

In the past decade the Responsible Innovation approach has been developed as a way to go beyond ethical principles (Owen and Pansera 2019). It has been applied in particular in Europe over a range of emerging technosciences such as nanotechnology, geoengineering and synthetic biology. While there are many definitions of Responsible Innovation, it is broadly understood that

> *responsible* forms of innovation should be aligned to social needs, be responsive to changes in ethical, social and environmental impacts as a research programme develops, and include the public as well as traditionally defined stakeholders in two-way consultation. (De Saille 2015, 153)

The alignment of innovation to societal needs includes processes and practices of anticipation of potential future developments and impacts of science and innovation, inclusion of diverse stakeholders, reflexivity and responsiveness

(Stilgoe, Owen, and Macnaghten 2013). While the specific term of Responsible Innovation has emerged only about a decade ago, it builds on long-established practices such as technology assessment and public engagement (e.g. Jasanoff 2016). What is novel about the RRI approach is that it aims to bring such practices into a more systematic framework along the four interconnected dimensions of anticipation, inclusion, reflexivity and responsiveness (Stilgoe, Owen, and Macnaghten 2013). Each of these four dimensions offers of a number of techniques to address and implement societal responses into research and innovation. Anticipation dimension includes techniques of foresight, technology and risk assessment, inclusion dimension encompasses approaches such as citizen conferences and focus groups, reflexivity dimension includes multidisciplinary training and collaboration as well as embedded social scientists and ethicists in laboratories, while responsiveness dimension draws on the insights from anticipation, inclusion and reflexivity exercises to design appropriate measures from regulation and funding programmes to moratoriums if need be. This can be a powerful approach to address societal needs by systematically integrating anticipation, inclusion, reflexivity and responsiveness activities and involving a broad range of stakeholders in them.

Some AI policy documents mention Responsible AI without explaining it and making explicit links to the Responsible Innovation approach. Several documents mention Responsible Innovation in passing with one exception (IEEE 2017) of engaging with Responsible Innovation work and its relevance for AI. More in-depth learning from a decade of Responsible Innovation advances and limitations in aligning emerging technosciences with societal needs could contribute to developing systematic frameworks for building socially beneficial AI. So far in other fields Responsible Innovation approach has been mostly applied at the level of research projects and laboratories where, for example, social scientists and ethicists collaborate with technology developers in knowledge co-production (see e.g. Aicardi et al. 2020) but it also has a potential to consider the role of politics and power to further democratize technology development and use (Van Oudheusden 2014). In that way, project and laboratory level Responsible Innovation practices can benefit from being part of a broader governance processes described in previous lessons such as diverse roles of state at national level, new policy paradigms of societal challenges and international cooperation arrangements.

To summarize, the six above-discussed lessons from the literature on governance, policy and ethics of emerging technologies are complementary and reinforcing each other. Insights from the governance literature focus on involving diverse stakeholders in decision-making, which can be facilitated by the state and organized along the dimensions of Responsible Innovation which focus on anticipation, inclusion, reflexivity and responsiveness. These can be geared towards developing and using AI in a way that addresses societal challenges locally and globally. They offer plenty of opportunities to develop many

concrete recommendations for governance of socially beneficial AI. However, concrete recommendations cannot be produced in a universal 'one size fits all' manner. Rather they should be developed in a reflexive, collaborative and inclusive way considering specific contexts, cultures and traditions.

Conclusions: some lessons for a good governance for AI

Recent advances in AI have triggered intensive public debates about potential impact of this technology. Against this background, AI policy documents launched by national governments, international organizations and various stakeholders focus on wide-ranging and long-lasting effects on jobs, politics, economy and society. These effects include positive expectations as well as major concerns about AI effects on individual and societal level. While some of these concerns, such as effects of automation, might seem novel, others related to risks and inequalities are well-known. Typical solutions for dealing with these concerns suggested in AI policy documents focus on ethics and regulation. The role of both of these suggested solutions remains contested.

