

Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach

Corinne Cath^{1,2} · Sandra Wachter^{1,2} ·
Brent Mittelstadt^{1,2} · Mariarosaria Taddeo^{1,2} ·
Luciano Floridi^{1,2}

Received: 20 January 2017 / Accepted: 19 March 2017 / Published online: 28 March 2017
© Springer Science+Business Media Dordrecht 2017

Abstract In October 2016, the White House, the European Parliament, and the UK House of Commons each issued a report outlining their visions on how to prepare society for the widespread use of artificial intelligence (AI). In this article, we provide a comparative assessment of these three reports in order to facilitate the design of policies favourable to the development of a ‘good AI society’. To do so, we examine how each report addresses the following three topics: (a) the development of a ‘good AI society’; (b) the role and responsibility of the government, the private sector, and the research community (including academia) in pursuing such a development; and (c) where the recommendations to support such a development may be in need of improvement. Our analysis concludes that the reports address adequately various ethical, social, and economic topics, but come short of providing an overarching political vision and long-term strategy for the development of a ‘good AI society’. In order to contribute to fill this gap, in the conclusion we suggest a two-pronged approach.

✉ Corinne Cath
ccath@turing.ac.uk

Sandra Wachter
sandra.wachter@oii.ox.ac.uk

Brent Mittelstadt
brent.mittelstadt@oii.ox.ac.uk

Mariarosaria Taddeo
mariarosaria.taddeo@oii.ox.ac.uk

Luciano Floridi
luciano.floridi@oii.ox.ac.uk

¹ Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

² The Alan Turing Institute, Headquartered at the British Library, 96 Euston Road, London NW1 2DB, UK

Keywords Algorithms · Artificial intelligence · Data ethics · Good society · Human dignity

Introduction

Artificial intelligence (AI) is no longer sci-fi. From driverless cars to the use of machine learning to improve healthcare services¹ and the financial industry,² AI is shaping our daily practices as well as a fast-growing number of fundamental aspects of our societies. Admittedly, the hype around AI has gone through several cycles of boom-and-bust since its beginning in the late 1950s. However, the renewed focus on AI in recent years is unlikely to be fleeting because of the very robust and rapid development of four self-reinforcing trends: ever more sophisticated statistical and probabilistic methods; the availability of increasingly large amounts of data; the accessibility of cheap, enormous computational power; and the transformation of ever more places into IT-friendly environments (e.g. domotics, and smart cities).³ Steady progress and cross pollination in these areas has reinvigorated the feasibility, importance, and scalability of AI. Which is why recently there has also been increasing concern about the impact that AI is having on our societies and about who should be responsible for ensuring that AI will be a force for good.

Because AI poses fundamental questions concerning its ethical, social, and economic impact⁴ in October 2016, the White House Office of Science and Technology Policy (OSTP), the European Parliament's Committee on Legal Affairs, and, in the UK, the House of Commons' Science and Technology Committee released their initial reports on how to prepare for the future of AI.⁵ To the best of our understanding, the three documents might have been prepared independently of each other. Regardless of whether this is the case, their release indicates how timely and synchronised efforts are becoming to deal with the challenges posed by AI. In this article, we provide a comparative evaluation of these three reports,⁶ by examining how well each of them addresses the following three topics:

- (a) the development of a 'good AI society';
- (b) the role and responsibility of the government, the private sector, and the research community (including academia), in pursuing such a development; and

¹ Furlow (2016).

² Fleury (2015).

³ National Science and Technology Council Networking and Information Technology, Research and Development Subcommittee, and National Science and Technology Council Networking and Information Technology (2016). On the IT-friendly trend see Floridi (2014).

⁴ Mittelstadt et al. (2016).

⁵ Our focus is solely on the initial reports coming out in the fall and winter of 2016. This choice was made to ensure that the comparison would be focused on the specifics of the first round of reports of these governments, as opposed to on the ensuing responses and follow up reports.

⁶ We also mention the US R&D compendium and the adjoining Economic Report, as they are integral to initial US report.

- (c) whether the recommendations to support such a development may be in need of improvement.

Each report focuses on specific, pressing challenges. We shall see that each report seems to have an implicit, overarching understanding of AI’s role in society and a view of how that may best be dealt with. However, none appears to deliver a comprehensive explicit vision of the role that AI should play in “mature information societies”.⁷ Arguably, this might not have been the goal of any of the three reports. However, as we shall indicate in the conclusion, from an ethical perspective, AI’s potential contribution to social good should include an in-depth plan for linking in a comprehensive socio-political design questions of responsibility of the different stakeholders, of cooperation between them, and of shareable values that underpin our understanding of a ‘good AI society’. Such a design needs to be forward looking, and capable of addressing current problems as well as being able to adapt to new challenges put forward in the “mature information societies” to follow in the next decades. In short, we need a social strategy for AI, not mere tactics.

Mature information societies are societies in which digital affordances are the expected backdrop to all aspects of society, as opposed to societies in which such affordances are new or unexpected (Floridi 2016a). The notion of mature information societies is introduced to stress the importance of addressing the current ethical challenges that AI poses in a comprehensive fashion. As societies become more “information mature”, their reliance on AI technologies will increase. And as the scale of such reliance increases, so will the impact of AI technologies on our shared values. However, we might very likely be less inclined to notice the fundamental impact of these technologies, because their existence and influence is increasingly rendered opaque by the level of maturity reached in an information society (ibid). Paradoxically, the more AI matters the less one may be able to realise how much it does.

Digital technologies and AI in particular are developing very rapidly. The direction of such fast innovations needs to be steered socio-politically, in terms of where we want to go, rather than how quickly we may get there. The risk is that, a lack of vision and strategy will lead the private sector—and sometimes academia—to continue to fill the vacuum by *de facto* setting the standard for what may be considered ‘the good AI society’, while governments are currently unwilling or unable to do so.

The current situation is both understandable and unacceptable. It is understandable because corporate R&D is driving AI-based innovation and, for the past decade, the private sector, sometimes together with academia, has led the discussion on how AI could best be applied for the good of society. Nevertheless, leaving such tasks to private or academic actors remains unacceptable because of a deficit of social and political accountability and long-term planning, which has the goal of fair sharing of benefits and opportunities for all. In the conclusion, we shall argue that a multi-stakeholder effort, in which governments play a leading role, may be the best way to steer the fast development and widespread dissemination of AI in the right

⁷ Floridi (2016a).

direction, and hence to ensure that the ‘good AI society’ will have the most positive influence on all individuals, societies, cultures, and environments.

AI is not merely another utility that needs to be regulated only once it is mature; it is a powerful force that is reshaping our lives, our interactions, and our environments. It is part of a profound transformation of our habitat into an infosphere. It has a deep ecological nature. As such, its future must be supported by a clear socio-political design, a regulative ideal, to put in Kantian terms. We are creating the digital world in which future generations will spend most of their time. This is why we shall suggest in this article that the design of a ‘good AI society’ should be based on the holistic respect (i.e., a respect that considers the whole context of human flourishing) and nurturing of human dignity as the grounding foundation of a better world. The best future of a ‘good AI society’ is one in which it helps the infosphere and the biosphere to prosper together.

The U.S. Report: Letting a Thousand Flowers Bloom

On October 12th, 2016 the White House Office of Science and Technology Policy (OSTP) released the US report on AI, entitled ‘Preparing for the Future of Artificial Intelligence’.⁸ The report follows five public workshops,⁹ and an official Request for Information on AI.¹⁰ All these inputs were used to guide the recommendations in the report. The report’s overall tone is confident, and reflects the positive view of technology reminiscent of that found in Silicon Valley. It is aimed at the tech-sector and the general public. The report defines AI as a technology that—when used thoughtfully—can help to augment¹¹ human capabilities, instead of replacing them. It lays out an image of what we have labelled so far a ‘good AI society’ as one in which AI is applied for “the public good (...) [and to tackle] some of the world’s greatest challenges and inefficiencies”.¹²

The thread that holds together the OSTP’s approach to AI is *innovation*. In a nutshell, AI is good for innovation and economic growth, and this is good for society, especially because commercially developed¹³ AI can be leveraged in new

⁸ Executive Office of the President National Science and Technology Council Committee on Technology (2016).

