






## Framing governance for a contested emerging technology: insights from AI policy

Inga Ulnicane , William Knight , Tonii Leach , Bernd Carsten Stahl   
and Winter-Gladys Wanjiku 

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

### ABSTRACT

This paper examines how the governance in AI policy documents have been framed as way to resolve public controversies surrounding AI. It draws on the studies of governance of emerging technologies, the concept of policy framing, and analysis of 49 recent policy documents dedicated to AI which have been prepared in the context of technological hype expecting fast advances of AI that will fundamentally change economy and society. The hype about AI is accompanied by major public controversy about positive and negative effects of AI. Against the backdrop of this policy controversy, governance emerges as one of the frames that diagnoses the problems and offers prescriptions. Accordingly, the current governance characterized by oligopoly of a small number of large companies is indicated as one of the reasons for problems such as lack of consideration of societal needs and concerns. To address these problems, governance frame in AI policy documents assigns more active and collaborative roles to the state and society. Amid public controversies, the state is assigned the roles of promoting and facilitating AI development while at the same time being a guarantor of risk mitigation and enabler of societal engagement. High expectations are assigned to public engagement with multiple publics as a way to increase diversity, representation and equality in AI development and use. While this governance frame might have a normative appeal, it is not specific about addressing some well-known challenges of the proposed governance mode such as risks of capture by vested interests or difficulties to achieve consensus.

### KEYWORDS

Artificial Intelligence; governance; policy; emerging technology

## Introduction

Recent advances in Artificial Intelligence (AI) driven by increases in availability of data and computational power have led to widespread applications of AI in numerous sectors from healthcare and social services to education and security (see e.g. Coeckelbergh, 2020; Dignum, 2019; Marcus & Davis, 2019). New AI advances and applications have sparked major public controversies about benefits and concerns related to AI development and use. While AI is expected to bring major societal and economic benefits, its development and use has also raised major concerns about concentration of power, inequality and discrimination (see e.g. Broussard, 2018; Fry, 2018; O'Neill, 2016;

**CONTACT** Inga Ulnicane  [inga.ulnicane@dmu.ac.uk](mailto:inga.ulnicane@dmu.ac.uk)

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Taeihagh, 2020). In recent years, controversies surrounding AI have featured prominently in popular media fuelled by rhetoric of AI as ‘the next space race’ (see e.g. Allen & Husain, 2017; Lee, 2018), alarmist statements about AI risks from popular figures such as Stephen Hawking and Elon Musk (Galanos, 2019) as well as high profile scandals such as Cambridge Analytica.

Public discourse surrounding AI controversies presents a number of frames that present solutions and ways to overcome conflicting views (Rein & Schon, 1996). One often invoked frame focusses on ethics. According to ethics frame, adoption of guidelines outlining ethical principles will allow to mitigate risks and increase benefits associated with AI use. Over recent years, governments, companies, international and other organizations around the world have actively adopted ethics guidelines focusing on principles such as transparency, privacy, justice and fairness (see e.g. Floridi et al., 2018; Jobin, Ienca, & Vayena, 2019; Vesnic-Alujevic, Nascimento, & Polvora, 2020). Widespread enthusiasm about AI ethics as a solution to concerns and controversies have received some criticism which has not only highlighted vagueness of and difficulties to implement ethical principles but also led to suggestions that it might serve vested interests which prefer to avoid or delay binding regulation (Hagendorff, 2020; Mittelstadt, 2019).

This article focuses on another frame invoked to resolve AI controversies, namely, that of governance. This frame can be found in policy rhetoric but so far has received less attention from researchers. In the context of the special issue ‘Governance of AI and Robotics’, this article examines how governance frame is used in policy rhetoric as a way to overcome controversies surrounding AI development and use. To do that, this article draws on data from a larger study on AI policy (Ulnicane, 2021; Ulnicane, Eke, Knight, Ogoh, & Stahl, 2021; Ulnicane, Knight, Leach, Stahl & Wanjiku *forthcoming*) which analysed 49 AI policy documents published in recent years (2016–2018) by national governments, international organisations, consultancies, think tanks and civil society organisations in the European Union and the United States.

In these policy documents, AI is often used as a broad umbrella term. As pointed out by the European Economic and Social Committee (EESC), ‘there is no single accepted and rigid definition of AI. AI is a catch-all term for a large number of sub(fields) such as: cognitive computing (algorithms that reasons and understand at a higher (more human) level), machine-learning (algorithms that can teach themselves tasks), augmented intelligence (cooperation between human and machine) and AI robotics (AI embedded in robots)’ (EESC 2017). The report ‘Artificial Intelligence and life in 2030’ views AI ‘as a branch of computer science that studies the properties of intelligence by synthesizing intelligence’ (The 2015 panel 2016), while the Swedish report defines AI as ‘the ability of a machine to imitate intelligence human behaviour’ (Vinnova 2018). According to the European Commission (2018b), AI ‘refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)’.

To examine the question – how governance frame is used in policy rhetoric as a way to overcome controversies surrounding AI development and use, this article proceeds as follows. First, insights from the studies of governance of technologies are introduced to

inform the document analysis afterwards. Second, concept of framing as well as methods and data sources used for analysis of policy documents are discussed. Third, empirical findings from the document analysis of emerging policy frame of AI governance are discussed. Finally, conclusions and some questions for future research are presented.

