

Governing in the Digital Age

A PGI Working Paper



Prepared by Marie Blythe-Hallman
Edited by Jocelyne Bourgon P.C., O.C.
March 2019

NS is an international co-operation initiative led by the Honourable Jocelyne Bourgon P.C., O.C.

PGI Working Paper: Governing in the Digital Age

By: Marie Blythe-Hallman
Edited By: Jocelyne Bourgon P.C., O.C.

© Public Governance International, 2019

Published by Public Governance International (PGI)

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author.

The content of the document is property of Public Governance International. The use of the document is restricted to participants in NS Labs, NS events and to NS partners.

April, 2019

ISBN: 978-1-989294-06-2

Table of Contents

I.	<i>A New Synthesis Initiative</i>	1
II.	<i>New Technologies: The Promises, Perils & Increasing Velocity of Change</i>	5
III.	<i>Social and Ethical Concerns</i>	13
	• <i>A Lack of Public Awareness and Concern</i>	14
	• <i>Data and Algorithmic Bias</i>	15
	• <i>The “Black Box” problem</i>	16
	• <i>The Value and Ownership of Data</i>	17
	• <i>The Ethical Application of AI</i>	18
IV.	<i>Learning from Practice: Government Led Initiatives</i>	20
	• <i>Industry Self-Regulation</i>	21
	• <i>Government Declaration of Principles</i>	24
	• <i>Regulation</i>	26
	• <i>Comprehensive Approach to Building a Digital Society</i>	28
V.	<i>Conclusion</i>	33
	<i>Annex</i>	34
	<i>Bibliography</i>	37

I. The New Synthesis Initiative

Governing in a Digital Age is one in a series of working papers produced for the Third Phase of the New Synthesis (NS) Initiative.

The NS Initiative is a *collaborative international research initiative* that was launched in 2009 with the explicit purpose of exploring the *new frontiers* of public administration to provide practitioners with a mental map adapted to the challenges of serving in the 21st century. Seeking insights from theory and practice, and testing ideas in a diversity of environments are a trademark of the NS Initiative¹.

Public administration has been lacking a *New Synthesis* able to integrate past practices of enduring value with new ones better aligned to the challenges of serving in a global, hyper-connected world and in the midst of a technological revolution. *A broader mental map and a different approach to problem-solving* are needed to invent solutions to the increasingly complex problems governments are facing in practice.

The work of the Initial Phase of the NS Initiative revealed that people serving in government today are facing a combination of factors that is significantly different from those prevailing during the post-World War II period to the early 2000s; increasing complexity, hyper-connectivity, high uncertainty, a technological revolution, a digital and biological revolution, the acceleration of environmental changes, rapid changes to the nature of work, the impact of social media, and the like. These factors are transforming the economic,

social, and political spheres of life in society. The pace of change is increasing and there is every reason to believe that *the velocity of change will continue to accelerate*.

The initial phase generated a conceptual framework that brings together the role of government, citizens and society in a dynamic and interactive system. The NS Framework:

- Expands the range of options open to government;
- Improves the likelihood of success of government actions and interventions;
- Brings special attention to society's resilience and adaptive capacity building; and
- Encourages system thinking and collective problem solving.

The Second Phase generated the *NS Exploratory Cycle*. This phase focused on *what can be done to*

¹ Jocelyne Bourgon, 2011, *A New Synthesis of Public Administration: Serving in the 21st Century*, Kingston: Queen's Policy Studies; Jocelyne Bourgon, 2019, "NS Research Program 2019" (unpublished). Also see website: <https://www.pgionline.com/>

ensure that the capacity of government to invent solutions will keep pace with the increasing complexity of the problems we are facing as a society. Based on the work of 1,000 practitioners in a diversity of contexts and circumstances, this phase confirmed the importance of a broader mental map and of dynamic systems thinking to invent solutions, encourage collective problem solving and build the resilience of society. More than ever, *governing in the 21st century is a process of invention*; it is not a process of replication.

The second phase underscored the need to explore more deeply the importance of civic results to propel society forward in a period of unprecedented changes². Civic results include but are not limited to:

Civic capacity: The capacity of people, families and communities to take charge of issues and to initiate actions with others and with government in a manner that addresses their concerns and promotes the overall interest of society.

Civic will: The will to deploy capabilities to build and share a better future and to contribute to collective problem solving as member of a broader human community.

Civic values (norms): Shared values and normative behaviours that

contribute to harmonious living and making society governable.

This is the focus of the *Third Phase* - to dive deeply into civic results and how they affect the overall functioning of a governing system.³ The aim of the research is to generate a *coherent and cogent synthesis* of ideas and principles about what government can do to accelerate the adaptive capacity and resilience of society and the capacity for collective problem-solving. The NS 2019 Research Agenda directs us to four questions in particular:

- What can government do to build the *collective capacity of society* to invent and share a better future together?
- What can government do to ensure that the *adaptive capacity of society* will keep pace with the increasing velocity of change?
- What can government do to enhance the *resilience of society* to adapt, evolve and prosper in unforeseen and unpredictable circumstances?
- What can government do to ensure that public institutions have the *capabilities* to successfully steer society through

² Bourgon, 2011.

³ NS Research Program 2019.

an unprecedented period of change?

The Digital Revolution

“Technological innovations have periodically transformed the world we live in and have played a key role in human history...Today we are witnessing the early signs of another technological revolution. Countries are navigating through a period of transformation that will be as deep and steep as what was experienced during the Industrial Revolution.”⁴

The first and second phases of the NS Initiative identified several important factors at play in accelerating the velocity of change that is transforming the economic, social and political spheres. Among these *accelerators of change* is the digital technological revolution. In its 2016 report on the future of jobs, the World Economic Forum (WEF) concluded that we are at the beginning of a Fourth Industrial Revolution.⁵ Citing advancements in artificial intelligence and machine learning, robotics, nanotechnology, 3D printing, genetics and biotechnology, which are “all building

on and amplifying one another,” the WEC cautioned that these developments will “lay the foundation for a revolution more comprehensive and all-encompassing than anything we have ever seen.”⁶ In a 2018 update to its initial report, the WEC concluded that the “fundamental pace of change has only accelerated.”⁷

As the work of the NS Initiative notes, technological developments *evolve at the pace of scientific discoveries* while the social sphere and governance systems do so at a much *slower pace*. Indeed, the digital revolution is generating disruptions and dislocations that exceed the absorptive and adaptive capacity of many governments. At the same time, it is transforming what it means to be a citizen and the relationship between the state and citizens. In sum, this raises questions about *what it means to govern in the digital age*.

A period of accelerating disruptive changes brings to prominence the need to *accelerate the adaptive capacity* of government and to build a *resilient society*. This paper, *Governing in a Digital Age*, explores some of the literature around the challenges of governing in this context and what can be learned from the early initiatives of governments to steer their societies

⁴ Jocelyne Bourgon, 2017, *The New Synthesis of Public Administration Fieldbook*, Denmark: Dansk Psykologisk Forlag A/S, pp.32-3.

⁵ World Economic Forum, 2016, *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*, World Economic Forum, http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf, accessed Feb.2, 2019.

⁶ WEC, p.v.

⁷ World Economic Forum, 2018, *The Future of Jobs Report 2018*, World Economic Forum, (January), p.v, http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf, accessed Feb.2, 2019.

through an unprecedented period of change.

The first section of this paper provides an overview of the literature around governing in the digital age; it promises and the perils. Section two highlights the growth and global nature of artificial intelligence (AI), a technology more than any other that is causing industry, governments, academe, and other civil society actors to sit up and take notice. The third section explores more deeply the social and ethical concerns associated with data, algorithms and AI. Finally, section four looks at some recent approaches, policies and practices adopted by governments to respond to the ethical and societal challenges AI technologies are posing, among them the United States, South Korea, Estonia, Singapore, Dubai, Germany, Barcelona, and the European Union.

II. New Technologies – The Promises, Perils and Increasing Velocity of Change

One of the challenges of navigating through the literature on governing in the digital age is its abundance. As governments and societies struggle to keep pace with the velocity of technological innovation, so too do those trying to grasp the promises and perils of this unprecedented period of change, and to discern what governments can do to guide their societies through. While this growing body of literature is united in the view that we have indeed entered a Fourth Industrial Revolution, it differs in the assessment of the opportunities and risks the digital revolution presents: that which is guided by ‘the excitement of unprecedented potential,’ and that which cautions against “the capacity to absorb the dislocation associated with this transformation.”⁸

Work underway in government and the private sector about the use of digital technologies is generally optimistic about what digital has to offer; the challenge is to figure out how to harness its potential while mitigating or preventing problems that may be difficult to correct at a later stage.