To look for systematic ways to address concerns associated with AI and facilitate its development and use in socially beneficial ways, in this article we reviewed some recent lessons from literature on governance, policy and ethics of emerging technologies. Such lessons can be relevant for AI which share typical features of emerging technologies such as fast growth, radical novelty, prominent impact and uncertainty. We derive six complementary and reinforcing lessons. First, rather than focus on top-down government decisions, consider governance arrangements that bring together government and diverse non-state groups from civil society and private sector in a balanced and transparent way. Second, rather than assume limited role of the state in market correction, think about diverse roles of the state in mitigating risks, enabling participation of diverse groups and mediating various needs and interests. Third, go beyond traditional goals of technology to support economic growth and productivity, and focus on how technology can address societal challenges. Fourth, rather than being a global race where one country wins and others loose, technology development can be based on international research collaboration supported by science diplomacy efforts. Fifth, rather than develop AI ethics based on principles of medical ethics, learn from a related field of computing ethics that has accumulated extensive knowledge on issues such as privacy, autonomy, agency, trust and inclusion. Sixth, one way to go beyond ethical principles is to learn from the Responsible Innovation approach on how to systematically address societal needs in the development and use of emerging technologies.

Governance of emerging technologies is a highly complex endeavour and there are no magic solutions. However, to address concerns associated with AI and shape its development in socially beneficial ways, recent lessons from

other emerging technologies can offer ideas, concepts and approaches that broaden a range of available options and allow to imagine various ways of achieving desirable objectives. Further work should focus on examining governance needs and arrangements of specific AI applications in diverse contexts at different but interconnected levels from technology development projects and laboratories to national policies and international cooperation arrangements.

Acknowledgements

This paper has benefited from the comments and suggestions on an earlier version presented at the Science in Public 2018 conference in Cardiff, Wales. The authors are grateful to Tonii Leach, Dinesh Mothi and Winter-Gladys Wanjiku who contributed to the document analysis as well as to Juliana Nnadi and Iffat Islam who participated in early discussions on the framing of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 720270 (HBP SGA1), No. 785907 (HBP SGA2) and No. 945539 (HBP SGA3).

Notes on contributors

Dr Inga Ulnicane has extensive international and interdisciplinary experience of research, teaching and engagement in the field of science, technology and innovation governance. Her scientific publications and commissioned reports focus on topics such as research and innovation policies, international research collaboration, Grand societal challenges, and dual use. She has worked at University of Vienna (Austria), University of Twente (Netherlands), University of Latvia and Latvian Academy of Sciences, and has been visiting scientist at University of Manchester (UK) and Georgia Institute of Technology (US). Currently she is at De Montfort University (UK).

Damian Okaibedi Eke (PhD) is a Research Fellow in the EU Human Brain Project in the Centre for Computing and Social Responsibility at De Montfort University, UK. Damian has Computer Ethics background and his current research includes work on responsible data governance of biomedical data, Data Ethics, Ethics of Emerging technologies including AI and ICT4D.

Dr William Knight is a research fellow at the Centre for Computing and Social Responsibility at De Montfort University, Leicester, UK. His research interests include health research, hacking and online activism, research management and ethics compliance, data governance and data protection. Dr Knight is the ethics compliance manager for EU funded Future Emerging Technology flagship: Human Brain Project.

George Ogoh is a Research Fellow at the Centre for Computing and Social Responsibility, De Montfort University. His current role is in the Human Brain Project (HBP) – an EU funded Future Emerging Technology flagship. His research interests cover a wide range of topics in emerging technology ethics including responsible innovation, data governance and data protection.

Bernd Carsten Stahl is Professor of Critical Research in Technology and Director of the Centre for Computing and Social Responsibility at De Montfort University, Leicester, UK. His interests cover philosophical issues arising from the intersections of business, technology, and information. This includes ethical questions of current and emerging information and communication technologies, critical approaches to information systems and issues related to responsible research and innovation.