⁹ Felten and Lyons (2016).

¹⁰ Request for Information on Artificial Intelligence (2016).

¹¹ The OSTP report states that: “Developing and studying machine intelligence can help us better understand and appreciate our human intelligence. Used thoughtfully, AI can augment our intelligence, helping us chart a better and wiser path forward.” (2016, pp. 14, 49).

¹² Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 1).

¹³ The OSTP report states that in certain cases this can be achieved by working together with public institutes, or supported by public funding: “Private and public institutions are encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit society. Social justice and public policy institutions that do not typically engage with advanced technologies and data science in their work should consider partnerships with AI researchers and practitioners that can help apply AI tactics to the broad social problems these institutions already address in other ways.” 2016, pp. 14, 40.

ways to address societal issues. As such, it comes as no surprise that the US government’s vision of its own role, as a regulator, is limited. The US government is focused on ensuring that it does not hinder the development of AI technologies, “allowing a thousand flowers to bloom”.¹⁴ Regulation of AI should happen in a light-handed fashion, and, where applicable, the government should aim to fit AI into existing regulatory schemes, for example in the automotive and aviation industries.¹⁵ However, the report also calls upon the relevant agencies to ensure that—in evolving regulation on the basis of existing schemes—they “remain mindful of the fundamental purposes and goals of regulation to safeguard the public good, while creating space for innovation and growth in AI”.¹⁶ These regulatory schemes, in particular for transport, will be evolving on the basis of on-going experiments and understanding of what constitutes safe operations.¹⁷ The general vision is one in which the government manages the tasks of defining the outer parameters of what AI should be used for, and of collecting data to further inform policy making.¹⁸ The private sector developing AI should continue to innovate within a broad risk management regulatory framework set by the government. This approach suggests that the US government implicit understanding of AI is one that relies heavily on the liberal notion of the free market.

The report emphasises the importance of research. Not only in order to monitor on-going developments in AI, but also to ask the research community to focus its efforts on ensuring that AI is accountable, transparent, and “[its] operation will remain consistent with human values and aspirations”.¹⁹ It also calls upon researchers to collaborate with industry and the government to enable the emergence of new industries that could support workforce development. In the recommendations section of the report, the OSTP focuses on the need for basic and long-term research for the development and application of AI.²⁰

¹⁴ Finley (2016).

¹⁵ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 17).

¹⁶ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 17).

¹⁷ The OSTP report for instance mentions the approach to evolving regulatory frameworks on the basis of ongoing experimentation: “The Department of Transportation (DOT) is using an approach to evolving the relevant regulations that is based on building expertise in the Department, creating safe spaces and test-beds for experimentation, and working with industry and civil society to evolve performance-based regulations that will enable more uses as evidence of safe operation accumulates.” (2016, p. 1).

¹⁸ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 20).

¹⁹ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 4).

²⁰ The OSTP report makes the following recommendation: “Recommendation 13: The Federal government should prioritize basic and long-term AI research. The Nation as a whole would benefit from a steady increase in Federal and private-sector AI R&D, with a particular emphasis on basic research and long-term, high-risk research initiatives. Because basic and long-term research especially are areas where the private sector is not likely to invest, Federal investments will be important for R&D in these areas.”, 2016, pp. 26, 41.

The OSTP's report also addresses the very sensitive issue of economic impact on jobs by stating that AI will both create and reduce jobs.²¹ It hypothesises that low-wage and middle income workers²² are most likely to be negatively impacted by AI, and that the government should develop public policy to ensure that AI does not increase economic inequality.²³ The impact on the job market is interpreted to be so significant that, in December 2016, a report entitled "Artificial Intelligence, Automation and the Economy"²⁴ was published by a team from the White House Executive Office of the President including staff from the Council of Economic Advisers, Domestic Policy Council, National Economic Council, Office of Management and Budget, and Office of Science and Technology Policy. The report focuses on the impacts of AI-driven automation on the US job market and economy.²⁵ It presents three specific policy responses to the perceived impact of AI on the US economy:

- (1) Invest in and develop AI for its many benefits;
- (2) Educate and train Americans for the jobs of the future; and
- (3) Aid workers in the transition and empower workers to ensure broadly shared growth.²⁶ These are all shareable and laudable suggestions.

Although the OSTP report does not lay out a comprehensive vision on how to achieve socially acceptable policies for the development of AI, its companion document, entitled the "National Artificial Intelligence Research and Development Strategic Plan", does provide a more detailed description on how to use R&D investments in order to guide the "long term transformational impact of AI on society and the world".²⁷ The R&D compendium, although very detailed, should not be mistaken for providing a comprehensive socio-political design. Its specific descriptions and ambitious goals are, in the end, only objectives for federally funded AI research. As such, we cannot take the R&D document to be more than an outline of interesting research goals.²⁸ Similarly, the later report on AI and the economy convincingly stipulates a specific and important role for the government in guiding

²¹ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 2).

²² Executive Office of the President National Science and Technology Council Committee on Technology (2016, pp. 2, 29).

²³ The OSTP report states that: "Public policy can [also] ensure that the economic benefits created by AI are shared broadly, and assure that AI responsibly ushers in a new age in the global economy.", 2016, p. 2.

²⁴ Executive Office of the President (2016). Here after referred to as Executive Office of the President (2016).

²⁵ Executive Office of the President (2016, p. 3).

²⁶ Executive Office of the President (2016, p. 27).

²⁷ The OSTP's companion document, entitled the "National Artificial Intelligence Research and Development Strategic Plan", details how R&D investments can be used to advance policies that have a positive long term impact on society and the world (2016, pp. 7–10). The plan is available at: https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf. Hereafter referred to as Networking and Information Technology Research and Development Subcommittee (2016).

²⁸ Networking and Information Technology Research and Development Subcommittee (2016, p. 7).

the economic impact of AI through its policies and institutions. However, many of the policy recommendations made are, as the report itself states, “important regardless of AI-driven automation”.²⁹ More could have been done to tailor further its important recommendations, like the need to improve access to education³⁰ and social safety nets,³¹ to the explicit challenges posed and affordances brought by AI.

The OSTP report misses the opportunity to consider how to spur the specific values that should steer and shape the development of our AI-powered societies. This is clear, for example, when considering the position on the deployment of *Lethal Autonomous Weapons* (LAWs) based on AI. While the report recognises that applying AI to national security brings with it many unresolved ethical dilemmas, the proposed solution—defining policies for the use of LAWs consistent with international humanitarian law³²—falls short of being fully satisfactory.

The future of AI-influenced cyber conflicts needs more than just the application of current and past solutions in order to ensure security and stability of societies, and avoid risks of escalation.³³ To achieve this end, efforts to regulate cyber conflicts require an in-depth understanding of this new phenomenon, identify the changes brought about by cyber conflicts and the information revolution, and define a set of shared values that will guide the stakeholders operating in the international arena. This becomes clear when considering, for example, cyber deterrence. Deploying conventional (Cold War) strategies to deter AI-influenced cyber conflicts proves highly problematic and unveils the urgent need to foster and coordinate new solutions able to account for the peculiarities of these kinds of conflicts, of the cyber domain, and of mature information societies.³⁴ We hope that in the on-going conversations and reviews³⁵ the US government will further specify how LAWs fit into their vision of the future of society, in this case the future of war and conflict.

The OSTP report suggests that many of the ethical issues related to AI—like fairness, accountability, and social justice³⁶—can be addressed through increasing transparency.³⁷ While this is an excellent step forward, the report does not appear to recommend specific methods for enabling transparency and understandability,³⁸

²⁹ Executive Office of the President (2016, p. 3).

³⁰ Executive Office of the President (2016, pp. 32–34).

³¹ Executive Office of the President (2016, p. 35).

³² The OSTP report states: “Agencies across the U.S. Government are working to develop a single, government-wide policy, consistent with international humanitarian law, on autonomous and semi-autonomous weapons.”, 2016, p. 3.