For many governments, the question of governing in a digital age is one of *transformation*,⁹ by which they mean “a significant step change” in service delivery and internal operations.¹⁰ Technology is regarded as an *enabler of transformation*, or as the Head of Digital and Data in the UK Government, Matthew Cain, describes, a way to “build a better society.”¹¹ The focus is how to benefit from the *productivity*

and *efficiency gains* that technology promises in a way that serves the overall interest of society. The view is that digital transformation stands to benefit governments and citizens by opening up opportunities for improving the service delivery experience, revamping administrative processes and extending self-service via chatbots, automated application process, and the like.¹²

A sense of urgency about going digital has started to emerge; harnessing digital technology is seen as a *requirement* for governments to remain relevant to their citizens. The Honourable Scott Brison, former President of the Treasury of Canada, shares this view: “In the 21st century,

8 Bourgon, 2011, p.33.

9 Estonia is perhaps the best and most well-known example of digital transformation. In addition to Canada and the U.K., other examples include Singapore, Australia, New Zealand, and the Netherlands.

10 Gov.UK, Cabinet Office, Government Digital Service, 2017, *Policy Paper: Government Transformation Strategy: Background*, (Feb 9), <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-background>, accessed Nov. 16, 2018.

11 Gov.UK.

12 See, for example, Ray Briggs, Ed Dobner, Jennifer Dul, 2018, “Digital Reality in Government: How AR and VR can enhance Government Services,” (Aug 24), Deloitte, <https://www2.deloitte.com/insights/us/en/industry/public-sector/augmented-virtual-reality-government-services.html>, accessed Nov 16, 2018.

you're either digital or you're dead. If a company fails to get digital right, it's out of business. If a government fails to get digital right, it's out of touch with its citizens...Right now we are a Blockbuster Government serving a Netflix citizenry.”¹³

For their part, industry experts like Deloitte, Forrester and Gartner are providing advice and offering solutions to government about how best to go digital. They package webinars, workshops and conferences about *digital disruption* targeted to public servants and leaders working in the fields of information technology, information management, and cybersecurity.

By focusing on service delivery and operations, however, others caution that governments risk missing out on what the new digital era truly has to offer as a means of serving the *public good*. Recent work by McKinsey (2018), for instance, offers an inventory of applications for new technologies like AI across a broad range of social domains.¹⁴

Others within academia and think tanks warn against the perils for governments and societies, seeing governments as ill-focused and the work led by industry experts as only skimming the surface of technology's murky waters. They underscore the “breakneck speed”¹⁵ with which technological advances are leaving the public and governments ill-equipped to deal with the disruption, and ultimately leaving them behind.

As Wylie (2018) notes, there is “a cultural realization that technology may be going too far, too fast and that we are unclear on how to address it.”¹⁶ While new technologies rally investors, developers and engineers, the public, unaware or lacking the expertise to understand them and their potential impact for societies, is being left behind. As Klugman (2018) describes, “this torrent of change often feels as if we're perched in the middle of a rapidly moving river on slippery stones that are being shunted around by the current.”¹⁷ Much of this worry is galvanized around the growth and global nature of artificial intelligence (AI).

13 Government of Canada, 2017, Speaking Notes for The Honourable Scott Brison, President of the Treasury Board of Canada, at FWD50, (Nov 2, 2017), https://www.canada.ca/en/treasury-board-secretariat/news/2017/11/speaking_notes_forthehonourablescottbrisonpresidentofthetreasury.html, accessed Dec 5, 2018.

14 M. Chui, M. Harrysson, J. Manyika, R. Roberts, R. Chung, P. Nel, & A.V. Heteren, 2018, *Applying artificial intelligence for social good*, McKinsey & Co. <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good>, accessed Dec 5, 2018.

15 Diane Francis, 2018, “The Biggest New Laws to Regulate Tech Giants—and Why They Matter,” (June 6), Singularity Hub, <https://singularityhub.com/2018/06/06/the-biggest-new-laws-to-regulate-tech-giants-and-why-they-matter/#sm.0000irpkce1dbrdswqo9oph18y2a0>, accessed Jan. 30, 2019.

16 Bianca Wylie, 2018, “Governance Vacuums and How Code is Becoming Law,” in *Data Governance in the Digital Age*, edited by Centre for International Governance Innovation (May 2), p. 90, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf> accessed Nov. 8, 2018.

17 Ian Klugman, 2018, “Shift Happens: Governments and the Fourth Industrial Revolution,” in *Government Digital: The Quest to Regain Public Trust*, edited by Alex Benay, Toronto: Dundurn, p.38.

The Growth of Artificial Intelligence

A number of factors are converging to drive the digital revolution, among them the exponential growth of computing power, advancements in AI (including machine learning and probabilistic reasoning), the Internet of Things and blockchain.¹⁸ However, no technology more than AI has captured the attention of governments, industry, academe and civil society, its impact being widely regarded as bringing about a wholesale shift in society as we know it today.¹⁹

AI is global and growing. A Price Waterhouse Coopers study (2017) estimates the economic potential of AI as contributing \$US 15.7 trillion to the global economy by 2030.²⁰ The findings of the 2017 AI Now Index, the first report of its kind to track progress in the field of AI in a comprehensive way, confirm that *investment and work in AI are accelerating at an unprecedented rate*. Its follow-up report in 2018 illustrates both the scope of growth and extent of its global reach.²¹

On the ground, public interest in AI is on the rise, both by governments and the media, with coverage in the popular media more positive since 2016.²² In the area of R&D, interest and activity have been growing worldwide. Research has seen dramatic growth; publications on Scopus, the largest database of peer-reviewed papers, have increased eight-fold since 1996.²³

18 Mina Down, 2019, "How Blockchain is Driving the Fourth Industrial Revolution," *The Startup*, (Jan. 17), <https://medium.com/swlh/blockchain-ai-internet-of-things-industrial-revolution-33e99a5ab3db>, accessed Feb. 2, 2019.

19 See, for example, D2L, 2018, "The Future of Work and Learning in the Age of the 4th Industrial Revolution," <https://www.d2l.com/wp-content/uploads/2018/01/The-Future-of-Work-and-Learning-D2L.pdf>, accessed November 27, 2018.

20 Price Waterhouse Coopers, 2017, "Sizing the prize: What's the real value of AI for your business and how can you capitalize?" p.5, <https://www.pwc.com.au/government/pwc-ai-analysis-sizing-the-prize-report.pdf>, p.6, accessed Feb. 21, 2019.

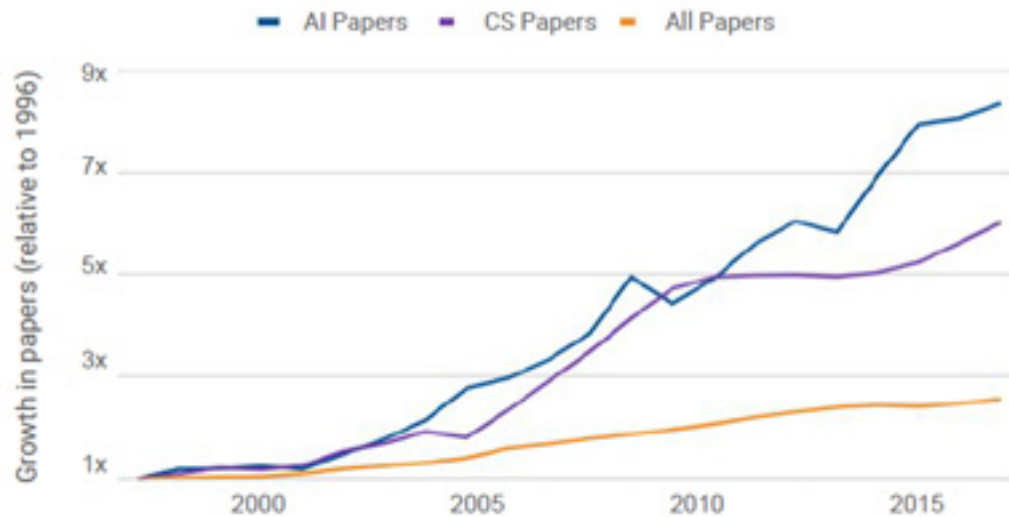
21 The AI Index was established by the AI Now Institute to track progress in the field of AI. Membership includes experts from Harvard, MIT, Stanford, OpenAI, the Partnership on AI and others.

22 Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Nieves, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, December 2018, "The AI Index 2018 Annual Report," p.9, AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>, accessed Dec. 12, 2018.

23 Shoham et al, pp.43-4.

Growth of annually published papers by topic (1996–2017)

Source: Scopus



Much of this activity is taking place in Europe and Asia, with China, Japan, and South Korea leading Asian countries in AI research papers, university enrolment and patent applications.