## Governance of technologies

Interdisciplinary studies of governance and policy of science, technology and innovation provide important insights about governance of technology which are relevant for examination of AI governance framing. While governance of technologies is a diverse research field drawing on concepts and ideas from Science and Technology Studies, public policy, innovation studies, economics and political science, one of its key features is recognition of the mutual shaping and co-creation of technology and society. As eloquently formulated by Sheila Jasanoff (2016), technologies

‘are not merely tools for achieving practical ends but devices with which modern societies explore and create potentially more liberating and meaningful designs for future living. Through technology, human societies articulate their hopes, dreams, and desires while also making material instruments for accomplishing them. Collective visions and aspirations, moreover, change and evolve as societies become habituated to new technologies and use them to pursue altered understandings and purposes. Technological choices are, as well, intrinsically political: they order society, distribute benefits and burdens, and channel power’ (Jasanoff, 2016, pp. 242–243).

To emphasize the idea that technologies are not operating in isolation but are co-created with and deeply embedded in society, terms such as ‘socio-technical systems’ are used (Borras and Edler 2014). Definitions of governance of technology and socio-technical systems often build on the idea of a shift from government to governance (e.g. Pierre & Peters, 2000) characterised by participation of a wider range of interacting and interdependent actors, such as civil society organisations, as well as complementing traditional hierarchies with coordination, for example, via networks.

In their work on the governance of socio-technical systems, Borras and Edler (2014, p. 13) highlight blurring boundaries between state and non-state action as a key characteristic of governance that resonate with Radu’s notion on hybrid governance that involves ‘mutual dependency’ between the public and private sectors (Radu, this issue). According to Borras and Edler, ‘in contrast to the approach on “steering” by the state, the governance approach of the early 1980s pointed at the fact that an increasing number of empirical cases showed that political institutions have limited capabilities to “steer” because social systems have dynamics that are determined by all kinds of institutional, cultural, technological and other factors. This observation led to another commonality of governance approaches in political science, namely, that systems are not influenced or changed through political institutions alone (as in “steering”), but by the interplay of societal and state actors’ (Borras and Edler 2014, pp. 12–13). They define governance ‘as the mechanisms whereby societal actors and state actors interact and coordinate to regulate issues of societal concern’ (Borras and Edler 2014, pp. 13–14) and ‘as intentional interaction (and coordination) towards some end’ (Borras and Edler 2014, p. 15). They distinguish two dimensions of governance, namely nature of the actors (state or non-state actors) and the nature of the coordination (ranging from hierarchical to non-

hierarchical). This leads to four governance modes: command and control (driven by state actors, hierarchical), state as *primus inter pares* (driven by state actors, non-hierarchical coordination), oligopoly (driven by societal actors, hierarchical) and self-regulation (driven by societal actors, non-hierarchical coordination) (Borras and Edler 2014, p. 13).

Accordingly, the two overarching themes in the studies of governance of technology and socio-technical systems are the role of the state and the role of society. Traditionally, the role of the state in socio-technical systems has been seen as rather limited focusing on market correction. Recent literature puts forward novel ideas of ‘the entrepreneurial state’ (Mazzucato, 2013) and ‘the collaborative state’ (Stilgoe, 2020). Borras and Edler (2020) highlight multiple and diverse roles of the state in the governance of socio-technical systems. Understanding ‘state’ as governmental action including notions of ‘government’ and ‘public policy’, they identify and define 13 roles of the state ranging from ‘observer’ monitoring the developments and trends and ‘watchdog’ ensuring compliance with norms to ‘enabler of societal engagement’ and ‘initiator’ of transformation of the socio-technical systems (Borras & Edler, 2020). They highlight the complexity of the role of the state, which is not a unitary actor and assumes several roles at once.

While a broader understanding of the role of the state in the governance of technology is relatively recent, the field Science and Technology Studies has traditionally understood governance of science and technology as public engagement (see e.g. Lyall & Tait, 2019; Stilgoe, Lock, & Wilsdon, 2014). In established approaches of public engagement with science and technology, stakeholder dialogue and inclusion has been assigned an important role in opening-up, democratizing and legitimizing scientific and technological development and deployment. Recent focus on the concept of responsible innovation (Stilgoe, 2015; Stilgoe, Owen, & Macnaghten, 2013) has reinforced the idea that science and technology should be aligned to societal needs and put forward an approach organized along four dimensions of anticipation, reflexivity, inclusion and responsiveness to incorporate public interests in shaping research and innovation. While importance of aligning scientific and technological development and use to societal values is widely acknowledged, the current practices are often seen as inadequate in achieving the stated goals (e.g. Jasanoff, 2016) and the need to recognize power dynamics, diversity of societal interests, the role of vested interests as well as policy and political aspects of public deliberations is emphasized.

The broadening of the roles of the state and society in the governance of science and technology is reflected in the shifting paradigm of science, technology and innovation policy. If traditional science, technology and innovation policy has a rather narrow focus on economic competitiveness, growth and jobs as well as national prestige and strategic priorities, then according to an emerging paradigm, the main policy objective is to address societal challenges in areas such as climate change and health (Diercks, Larsen, & Steward, 2019). This broader paradigm also incorporates shift towards including a wide variety of actors in innovation process e.g. from civil society (rather than just focus on firms, academia and government) and focus on demand-side of technology in addition to supply-side. This paradigm shift means that governance of technology is changing in terms of its objectives, actors involved and activities undertaken.

An important element of understanding the current framing of AI governance is seeing it as an emerging technology. While the term AI was coined already in the 1950s

(i.e. at the 1956 Dartmouth conference), major advances have taken place more recently due to considerable increase in computational power and availability of data (see e.g. Coeckelbergh, 2020; Dignum, 2019; Marcus & Davis, 2019). Thus, AI today exhibits typical characteristics of an emerging technology such as radical novelty, relatively fast growth, prominent impact, and uncertainty and ambiguity (Rotolo, Hicks, & Martin, 2015).