³³ Taddeo (2016a, b).

³⁴ Libicki (2009), Quackenbush (2011), Floridi (2016a, b).

³⁵ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 38).

³⁶ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 2).

³⁷ The OSTP report defines transparency as consisting of two parts: “The data and algorithms involved, and the potential to have some form of explanation for any AI-based determination”, 2016, p. 2.

³⁸ Transparency is covered in some greater detail in the R&D strategy compendium of the OSTP report: “A key research challenge is increasing the “explainability” or “transparency” of AI. Many algorithms, including those based on deep learning, are opaque to users, with few existing mechanisms for explaining

other than the on-going work³⁹ in industry-led voluntary standards, and further research.⁴⁰ Such further research is already indicating that additional and novel approaches are needed that go beyond transparency-based regulation,^{41,42,43} and that the creation of a new federal body focused on robotics⁴⁴ and related AI developments, to provide advice on the policy, legal and consumer protection issues arising in these fields,^{45,46} should be considered.

Note that the OSTP report and the two additional reports, should be commended for explicitly referring to the need for more diversity⁴⁷ in the AI workforce and more inclusivity of various voices influencing the development of AI, a point that has been made on several occasions by leading AI scholars.⁴⁸

The report states that there is a need for having more openly available and unbiased data sets,⁴⁹ privacy considerations, and ethical training for engineers.⁵⁰ The onus of getting all these different solutions in place will be shared by the private sector and the government, with the latter doing this through its R&D strategy.⁵¹ The report indicates that these suggestions are necessary, but not sufficient.⁵² Ethical training of staff and ethical education of the public is certainly important. Yet, it may also be a mechanism for the governments to delegate and transfer

Footnote 38 continued

their results. This is especially problematic for domains such as healthcare, where doctors need explanations to justify a particular diagnosis or a course of treatment. AI techniques such as decision-tree induction provide built-in explanations but are generally less accurate. Thus, researchers must develop systems that are transparent, and intrinsically capable of explaining the reasons for their results to users.”, See Networking and Information Technology Research and Development Subcommittee (2016, p. 28).

³⁹ United States Standards Strategy Committee (2015).

⁴⁰ Networking and Information Technology Research and Development Subcommittee (2016, pp. 14, 26).

⁴¹ Kroll et al. (2017), Annany and Crawford (2016).

⁴² Crawford and Calo (2016).

⁴³ Wachter et al. (Forthcoming).

⁴⁴ Calo (2014).

⁴⁵ Tutt (2016).

⁴⁶ Scherer (2016).

⁴⁷ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 27), Executive Office of the President (2016, pp. 3, 28–29); Executive Office of the President National Science and Technology Council Committee on Technology (2016, pp. 35–36).

⁴⁸ Crawford (2016).

⁴⁹ The OSTP report emphasis the problems of the lack of quality data, especially in the context of the criminal justice system (2016, p. 30).

⁵⁰ Executive Office of the President National Science and Technology Council Committee on Technology (2016, p. 32).

⁵¹ National Science and Technology Council Networking and Information Technology. Networking and Information Technology Research and Development Subcommittee (2016).

⁵² The OSTP report specifically mentions that: “Ethical training for AI practitioners and students is a necessary part of the solution. Ideally, every student learning AI, computer science, or data science would be exposed to curriculum and discussion on related ethics and security topics. However, ethics alone is not sufficient. Ethics can help practitioners understand their responsibilities to all stakeholders, but ethical training needs to be augmented with the technical capability to put good intentions into practice by taking technical precautions as a system is built and tested.” (2016, p. 32).

responsibility for ethical behaviour and design to the private sector and the citizens. Notably, this is a potentially risky aspect that unites all three reports. In relation to the open and unbiased data sets,⁵³ the OSTP report leaves unspecified who may have the authority and legitimacy to set the bar for what is unbiased and declare that something is unbiased, and what certification schemes exist or may need to be created to ensure that the vetting of data sets is standardized. The Economic Report mentions that in certain sector limits on the use of consumer data are imposed, but more protections are needed in this space.⁵⁴

Self-regulatory partnerships—although not explicitly mentioned in the report, like the *Partnership on AI to Benefit People and Society*, which was launched in September⁵⁵—have been a staple of the US’s regulatory approach to AI. They are a good step forward, for they indicate that the private sector is starting to institutionalise and operationalise the discussion of important ethical and social questions. However, self-regulation as a key-strategic approach to AI seems too limited. Quite reasonably, it will tend to favour the goals of industry over those of other stakeholders. And by suggesting that AI should be incorporated into existing regulatory schemes,⁵⁶ even when these are so-called “evolving frameworks”,⁵⁷ the report seems to be trying to fit new round pegs into old square holes. A bolder strategy is needed, with a clearer role for the government and other stakeholders, such that the full spectrum of unique challenges that AI brings to society in terms of fairness, social equity, and accountability are addressed.

The fact that the report was officially issued by the White House OSTP and that much of the report’s content came from a set of public workshops⁵⁸ indicates that the issue is taken seriously at the highest levels of government, at least by the former Obama administration. Yet, the heavy focus on private sector initiative—both for development of the AI technology and defining good AI—remains problematic. In particular, the government’s innovation driven approach to defining the potential,

⁵³ The OSTP report states on this topic: “AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias. It is important that anyone using AI in the criminal justice context is aware of the limitations of current data.” (2016, p. 30). The R&D compendium focuses on the need for establishing “AI technology benchmarks” and ensuring coordination between the different partners in the AI community. It warns that current examples are sector specific and that many questions remain unanswered surrounding the development, use and availability of datasets that produce reliable outcomes (2016, p. 30–33).

⁵⁴ Executive Office of the President (2016 p. 29).

⁵⁵ Partnership on AI (2016).

⁵⁶ In the OSTP report it is stated that: “The general consensus of the RFI commenters was that broad regulation of AI research or practice would be inadvisable at this time. Instead, commenters said that the goals and structure of existing regulations were sufficient, and commenters called for existing regulation to be adapted as necessary to account for the effects of AI. For example, commenters suggested that motor vehicle regulation should evolve to account for the anticipated arrival of autonomous vehicles, and that the necessary evolution could be carried out within the current structure of vehicle safety regulation. In doing so, agencies must remain mindful of the fundamental purposes and goals of regulation to safeguard the public good, while creating space for innovation and growth in AI.” 2016, p. 17.

⁵⁷ The OSTP report mentions the example of the Department of Transportation (DOT) which: “[Is] using an approach to evolving the relevant regulations that is based on building expertise in the Department (...).” 2016, p. 1.

⁵⁸ Felten (2016).

positive impact of AI shows that more could be done to ensure that the opportunities and advantages brought about by AI are shared by all society. It is also important to take into account that it is unclear what will happen to this report and its findings under the Trump administration. Currently, it seems that there is little attention to the topic and limited resources and manpower to carry forward the implementation of the report's recommendations after 2016. The ambitious R&D strategy may remain only aspirational.

Summarizing, the OSTP report is an extensive review of the different ways in which AI will impact the economy and social structure of society. It provides a good overview of the various conundrums, ethical and otherwise. Yet, the US report could have acknowledged more clearly its underlying reliance on economic and political notions of free market trade, and market capitalism. To be fair, we shall see that this is a criticism that can be levied against all three reports. The use of public workshops and a formal "Request for Information" leveraged existing communities of knowledge, and encouraged public debate on the topic. The OSTP report also presents a way forward for implementing its various recommendations through its R&D strategy. The R&D strategy and the Economic Report gets closer than the OSTP report itself to formulating a larger vision of what a good AI society might be, by looking towards the "longer-term transformational impacts of AI on society and the world"⁵⁹ and arguing for "aggressive policy action (...) to help Americans who are disadvantaged by these [AI driven] changes and to ensure that the enormous benefits of AI and automation are developed by and available to all".⁶⁰

Together, these reports emphasize increased economic prosperity, improved educational opportunity, social security and quality of life, and enhanced national and homeland security. However, although important, these issues are approached in a way that can best be summarized as trying to fit AI into the specific vision of US national priorities, instead of seeing the new features of AI as a good opportunity to revisit these priorities, both nationally and internationally.