While Europe is emerging as a centre of research in AI - it was the largest publisher of AI papers last year (28%) - China is a rising competitor, following closely on its heels (25%).²⁴ In 2017, the Chinese government released *A Next Generation Artificial Intelligence Development Plan*, an ambitious plan to become the world leader in AI by 2030, setting aside US\$2 billion alone for an AI research park, a figure topping the EU Commission's entire investment in AI to 2020 (\$US 1.75 billion).²⁵ It is the world's largest producer of government research in AI, its activity having increased 400% since 2007.²⁶ And as university course enrolment in AI and Machine Learning (ML) increased worldwide, none has more so than at Tsinghua in China, whose combined AI and ML 2017 course enrolment was 16x greater than in 2010.²⁷

²⁴ Shoham et al, p.10.

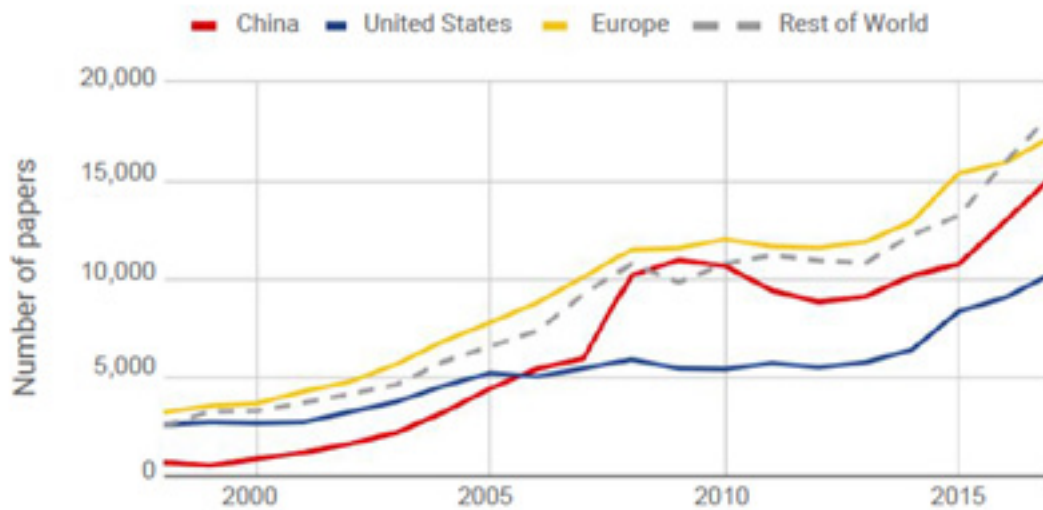
²⁵ Tim Dutton, 2018, "Building an AI World: Report on National and Regional AI Strategies," CIFAR: Ottawa, <https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies>, accessed Feb 10, 2019.

²⁶ Shoham et al, pp.13-14.

²⁷ Shoham et al, p.23.

Annually published AI papers on Scopus by region (1998–2017)

Source: Elsevier



The story of the U.S. is an equally important one to tell. The U.S. produced 17% of all AI research papers last year, with research dominated by industry. By way of comparison, the proportion of corporate AI papers produced in the U.S. was almost seven times greater than that in China.²⁸ On the ground, venture-back AI startups in the U.S. showed exponential growth, with the number more than doubling since 2015.²⁹ While it is unclear how much the U.S. government invests in AI R&D, government documents report \$1.1 billion in spending on unclassified AI-related R&D in 2015, with that figure growing by over 40% in 2018.³⁰

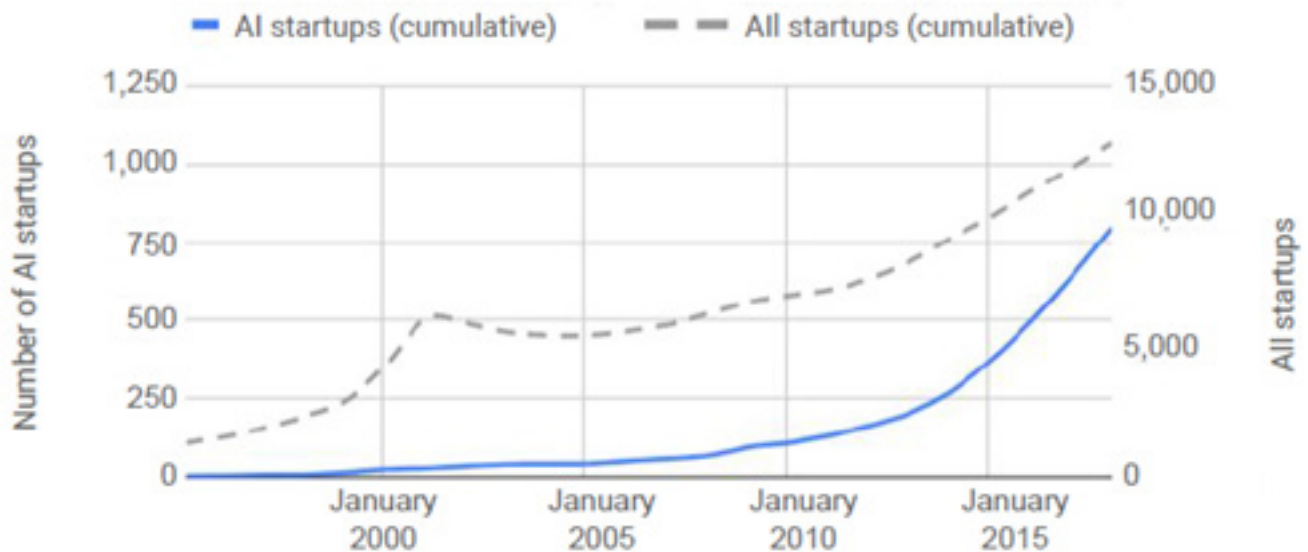
28 Shoham et al, p.15.

29 Shoham et al, p.31.

30 Tim Dutton, 2018, "An Overview of National AI Strategies," <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>, accessed Jan 27, 2019.

AI startups (U.S., January 1995 – January 2018)

Source: Sand Hill Econometrics



Note: The majority of the AI startups above develop AI systems. A minority use AI as an integral part of business, but do not develop the systems themselves. See appendix for more details.

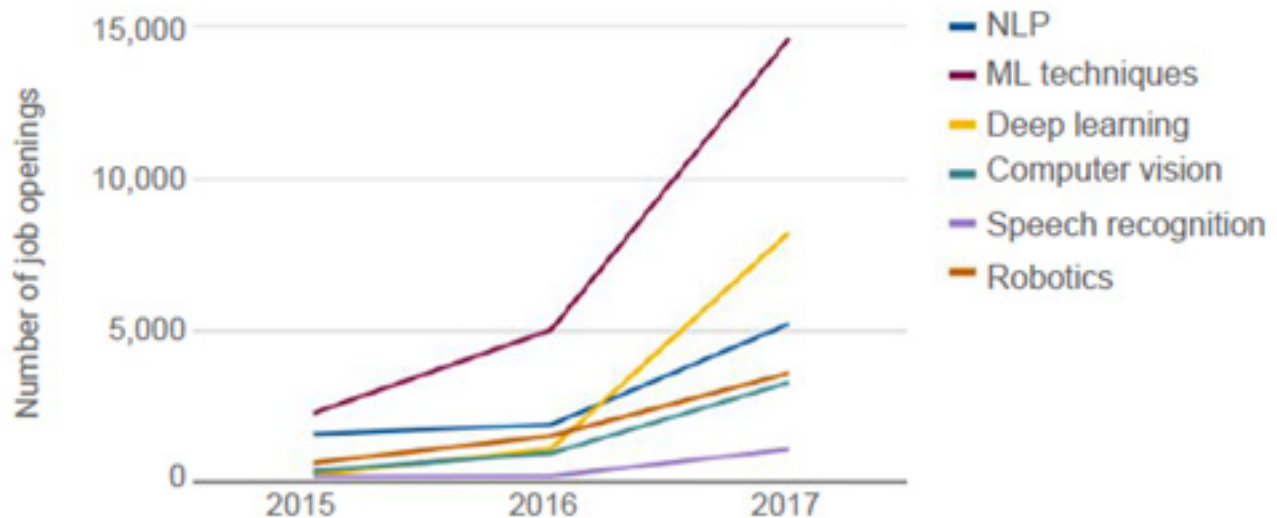
Within the AI field, interest in Machine Learning and Deep Learning is rising rapidly. The majority of AI research focused on Machine Learning and Probabilistic Reasoning (56%).³¹ Interest in Machine Learning was likewise matched by growth in university study, with the number of students enrolled in introductory Machine Learning courses growing at a faster rate (5x larger than in 2012) than those enrolled in AI courses (3.4x larger in 2017 than in 2012). Meanwhile, job openings in Machine Learning tripled over the last three years, from fewer than 5,000 to roughly 15,000, and the number of jobs requiring Deep Learning increased 34-fold.³²

³¹ Shoham et al, p.11.