ORCID

Inga Ulnicane http://orcid.org/0000-0003-2051-1265
Damian Okaibedi Eke http://orcid.org/0000-0002-6210-1283
William Knight http://orcid.org/0000-0001-9818-6277
George Ogoh http://orcid.org/0000-0002-5287-408X
Bernd Carsten Stahl http://orcid.org/0000-0002-4058-4456

References

The 2015 panel. 2016. "Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence." Report of the 2015 Study Panel.

Accenture. 2017. "Embracing Artificial Intelligence." Enabling Strong and Inclusive AI Driven Growth.

Adams, Rachel. 2021. "Decolonising Artificial Intelligence: A Theoretical Critique." *Interdisciplinary Science Reviews* 46 (1).

Aggarwal, Nikita. 2020. "Introduction to the Special Issue on Intercultural Digital Ethics." *Philosophy & Technology* 33 (4): 547–550. doi:10.1007/s13347-020-00428-1.

Aicardi, Christine, Simisola Akintoye, B. Tyr Fothergill, Manuel Guerrero, Gudrun Klinker, William Knight, Lars Klüver, et al. 2020. "Ethical and Social Aspects of Neurorobotics." *Science and Engineering Ethics* 26 (5): 2533–2546. doi:10.1007/s11948-020-00248-8.

Beauchamp, Tom L., and James F. Childress. 2009. *Principles of Biomedical Ethics*. 6th ed. New York: OUP USA.

Bernal, John D. (1939) 1967. *The Social Function of Science*. Cambridge: The MIT Press.

BIC/APPGAI (Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence). 2017a. Governance, Social and Organisational Perspective for AI. 11 September 2017.

BIC/APPGAI (Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence). 2017b. International Perspective and Exemplars. 30 October 2017.

Blackwell, Alan. 2021. "Ethnographic Artificial Intelligence." *Interdisciplinary Science Reviews* 46 (1).

Borras, Susana, and Jakob Edler, eds. 2014. *The Governance of Socio-Technical Systems: Explaining Change*. Cheltenham: Edward Elgar.

Borras, Susana, and Jakob Edler. 2020. "The Roles of the State in the Governance of Socio-Technical Systems' Transformation." *Research Policy* 49 (5): 103971.

Bowser, Anne, Michael Sloan, Pietro Michelucci, and Eleonore Pauwels. 2017. *Artificial Intelligence: A Policy-Oriented Introduction*. Washington, DC: Wilson Briefs. Wilson Center.

Campolo, Alex, Medelyn Sanfilippo, Meredith Whittaker, and Kate Crawford. 2017. *AI Now 2017 Report. AI Now Institute*. New York: New York University.

Clouser, K. Danner, and Bernard, Gert. 1990. "A Critique of Principlism." *Journal of Medicine and Philosophy* 15: 219–236. doi:10.1093/jmp/15.2.219.

Coeckelbergh, Mark, and Thomas Metzinger. 2020. "Europe Needs More Guts When it Comes to AI Ethics." *Tagesspiegel*, April 14.

Crawford, Kate, and Meredith Whittaker. 2016. *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. New York: AI Now Institute.

Datta Burton, Saheli, Tara Mahfoud, Christine Aicardi, and Nikolas Rose. 2021. "Clinical translation of computational brain models: understanding the salience of trust in clinician-researcher relationships." *Interdisciplinary Science Reviews* 46 (1).

De Saille, Stevienna. 2015. "Innovating Innovation Policy: The Emergence of 'Responsible Research and Innovation'." *Journal of Responsible Innovation* 2 (2): 152–168. doi:10.1080/23299460.2015.1045280.

De Saille, Stevienna, Fabien Medvecky, Michiel van Oudheusden, Kevin Albertson, Effie Amanatudou, Timothy Birabi, and Mario Pansera. 2020. *Responsibility Beyond Growth. A Case for Responsible Stagnation*. Bristol: Bristol University Press.