Emerging technologies are often associated with hypes and expectations, which are important also for understanding dynamics surrounding AI today. One useful approach to hypes and expectations focuses on the performative function of hypes and expectations (e.g. Van Lente, Spitters, & Peine, 2013). This approach is ‘not interested in hypes as more or less accurate forecasts, but as collectively pursued explorations of the future that affect activities in the present. While the early and high-rising expectations that characterize hype hardly ever materialize precisely as foreseen, they structure and shape the materializations that eventually occur’ (Van Lente et al, 2013, p. 1616). Hypes are closely related to expectations, as ‘hypes are constituted by expectations at different levels’ (Van Lente et al, 2013, p. 1616) and expectations – both positive as well as negative – shape emerging technologies. According to Van Lente et al. (2013, p. 1616), expectations ‘guide the activities of innovative actors by setting agendas; they provide legitimacy and thus help to attract financing and enrol actors; and they, while often spread through spoken and written words, may materialize in experiments and prototypes’. Importantly for governance ‘when more and more actors share similar expectations, the promises inherent to these expectations are gradually translated into requirements, guidelines and specifications regarding the new technology’ (Van Lente et al, 2013, p. 1616).

Governance of emerging sciences and technologies faces special challenges due to uncertainties of their future developments, the large number of stakeholders involved and their interdependencies, societal benefits and risks (Kuhlmann, Stegmaier, & Konrad, 2019; Radu, 2020). To address the specific needs of emerging sciences and technologies, Kuhlmann et al. (2019, p. 1091) suggest the notion of tentative governance ‘when public and private interventions are designed as a dynamic process that is prudent and preliminary rather than assertive and persistent. Tentative governance typically aims at creating spaces for probing and learning instead of stipulating definitive targets’. As opposed to more definitive modes of governance, tentative governance maintains flexibility and is open to experimentation, learning and reflexivity.

This short review of the studies of governance of technologies suggests the key elements to be examined in empirical analysis of AI policy documents: the roles of government and societal actors. For further insights on AI governance see other articles in this special issue: Dickinson, Smith, Carey, and Carey (2020), Gahnberg (2020), Leiman (2020), Radu (2020), and Tan and Taeihagh (2020).

## Methods, data and approach

To study the role that the framing of governance plays in overcoming controversies surrounding AI, this paper analyses AI policy documents. Policy documents here ‘are treated as vehicles of messages, communicating or reflecting official intentions, objectives, commitments, proposals, “thinking”, ideology and responses to external events’

(Freeman & Maybin, 2011, p. 157). Thus, AI policy documents reveal the rhetorical frame of AI governance.

To analyse controversies surrounding AI, this paper uses a policy framing approach. According to Rein and Schon (1996) policy frames are narratives, namely ‘diagnostic/prescriptive stories that tell, within a given issue terrain, what needs fixing and how it might be fixed’ (Rein & Schon, 1996, p. 89). In that way framing plays a crucial role in understanding and overcoming controversies. When thinking about framing as diagnosis of problems and suggested prescriptions (Rein & Schon, 1996), it is useful ‘to recognize the non-innocence of how “problems” get framed within policy proposals, how the frames will affect what can be thought about and how this affects possibilities for action’ (Bacchi, 2000, p. 50). Thus, according to Carol Bacchi (2000), analysis of framing also includes revealing silences and kinds of politics are hidden in the framing.

Of particular relevance for this article are ‘rhetorical frames’, which Rein and Schon describe as ‘constructed from the policy-relevant texts that play important roles in policy discourse, where the context is one of debate, persuasion, or justification’ (Rein & Schon, 1996, p. 90). For them, ‘framing is a way of selecting, organizing, interpreting, and making sense of a complex reality to provide guideposts for knowing, analysing, persuading, and acting. A frame is a perspective from which an amorphous, ill-defined, problematic situation can be made sense of and acted on’ (Rein & Schon, 1993, p. 146) and in such frames ‘facts, values, theories, and interests are integrated’ (Rein & Schon, 1993, p. 145).

To analyse framing of AI governance, the authors of this paper draw on their dataset of 49 AI policy documents (see Annex 1) published by a range of governance actors: national governments, international organisations, civil society organisations, think tanks and consultancies. Due to feasibility reasons, it was decided to limit the geographical focus of this paper to documents available in English (either original or official translation) from the European Union (EU) and the United States (US) as well as international and regional organisations from these regions. The first AI policy documents were published in 2016 (see e.g. Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018; Galanos, 2019). This dataset contains documents published from 2016 until the end of 2018; the cut-off date was chosen for pragmatic reasons i.e., to allow time for preparing this paper.

What is a policy document is far from obvious. There are no centralised databases of policy documents covering all relevant items, as it is the case with the Web of Science and Scopus for scientific articles. The authors developed a set of criteria for constructing AI policy document dataset for this article:

- Policy document should have a strong focus on AI, i.e. AI should be mentioned in the title and/or abstract/summary/preface of the document;
- Overall focus of document is on overarching AI policy rather than AI in a specific sector/issue/domain, e.g. jobs, education, financial services, health;
- It should be an independent, stand-alone, self-contained document rather than e.g. submission for/response to a public consultation for a larger document;
- Policy document should address national, regional and/or global level policy and governance questions and thus should be different from ethics guidelines that focus

on ethical principles and are analysed elsewhere (see e.g. Floridi et al., 2018; Jobin et al., 2019).

Search methods for policy documents in this study included following AI policy debates and events, identifying the main actors and snowballing method of reading relevant documents and articles to identify further references. Additionally, existing online sites covering AI policy were perused including

- AI watch (EU) [https://ec.europa.eu/knowledge4policy/ai-watch\\_en](https://ec.europa.eu/knowledge4policy/ai-watch_en)
- Charlotte Stix (Cambridge, CFI) <https://www.charlottestix.com/ai-policy-resources>
- Future of Life Institute <https://futureoflife.org/ai-policy/>
- Politics + AI <https://medium.com/politics-ai>

These sites provide useful resources but they focus more on national governments and international/regional organisations and less on other important governance actors such as think tanks, civil society organisations and consultancies.