³² Shoham et al, p.33.

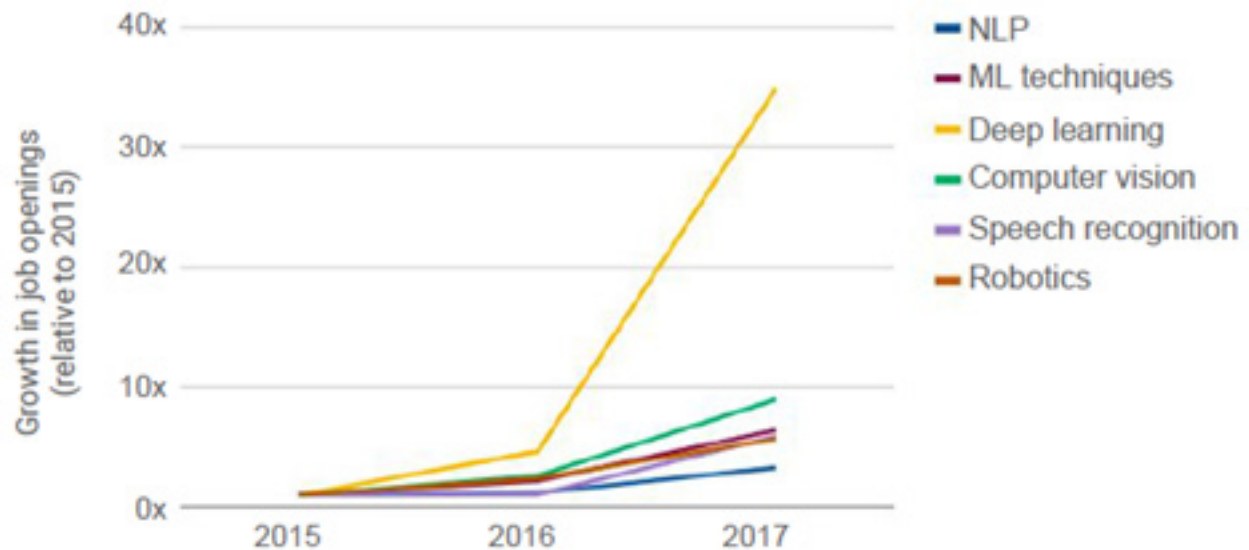
Job openings by AI skills required (2015 – 2017)

Source: Monster.com



Growth of job openings by AI skills required (2015 – 2017)

Source: Monster.com



Note: While AI job openings are increasing across the board, they still represent a very low number of job openings for computer engineers.

AI technology is also showing swift improvements. In the field of computer vision, the accuracy of image recognition increased from around 72% in 2010 to roughly 97% in 2017, surpassing human performance (95%).³³ Likewise, the

time needed to train a network to classify pictures from the ImageNet corpus (an image training database) fell from roughly one hour to four minutes in the space of 18 months.³⁴

As research proceeds, AI technologies are being adopted across all sectors and business functions around the world.³⁵ A McKinsey (2018) survey found that while the pace and extent of adoption varies across sectors - the telecom, high-tech, and financial-services firms are leading the way in overall adoption - many organizations (50%) have begun to embed AI in their business processes or pilot AI (30%), with the most common AI applications being robotics process automation (RPA), computer vision, and machine learning. Generally speaking, industries are adopting AI where it is of most value to them; for instance, the retail industry is using AI primarily in marketing and sales, the automotive sector in manufacturing, and financial services in risk and fraud detection.³⁶

Within the public sector, cities are adopting Smart City applications that use AI for service delivery, resource management, and energy utilization. Although there have also been controversial applications in the areas of surveillance and law enforcement;

such as the use of facial recognition technology in China to monitor citizen behaviour, like jay walking, or in public schools to monitor student's facial expressions to gauge their level of classroom engagement, or to support the development of a social credit system. Likewise, the U.S. military's Project Maven has drawn strong criticism for its use of AI - with the help of Google computer vision experts - to sift through mountains of data and video to look for patterns of abnormal or suspicious activity.³⁷

Comment:

The story of AI is its scope and scale. AI is global and growing at rapid rate. It is being researched, adopted, and advanced at a frenetic pace around the world and across industries. As this activity charts unfamiliar waters, it also raises concerns about the ethical and societal impact of these new discoveries. The challenge for governments is to keep pace with these technological advances, and in the process ensure that societies benefit from the best that AI has to offer while preventing the worst. The next section describes some of the ethical and societal challenges that advances in AI are bringing along with it.

33 Shoham et al, p.47.

34 Shoham et al, p.48.

35 Shoham et al, p.36.

36 McKinsey Analytics, 2018, "Notes from the AI frontier: AI Adoption advances, but foundational barriers remain," (Nov), <https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain>, accessed Feb. 21, 2019.

37 See for example, Louise Lucas and Emily Feng, 2018, "Inside China's surveillance state," Financial Times, (July 20), <https://www.ft.com/content/2182eebe-8a17-11e8-bf9e-8771d5404543>, accessed Feb. 21, 2019; and Darrell West and John Allen, 2018, "How artificial intelligence is transforming the world," (April 24), <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/>, accessed Feb. 21, 2019.

III. Social and Ethical Concerns

In 2009, on the occasion of the celebration of the 150th anniversary of Darwin's *On the Origin of Species*, noted sociobiologist E.O Wilson (2009) was asked whether humanity would solve the crises of the next hundred years. He replied,

*“[y]es, if we are honest and smart...The real problem of humanity is the following: we have paleolithic emotions; medieval institutions; and god-like technology. And it is terrifically dangerous, and it is now approaching a point of crisis overall.”*³⁸

The nature of technological change is such that innovation often precedes regulation. Indeed, as Francis (2018) notes about the invention of the automobile, “[t]he world’s first stop sign appeared in Michigan around 1915, decades after the first privately owned passenger car.”³⁹ What is different about governing in the 21st century, however, is the rapidity with which innovation is outpacing current ethical guidelines and rules, leaving societies vulnerable to the exigencies of technological advancement. As Turner (2017) describes, “AI is already being given ownership of difficult

decisions that have until now rested on human intuition or principles—actions and doctrines that have been legally codified...If a human were to make these decisions, they would be held to a legal or moral standard. No such rules exist in the wild west of AI.”⁴⁰

Advancements in AI technologies are introducing fundamental questions about the ethical and social impact on our societies, and “about who should be responsible for ensuring that AI will be a force for good.”⁴¹

A body of work is emerging that shares an interest in distilling the broader impact of disruptive technologies like AI on society, and harnessing it for the benefit of humanity. Think tanks and not-for profit organizations, papers, opinion pieces and interviews by academics and ethicists raise concerns about the ethical values embedded within AI, and the “digital sovereignty” of corporations and their coders making decisions on behalf of humanity.⁴² This

38- 2009, “An Intellectual Entente,” *Harvard Magazine*, (Oct. 9), <https://harvardmagazine.com/breaking-news/james-watson-edward-o-wilson-intellectual-entente>, accessed Dec. 12, 2018.

39 Francis, 2018.

40 Jacob Turner, 2017, “Letting Facebook control AI regulation is like letting the NRA control gun laws,” (Dec 6), <https://qz.com/1145668/letting-facebook-control-ai-regulation-is-like-letting-the-nra-control-gun-laws/>, accessed Jan. 27, 2018.

41 Corinne Cath, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo, and Luciano Floridi, 2016, “Artificial Intelligence and the ‘Good Society’: The US, EU, and UK Approach” (Dec. 23), p.2. <https://ssrn.com/abstract=2906249>, accessed Jan. 25, 2019.

42 Darrell West, 2018, “The role of corporations in addressing AI’s ethical dilemmas,” (Sept. 13), <https://www.brookings.edu/research/how-to-address-ai-ethical-dilemmas/>, accessed Feb. 8, 2019.

section of the paper highlights some of the ethical and social challenges this work brings to light around AI and the data and algorithms that it relies on.