Diercks, Gijs, Henrik Larsen, and Fred Steward. 2019. "Transformative Innovation Policy: Addressing Variety in an Emerging Policy Paradigm." *Research Policy* 48 (4): 880–894.

EESC (European Economic and Social Committee). 2017. "Artificial Intelligence - the Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society." *Opinion*.

EGE (European Group on Ethics in Science and New Technologies). 2018. "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems."

European Commission. 2017. *AI Policy Seminar: Towards and EU Strategic Plan for AI*. Brussels: Digital Transformation Monitor.

European Commission. 2018a. *Artificial Intelligence: A European Perspective*. Luxembourg: Publications Office of the European Union.

European Commission. 2018b. "Coordinated Plan on Artificial Intelligence." Communication COM(2018) 795 final. Brussels 7.12.2018.

European Commission. 2018c. "Artificial Intelligence for Europe. Communication." COM (2018) 237 final. Brussels 25.4.2018.

European Commission. 2019. "Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission." Brussels. 8.4.2019.

European Commission. 2020. "On Artificial Intelligence – A European Approach to Excellence and Trust." White Paper. COM(2020) 65 final. Brussels 19.2.2020.

European Parliament. 2016. "European Civil Law Rules in Robotics." Study for the JURI Committee.

Executive Office of the President. 2016a. *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, DC: National Science and Technology Council.

Flink, Tim, and Nicolas Rüffin. 2019. "The Current State of the Art of Science Diplomacy." In *Handbook on Science and Public Policy*, edited by Dagmar Simon, Stefan Kuhlmann, Julia Stamm, and Weert Canzler, 104–121. Cheltenham: Edward Elgar.

Floridi, Luciano. 1999. "Information Ethics: On the Philosophical Foundation of Computer Ethics." *Ethics and Information Technology* 1: 33–52.

Freyhofer, Horst H. 2004. *The Nuremberg Medical Trial: The Holocaust and the Origin of the Nuremberg Medical Code*, 2nd Revised ed. New York: Peter Lang.
Garvey, Colin Shunryu. 2021. "Unsavory Medicine for Technoscientific Civilization: Introduction to AI & its Discontents." *Interdisciplinary Science Reviews* 46 (1).
Gotterbarn, Donald. 1995. "Computer Ethics – Responsibility Regained." In *Computers, Ethics and Social Values*, edited by Deborah G. Johnson and Helen Nissenbaum, 18–24. Upper Saddle River: Prentice Hall.
Hagendorff, Thilo. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30: 99–120. doi:10.1007/s11023-020-09517-8.
Holzmeyer Cheryl. 2021. "Beyond "AI for Social Good" (AI4SG): Social Transformations - Not Tech-Fixes - for Health Equity." *Interdisciplinary Science Reviews* 46 (1).
IEEE. 2017. "Ethically Aligned Design. A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems." Version 2 – for Public Discussion.
ITU (International Telecommunication Union). 2017. *AI for Good Global Summit Report 2017*, Geneva, 7–9 June 2017.
Jasanoff, Sheila. 2016. *The Ethics of Invention: Technology and the Human Future*. New York: W.W. Norton.
Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1: 389–399. doi:10.1038/s42256-019-0088-2.
Johnson, Deborah G. 2001. *Computer Ethics*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.
Juma, Calestous. 2016. *Innovation and Its Enemies. Why People Resist New Technologies*. Oxford: Oxford University Press.
Klitzman, Robert. 2015. *The Ethics Police?: The Struggle to Make Human Research Safe*. 1st ed. Oxford; New York: OUP USA.
Krugman, Paul. 1994. "Competitiveness: A Dangerous Obsession." *Foreign Affairs* 73 (2): 28–44.
Kuhlmann, Stefan, Peter Stegmaier, and Kornelia Konrad. 2019. "The Tentative Governance of Emerging Science and Technology – a Conceptual Introduction." *Research Policy* 48 (5): 1091–1097. doi:10.1016/j.respol.2019.01.006.
Loeb, Zachary. 2021. "The Lamp and the Lighthouse: Joseph Weizenbaum, contextualizing the critic." *Interdisciplinary Science Reviews* 46 (1).
Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence we Can Trust*. New York: Pantheon Books.
Metzinger, Thomas. 2019. "Ethics Washing Made in Europe." *Tagesspiegel*, April 8.
Mittelstadt, Brent. 2019. "Principles Alone Cannot Guarantee Ethical AI." *Nature Machine Intelligence* 1 (11): 501–507. doi:10.1038/s42256-019-0114-4.
Moor, James H. 1985. "What is Computer Ethics." *Metaphilosophy* 16: 266–275.
The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont Report - Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Department of Health, Education, and Welfare.
Nelson, R. Richard. 1977. *The Moon and the Ghetto. An Essay on Public Policy Analysis*. New York: W.W. Norton & Company.
Nelson, R. Richard. 2011. "*The Moon and The Ghetto* Revisited." *Science and Public Policy* 38 (9): 681–690.
Ochigame, Rodrigo. 2019. "How Big Tech Manipulates Academia to Avoid Regulation." *The Intercept*, December 20. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.
Olson, Mancur. 1974. *The Logic of Collective Action: Public Goods and the Theory of Groups*. 2nd ed. Cambridge, MA: Harvard University Press.