The advantage of the dataset used for this article is inclusion of documents from a range of governance actors from national governments to think tanks. However, due to the lack of a centralised database of the relevant items, there is a risk of overlooking further documents. To mitigate this risk, the authors followed AI policy discussions closely aiming to cover diverse relevant actors.

The AI policy documents included in the dataset (Annex 1) share some commonalities (e.g. launched approximately at the same time, focusing on similar opportunities and concerns associated with AI) but also have a number of differences, e.g. in their length and format. While some of them draw on lengthy consultation processes (e.g. House of Lords 2018), others are outputs from single consultation meetings (e.g. Big Innovation Centre/All-party Parliamentary Group on Artificial Intelligence 2017b; 2017c; 2017d; 2017e). Authorship of policy documents differs as well. While many documents only indicate a name of organisation as an author, others explicitly state name/s of expert/s who have prepared these documents for governments (e.g. Hall and Pesenti 2017; Villani 2018) and think tanks (e.g. Bowser et al 2017; Ponce Del Castillo 2017; Whittaker et al 2018). For future studies, it would be interesting to analyse what different types of authorship – organisational and individual – mean for accountability and legitimacy of policy priorities and recommendations.

This article draws on qualitative content analysis of these documents. While the documents address a broad range of topics including the definition of AI, associated opportunities and threats and policy recommendations (Ulnicane, 2021; Ulnicane et al., 2021; forthcoming), this article focusses specifically on the framing of governance. This analysis is complementary to the article by Radu (2020) in this special issue which focusses on how AI national strategies address private–public partnerships with a more specific focus on the elements of hybrid governance to analyse the relationship between the public and private sector in AI governance, which are reflected in the AI national strategies through new institutional dynamics and redefinition of roles between these entities.

The following analysis of AI policy documents is informed by the review on governance of technologies above: the role of government and societal actors. Building our

document analysis on the insights from studies on governance of technologies is necessary because while the term governance is used in the policy documents, there it is hardly ever explained or defined. Typically, policy documents mention ‘governance’ vaguely, inconsistently and imprecisely, for example, as something related either to government, ethics or regulation. Thus, elements identified in scientific literature help to analyse framing of AI governance in policy documents.

NVivo software (version Server 11) was used to help with coding relevant passages in policy documents focusing on these topics. The documents were coded by six team members. To develop a common understanding, the team met regularly to discuss research framework, compare results and discuss and clarify any questions. Coded passages were qualitatively analysed to identify emerging governance patterns, which are discussed below.

## Framing governance in AI policy documents

We will discuss the insights from our empirical analysis in three steps. First, we introduce the context in which AI governance is framed highlighting the performative role of hype, public controversies and launch of dedicated AI policy documents. Second, we discuss the framing of the role of the state. And third, we discuss the framing of the public engagement.

### Context

Hype surrounding AI development and use has been an important driving force for launching dedicated AI policy documents around the world. Since 2016 national governments, international organizations, civil society organizations, consultancies and think tanks have almost simultaneously launched their AI strategies. As suggested by research on performative function of hypes (Van Lente et al., 2013), irrespective of how accurate predictions surrounding AI technological hype are, they are influencing agenda setting and contributing to a policy hype (Ulnicane et al., *forthcoming*). Due to the work on AI documents happening at the same time and being discussed at international forums such as the World Economic Forum and the Organisation for Economic Cooperation and Development, countries and organizations have been inspired by and learned from each other (Dolowitz & Marsh, 2000). A number of policy documents acknowledge that when preparing their AI policies, they have reviewed policies and documents from other countries (e.g. BIC/APPGAI 2017d; European Commission 2017).

In the context of a global trend of producing policy documents dedicated to AI, it is helpful to recognize, as Carol Bacchi has suggested, ‘the non-innocence of how problems get framed’ (Bacchi, 2000, p. 50). Alternative approaches of addressing AI in public policy are possible. In line with the shifting paradigm of technology policy towards addressing societal problems (Diercks et al., 2019) AI could have been included as one of relevant technologies in policy documents dedicated to health, climate change and the Sustainable Development Goals (SDGs) rather than launching documents dedicated to AI which give impression of traditional technology-push and supply side approach.

AI policy documents reflect controversies surrounding AI as a highly contested technology that raises hopes as well as fears. To different degrees, the documents outline



positive as well as negative expectations towards AI. Positive expectations are framed in superlatives describing AI as ‘a transformative technology that holds promise for tremendous societal and economic benefit. AI has the potential to revolutionize how we live, work, learn and communicate’ (Executive Office of the President 2016c). While AI hype is based on high positive expectations towards AI contribution to economic growth and addressing societal problems, AI development and use also raise major concerns about its negative effects on individuals and societies across multiple areas including jobs, welfare state and political systems.

Important concerns discussed in the AI policy documents include concentration of power and potential increases in social and economic inequality. This has been highlighted as a key issue in the academic and policy discourse on AI (see also Taeihagh, 2020). The IPPR (2017) report warns of a ‘paradox of plenty’ and explains that ‘managed poorly, automation could create a “paradox of plenty”: a society would be far richer in aggregate, but, for many individuals and communities, technological change could reinforce inequalities of power and reward’. Potentially dangerous power shifts are also outlined in the AI Now Report (Crawford and Whittaker 2016) which highlights

‘the risk that AI systems will not only aggregate power by reducing the ability of the weakest to contest their own treatment, but also that those who design the systems will be able to define what counts as ethical behaviour. Such power can take very subtle forms. For instance, we see that various kinds of automated systems are often used to influence or ‘nudge’ individuals in particular directions, largely dictated by those who design and deploy (and profit from) such systems’ (Crawford and Whittaker 2016).