The EU Report: European Standards for Robotics and AI

On 31st May, 2016, the European Parliament's Committee on Legal Affairs (JURI) published the draft report⁶¹ on Civil Law Rules on Robotics with recommendations to the European Commission.^{62,63} Compared to the US report, the EU report is shorter and focuses more on robotics than AI, with immediate attention called to

⁵⁹ There is a need for further investment in research and the development of systems to make algorithms more transparent and understandable. Networking and Information Technology Research and Development Subcommittee (2016, p. 7).

⁶⁰ Executive Office of the President (2016), *Introduction*.

⁶¹ European Parliament Committee on Legal Affairs (2016).

⁶² For further information on the history e.g. the working group established by the committee and its members see European Parliament Committee on Legal Affairs (2016) p. 20.

⁶³ This report was adopted in a modified form by the European Parliament on the 16th of February 2017. <http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0051&format=XML&language=EN>.

autonomous vehicles, drones, and medical-care robots,⁶⁴ and the suggestion that specific rules might be required in these areas.^{65,66}

The treatment of AI in the EU report also reflects a different understanding of the technology. Rather than a standalone technology, AI is approached as an underlying component⁶⁷ of “smart autonomous robots”.⁶⁸ AI is thus thought of as something that enables autonomy in other technological systems. However, the decision not to include unembodied AI sets the report apart from the other two, and has distinct political and legal consequences.⁶⁹

One of the biggest concerns of the report is the impact of robotics and AI on the workforce.⁷⁰ The report urges to implement employment forecast mechanisms to monitor job trends.⁷¹ It also calls for refocusing educational goals in order to equip the workforce, especially women, with the necessary digital skills to compete on the free market.⁷² It even considers a new tax to cater for the negative effects under current tax regimes, insofar as automation can decrease tax revenues (less tax payers employed), undermine the viability of social security, and increase inequality in wealth and influence.⁷³ It proposes to make it obligatory for undertakings (i.e. businesses or ventures using robotics) to disclose savings made in social security contributions due to automation.⁷⁴

The report calls for the creation of a “European Agency for Robotics and AI” consisting of regulators and external technical and ethical experts, who can monitor AI and robotics-based trends, identify standards for best practice, recommend regulatory measures, define new principles, and address potential consumer protection issues. The Agency will provide advice both at the EU and at Member State level, including annual reporting to the European Commission, to help to harness the potential of these technologies and mitigate possible risks.⁷⁵ It will also provide the public sector with technical, ethical, and regulatory advice. Further, the agency will manage an EU-wide registration system for all smart robots.⁷⁶

⁶⁴ The EP report specifically “Asks for the establishment of committees on robot ethics in hospitals and other health care institutions tasked with considering and assisting in resolving unusual, complicated ethical problems involving issues that affect the care and treatment of patient” European Parliament Committee on Legal Affairs (2016, pp. 8–9).

⁶⁵ European Parliament Committee on Legal Affairs (2016, p. 22).

⁶⁶ The European focus on robotics can best be understood taking into account that the RoboLaws project: Palmerini et al. (2016) and the Green Paper on legal issues in robotics by Leroux and Labruto (2013). This research played a crucial in defining the framing and focus of the European debate.

⁶⁷ European Parliament Committee on Legal Affairs (2016, pp. 3, 5), 10ff, 22.

⁶⁸ European Parliament Committee on Legal Affairs (2016, pp. 11, 21).

⁶⁹ Schafer (2016).

⁷⁰ European Parliament Committee on Legal Affairs (2016, pp. 3, 9–10, 22).

⁷¹ European Parliament Committee on Legal Affairs (2016, p. 10).

⁷² Ibid.

⁷³ Ibid.

⁷⁴ European Parliament Committee on Legal Affairs (2016, p. 14).

⁷⁵ European Parliament Committee on Legal Affairs (2016, p. 7ff.

⁷⁶ European Parliament Committee on Legal Affairs (2016, p. 13).

The report envisions a combination⁷⁷ of hard and soft laws to guard against possible risks. This is particularly welcome considering that in complex, or, in our conceptualization, in mature information societies a double-pronged approach that includes both primary and secondary legal rules will be necessary (Pagallo 2016a). A need is recognised for regulatory action at a European level to avoid fragmentation of standards in the single market, and it is urged to evaluate current EU legislation for required adaptations.⁷⁸ The report does not want European industry to be dominated by standards set outside Europe,⁷⁹ and calls for clear rules for the development and deployment of AI and robotics. The mixed approach can also be seen in the call for evaluation of current EU⁸⁰ (e.g. intellectual property law) and international⁸¹ frameworks (e.g. on road traffic), and for the possible adoption of new legislation.⁸²

The committee calls on the European Commission to carry out an impact assessment of new possible legal tools, mainly focusing on liability issues regarding smart robots. The report stresses that ‘testing robots in real-life scenarios is essential for the identification and assessment of the risks they might entail, as well as of their technological development beyond a pure experimental laboratory phase; underlines, in this regard, that testing of robots in real-life scenarios, in particular in cities and on roads, raises numerous problems and requires an effective monitoring mechanism’.⁸³ However, in its call to the European Commission to ‘draw up uniform criteria across all Member States (...) to identify areas where experiments with robots are permitted’⁸⁴ the committee does not address the nature of such rules. It leaves open the tensions between developing such criteria based on an infrastructure ethics, or infraethics (Floridi 2013), the ethical method which argues that ethical behaviour can be instantiated through the creation of soft rules, affordances, and constraints, or by using traditional legal tools (Hart 1961) (Pagallo 2016b) of governance. And second, the text remains silent on the ‘basic difference among rules of the legal system, e.g. between the primary rules that aim to govern social and individual behaviour, and the secondary rules of change, namely, the rules of the law that create, modify, or suppress the primary rules of the system.’ (Pagallo 2016a, p. 13).

Both issues need to be further addressed before the risks of robots can be tested in real-life scenarios. The report further proposes possible entry points for governance of robotics and AI, such as mandatory insurance or robotic registration schemes.⁸⁵ Additionally, the report calls for the creation of “a guiding ethical framework for

⁷⁷ European Parliament Committee on Legal Affairs (2016, pp. 5, 10ff, 14).

⁷⁸ European Parliament Committee on Legal Affairs (2016, p. 8).

⁷⁹ European Parliament Committee on Legal Affairs (2016, p. 4).

⁸⁰ European Parliament Committee on Legal Affairs (2016, p. 8).

⁸¹ European Parliament Committee on Legal Affairs (2016, pp. 12, 22).

⁸² European Parliament Committee on Legal Affairs (2016, p. 11ff).

⁸³ European Parliament Committee on Legal Affairs (2016, p. 8).

⁸⁴ Ibid.

⁸⁵ European Parliament Committee on Legal Affairs (2016, pp. 10–11).

the design, production and use”⁸⁶ of AI and robotics, “based on the principles of beneficence, non-maleficence and autonomy, as well as on principles enshrined in the EU Charter of Fundamental Rights, such as human dignity and human rights, equality, justice and equity, non-discrimination and non-stigmatisation, autonomy and individual responsibility, informed consent, privacy and social responsibility.”⁸⁷

An initial “Charter on Robotics”,⁸⁸ based upon the aforementioned ethical framework and guiding principles, is proposed. It should be complementary to legislation and comprise ethical codes of conduct for robotics researchers and designers, codes for research ethics committees, as well as licenses (rights and duties) for designers and users. It also states that the European Commission shall take the aforementioned principles into account when proposing new legislation.⁸⁹ This approach is important, as it clearly envisions a role for governments and policy-makers in setting a long-term strategy for the ‘good AI society’, instead of leaving it to industry and the research sector. It remains to be seen how the European Commission will translate these values and regulatory proposals into governance of AI and robotics.