A Lack of Public Awareness and Concern

The well-publicized cases of Cambridge Analytica interference in the UK and US elections, Facebook hosting fake Russian accounts, and the co-opting of the Yellow Vest movement in France by outsiders to push their political agenda on Twitter - provides context to some of the concerns raised by authors about the ease with which technology can be deployed to spread disinformation, ignite political unrest, and influence the shape and content of digital citizenship.⁴³ Some former tech executives and tech ethicists advocate for the coming together of citizens, families and civil society more broadly toward the ethical and humane design and use of technology. For instance, the Center for Humane Technology, founded by former Google ethicist Tristan Harris, seeks to raise public awareness about the intentional

design of tech products to steer our behaviour with the use of AI-driven content that continuously learns from us and keeps us engaged, and the impact of this on our mental, social and democratic wellbeing. It calls on Apple, Samsung and Microsoft to redesign their devices, tech employees to advocate for non-extraction based design decisions and business models, and governments to update consumer protections.⁴⁴

Some advocates worry that corporate R&D is steering the rapid development of digital technologies like AI, and what they see as citizens' unquestioning acceptance of it. They call for thoughtful citizen deliberation and engagement around the purposes to which emerging technologies could be put,⁴⁵ and renegotiating the social contract.⁴⁶ Not doing so is the same as giving permission to "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'."⁴⁷

Concerned about the limited research effort toward understanding the implications of AI for society, several

43 See for example, Paris Martineau, 2018, "The Co-opting of French Unrest to Spread Disinformation," *Wired*, (Dec. 11), <https://www.wired.com/story/co-opting-french-unrest-spread-disinformation/>, accessed Dec. 12, 2018; Kevin Roose, 2018, "Is Tech too Easy to Use?" *The New York Times*, (Dec.12), <https://www.nytimes.com/2018/12/12/technology/tech-friction-frictionless.html>, accessed Dec. 12, 2018.

44 See the Center for Humane Technology and All Tech is Human Initiative.

45 See for example, Adam Greenfield, 2017, *Radical Technologies: The Design of Everyday Life*. Verso: N.Y. Greenfield is a former head of user interface design at Nokia, and presently senior fellow in the London School of Economics' urban studies centre. See also the All Tech is Human Initiative, launched by tech ethicist David Polgar, which seeks "to better align technology with societal interests" by bringing together a diverse range of people, including citizens and representatives from government and industry, to influence the ethical development of technology.

46 Wylie, p.90.

47 Cath et al, p.3.

academic research institutes have emerged dedicated to examining the challenges associated with AI.⁴⁸ Chief among their concerns are:

- A lack of transparency;
- Poor accountability for decision-making;
- Bias that may result from the use of automated tools, algorithms and computer learning; and
- Unethical applications of AI.

These concerns are all the more urgent given the potential for AI to be used in the public domain, in areas such as criminal justice, law enforcement, healthcare, education, and even warfare.

Data and Algorithmic Bias

Several authors have raised concerns about the risk that machine learning algorithms introduce and amplify social biases, and thus that they

themselves become a source of discrimination. Many have noted that data is not necessarily objective, but rather reveal human biases. As Obar and McPhail (2018) note, a common misconception is thinking that since the technology is new, the data must also be new. The reality is that many historical data sets are integrated into new Big Data systems, and along with them, the biases of the past. At the same time, new data may perpetuate and amplify existing biases.⁴⁹ The result is that automated data-driven decision-making can (re-)produce inaccurate, unfair, or discriminatory decisions. As Hume et al (2018) explain, “[e]ven subtle and unconscious bias can produce data that steers systems in directions their designers would never choose... Algorithms powered by that data are not objective oracles, but mathematical tools that may pick up, refract and amplify the biases that exist in society.”⁵⁰ These concerns will only grow “as algorithms grow more complex, autonomous and powerful.”⁵¹

Related to these concerns, is the extent

48 See for example, the [AI Now Institute](#), established in 2017, as an interdisciplinary research institute at New York University that produces research on the social implications of AI and serves as a hub for this emerging field; the [Centre for Advancing Responsible and Ethical Artificial Intelligence \(CARE-AI\)](#) launched in December 2018 at the University of Guelph as a multidisciplinary research and teaching institute to integrate human and ethical considerations into the development and implementation of AI, influence public policy and regulations around AI, and be active on the ground, applying machine learning and AI to the university’s current research areas; the [Stanford Human-Centered AI Initiative \(HAI\)](#), established in October 2018 to foster a dialogue among academia, industry, government, and civil society toward the responsible development of AI.

49 Jonathan Obar and Brenda McPhail, 2018, “Preventing Big Data Discrimination in Canada,” in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, p.58, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

50 Kathryn Hume, Daniel Moore and Michael Zerbs, 2018, “The AI revolution needs a rulebook. Here’s a beginning” in the *Globe and Mail*, (Dec. 6), accessed Dec.7, 2018.

51 Nicolas Mialhe and Cyrus Hodes, 2017, “Making the AI Revolution Work for Everyone: A Report to the OECD” (March), The Future Society, AI Initiative, p.21, <https://www.tuftsgloballeadership.org/sites/default/files/images/resources/Mialhe%20Reading.pdf>, accessed Dec.7, 2018.

to which, programmers either knowingly or otherwise build in misinformation and bias.⁵² Some suggest that the AI industry itself is unrepresentative of the population, which may act as a source of unconscious bias.⁵³ Some authors call for ethical training for students in AI, computer science, and data science programs, as well as an interdisciplinary approach to the development and application of AI, including people with diverse experiences as well as those with specialized knowledge, including mathematics, data science, risk, social sciences, ethics and law.⁵⁴

The “Black Box Problem”

The “Black Box Problem,” also known as the “Responsibility Gap,” refers to the inability to explain how an algorithm arrived at a particular response using a given data set. Since “many deep learning systems function as ‘black boxes’,...their behaviour can be difficult to interpret and explain, thus raising concerns over explainability, transparency and

human control.”⁵⁵

While “the black box problem is not new to computer science,” as Mialhe and Hodes (2017) explain, “the rise of advanced AI in the age of big data has caused a cardinal shift in its manifestation. According to the most renowned experts, tracing and understanding in detail the complex decision-making mechanisms of AI algorithms will be difficult.”⁵⁶

Geoff Hinton, distinguished computer scientist, head of Google Brain, and renowned for his work on artificial neural networks shares this view: just as “people can’t explain how they work, for most of the things they do,” “[n]eural networks have a similar problem.”⁵⁷ As Hinton (2018) explains, “[w]hen you train a neural net, it will learn a billion numbers that represent the knowledge it has extracted from the training data. If you put in an image, out comes the right decision, say, whether this was a pedestrian or not. But if you ask ‘Why did it think that?’ well if there were any simple rules for deciding whether an image contains a pedestrian or not, it would have been a solved problem

52 Obar and McPhail, p.59.

53 See, for example, Kate Crawford, 2016, “Artificial intelligence’s white guy problem,” (June 25), http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=1, accessed Jan. 25, 2019.

54 See for example the Human-centred Artificial Intelligence Institute at Stanford University, <https://hai.stanford.edu/>.

55 Dr. Jason Millar, Brent Barron, Dr. Koichi Hori, Rebecca Finlay, Kentaro Kotsuki, and Dr. Ian Kerr, 2018, “Discussion Paper for G7 Multistakeholder Conference on Artificial Intelligence, “Theme 3: Accountability in AI Promoting Greater Societal Trust,” Dec. 6, Montreal, Canada, p. 3, [https://www.ic.gc.ca/eic/site/133.nsf/vwapj/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf/\\$FILE/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf](https://www.ic.gc.ca/eic/site/133.nsf/vwapj/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf/$FILE/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf), accessed Feb. 6, 2018.

56 Mialhe and Hodes, p.19.

57 Quoted in Hesse Jones, “Geoff Hinton Dismissed the Need for Explainable AI: 8 Experts Explain why he’s Wrong,” *Forbes*, (Dec. 20), <https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/#ccc0b8d756d3>, accessed Dec. 20, 2018.

ages ago.”⁵⁸

Adding to this complexity is what some refer to as the *industrial culture* that creates these black boxes. As Whittaker et al (2018) explain, “[m] any of the fundamental building blocks required to understand AI systems and to ensure certain forms of accountability - from training data, data models, to the code dictating algorithmic functions, to implementation guidelines and software, to the business decisions that direct design and development” are protected by corporate secrecy laws.⁵⁹

Some suggest the need for a broader understanding of explainability. For instance, Alejandro Saucedo, Chief Scientist at The Institute for Ethical AI & Machine Learning introduces the notion of a reasonable level of explainability, which considers the processes, infrastructure and humans operating the algorithms, and depends on cross-functional collaboration across technology, industry and public policy domains. As Saucedo explains, “[i]t is possible to reach a reasonable level of explainability and accountability by ensuring the right

touchpoints with domain experts are in place throughout the development and operation of AI systems. Sometimes this may involve a trade-off between explainability and accuracy, but it may be required depending on the critical nature of the project.”⁶⁰

The Value and Ownership of Data

Data holds great economic potential, and AI is the key to unlocking its potential. Often referred to as “Big Data” and described as the 21st century’s *oil or gold*, data is the new commodity that can be tapped for profit.⁶¹

The marketization of raw data - from heartbeats to “likes” - that is captured, held and sold is often referred to as the *data-driven economy*. As Ciuriak (2018) explains, “[d]igital transformation is creating a new kind of economy based on the “datafication” of virtually any aspect of human social, political and economic activity as a result of the information generated by the myriad daily routines of digitally connected individuals and machines.”⁶²

⁵⁸ Quoted in Jones, 2019.