In this context, several studies have highlighted the discriminatory treatment displayed by digital platforms’ algorithms (Di Porto & Zuppetta, 2020) as well as susceptibility of algorithms to biases in the data and of its designers, such as autonomous vehicles that can be designed to prioritise passenger safety and maximise profit rather than ensure overall public safety, which is difficult to correct in complex and opaque machine-learning systems (Lim & Taeihagh, 2019; Taeihagh & Lim, 2019).

Against the background of major controversies surrounding AI, policy documents tell a story of new types of challenges posed by AI and frame governance as a prescription to overcome controversies by addressing concerns and facilitating benefits. Diagnosis of the problem includes current dominance of a small number of large multinational companies in the field of the development of AI, which the EESC (2017) describes as follows: ‘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’.

### ***The role of the state***

If part of diagnosis of the problems and controversies surrounding AI is an oligopoly as a governance mode (i.e. driven by a small number of large and powerful companies), then governance framing in AI documents prescribes more active and collaborative roles to the state in AI development and use. A range of roles of the state in socio-technical system identified by Borrás and Edler (2020) can be found in AI policy documents.

The state is often framed as a ‘promoter’ acting as ‘a champion, proponent and exponent of change in the socio-technical system’ (Borrás & Edler, 2020, p. 7). This can be seen in the

French document, which states that ‘it is vital for the State to be a key driver in these various areas of transformation. Public authorities must ensure that they adopt the necessary material and human resources to factor AI into the way they address public policy, with the aim of both pursuing modernization and acting as an example to be followed.’ (Villany 2018). Similarly, the US document states that ‘government leadership and coordination is needed to drive standardization and encourage its widespread use in government, academia, and industry’ (Executive Office of the President 2016c). In line with the shifting paradigm of technology policy towards addressing societal challenges, the Swedish government agency Vinnova highlights that the government should ensure that AI contributes to addressing societal challenges and Sustainable Development Goals (SDGs): ‘Developed state governance will be crucial for the utilisation of AI’s transformative potential in the development of systems solutions to address important social challenges and to achieve the targets in the 2030 Agenda for Sustainable Development’ (Vinnova 2018).

Against the background of AI controversies, the role of the state as a ‘promoter’ is closely calibrated and balanced with its role as a ‘guarantor’ of actively and directly securing against security and safety risks (Borras & Edler, 2020, p. 7). This quote from the report from the British Parliament demonstrates this balancing act between harnessing benefits and safeguarding against the risks:

‘Government must set the boundaries for what AI can and cannot do. Forward-looking policies, prioritising the good of the wider public, should be adapted. Taking into account the remarkable speed technology is progressing, Government must act quick in guiding the norms and standards for AI and set the appropriate regulation where need be. Furthermore, Government must evolve with technology to harness the opportunities and protect society from the risks.’ (Big Innovation Centre/APPG AI 2017a).

Similarly, the ‘AI and Life in 2030’ report emphasises the importance of ensuring the use of AI for social good and avoiding its misuse stating that ‘public policies should help ease society’s adaptation to AI applications, extend their benefits, and mitigate their inevitable errors and failures’ (The 2015 panel 2016).

Framing of the role of the state in AI policy documents includes a number of roles where the state shapes activities of other stakeholders. The state is expected to act as a ‘facilitator’ seeking ‘to make a process easier by supporting specific dynamics of other agents’ change initiatives’ (Borras & Edler, 2020, p. 7) when its role is seen as ‘laying the foundation for innovation and providing stakeholders with the means and the resources for breaking new ground’ (Villani 2018) or as providing funding to incentivize ‘collaboration by bringing industry, academia, and government interests together in pre-competitive research’ (Bowser et al 2017). The state is also framed as a ‘moderator’, namely, ‘an arbitrator or negotiator between different social and political positions among agents’ (Borras & Edler, 2020, p. 7), as can be seen in the statements about the importance of cooperation with stakeholders (e.g. Vinnova 2018).

The policy documents not only suggest that the state is not a unitary actor and can undertake a number of roles simultaneously but also indicate that the role of the state in the development of AI has changed over time. In the US, the role of the state has shifted from a ‘promoter’ to a ‘facilitator’, as explained in this quote:

‘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).

Mention of collaborative approach between state agencies and private corporations in this quote points to some of the key questions related to governance, public engagement and AI controversies, namely about the relationship between the state and the private sector (see Radu, 2020) and distribution of power among different types of non-state actors. These questions are also important for the state as an 'enabler of societal engagement' actively encouraging 'the involvement of stakeholders in participatory processes' (Borras & Edler, 2020, p. 7). A number of policy documents call governments to engage with stakeholders and stimulate cooperation between all significant stakeholders (e.g. Vinnova 2018). For example, IEEE (Institute of Electrical and Electronics Engineers) Europe Public Policy Initiative (2017) suggests that 'the European institutions and member states need to collaborate with multi-stakeholder bodies, such as IEEE, the European standardisation organisations and other relevant global standardisation organisations, and the European Association of Artificial Intelligence, in order to harmonise technical standards for the European market as a means of consumer protection'. A concrete example of government institutions enabling stakeholder participation can be found in the European Commission's AI strategy declaring intent to facilitate 'the creation and operation of a broad multi-stakeholder platform, the European AI Alliance, to work on all aspects of AI'.<sup>1</sup>

To summarize, different roles of the state outlined in the AI policy documents can be seen as an attempt to appeal to the competing views in AI controversies. As a 'promoter' and 'facilitator', the state is aiming to harness the benefits of AI for society and economy. As a 'guarantor', it is expected to mitigate risks associated with AI, while as a 'moderator' and an 'enabler of societal engagement' it has a role of enabling and moderating diverse views on AI benefits and concerns.