The guidance contained in the Charter would be non-binding, challenging its actual strength. Nonetheless, the report on the Charter states that “special emphasis should be placed on the research and development phases of the relevant technological trajectory (design process,⁹⁰ ethics review, audit controls, etc.).”⁹¹ Further, researchers and designers are invited to consider values such as “dignity, privacy and safety.”⁹² This clearly shows that ethical foresight is desired. Along with these principles, researchers are also called upon to keep in mind at all stages of their research that “people should not be exposed to risks greater than or additional to those to which they are exposed in their normal lifestyles.”⁹³ Additionally, designers are specifically invited to be guided by European values, such as “dignity, freedom and justice.”⁹⁴ As mentioned above, the report also suggests further research to assess current and possible new legislation to determine where further adjustments are required in the application to AI and robotics, meaning that binding European legislation may be forthcoming.

Beyond the proposed Agency and Charter, the report addresses several other aspects of the relationship between government, industry, and the research sector.

⁸⁶ European Parliament Committee on Legal Affairs (2016, pp. 7), in more depth 14.

⁸⁷ European Parliament Committee on Legal Affairs (2016, p. 7).

⁸⁸ European Parliament Committee on Legal Affairs (2016, p. 14).

⁸⁹ European Parliament Committee on Legal Affairs (2016, p. 14).

⁹⁰ E.g. European Parliament Committee on Legal Affairs (2016, p. 18). On the relation to licences for designers: “You should develop tracing tools at the robot’s design stage. These tools will facilitate accounting and explanation of robotic behaviour, even if limited, at the various levels intended for experts, operators and users.”

⁹¹ European Parliament Committee on Legal Affairs (2016, p. 14).

⁹² European Parliament Committee on Legal Affairs (2016, pp. 14–15).

⁹³ European Parliament Committee on Legal Affairs (2016, p. 16).

⁹⁴ European Parliament Committee on Legal Affairs (2016, p. 17f).

The European Commission is asked to work with multi-stakeholder bodies—such as the European Standardisation Organisations and the International Standardisation Organisation—in order to harmonise technical standards for the European market as a means of consumer protection.⁹⁵ The European Commission is also called upon to clarify the liability of industry and autonomous robots⁹⁶ when harms or damages occur,⁹⁷ and to consider the principles of the Charter when adopting new legislation. A much-needed call is placed for the Commission and Member States to provide significant funding for R&D of AI and robotics, also in order to enable industry and the research sector to explore the risks and opportunities raised by their dissemination of AI-based technologies and solutions.⁹⁸ Social and ethical challenges around “human safety, privacy, integrity, dignity, autonomy, and data ownership”⁹⁹ are highlighted as especially pressing. In particular, AI and robotics are seen as potential factors in the erosion of privacy through generation of large amounts of personal data that can be used as a “currency” to purchase “services.”¹⁰⁰

One of the major differences occurring between this report and the other two is that the EU report does not explicitly include unembodied AI, nor does it include accountability or transparency as guiding ethical principles. In part, this may be explained by the narrower, primary focus on robotics and civil liability.¹⁰¹ As mentioned, a major concern of the report is civil liability¹⁰² for robotics, which is implicitly related to accountability. Therefore, in terms of consistency, one may consider accountability as an implicit, underlying guiding principle of the EU report. Furthermore, transparency is only addressed in the proposal for licensing robotics designers in the Charter, which includes requirements for the behaviours of robots to be traceable, transparent, predictable, reversible, and explainable.¹⁰³ As with the treatment of accountability, the omission of transparency as an explicit guiding ethical principle may merely reflect the narrower focus of the report. Unfortunately, the situation that both accountability and transparency were not placed in the foreground informing the EU report represents a missed opportunity when it comes to outlining the ethical framework for a ‘good AI society’.

⁹⁵ European Parliament Committee on Legal Affairs (2016, p. 8).

⁹⁶ This was mentioned in the context of possibly assigning electronic personhood to robots. See European Parliament Committee on Legal Affairs (2016, p. 12).

⁹⁷ European Parliament Committee on Legal Affairs (2016, p. 10ff).

⁹⁸ European Parliament Committee on Legal Affairs (2016, p. 7).

⁹⁹ European Parliament Committee on Legal Affairs (2016, p. 7).

¹⁰⁰ European Parliament Committee on Legal Affairs (2016, p. 8).

¹⁰¹ However, despite the narrow focus the report does cover a wider set of issues that make this report comparable to its British and American equivalents.

¹⁰² The specific focus on civil liability rules of the EU report comes from its ability to regulate this particular area, whereas in some of the other recommendation areas it may or may not be able to promote the proposals made. The report’s focus on what is clearly a rather narrow competency, allows it to also venture out into more ambitious proposals. Yet, the EU does not have the flexibility of its counterparts to suggest broad more generic approaches that deal with all aspects of AI.

¹⁰³ European Parliament Committee on Legal Affairs (2016, p. 18).

The UK Report: Keep Calm and Commission on

On October 13th 2016, the House of Commons' Science and Technology Committee released the UK report on AI.¹⁰⁴ The report aimed to identify "the potential value and capabilities [of AI and robotics], as well as examining prospective problems, and adverse consequences, that may require prevention, mitigation and governance".¹⁰⁵ The language used in the report implies a sense of urgency. This may be explained by the fact that its intended audience is the British government. The report holds that the UK's position at the forefront of AI development comes, amongst other factors, from the work done in academia and that the government needs to ensure that funding remains available for AI research.¹⁰⁶ However, the committee asserts that the UK government is trailing behind and running the risk of losing its competitive edge as a thought leader on AI. The sense of urgency might also have been further increased by the uncertainty surrounding the impact that the United Kingdom's withdrawal from the European Union (Brexit) will have on research funding.

Much like the US government, the UK committee suggests that the UK government should maintain its light-touch regulation¹⁰⁷ of the AI sector. This is seen as one of the main reasons why the UK and especially London, is a hub for the tech-industry in Europe and the intellectual home of start-ups like DeepMind,¹⁰⁸ the leading AI company now part of Google. It follows that the critique leveraged against this approach in the analysis of the US report also applies to the UK report, especially as the report does not make explicit its assumptions about the importance of the free market economy. It must be remarked that the committee seems to have visited only one AI company, namely Google DeepMind, which is based in King's Cross, London.¹⁰⁹ This gives the impression that special preference to a particular private sector player might have been given, over other concerns. On a more positive note, the UK report does establish a clear role for the government to play in the development of AI, mainly through "careful scrutiny of its ethical, legal and societal dimensions".¹¹⁰ This can be seen most clearly in the call for the

¹⁰⁴ House of Commons Science and Technology Committee (2016a).

¹⁰⁵ Ibid, p. 7.

¹⁰⁶ The report states that: "There is not a Government strategy for developing the skills, and securing the critical investment, that is needed to create future growth in robotics and AI. Nor is there any sign of the Government delivering on its promise to establish a 'RAS Leadership Council' to provide much needed coordination and direction. Without a Government strategy for the sector, the productivity gains that could be achieved through greater uptake of the technologies across the UK will remain unrealised. (Paragraph 98)" (2016, p. 37).

¹⁰⁷ The report states that: "While it is too soon to set down sector-wide regulations for this nascent field, it is vital that careful scrutiny of the ethical, legal and societal dimensions of artificially intelligent systems begins now." 2016, pp. 25, 36.

¹⁰⁸ "DeepMind." (2016).

¹⁰⁹ House of Commons Science and Technology Committee (2016b, pp. 34–35).