⁵⁹ Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schulz, Oscar Swartz, 2018, “AI Now Report 2018,” AI Now Institute, https://ainowinstitute.org/AI_Now_2018_Report.pdf, accessed Dec.20, 2018.

⁶⁰ Quoted in Jones, 2019. For example, the Institute for Ethical AI developed an AI Procurement Framework for professionals to evaluate the maturity of their machine learning systems through a checklist, which highlights red flags around processes and infrastructure.

⁶¹ Dirk Helbing, 2015, “Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies,” <https://arxiv.org/abs/1504.03751>, accessed Jan.24, 2019.

⁶² Dan Ciuriak, 2018, “The Economics of Data: Implications for the Data-driven Economy,” in *Data Governance in the Digital Age*, edited by the Centre for International Governance Innovation, (May 2), p. 12, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

Supporting the data-driven economy is the development of a data analytics industry. Public and private sector leaders are urged to build data analytics into their organizations, and hire Chief Data Officers. They are warned that organizations that rely solely on experience, intuition and judgement in decision-making are at risk of being left behind.⁶³ Private, public and academic institutions have begun to establish data analytics academies to turn out the “unicorns” - data analysts with a combination of business and analytics skills - to harvest, analyze, interpret, and make predictions about the data.

The economic value of data is “galvanizing entrepreneurs and investors.”⁶⁴ Some authors warn that the prospect of “extracting lucrative insights”⁶⁵ coupled with the lack of ethical and regulatory frameworks, is at risk of contributing to inequalities, marginalization, and abuses of power. Much like the Wild West before there was a sheriff in town, this is an unregulated space with implications for the distribution of wealth and power, and the governance of society more generally. There is little doubt that data collection, intellectual property, and AI are contributing to imbalances in wealth and power in

favour of the most lucrative companies - Apple, Google, Facebook, Amazon, and Microsoft. Some authors raise probing questions about who owns, and who should own the data, and urge the extension of property rights to personal data.⁶⁶

The question of ownership of data circles around the role of government in ensuring that society benefits from all that the digital era has to offer, while at the same time ensuring that it is contributing to the collective interest. This raises important questions about the stewardship role of the state, which includes, but is not limited to, regulatory power; how to anticipate risks and proactively prevent harm; how to regulate without losing the benefit of enhanced knowledge; and how to define the boundaries of data ownership and use between the public, private, civic and individual spheres of life.

The Ethical Application of AI

The lack of accountability for how AI technology is procured, deployed and used raises a variety of concerns. Some authors suggest the need for *Algorithmic Impact Assessments*, similar to Privacy Impact Assessments, as a means of

63 See for example Forbes Insights, 2015, *Analytics: Don't Forget the Human Element*, (Nov), https://www.forbes.com/forbesinsights/ey_data_analytics_2015/index.html, accessed June 26, 2018.

64 Rohinton Medhora, 2018, “Data Governance in the Digital Age,” in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, p.2, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

65 Medhora, 2018, p.2.

66 See, for example, Dan Breznitz, 2018, “Data and the Future of Growth,” in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, pp.66-72, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

introducing accountability into decision-making around the broader application and adoption of AI.⁶⁷

Of particular concern for some is the growing use of AI for surveillance via sensor networks, social media tracking, facial recognition, and affect recognition. The technology, often used without people's consent or knowledge, has the potential for unethical and discriminatory purposes, thus raising concerns around the violation of human rights. Some oppose the use of facial recognition by governments, while others call for strict regulations around its use.

Authors further raise ethical concerns around the acceptability of lethal autonomous weapon systems. Concerns centre on the intent to kill or injure, the absence of human agency, and the loss of human dignity, with many authors calling for an outright ban.⁶⁸

Comments:

This work draws attention to some of the challenges for governments in the development and adoption of AI, and the algorithms and data that fuel it; among them accountability and transparency for AI decision-making,

the “blackbox” and the associated risks of machine learning bias, and the ethical use of AI technologies. Some voices argue that in a field that is evolving, the certitude of these concerns is unproven, that regulating risks stifling innovation, or missing the true challenges of AI. Others call for governments to take action now: to regulate while the industry is still in its infancy, noting that the risks of waiting for the field to advance are too grave, and the challenges of regulating down the road too great.

While governments must decide on a course of action that is best suited to their particular context and circumstances, the challenges for governments in doing so are compounded by the transnational sweep and nature of AI technological development. While cross-jurisdictional issues are not unique, what is new is the speed with which AI technologies are “developed and deployed in multiple jurisdictions,” “cross international and cultural boundaries,” and may or may not be built “to respect local laws and cultural norms.”⁶⁹ Governments are responding to these challenges in a variety of ways. The next section highlights some of the approaches they are taking to do so.

67 Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, 2018, “Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,” AI Now Institute, <https://ainowinstitute.org/aiareport2018.pdf>, accessed Feb, 6, 2018.

68 See Samuel Gibbs, 2017, “Elon Musk leads 116 experts calling for outright ban of killer robots,” *The Guardian*, Aug. 20, <https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war>, accessed Feb. 28, 2019; Ian Sample, 2018, “Thousands of leading AI Researchers sign pledge against killer robots,” (July 18), *The Guardian* <https://www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots>, accessed Feb. 27, 2019.

69 Millar et al, p. 3.

IV. Learning from Practice: Government-led Initiatives

In exploring the challenges of governing in the digital era, it is helpful to return to the four research questions posed by the *New Synthesis 2019* research program:

- *What can government do to build the collective capacity of society to invent and share a better future together in the digital era?*
- *What can government do to ensure that the adaptive capacity of society will keep pace with the increasing velocity of change?*
- *What can government do to enhance the resilience of society to adapt, evolve and prosper in the unforeseen and unpredictable circumstances that a digital world entails?*
- *What must be done to ensure that public institutions have the capabilities to successfully steer society through a technological revolution and an unprecedented period of change?*

The third phase of the NS Initiative is intended to generate useful and usable insights to help practitioners think through challenges, and set a course adapted to their context and circumstances.

Considering the challenges mentioned in the previous section, governments

do not have the luxury of waiting to act. They must set a course with imperfect knowledge in a context with a high degree of uncertainty about the shape that the digital world may ultimately take. Governments are searching for a mix of actions that taken together may help their society to adapt to a fast-changing landscape, reap the benefits that a digital world has to offer and prevent, to the extent possible, some of the most detrimental impacts for society.

A summary review of government initiatives in various countries indicates that the actions to date tend to follow four broad trends:

- Industry based self-regulation;
- Governmental declarations of principles and norms;
- Regulatory actions; and
- Comprehensive approaches to building a digital society.

Countries give different weight to the importance of these measures and use a combination of measures that set a distinctive trajectory.

This section of the paper presents actions taken by some governments to harness the potential of digital

technologies and mitigate adverse or unintended consequences. Some rely primarily on self-regulation by the main industry leaders, some adopt principles to guide developers and users, others use regulation to protect privacy and human rights, and most take proactive measures to encourage the development of the digital economy. The examples noted in this section draw from the approaches, policies and practices initiated by governments in the United States, South Korea, Estonia, Singapore, Germany, Dubai, the European Union and the city of Barcelona.

Reliance on Industry Self-Regulation

The American approach is an example of self-regulation. Not surprisingly, it enjoys strong support from the biggest and most important industry leaders located inside the U.S.

In the final months of the Obama presidency, the White House laid the groundwork for a national AI strategy aimed at increasing investment and responding to some of the challenges the digital economy may generate for society. The report [Preparing for the Future of Artificial Intelligence](#),

drafted with input from the public, put forward a view that AI should serve the public good, helping to solve “some of the world’s greatest challenges and inefficiencies,”⁷¹ and be “a major driver of economic growth and social progress.” It proposed a limited role for government in regulating AI, suggesting that government provide the infrastructure and support for R&D, and develop policy to ensure “the economic benefits are shared broadly.”⁷² It called on government agencies to fit AI into existing regulatory schemes. It encouraged the research community and practitioners to address the questions of openness, transparency, and understandability. As President Obama described following the report’s release, the role of Government was to allow “*a thousand flowers to bloom*,” applying “*a relatively light touch, investing heavily in research and making sure there’s a conversation between basic research and applied research*.”⁷³

The approach taken by the Trump administration thus far further limits the scope of government in regulating AI, endorsing the “*free market approach to scientific discovery*,” and removing “*barriers*” to innovation. As Deputy Assistant to the President

70 Executive Office of the President National Science and Technology Council Committee on Technology, 2016, *Preparing for the future of artificial intelligence*, Washington, DC, USA, p.1, https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf, accessed January 25, 2019.