### ***Public engagement and dialogue***

Public engagement is another major prescription for governance to help to overcome controversies surrounding AI. It is recognized that so far the AI development has been largely closed to the public and 'those directly impacted by the development of AI systems within core social and economic domains rarely have a role in their design, or means to alter their assumptions and uses' (Crawford and Whittaker 2016). To engage society in shaping AI development and use, policy documents suggest 'multi-stakeholder' approaches, ecosystems, partnerships and dialogues.

An important consideration in policy documents is given to one of the key questions related to public engagement, namely, which are relevant publics. This question is particularly relevant for AI controversies with their concerns about the AI development being driven by white male programmers working for large companies. While some policy documents value the involvement of academic, industry and government stakeholders due to their expertise (IEEE-USA 2017), the ITU document contrasts an inclusive

<sup>1</sup>For more information on the European AI alliance, see the relevant webpage of the European Commission <https://ec.europa.eu/digital-single-market/en/european-ai-alliance> (Accessed 27 May 2019).

multi-stakeholder approach to a more technocratic expert-dominated approach stating that ‘AI experts have said themselves that we cannot leave AI to just the experts – a multi-stakeholder approach is needed’. This is reasserted by Radu’s (2020) findings that, while governments’ AI strategies have tended to depend on ‘trusting the private sector’ to self-regulate based on voluntary guidelines that suggest the dominance of industry in AI governance, there is emerging ‘scepticism’ about potentially conflicting interests (Radu, 2020). IEEE (2017) values interests, insights, on the ground expertise and wisdom of stakeholders and suggests to ‘account for the interests of the full range of stakeholders or practitioners who will be working alongside A/IS, incorporating their insights. Build upon, rather than circumvent or ignore, the social and practical wisdom of involved practitioners and other stakeholders’. In particular, the IEEE highlights the requirement for designers to engage with users of AI. Similarly, the US document emphasises the need to bring together developer- and user-driven standards and sharing common visions (Executive Office of the President 2016c).

A call to engage multiple publics can be found, for example, in the document from the EESC, which suggest to involve policy-makers, industry, the social partners, consumers, non-governmental organisations, educational and health institutions, and academics and specialists from various disciplines including applied AI, ethics, safety, economics, occupational science, law, behavioural science, psychology and philosophy (EESC, 2017). The World Economic Forum (WEF) document contrasts such a multi-stakeholder approach to the governance ‘merely dominated by leading transnational companies’ and, importantly, admit that ensuring the former ‘will not be easy’ (WEF, 2018).

Involving multiple publics is presented as a prescription against some of the key concerns about AI such as lack of diversity and bias. To increase transparency about diversity situation, the AI Now 2017 report suggests that ‘companies, universities, conferences and other stakeholders in the AI field should release data on the participation of women, minorities and other marginalized groups within AI research and development’ (Campolo et al 2017). This report argues that ‘more diversity within the fields building these systems will help ensure that they reflect a broader variety of viewpoints’ and that ‘there is an urgent need to expand cultural, disciplinary and ethnic diversity within the AI field to diminish groupthink, mitigate bias and broaden intellectual frames of reference beyond the purely technical’ (Campolo et al, 2017). The report warns that if AI development is not inclusive, its success is doubtful’ (Campolo et al 2017). Similarly, the UK’s AI Sector Deal (HM Government 2018) emphasise the need to increase the diversity in the AI workforce as essential to ensure that AI developed in the UK reflects the needs and make-up of society.

Furthermore, prescriptions include not only increasing diversity of AI developers but also giving a voice to diverse views and raising awareness of diverse needs. The European Group on Ethics in Science and New Technologies (EGE) (2017) suggests ‘a wide, inclusive and far-reaching societal debate, drawing upon the input of diverse perspectives, where those with different expertise and values can be heard’. It is suggested that designers and developers of AI should remain aware of diversity of cultural norms among users including those from vulnerable and underrepresented groups and even individuals such as disabled persons (IEEE 2017) and ‘work with representatives and members of communities impacted by the application of automated decision-making and AI systems

to co-design AI with accountability, in collaboration with those developing and deploying such systems' (Crawford and Whittaker 2016).

The rationales for a broad-ranging public engagement indicated in the policy documents include aligning policies to societal needs and goals such as equality. According to ITU (2017) 'it is vital that the needs of a diverse range of people, including the most vulnerable, guide the design and development of AI systems. Those who are furthest behind in terms of social and economic development, are at the centre of the SDGs and need to be at the centre of design and application of technologies such as AI'. The House of Lords (2018) report sees the participation of all groups of society in AI development as a way to avoid an increase in inequality. The French document (Villani 2018) goes even further suggesting that a policy of inclusion should fulfil a dual objective, namely not only ensure that AI does not reinforce social and economic inequality but also using AI to help to reduce the problems of exclusion and the concentration of wealth and resources.

However, what are societal needs and interests are far from obvious. Therefore, public engagement provides a way to open up debates and to formulate societal needs and goals, which in the case of AI controversies is of major importance for making-decisions, for example, about social and ethical issues (e.g. IEEE-USA 2018; WEF 2018) as well as the future of work, jobs and skills (Villani 2018). Moreover, having debate about societal preferences is only valuable if the results are taken up by decision-makers and it is a common understanding that the 'key decisions on the regulation of AI development and application should be the result of democratic debate and public engagement' (EGE 2017).