¹¹⁰ House of Commons Science and Technology Committee (2016b, p. 36).

development of novel regulatory frameworks¹¹¹ and principles, which can address the unique legal and ethical issues raised by AI and robotics. The committee states that this should not be done solely by the government, but rather through establishing “a standing Commission on Artificial Intelligence”¹¹² at the Alan Turing Institute¹¹³ and broad stakeholder collaboration.¹¹⁴ This Commission will be tasked with developing principles to govern the development and application of AI. As acknowledged in the document, the same Committee previously also recommended the creation of a Council for Data Ethics.¹¹⁵ The report recommends coordination, but the impression is that the two bodies would need to be firmly connected and probably unified into a single entity, because the ethics of AI, robotics, and machine-learning need more than just being “closely coordinated”¹¹⁶ with the field of the ethics of data, algorithms, and practices (e.g. responsible innovation). Hardware, software, and data constitute a single ecosystem, which needs a comprehensive and systematic normative approach.¹¹⁷ Any fragmentation would be hugely detrimental in terms of efficiency and efficacy. For example, algorithms may be biased because of the data on which they are trained or because of the low-quality data that they are fed, or they may indeed not be biased but produce biased data that go on making an AI-application unfair. The recent scandal affecting Amazon’s Prime Free Same-Day Delivery is a good illustration.¹¹⁸ A comprehensive approach seems to be the only reasonable way forward.

Returning to the report, it suggests that the new standing commission will be tasked with providing leadership on the ethical, legal, and social implications (ELSI) of AI, currently perceived to be lacking at the government level. It would be a watchdog, and the launch platform for the development of next steps, like regulatory frameworks or bodies.

The Science and Technology Committee seems unwilling to assert in the report whether AI will replace or augment human labour.¹¹⁹ However, it does state that the

¹¹¹ The report holds: “Though some of the more transformational impacts of AI might still be decades away, others—like driverless cars and supercomputers that assist with cancer prediction and prognosis—have already arrived. The ethical and legal issues discussed in this chapter, however, are cross-cutting and will arise in other areas as AI is applied in more and more fields. For these reasons, witnesses were clear that the ethical and legal matters raised by AI deserved attention now and that suitable governance frameworks were needed.”, 2016, p. 22.

¹¹² House of Commons Science and Technology Committee (2016b, pp. 3, 26, 36).

¹¹³ The Alan Turing Institute (2016).

¹¹⁴ The report holds that: “Membership of the Commission should be broad and include those with expertise in law, social science and philosophy, as well as computer scientists, natural scientists, mathematicians and engineers. Members drawn from industry, NGOs and the public, should also be included and a programme of wide ranging public dialogue instituted.” 2016, p. 37.

¹¹⁵ House of Commons Science and Technology Committee (2016b). pp. 7, 11, 12.

¹¹⁶ House of Commons Science and Technology Committee (2016b), pp. 26, 36.

¹¹⁷ Floridi and Taddeo (2016).

¹¹⁸ Ingold and Soper (2016).

¹¹⁹ The report states that: “Advances in robotics and AI hold the potential to reshape fundamentally the way we live and work. While we cannot yet foresee exactly how this ‘fourth industrial revolution’ will play out, we know that gains in productivity and efficiency, new services and jobs, and improved support

government should take the lead in ensuring the UK is ready for the changes brought by AI, by addressing the digital skills gap in the current population.¹²⁰

Clearly, there are important similarities between this and the other two reports. The US's emphasis on transparency, minimizing bias, accountability and adjusting the educational system is shared by the UK report as well. The specific focus on robotics is central to the UK and EU reports. The UK report also considers various national security questions. It does not mention predictive policing, but it does suggest that some types of profiling on the basis of AI can lead to discrimination.¹²¹ On the subject of autonomous weapons and, more specifically, "lethal autonomous weapons systems"¹²² (LAWS) the committee calls for additional accountability measures, as these technologies have the potential to kill without human intervention.¹²³ Like the US report, the UK report's position is that "international humanitarian law remains the appropriate legal basis and framework for the assessment of the use of all weapons systems in armed conflict".¹²⁴ However, the recommendations in the UK report go a bit further by suggesting that the government could do more to explain how humans will remain part of the control system of these weapons.¹²⁵

On education, although the recommendation on closing the digital skills gap¹²⁶ is important, it should be seen as more than an aim in itself but also as an opportunity for the government to develop an explicit vision of the role of AI in society. As long as that remains unclear, attempts to upscale education efforts are probably aimless and may end up being mere palliative. The same holds true for any measure a government may take in socialising the costs of AI-induced unemployment through a welfare mechanism. 'Play it by ear' is, in the case of AI, an unsatisfactory tactic.

The British committee makes a concrete recommendation about how to regulate transparency that goes beyond any of the suggestions present in the US report. The committee suggests that, regarding protections against automated decision-making, the government should take note of European General Data Protection Regulation (GDPR),¹²⁷ which is poised to come into effect in 2018. Admittedly, it is unclear how the GDPR will apply to the UK, as the Brexit negotiations are expected to be finalised around the time the GDPR will come into force. So, the reluctance¹²⁸ of the report to make bolder statements on this subject is understandable. And yet, a more explicit and substantive ethical position on transparency and accountability

Footnote 119 continued

in existing roles are all on the horizon, alongside the potential loss of well-established occupations. Such transitions will be challenging." 2016, p. 15.

¹²⁰ House of Commons Science and Technology Committee (2016b, pp. 5, 13, 36).

¹²¹ House of Commons Science and Technology Committee (2016b, p. 18).

¹²² House of Commons Science and Technology Committee (2016b, p. 21).

¹²³ Ibid.

¹²⁴ Ibid.

¹²⁵ House of Commons Science and Technology Committee (2016b, p. 22).

¹²⁶ House of Commons Science and Technology Committee (2016b, p. 13).

¹²⁷ European Union (2016).

¹²⁸ House of Commons Science and Technology Committee (2016b, p. 18).

remains a missed opportunity, given that it is especially in this area that strong leadership is most needed and could have been exercised.

As already mentioned, the committee recommends that a “standing Commission on Artificial Intelligence” should be set up at the Alan Turing Institute to provide advice and encourage public discussion on the application and development of AI. This is a completely original point, unparalleled by the other reports, which do not make any comparable suggestion for a Commission paired with public debate. The standing Commission would be made up of a diverse and interdisciplinary group of individuals covering the fields of computer science, engineering, law, math, social science, and philosophy.¹²⁹ This is an excellent suggestion.¹³⁰ However, it will have the intended effect only if the government, industry, and the research sector will rely on the advice given by such a Commission, to devise a view of what the good AI society should look like. The committee does recommend that more public dialogue on AI should be held. This suggestion too is to be welcomed. Especially considering how fruitful this approach proved to be in the US context. But it is also to be hoped that the onus of making such a dialogue happen will not be placed solely on industry or academia.

The UK report gives an overview of the various issues related to AI as they play out in the UK context. It takes a less definitive stance on how to start preparing for the future of AI, providing more of an overview of the arguments made by the various experts consulted. This is partly due to the very nature of the committee’s reports, which are purposefully based on expert consultations. But it may also be partly due to the committee having an implicit view on how AI should fit into society. Yet this never solidifies into an explicit and clear strategy for the good AI society. Even though the focus on free market principles echoes through the report. Nor does the report offer a strategic plan, based on R&D or otherwise, to follow up on the recommendations made, differently from the US report. Yet one should recall that its nature is that of a proposal to the government. The government then usually has 60 days to reply to the committee’s recommendations, specifying how far they are taken on board and how they may be implemented.

It seems clear from the original report, and from a more recent briefing from the Government Office for Sciences,¹³¹ that the underlying view is that the UK government, its private sector and its academic institutions should collaborate in driving the creation of a framework for the regulation of AI. This is commendable. There is simply too much overlap between social, political, commercial, and research interests for a single actor to have a monopoly on the ethics of AI and dominate the whole agenda. Rather, as we shall argue in the next, concluding section, the right recommendations and policies should be developed through an independent multi-stakeholder process driven by governments that brings together *all* those impacted by AI, including civil society and Non-Governmental Organizations (NGOs) for example, in order to bring about the best framework to deliver a ‘good AI society’.

¹²⁹ House of Commons Science and Technology Committee (2016b, p. 22).

¹³⁰ Disclosure: please note that please note that ATI.

¹³¹ UK Government Office for Science (2016).