Owen, Richard., and Mario. Pansera. 2019. "Responsible Innovation and Responsible Research and Innovation." In *Handbook on Science and Public Policy*, edited by Dagmar. Simon, Stefan. Kuhlmann, Julia. Stamm, and Weert. Canzler, 26–48. Cheltenham: Edward Elgar.

Pierre, Jon, and B. Guy Peters. 2000. *Governance, Politics and the State*. Houndmills: Macmillan.

Rathenau Institute. 2017. "Human Rights in the Robot Age. Challenges Arising from the Use of Robotics, Artificial Intelligence, and Virtual and Augmented Reality." Report for the Parliamentary Assembly of the Council of Europe.

Rotolo, Daniele, Diana Hicks, and Ben R. Martin. 2015. "What is an Emerging Technology?" *Research Policy* 44 (10): 1827–1843. doi:10.1016/j.respol.2015.06.006.

Schrag, Zachary. M. 2010. *Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965–2009*. 1st ed. Baltimore: Johns Hopkins University Press.

Schwab, Klaus. 2017. *The Fourth Industrial Revolution*. London: Penguin.

Stahl, Bernd Carsten, Job Timmermans, and Brent Daniel Mittelstadt. 2016. "The Ethics of Computing: A Survey of the Computing-Oriented Literature." *ACM Computing Surveys* 48 (4): 55:1–55:38. doi:10.1145/2871196.

Stark, Laura. 2011. *Behind Closed Doors: IRBs and the Making of Ethical Research*. 1st ed. Chicago: University of Chicago Press.

Stilgoe, Jack, Richard Owen, and Phil Macnaghten. 2013. "Developing a Framework for Responsible Innovation." *Research Policy* 42 (9): 1568–1580. doi:10.1016/j.respol.2013.05.008.

Thierer, Adam, Andrea Castillo O'Sullivan, and Raymond Russell. 2017. *Artificial Intelligence and Public Policy. Report*. Arlington: Mercatus Center, George Mason University.

Ulnicane, Inga, William Knight, Tonii Leach, Bernd Carsten Stahl, and Winter-Gladys Wanjiku. 2020. Framing Governance for a Contested Emerging Technology: Insights from AI Policy. *Policy and Society*. doi:10.1080/14494035.2020.1855800.

Ulnicane, Inga. 2021. "Artificial Intelligence in the European Union: Policy, Ethics and Regulation." In *Routledge Handbook of European Integrations*, edited by Thomas Hoerber, Ignazio Cabras, and Gabriel Weber. Routledge.