The policy documents suggest a range of formats for public engagement including ones that should take place globally (Big Innovation Centre/APPG AI 2017d; UNI Global Union 2017). Villani (2018) recommends that 'given the important nature of the ethical questions that confront future developments in AI, it would be prudent to create a genuinely diverse and inclusive social forum for discussion, to enable us to democratically determine which forms of AI are appropriate for our society'. Additionally, some documents mention the requirement for a 'high quality-informed dialogue with the wider public about the impacts of AI' (Big Innovation Centre/APPG AI 2017b). IPPR (2017) presents a broad-ranging debate as an alternative to technocratic decision-making stating that important questions 'must be addressed by society through collective debate and decision, and not simply left to the developers of the technologies themselves'. Stakeholders themselves play an important role in shaping public engagement, for example, the EESC declares that 'as the representative of European civil society, the EESC will shape, focus and promote the public debate on AI in the coming period, involving all relevant stakeholders' (EESC 2017).

To summarize, in policy documents, public engagement is prescribed as a solution to address concerns about concentration of power, increases in inequality, lack of diversity and bias. Governance framing indicates high hopes and expectations towards opening up AI development and use to diverse stakeholders. This is in line with normative commitment to public engagement widely shared in studies of governance of technology. However, framing of public engagement as a prescription of addressing major concerns about AI so far gives little consideration to some of the challenges well-known for researchers and practitioners of public engagement with science and technology such

as difficulty of achieving consensus among diverse societal views, high resource requirements for public engagement exercises and risks of capture by vested interests.

The case of the High-Level Expert Group (HLEG) on AI launched by the European Commission and closely linked to its multi-stakeholder forum AI Alliance exemplifies some of these challenges (Ulnicane, 2021; Ulnicane et al., 2021). This expert group of 52 members, which was formed by the European Commission in an open selection process, ended up being largely dominated by industry representatives with limited representation from civil society and academia; the ethics guidelines the group developed has been criticized as ‘ethics washing’ where industry talks about ethics to delay or avoid binding regulation (Metzinger, 2019).

## Conclusions

In this paper, we examine how the governance in AI policy documents have been framed as way to resolve public controversies surrounding AI. We draw on analysis of 49 recent policy documents dedicated to AI which have been prepared in the context of technological hype expecting fast advances of AI that will fundamentally change economy and society. As suggested by work on performative function of hypes, irrespective of how accurate predictions about AI are, they influence agenda setting including assigning high policy priority to AI. The hype about AI is accompanied by major public controversy about positive and negative effects of AI.

Against the backdrop of this policy controversy, governance emerges as one of the frames that diagnoses the problems and offers prescriptions. Accordingly, the current governance characterized by oligopoly of a small number of large companies is indicated as one of the reasons for problems such as lack of consideration of societal needs and concerns. To address these problems, governance frame assigns more active and collaborative roles to the state and society. Amid public controversies, the state is assigned the roles of promoting and facilitating AI development while at the same time being a guarantor of risk mitigation and enabler of societal engagement. High expectations are assigned to public engagement with multiple publics as a way to increase diversity, representation and equality in AI development and use. While this governance frame might have a normative appeal, it is not specific about addressing some well-known challenges of the proposed governance mode such as risks of capture by vested interests or difficulties to achieve consensus.

Although governance frame examined here prioritizes AI, alternative frames are possible, for example, prioritizing tackling societal problems where AI is one of means addressing them or reforming the current economic system which might be the root cause of some of concerns about AI development and use.

This paper had focussed on a ‘rhetorical frame’ of governance constructed from recent AI policy documents. To follow the further development of AI governance and to advance our understanding of it, future research could connect this ‘rhetorical frame’ with an emerging ‘action frame’ – ‘constructed from the evidence provided by observation of patterns of action inherent in the practice of policy practitioners’ (Rein & Schon, 1996, p. 91).

The focus of this paper has been on some common themes in AI policy documents. Similarities in discourses about AI have also been noticed in recent research on AI ethics

(e.g. Coeckelbergh, 2020; Jobin et al., 2019; Vesnic-Alujevic et al., 2020). Similarities in emerging AI discourse can be due to policy learning and inspiration (Dolowitz & Marsh, 2000) among countries and organizations which are developing these documents largely at the same time as well as because of activities of international and regional organizations such as the European Union and the World Economic Forum (Ulnicane, 2021; Ulnicane et al., *forthcoming*). However, existence of similar themes and suggestions in AI discourse does not immediately mean that there is total convergence and differences across countries and organizations have disappeared. Jobin et al. (2019) find that in AI ethics guidelines a global convergence emerging around five ethical principles coexists with substantive divergence in interpretation and implementation of these principles. To examine relationship between convergence and diversity in AI policies and governance, future studies should select a smaller set of documents and cases which would allow to undertake in-depth analysis and to compare not only across countries and regions but also across different types of organizations such as national governments, consultancies, think tanks and civil society organizations.

## Acknowledgements

This article has benefited from the comments and suggestions from the special issue editor Professor Araz Taeihagh, two anonymous reviewers, participants of the special issues workshop at the Lee Kuan Yew School of Public Policy at the National University of Singapore, August 2019 and the AsSIST-UK (UK Association for Studies in Innovation, Science and Technology) conference at the Manchester Institute of Innovation Research at University of Manchester, September 2019. The authors are grateful to Dinesh Mothi for his contribution to the document analysis. Support from the De Montfort University's Future Research Leaders Programme (Inga Ulnicane), Frontrunners internship (Tonii Leach and Dinesh Mothi) and student placement (Winter-Gladys Wanjiku) is acknowledged.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## 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, research infrastructures, 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).

*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.

*Tonii Leach*, contributed to this article within the framework of the Frontrunner in Responsible AI internship at De Montfort University.

*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.

*Winter-Gladys Wanjiku*, is a Law LLB graduate. She is currently undertaking her LLM in Legal Practice at De Montfort University (DMU). She contributed to this article as part of her student placement at DMU.