Conclusions

In the previous pages, we highlighted several common values found across the three reports in relation to AI, machine learning, algorithms, and robotics. In particular, *transparency*, *accountability*, and a ‘*positive impact*’ on the economy and society are among the key values indicative of the kind of view of a ‘good AI society’ that seems to underlie the three reports, even if a more encompassing and ambitious vision is not explicitly stated. The reports are especially valuable in identifying several of the most salient issues surrounding AI, like its impact on the economy, education, warfare, diversity, and national security. Some of the best practices suggested in the different reports are summarized below.

The US report is to be praised for being the only one to have an elaborate R&D strategy to support its recommendations. It also does an excellent job in including the work of experts and the public through the public workshops and the government’s “Request for Information”. The EU report helpfully recommends the creation of a “European Agency for Robotics and AI”, which would be tasked not only with monitoring the trends in AI but also with envisioning its future impact and with advising public players. The EU report also makes several useful recommendations for legislation, reflecting a ‘less light touch’ approach to governance of AI and robotics. The UK report rightly calls both for the development of novel regulatory frameworks and for relying on existing regulation like the GPDR. It is also the only report to suggest the creation of an independent standing, national Commission, which organizes public debate about the challenges brought about by AI.

Each report specifies the role and responsibility of the government, the private sector, and the research sector. Another common theme is the importance of cooperation between the different leading actors involved in the development of AI. All three reports appoint different actors to spearhead this cooperation. For the US it is the government with private industry. For the EU it is the European Commission and a new advisory agency. For the UK it is a “coordinated approach”¹³² between the government and a standing Commission.

The reports also have different ways of defining what specific values should guide the development of AI. The US report focuses on the “public good” and “fairness and safety”¹³³ as guiding principles. Its compendium R&D report describes a vision for the future,¹³⁴ focusing on the specific impact to be aimed for in different sectors.¹³⁵ The adjoining Economic Report identifies specific policies

¹³² House of Commons Science and Technology Committee (2016b, p. 3).

¹³³ Executive Office of the President National Science and Technology Council Committee on Technology (2016, pp. 2, 30–32).

¹³⁴ The report’s companion document, entitled the “National Artificial Intelligence Research and Development Strategic Plan”, details how AI should ideally impact various sectors, pp. 8–10.

¹³⁵ As said before, the vision laid out in the R&D report cannot be seen as indicative of the Government approach in the same way that the general report can, as the R&D report focuses specifically on: “defining a high-level framework that can be used to identify scientific and technological gaps in AI and track the Federal R&D investments that are designed to fill those gaps. The AI R&D Strategic Plan identifies strategic priorities for both near-term and long-term support of AI that address important

responses to “amplify the best and temper the worst impacts” of AI and automation.¹³⁶ The EU report calls for “intrinsically European and humanistic values” to ground “rules, governing in particular liability and ethics” of robotics and AI,¹³⁷ represented in a “guiding ethical framework for the design, production and use of robots.”¹³⁸ The UK report emphasises the importance of examining “the social, ethical, and legal implications of recent and potential developments in AI”¹³⁹ and developing “socially beneficial AI systems”.¹⁴⁰

That the different reports define various constellations of responsibility, emphasise the importance of cooperation, and mention specific areas of concern or even values to be upheld are all steps in the right direction. What is lacking in all the three reports is a tightly woven understanding of how responsibility, cooperation, and values fit together to design and steer the development of a ‘good AI society’. This is relevant not just for our societies today, but also, if not mainly, for the ‘mature information societies’ in which future generations will live. Although the three reports and compendiums clearly address some of the most important and thorny questions posed by current developments in AI, their impact would have been greater had they comprehensively integrated their ethical evaluations, already present in their implicit vision on AI, with a foresight analysis of the sort of society we would like to build. What is lacking is an ambitious and bold attempt to deal with the most difficult question behind the whole debate: what is the human project for the mature information societies of the twenty-first century? It is certainly not the task of this comparative analysis to answer such a momentous question. However, by way of conclusion, we would like to contribute to its debate by recommending a two-pronged approach¹⁴¹ to it, in order to steer the process of developing the “good AI society”, as well as suggest an ethical and legal principle to guide it.

On the one hand, policies should ensure that AI is steered fully towards promoting the public good. To this end, we need a clear and convincing understanding of what kind of ‘good AI society’ we wish to develop. Such understanding can best achieved through an independent, international, multi-stakeholder process of research and consultations on AI and Data Ethics. This process should bring together governments, the corporate sector, civil society, and

Footnote 135 continued

technical and societal challenges. The AI R&D Strategic Plan, however, does not define specific research agendas for individual Federal agencies. Instead, it sets objectives for the Executive Branch, within which agencies may pursue priorities consistent with their missions, capabilities, authorities, and budgets, so that the overall research portfolio is consistent with the AI R&D Strategic Plan. The AI R&D Strategic Plan also does not set policy on the research or use of AI technologies nor does it explore the broader concerns about the potential influence of AI on jobs and the economy.” 2016, p. 7.

¹³⁶ Executive Office of the President (2016, p. 22).

¹³⁷ European Parliament Committee on Legal Affairs (2016, p. 4).

¹³⁸ European Parliament Committee on Legal Affairs (2016, p. 7).

¹³⁹ House of Commons Science and Technology Committee (2016b, pp. 26, 36).

¹⁴⁰ House of Commons Science and Technology Committee (2016b, pp. 25, 36).

¹⁴¹ This approach is based on the work on digital ethics developed at the University of Oxford and at The Alan Turing Institute by our research group.

the research community in order to establish an international, independent, multi-stakeholder Council on AI and Data Ethics. This Council can then be instrumental in advising the various stakeholders, especially governments, organisations, and companies, on how to design comprehensive, socio-political strategies that support the widespread application of AI solutions that are environmentally friendly and socially preferable. Governments could take the lead in organizing this process, as they have the democratic mandate to develop regulation for AI and can be held accountable for their decisions, in a way that the private sector and the research community cannot. However, the Council, once established, should be independent. This will ensure that all the stakeholders impacted by AI can, on a rolling basis, equal footing and as the technology evolves, provide input that shapes our AI-powered society.

On the other hand, the 'good AI society' projects could fruitfully rely on the concept of human dignity as the lens through which to understand and design what a good AI society may look like. Of course, there are drawbacks to using the concept of dignity in this context. Many have argued it is an empty concept. And it certainly does not automatically mean the same thing to different sets of people (Floridi 2016b). That being said, our focus is specifically on human dignity as implicitly assumed in the new European General Data Protection Regulation (GDPR),¹⁴² and included in the 1948 Universal Declaration of Human Rights (Preamble and Article 1), and in the EU Charter of Fundamental Rights.¹⁴³ This approach to human dignity provides the much needed grounding in a well-established, ethical, legal, political, and social concept, which can help to ensure that tolerant care and fostering respect for people (both as individuals and as groups), their cultures and their environments, play a steering role in the assessments and planning for the future of an AI-driven world. By relying on human dignity as the pivotal concept, it should become less difficult to develop a comprehensive vision of how responsibility, cooperation, and sharable values can guide the design a 'good AI society'.

Digital technologies, practices, sciences, goods, and services can be enormously beneficial for human flourishing. AI plays a crucial role in such a wider trend. But we are fragile entities, delicate systems, vulnerable individuals and AI can easily become the elephant in the crystal room, if we do not pay attention to its development and application. Exposed to such extraordinary technologies, human life may easily be distorted, with humans adapting to inflexible technologies, following their predictive suggestions in self-generated bubbles, or being profiled into inescapable and generic categories, for example. We need to ensure that our new smart technologies will be at the service of the human project, not vice versa.

So, a first step for a future Council on AI and Data Ethics would be not so much to advice ethically and normatively about the world of AI innovation, but to provide foresight¹⁴⁴ by describing the future that, as a society, we would like to see AI

¹⁴² European Union (2016).

¹⁴³ <http://fra.europa.eu/en/charterpedia/article/1-human-dignity>.