71 Executive Office of the President National Science and Technology Council Committee on Technology, 2016, p.3.

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

73 Klint Finley, 2016, “Obama wants to help the government to develop AI,” (Oct.10), <https://www.wired.com/2016/10/obama-envisions-ai-new-apollo-program/>, accessed Feb 16, 2019.

for Technology Policy, Michael Kratsios, notes in the summary report of the 2018 White House Summit on Artificial Intelligence for American Industry, “Our Administration is not in the business of conquering imaginary beasts. We will not try to ‘solve’ problems that don’t exist. To the greatest degree possible, we will allow scientists and technologists to freely develop their next great inventions right here in the United States.”⁷⁴ To the extent that the administration will regulate AI, a 2019 Executive Order calls on leaders at the Office of Management and Budget, the Office of Science and Technology Policy, the Domestic Policy Council, and the National Economic Council to advise on “regulatory and non regulatory approaches...that advance American innovation while upholding civil liberties, privacy, and American values.”⁷⁵

This approach enjoys the support of big tech companies, like Google, Amazon, Facebook, Apple, Microsoft, and IBM (GAFAMI). These corporations are keen to keep the field as free from government regulation as possible for as long as possible, and to maintain leadership of the discussion around measures for ensuring ethical design and use of

digital technologies.

Some of these measures include hiring company ethicists, adopting codes of ethics, establishing ethics boards, and industry self-regulatory partnerships:

- *Ethicists* employed or retained by the company to help corporations question the benefit and impact of new technologies for societies
- *Formal and publicly-available codes of ethics* that lay out a corporation’s “principles, processes, and ways of handling ethical aspects of AI development.”⁷⁶ Some private sector corporations, including Google, SAP and Microsoft, have already introduced principles-based frameworks to guide AI development and adoption, which generally cover topics such as social benefit, bias, safety, accountability, and privacy. A detailed description of Google’s principles is available in Annex 1.
- *Ethics boards* may be used to perform a variety of functions, including facilitating discussion about the nature, purpose, and

74 Executive Office of the President of the United States, The White House Office of Science and Technology Policy, 2018, *Summary of the 2018 White House Summit on Artificial Intelligence for American Industry*, Washington, D.C., p.10, <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>, accessed Feb. 16, 2019.

75 The White House, 2019, “Executive Order on Maintaining American Leadership in Artificial Intelligence,” (Feb 11), <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>, accessed Feb. 17, 2019.

76 Darrell West, 2018, “The role of corporations in addressing AI’s ethical dilemmas,” <https://www.brookings.edu/research/how-to-address-ai-ethical-dilemmas/>, accessed Feb. 8, 2019.

consequences of artificial intelligence among those building AI, members of academe, civil society and the public more broadly; sharing best practices; and educating lawmakers about the progress of AI.⁷⁷ Private sector AI ethics boards are appearing on the scene in the U.S., among them, the Microsoft AI Ethics in Engineering and Research committee (2016), Google DeepMind Health (2016), and Google DeepMind Ethics and Society (2017). Most recently, Axon, manufacturer of the Taser, instituted an AI Ethics Board for Public Safety of external experts in 2018.

- *Self-regulatory partnerships*, comprised of members of corporations and academe, may operate in tandem with corporate ethics boards. The most notable, the Partnership on AI to Benefit People and Society, is a consortium of the big six technology companies - Apple, Amazon, Google and DeepMind, Microsoft, IBM, and Facebook - which launched in 2016.

Comments;

While these self-regulatory practices signal that the private sector is keen

to lead the discussion about ethical and social challenges, this approach has serious limitations. History is replete with examples of how self-regulation tends “to favour the goals of industry over those of other stakeholders.”⁷⁸

Coupled with this are issues of transparency and representativeness. For instance, Axon’s new board was criticized for not including “representatives from the heavily policed communities most likely to suffer the downsides of new police technology.”⁷⁹ As Wendell Wallach, a scholar at Yale University’s Interdisciplinary Centre for Bioethics argues “purely internal processes... are hard to trust, particularly when they are opaque to outsiders.”⁸⁰

Finding agreement across a diverse group of stakeholders is a challenge, and as a result the consensus is likely to represent the lowest common denominator, and where these boards can find agreement, they lack the means of enforcing agreed upon principles.⁸¹ As Turner (2017) notes, “rules formulated by corporate ethics boards will always lack the legitimacy that a government can provide.”⁸² The tobacco industry or indeed the 2008 global financial crisis are important reminders of the limits of

77 Klint Finley, 2016, “Tech Giants Team Up to Keep AI From Getting Out of Hand,” (Sept. 28), <https://www.wired.com/2016/09/google-facebook-microsoft-tackle-ethics-ai/>, accessed February 7, 2019.

78 Cath et al, p.9.

79 Tom Simonite, 2018, “Tech Firms Move to Put Ethical Guard Rails Around AI,” (May 16) <https://www.wired.com/story/tech-firms-move-to-put-ethical-guard-rails-around-ai/>, accessed Feb. 8, 2019.

80 Simonite, 2018.

81 Finley, 2016.

82 Turner, 2017.

self-regulation.⁸³ Relying on self-regulatory partnerships is unlikely to get us any closer to establishing a shared understanding of the scope and limits of AI; “too many private ethics boards could also lead to there being too many sets of rules. It would be chaotic and dangerous if every major company had its own code for AI, just as it would be if every private citizen could set his or her own legal statutes.”⁸⁴

Government Declarations of Principles

Several governments have drafted, or are in the process of drafting, non-binding ethical principles in consultation with stakeholders and the private sector. This approach is informed by the need for a *light touch* so as not to hinder the development of AI while still in its infancy. Dr. Aisha Bint Butti Bin Bishr, Director General of Smart Dubai, explains that, “AI regulation is needed, but that the field is not yet mature enough to devise fixed rules to govern it. However, organisations still require guidance, and regulators still need to begin to learn how to oversee this emerging technology, but without creating restrictions that could stifle

innovation.”⁸⁵

Government declarations generally cover five broad themes: *accountability, transparency, fairness, explainability, and human-centricity*. The government of Singapore and Smart Dubai provide examples of this approach.

Singapore

In 2018, the Singapore government announced the creation of an Advisory Council on the Ethical Use of AI and Data to assist in developing guidelines around the responsible development and adoption of AI. A discussion paper on responsible development and adoption of AI was developed in consultation with industry with the goal of encouraging the private sector “to develop voluntary governance frameworks, including voluntary codes of practice.”⁸⁶

The paper puts forward two key recommendations governing the development and adoption of AI: that decisions made by AI should be *explainable*, transparent and fair; and that the technology should be human-centric, *meaning that it is of benefit to humanity, and does no harm*:

- **Explainable:** Automated

⁸³ Turner, 2017.

⁸⁴ Turner, 2017.

⁸⁵ -, 2019, “Smart Dubai launches guidelines on ethical use of Artificial Intelligence,” (Jan 8), *Gulf News Tech*, <https://gulfnews.com/technology/smart-dubai-launches-guidelines-on-ethical-use-of-artificial-intelligence-1.61317752>, accessed Feb. 7, 2019.

⁸⁶ Government of Singapore, Personal Data Protection Commission, 2018, “Discussion Paper on Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI,” (June 5), p.2, <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf>, accessed Feb. 8, 2019.

algorithmic decisions and the data that drives these decisions *can be explained to end-users and other stakeholders in non-technical terms.*

- **Transparent:** In order to build trust in the entire AI ecosystem *AI developers, data scientists, application builders and user companies should be accountable for the AI algorithms, systems, applications and resultant decisions.*

- **Fair:** AI algorithms and models embedded in decision-making systems should incorporate *fairness at their core.* This could include the training dataset, AI engine and selection of model(s) for deployment in the intelligent system.

- **AI systems, robots and decisions made using AI should be human-centric:**

o Decisions should strive to *confer a benefit* on or provide individuals with assistance in the performance of a task;

o Decisions should *not cause foreseeable harm*⁸⁷ to an individual, or should at least minimise harm (in necessary circumstances, when weighed against the greater good);⁸⁸

o *Tangible benefits to individuals*

should be identified and communicated in order to build consumer understanding and confidence; and

AI systems and robots should be designed to *avoid causing bodily harm* or affecting the safety of individuals.⁸⁹

Smart Dubai: AI Ethical Principles and Guidelines

In January 2019, Smart Dubai released a set of AI ethical principles and guidelines in the form of an Ethical AI Toolkit. Modelled after Google's seven AI research principles and developed in consultation with Microsoft, IBM, and Google, the toolkit was designed to guide organizations delivering AI services.⁹⁰

The toolkit includes an AI Ethics Self-Assessment Tool for developers and operators to evaluate their AI systems in accordance with the principles and guidelines. The toolkit is built on the understanding that as the AI field evolves so too should the ethical principles and guidelines that prescribe its development and adoption. In keeping with this, the toolkit is iterative, encouraging ongoing feedback from citizens, service providers and developers to ensure the principles and guidelines keep pace with technological advancements.