## ORCID

Inga Ulnicane  <http://orcid.org/0000-0003-2051-1265>

William Knight  <http://orcid.org/0000-0001-9818-6277>

Tonii Leach  <http://orcid.org/0000-0001-9417-7955>

Bernd Carsten Stahl  <http://orcid.org/0000-0002-4058-4456>

Winter-Gladys Wanjiku  <http://orcid.org/0000-0001-8236-8957>

## References

- Allen, J., & Husain, A. (2017, November 3). The next space race is Artificial Intelligence. *Foreign Affairs*.
- Bacchi, C. (2000). Policy as discourse: What does it mean? Where does it get us? *Discourse: Studies in the Cultural Politics of Education*, 21(1), 45–57.
- Borras, S., & Edler, J. (eds.). (2014). *The governance of socio-technical systems: Explaining change*. Cheltenham: Edward Elgar.
- Borras, S., & Edler, J. (2020). The roles of the state in the governance of socio-technical systems' transformation. *Research Policy*, 49(5), 103971.
- Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. Cambridge, MA: The MIT Press.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, 24, 505–528.
- Coeckelbergh, M. (2020). *AI ethics*. Cambridge, MA: The MIT Press.
- Di Porto, F., & Zuppetta, M. N. (2020). Co-regulating algorithmic disclosure for digital platforms. *Policy & Society*, 1–22. this issue. doi:10.1080/14494035.2020.1809052.
- Dickinson, H., Smith, C., Carey, N., & Carey, G. (2020). Exploring governance dilemmas of disruptive technologies: The case of care robots in Australia and New Zealand. *Policy & Society*. this issue.
- Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy*, 48(4), 880–894.
- Dignum, V. (2019). *Responsible Artificial Intelligence. How to develop and use AI in a responsible way*. Cham: Springer.



- Dolowitz, D. P., & Marsh, D. (2000). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance: An International Journal of Policy and Administration*, 13(1), 5–24.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds & Machines*, 28(4), 689–707.
- Freeman, R., & Maybin, J. (2011). Documents, practices and policy. *Evidence & Policy*, 7(2), 155–170.
- Fry, H. (2018). *Hello world. How to be human in the age of the machine*. London: Black Swan.
- Gahnberg, C. (2020). The governance of artificial agency. *Policy & Society*. this issue.
- Galanos, V. (2019). Exploring expanding expertise: Artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. *Technology Analysis & Strategic Management*, 31(4), 421–432.
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120.
- Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. New York: W.W. Norton & Company.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Kuhlmann, S., Stegmaier, P., & Konrad, K. (2019). The tentative governance of emerging science and technology—A conceptual introduction. *Research Policy*, 48(5), 1091–1097.
- Lee, K. F. (2018). *AI superpowers. China, Silicon Valley, and the new world order*. Boston: Houghton Mifflin Harcourt.
- Leiman, T. (2020). Law and tech collide: Foreseeability, reasonableness and advanced driver assistance systems. *Policy & Society*, 1–22. this issue. doi:10.1080/14494035.2020.1787696.
- Lim, H. S. M., & Taihagh, A. (2019). Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. *Sustainability*, 11, 5791.
- Lyll, C., & Tait, J. (2019). Beyond the limits to governance: New rules of engagement for the tentative governance of the life sciences. *Research Policy*, 48(5), 1128–1137.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. New York: Pantheon Books.
- Mazzucato, M. (2013). *The entrepreneurial state. Debunking public vs. private sector myths*. London: Anthem Press.
- Metzinger, T. (2019). Ethics washing made in Europe. *Tagesspiegel*, 8 April.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507.
- O’Neill, C. (2016). *Weapons of math destruction. How big data increases inequality and threatens democracy*. London: Penguin Books.
- Pierre, J., & Peters, B. G. (2000). *Governance, politics and the state*. Houndmills: Macmillan.
- Radu, R. (2020). AI governance: National, hybrid, ambiguous. *Policy & Society*. this issue.
- Rein, M., & Schon, D. (1993). Reframing policy discourse. In F. Fischer & J. Forester (Eds.), *The argumentative turn in policy analysis and planning* (pp. 145–166). London: UCL Press.
- Rein, M., & Schon, D. (1996). Frame-critical policy analysis and frame-reflective policy practice. *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, 9(1), 85–104.
- Rotolo, D., Hicks, D., & Martin, B. (2015). What is an emerging technology? *Research Policy*, 44(10), 1827–1843.
- Stilgoe, J. (2015). *Experiment Earth. Responsible innovation in geoengineering*. Abingdon: Routledge.
- Stilgoe, J. (2020). *Who’s driving innovation? New technologies and the collaborative state*. Cham: Palgrave Macmillan.
- Stilgoe, J., Lock, S., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science*, 23(1), 4–15.
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.

- Taeihagh, A. (2020). The governance of Artificial Intelligence and Robotics. *Policy & Society*. this issue.
- Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, 39(1), 103–128.
- Tan, S., & Taeihagh, A. (2020). Governing the adoption of robotics and autonomous systems in long-term care in Singapore. *Policy & Society*, 1–21. this issue. doi:10.1080/14494035.2020.1782627.
- Ulnicane, I. (2021). Artificial Intelligence in the European Union: Policy, ethics and regulation. In T. Hoerber, I. Cabras, & G. Weber (Eds.), *Routledge handbook of European integrations*. Abingdon: Routledge.
- Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (forthcoming). Governance of Artificial Intelligence: Emerging international trends and policy frames. In M. Tinnirello (Ed.), *Global politics of Artificial Intelligence*. Boca Raton: CRC Press.
- Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews*, 46(1). <https://doi.org/10.1080/03080188.2020.1840220>
- Van Lente, H., Spitters, C., & Peine, A. (2013). Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change*, 80(8), 1615–1628.
- Vesnic-Alujevic, L., Nascimento, S., & Polvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6), 101961.

## Annex 1. Dataset of AI policy documents (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 (2017 c) 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 1<sup>st</sup> 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 (2018 c) 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 (2016 c) 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.