Ulnicane, Inga, William Knight, Tonii Leach, Bernd Carsten Stahl, and Winter-Gladys Wanjiku. Forthcoming. "Governance of Artificial Intelligence: Emerging International Trends and Policy Frames." In *Global Politics of Artificial Intelligence*, edited by Maurizio Tinnirello. Boca Raton: CRC Press.

UNI Global Union. 2017. "Top 10 Principles for Ethical Artificial Intelligence." The Future World of Work.

Van Oudheusden, Michiel. 2014. "Where Are the Politics in Responsible Innovation? European Governance, Technology Assessments, and Beyond." *Journal of Responsible Innovation* 1 (1): 67–86. doi:10.1080/23299460.2014.882097.

Vesnic-Alujevic, Lucia, Susana Nascimento, and Alexandre Polvora. 2020. "Societal and Ethical Impacts of Artificial Intelligence: Critical Notes on European Policy Frameworks." *Telecommunications Policy* 44 (6): 101961.

Villani, Cedric. 2018. *For a Meaningful Artificial Intelligence. Towards a French and European Strategy*. Paris.

Vinnova. 2018. *Artificial Intelligence in Swedish Business and Society*. Stockholm: Vinnova.

Wagner, Caroline, Travis Whetsell, and Satyam Mukherjee. 2019. "International Research Collaboration: Novelty, Conventionality, and Atypicality in Knowledge Recombination." *Research Policy* 48: 1260–1270.

Wanzenbock, Iris, Joeri H. Wesseling, Koen Frenken, Marko P. Hekkert, and K. Matthias Weber. 2020. "A Framework for Mission-Oriented Innovation Policy: Alternative Pathways Through the Problem-Solution Space." *Science and Public Policy*. doi:10.1093/scipol/scaa027.

Wiener, Norbert. 1954. *The Human Use of Human Beings*. New York: Doubleday.

Woodhouse, Edward J., and Daniel Sarewitz. 2007. "Science Policies for Reducing Societal Inequities." *Science and Public Policy* 34 (2): 139–150.

World Medical Association. 2008. "Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects."

Appendix

AI policy documents analysed (in alphabetical order)

1. Accenture (2017) Embracing artificial intelligence. Enabling strong and inclusive AI driven growth.
2. Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence (2017a) APPG AI Findings 2017.
3. Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence (2017b) Governance, Social and Organisational Perspective for AI. 11 September 2017.
4. Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence (2017c) Inequality, Education, Skills, and Jobs. 16 October 2017.
5. Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence (2017d) International Perspective and Exemplars. 30 October 2017.
6. Big Innovation Centre/All-Party Parliamentary Group on Artificial Intelligence (2017e) What is AI? A theme report based on the 1st meeting of the All-Party Parliamentary Group on Artificial Intelligence. 20 March 2017.
7. Bowser, A., M. Sloan, P. Michelucci and E. Pauwels (2017) Artificial Intelligence: A Policy-Oriented Introduction. Wilson Briefs. Wilson Center.
8. Campolo, A, M.Sanfilippo, M.Whittaker and K.Crawford (2017) AI Now 2017 Report. AI Now Institute, New York University.
9. CNIL (2017) Algorithms and artificial intelligence: CNIL's report on the ethical issues.
10. Crawford, K. and M.Whittaker (2016) The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term. AI Now Institute.
11. EDPS (2016) Artificial Intelligence, Robotics, Privacy and Data Protection. Room document for the 38th International Conference of Data Protection and Privacy Commissioners.
12. European Commission (2017) AI Policy Seminar: Towards and EU strategic plan for AI. Digital Transformation Monitor.
13. European Commission (2018a) Artificial Intelligence: A European Perspective.
14. European Commission (2018b) Artificial Intelligence for Europe. Communication.
15. European Commission (2018c) Coordinated Plan on Artificial Intelligence. Communication.
16. European Economic and Social Committee (2017) Artificial Intelligence - The consequences of Artificial intelligence on the (digital) single market, production, consumption, employment and society. Opinion.
17. European Group on Ethics in Science and New Technologies (2018) Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems.