¹⁴⁴ The importance of such foresight has been elaborately described by one of us: "The development of ICT has not only brought enormous benefits and opportunities but also greatly outpaced our understanding of its conceptual nature and implications, while raising problems whose complexity and global dimensions are rapidly expanding, evolving and becoming increasingly serious. A simple analogy

contribute to bringing about. This two-pronged approach is ambitious, but far from impossible. Similar initiatives can be found in the realm of Internet Governance, for example, where Internet standards setting bodies are run using such a bottom-up multi-stakeholder approach to governing and developing technology. A multi-stakeholder initiative, paired with an international Council is, we believe, the way forward to try to ensure that the development and impact of AI is kept on course to achieving the sort of good societies in which human dignity may flourish.

Acknowledgements We discussed multiple versions of this article on various conferences and mailing lists. Specifically, the first author discussed some of the ideas included in this article at the IEEE Global Initiative for Ethical Considerations in the Design of Autonomous Systems conferences in Brussels. We are deeply indebted for the feedback we received from these various communities and audiences. In particular, we wish to thank the three anonymous reviewers whose comments greatly improved the final version. We also want to thank John Havens, Greg Adamson and Inez De Beaufort for their insightful comments and for the time they put into discussing the ideas presented in this article.

References

- Annany, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, 1–17. <http://journals.sagepub.com/doi/pdf/10.1177/1461444816676645>.
- Calo, R. (2014). *The case for a federal robotics commission* Brookings Institution. Retrieved from <https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/>.
- Crawford, K. (2016). *Artificial intelligence's white guy problem*. Retrieved from http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=1.
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature News*, 538(7625), 311. doi:10.1038/538311a.
- DeepMind. (2016). *DeepMind*. November 15. <https://deepmind.com/about/>.
- European Parliament Committee on Legal Affairs. (2016). *Civil law rules on robotics (2015/2103 (INL))*. Brussels, Belgium: European Parliament. Retrieved from <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARI%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0/EN>.
- European Union. (2016). *European Union (EU) General Data Protection Regulation 2016/679*. Brussels, Belgium. Retrieved from http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.
- Executive Office of the President. (2016). *Artificial intelligence, automation and the economy*. Washington, DC, USA. Retrieved from <https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>.
- Executive Office of the President National Science and Technology Council Committee on Technology. (2016). *Preparing for the future of artificial intelligence*. Washington, DC, USA. Retrieved from https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
- Felten, E. W. (2016). Preparing for the future of artificial intelligence. *White House Website Blog*. Retrieved from <https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence>.

Footnote 144 continued

may help to make sense of the current situation. Our technological tree has been growing its far-reaching branches much more widely, rapidly and chaotically than its conceptual, ethical and cultural roots. (...) The risk is that, like a tree with weak roots, further and healthier growth at the top might be impaired by a fragile foundation at the bottom.” He also states that: “as a consequence, today, any advanced information society faces the pressing task of equipping itself with a viable philosophy and ethics of information”. We argue that this argument needs to be extended to the realm of governance, which equally needs a clear vision to root the tree of AI. See Floridi (2010).

- Felten, E. W., & Lyons, T. (2016). Public input and next steps on the future of artificial intelligence. *Medium*. Retrieved from <https://medium.com/@USCTO/public-input-and-next-steps-on-the-future-of-artificial-intelligence-458b82059fc3#.fj949abr5>.
- Finley, K. (2016). *Obama wants to help the government to develop AI*. Retrieved from <https://www.wired.com/2016/10/obama-envisions-ai-new-apollo-program/>.
- Fleury, M. (2015). *How artificial intelligence is transforming the financial industry*. Retrieved from <http://www.bbc.co.uk/news/business-34264380>.
- Floridi, L. (2010). Ethics after the information revolution. In L. Floridi (Ed.), *The Cambridge handbook of information and computer ethics* (pp. 3–19). Cambridge: Cambridge University Press. Retrieved from <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521888981>.
- Floridi, L. (2013). Infraethics. *Philosophers' Magazine*, 60(1), 26–27.
- Floridi, L. (2014). *The fourth revolution. How the infosphere is reshaping human reality*. Oxford, UK: Oxford University Press.
- Floridi, L. (2016a). Mature information societies—A matter of expectations. *Philosophy and Technology*, 29(1), 1–4. doi:10.1007/s13347-016-0214-6.
- Floridi, L. (2016b). On human dignity as a foundation for the right to privacy. *Philosophy and Technology*, 29(4), 307–312.
- Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society*, 374(2083), 1–4. doi:10.1098/rsta.2016.0360.
- Furlow, B. (2016). *IBM Watson collaboration aims to improve oncology decision support tools*. Retrieved from <http://www.cancernetwork.com/mbcc-2016/ibm-watson-collaboration-aims-improve-oncology-decision-support-tools>.
- Hart, A. (1961). *The concept of law*. Oxford: Clarendon.
- House of Commons Science and Technology Committee. (2016a). *Robotics and artificial intelligence* (No. Fifth Report of Session 2016–17). London, UK. Retrieved from <http://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>.
- House of Commons Science and Technology Committee. (2016b). *The Big Data dilemma: Government response to the Committee's fourth report of session 2015–16 contents*. <http://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/992/99204.htm>.
- Ingold, D., & Soper, S. (2016). Amazon doesn't consider the race of its customers. Should It? Retrieved from <http://www.bloomberg.com/graphics/2016-amazon-same-day/>.
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165, 1. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268.
- Leroux, C., & Labruto, R. (2013). A green paper on legal issues in robotics. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/310167745_A_green_paper_on_legal_issues_in_robotics.
- Libicki, M. C. (2009). Cyberdeterrence and cyberwar. *The RAND Corporation*. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*. doi:10.1177/2053951716679679.
- National Science and Technology Council Networking and Information Technology. Networking and Information Technology Research and Development Subcommittee. (2016). *The national artificial intelligence research and development strategic plan* (Washington DC, USA). Retrieved from https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.
- Pagallo, U. (2016a). Three lessons learned for intelligent transport systems that Abide by the law. In *Jusletter IT* 24.
- Pagallo, U. (2016b). Even angels need the rules: AI, roboethics, and the law. *ECAI*, 258, 209–215.
- Palmerini, E., Bertolini, A., Battaglia, F., Koops, B.-J., Carnevale, A., & Salvini, P. (2016). RoboLaw: Towards a European framework for robotics regulation. *Robotics and Autonomous Systems*, 86, 78–85. doi:10.1016/j.robot.2016.08.026.
- Partnership on AI. (2016). Retrieved from <https://www.partnershiponai.org/>.
- Quackenbush, S. L. (2011). Deterrence theory: Where do we stand? *Review of International Studies*, 37(2), 741–762.
- Request for Information on Artificial Intelligence. (2016). Science and technology policy office. Retrieved from <https://www.federalregister.gov/documents/2016/06/27/2016-15082/request-for-information-on-artificial-intelligence>.

- Schafer, B. (2016). *Closing Pandora's box?* (pp. 55–67). Law and Technology: The EU Proposal on the Regulation of Robots. Pandora's Box.
- Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law and Technology*, 29(2), 372. <http://dx.doi.org/10.2139/ssrn.2609777>.
- Taddeo, M. (2016a). Just information warfare. *Topoi*, 35(1), 213–224.
- Taddeo, M. (2016b). On the risks of relying on analogies to understand cyber conflicts. *Minds and Machines*, 26(4), 317–321.
- The Alan Turing Institute. (2016). Accessed September 1. <https://www.turing.ac.uk/>.
- Tutt, A. (2016). An FDA for algorithms. *Administrative Law Review*, 67, 18. Available at SSRN: <https://ssrn.com/abstract=2747994>.
- UK Government Office for Science. (2016). *Artificial intelligence: An overview for policy-makers*. Retrieved from <https://www.gov.uk/government/publications/artificial-intelligence-an-overview-for-policy-makers>.
- United States Standards Strategy Committee. (2015). *United States Standards Strategy*. Retrieved from https://share.ansi.org/shared%20documents/Standards%20Activities/NSSC/USSS_Third_edition/ANSI_USSS_2015.pdf.
- Wachter, S., Mittelstadt, B. D., & Floridi, L. (Forthcoming). *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*. Retrieved from Available at SSRN: <https://ssrn.com/abstract=2903469>.