18. European Parliament (2016) European Civil Law Rules in Robotics. Study for the JURI Committee.
19. European Parliament (2017) Report with recommendations to the Commission on Civil Law Rules on Robotics.
20. European Parliament (2018) Understanding Artificial Intelligence. Briefing EPRS.
21. Executive Office of the President (2016a) Artificial Intelligence, Automation, and Economy, Report.
22. Executive Office of the President (2016b) Preparing for the future of artificial intelligence. National Science and Technology Council Committee on Technology.
23. Executive Office of the President (2016c) The National Artificial Intelligence research and development Strategic Plan. National Science and Technology Council. Networking and Information Technology Research and Development Subcommittee.
24. Future of Humanity Institute et al (2018) The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation.
25. Government Office for Science (2016) Artificial Intelligence: opportunities and implications for the future of decision making.
26. Hall, W. and J.Pesenti (2017) Growing the Artificial Intelligence Industry in the UK.
27. HM Government (2018) Artificial Intelligence Sector Deal. 26 April 2018.
28. House of Commons Science and Technology Committee (2016) Robotics and artificial intelligence. Fifth report of session 2016-17.
29. House of Lords (2018) AI in the UK: ready, willing and able?
30. IEEE (2017) Ethically aligned design. A vision for prioritizing human well-being with autonomous and intelligent systems. Version 2 – for public discussion.
31. IEEE European Public Policy Initiative (2017) Artificial Intelligence: Calling on Policy--Makers to Take a Leading Role in Setting a Long-Term AI Strategy. Position Statement.
32. IEEE-USA (2017) Artificial Intelligence Research, Development & Regulation. Position Statement.
33. Information Commissioner's Office (2017) Big data, artificial intelligence, machine learning and data protection. Data Protection Act and General Data Protection Regulation.
34. International Telecommunication Union (2017) AI for Good Global Summit Report 2017, Geneva, 7-9 June 2017.
35. IPPR (2017) Managing automation: Employment, inequality and ethics in the digital age. Discussion Paper.
36. Ministry of Economic Affairs and Employment (2017) Finland's Age of Artificial Intelligence.
37. Ponce Del Castillo, A. (2017) A Law on Robotics and Artificial Intelligence in the EU? Foresight Brief. European Trade Union Institute ETUI.
38. Rathenau Institute (2017) Human Rights in the Robot Age. Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality. Report for the Parliamentary Assembly of the Council of Europe.
39. SGPAC (2017) Governance, Risk & Control: Artificial Intelligence. Effective Deployment, Management and Oversight of Artificial Intelligence (AI). Version 1.0. 22 March 2017. SGPAC Consulting & Advisory.
40. Tata Leading the Way with Artificial Intelligence: The Next Big Opportunity for Europe. TCS Global Trend Study – Europe. Tata Consultancy Services.
41. The 2015 panel (2016) Artificial Intelligence and life in 2030. One hundred year study on artificial intelligence. Report of the 2015 study panel.
42. The Federal Government (2018) Artificial Intelligence Strategy. November 2018
43. The Royal Society (2017) Machine learning: the power and promise of computers that learn by example.

44. Thierer, A., A. Castillo O'Sullivan, and R. Russell (2017) Artificial Intelligence and Public Policy. Report. Mercatus Center, George Mason University.
45. UNI Global Union (2017) Top 10 Principles for ethical artificial intelligence. The future world of work.
46. Villani, C. (2018) For a meaningful artificial intelligence. Towards a French and European Strategy.
47. Vinnova (2018) Artificial Intelligence in Swedish business and society.
48. Whittaker, M., K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. Myers West, R. Richardson, J. Schultz, O. Schwartz (2018) AI Now Report 2018.
49. World Economic Forum (2018) Artificial Intelligence for the Common Good. Sustainable, Inclusive and Trustworthy. White Paper for attendees of the WEF 2018 Annual Meeting.