⁸⁷ "Harm" includes physical, psychological, emotional and economic harm.

⁸⁸ Greater good -- Adapted from UNICEF's Humanitarian Principles

⁸⁹ Government of Singapore, p.5.

⁹⁰ "Smart Dubai launches guidelines on ethical use of Artificial Intelligence," 2019.

While Smart Dubai’s principles are similar to those of Singapore, they go a step beyond, spelling out in further detail what these principles mean in practice.

Smart Dubai Principles:

- **Ethics:** AI systems should be fair, transparent, accountable and understandable.
- **Security:** AI systems should be safe and secure, and should serve and protect humanity.
- **Humanity:** AI should be beneficial to humans and aligned with human values, in both the long and short term.
- **Inclusiveness:** AI should benefit all people in society, be governed globally, and respect dignity and people rights.

Smart Dubai Guidelines:

<p>Fairness</p> <ul style="list-style-type: none"> • Demographic fairness • Fairness in Design • Fairness in Data • Fairness in Algorithm • Fairness in Outcomes 	<p>Accountability</p> <ul style="list-style-type: none"> • Apportionment of accountabilities • Accountable measures for mitigating risks • Appeals procedures and contingency plans
<p>Explainability</p> <ul style="list-style-type: none"> • Process explainability • Outcomes explainability • Explainability in non-technical terms • Channels of explanation 	<p>Transparency</p> <ul style="list-style-type: none"> • Identifiable by humans • Traceability of cause of harm • Auditability by public

Comments:

A Declaration of Principles takes self-regulation a step further, providing a common reference for assessing private and public actions. It can be improved and made more robust as governments gain experience, or as new knowledge becomes available. Nonetheless, it is minimalist in its nature, offering less protection than what currently exists in some parts of the world, including the protections granted for human rights, especially privacy, freedom of expression, non-discrimination, and equality. As a result, it is not able to prevent the human rights risks that AI may pose in cases like facial recognition or autonomous weapons.

Regulation

Some governments consider it necessary to go beyond industry self-regulation and voluntary compliance through a declaration of principles. These governments are exploring how legislative frameworks for the development and adoption of AI could prevent harm that may be difficult to prevent through other means, or difficult to rectify later on. The most prominent example is provided by the EU with respect to privacy and personal data protection. The government of Estonia is going a step further, undertaking the first public discussion around data ownership and algorithmic liability.

The General Data Protection Regulation (GDPR) of the EU

The EU introduced the General Data Protection Regulation (GDPR) in 2018 to protect people's privacy rights. It is the most comprehensive data protection law to date, shifting control of data away from technology companies to individuals. It is also driving changes in other jurisdictions like Switzerland who are aligning their data protection laws to the GDPR to ensure the continued free flow of data across international borders.⁹¹

The GDPR defines personal data as “*all information that relates to any*

living individual who is identified or identifiable from that information, whether in isolation or in combination with any other available information.”

⁹²This means that anywhere information about people is handled, the GDPR will follow.⁹³

It *broadens individual* rights established in previous directives, including the right of individuals to gain access to their personal data, and the *right to explanation* when algorithmic decision-making occurs. It also grants new individual rights - the right to be *forgotten*, which gives individuals the right to have their personal data erased, and the *right to data portability*.⁹⁴ It provides for protection of personal data regardless of the technology used for processing it, whether automated or manually processed, or how the data is stored - in an IT system, video surveillance, or on paper.

The GDPR also has broad territorial scope, applying to businesses located inside the EU, and those on the outside interacting with individuals within the EU. *It raises the threshold for compliance* for businesses, requiring greater openness and transparency around their data processing activities, imposing stricter limits on the use of personal data and penalties for non-compliance than in the past. The impact for businesses will be an all-encompassing and

⁹¹ Detlev Gabel and Tim Hickman, 2018, “The Rapid Evolution of Data Protection Laws,” in *The International Comparative Legal Guide to Data Protection* 2018, London: Global Legal Group, p.4, https://dvaiaapp.org/media/pdf/resource_center/Legal_Guide_To_Data_Protection_2018.pdf, accessed Feb. 8, 2019.

⁹² Gabel and Hickman, p.1.

⁹³ Gabel and Hickman, p.1.

⁹⁴ Gabel and Hickman, p.2.

ongoing effort focused on assessing for, protecting against and mitigating risk.⁹⁵

The EU is presently finalising the *eprivacy Regulation*, which in its final form will impose tighter restrictions on the ability of businesses to track individuals using cookies, or to market to them via electronic means.

Estonia #KrattLaw

Estonia is keen to establish a legal framework around AI in order to enjoy its benefits. In 2016, it established a taskforce to study the implications of self-driving vehicles for society. Finding the study of the ethical, social and legal implications of AI within the context of traffic laws limiting, and lacking the benefit of public input, it broadened its scope. In 2017 Estonia introduced public discussion around *extending legal rights* to AI, and is the first county to do so. While different approaches are being discussed, the issue of liability in the case of accidents or malfunctions is at the core of the debate. #Kratt-law proposes the creation of a new category of legal entity comprising AI, algorithms and robots, while also clarifying the liability of owners, operators and manufacturers.⁹⁶ Public discussion is tending toward general algorithmic liability, “giving the algorithm a separate legal status, similar to companies.”⁹⁷ A draft bill is

expected to enter Parliament for debate in June 2019.

Comment:

The EU provides public sector leaders and decision-makers with the most important laboratory about regulating in the early stage of development of AI, while the field is not yet mature, and the shape of the digital economy and society are still emerging. The EU approach is inspired by the concept of *precautionary principle*, which recognizes the responsibility of government to act to mitigate potential risks in the face of high uncertainty, and even when it is not yet possible to devise a full regime of rules. In such cases, the approach is incremental and prudent, focusing primarily on preventing the erosion of rights. It will be important to monitor and learn from the experience as regulation take effect.

The Estonia project is even more ambitious. In providing legal clarity around an ethical question - *who is responsible when something goes wrong* - it would create a new set of rights that would profoundly transform the public-private-civic interfaces. Public involvement in these discussions is also a powerful reminder of the depth and reach of the technological revolution, the resolution of which paves the way for the development and adoption of ever more

⁹⁵ Gabel and Hickman, p.3.

⁹⁶ The Kratt, a creature in Estonian mythology that is brought to life from hay and old household items, serves as a metaphor for the complexities of AI.

⁹⁷ Invest in Estonia, 2018, “AI and Kratt momentum,” (Oct.18), <https://e-estonia.com/ai-and-the-kratt-momentum/>, accessed Feb 6, 2019.

sophisticated AI systems that promise to fundamentally alter the daily lives of Estonian citizens.

A Comprehensive Approach to Building a Digital Society

Most if not all governments are focusing on how to reap the benefits from a digital economy and how to prepare society to successfully navigate through the changes ahead. These measures range from building digital infrastructure, supporting research, investing in skills and re-skilling, to providing support for those workers in sectors most at risk of displacement. That said, to guide society through an unprecedented period of change, some fundamental questions will need to be addressed. What kind of society would people like to live in and what can governments do to advance a vision of society? How can AI contribute to building a desirable future that people aspire to share?

Germany serves as an example of a comprehensive approach applied to a given sector, autonomous vehicles. South Korea is an example of a comprehensive approach that includes measures outlined in the previous

sections as well as others. And while most countries explore avenues within the limits of existing public institutions, others, like Barcelona are rethinking traditional institutional frameworks, seeing cities as generators of technological sovereignty for citizens.

Germany

The German government has taken several actions to position Germany as a global leader in the development and use of AI technologies, a key component of which is transparent and ethical AI. In 2018, the government released a comprehensive €3 billion AI Strategy toward the development of an “AI made in Germany” label. The strategy includes investments in R&D, as well as initiatives related to developing transparent and ethical AI and responding to the changes that AI will bring to the workforce. The latter includes several measures that “[s]afeguard... the responsible development and use of AI that serves the good of society” and “[i]ntegrate AI in society in ethical, legal, cultural and institutional terms,”⁹⁸ such as:

- Establish guidelines for developing and using AI systems that are compatible with data protection rules;
- Developing a broad-based set

⁹⁸ Government of Germany, Federal Ministry for Economic Affairs and Energy, 2018, Press Release: “Federal Government adopts Artificial Intelligence Strategy,” (Nov. 16) <https://www.de.digital/DIGITAL/Redaktion/EN/Meldungen/2018/2018-11-16-federal-government-adopts-artificial-intelligence-strategy.html>, accessed Feb 9, 2018.

⁹⁹ Deutscher Bundestag, 2018, “Enquete Commission on Artificial Intelligence,” <https://www.bundestag.de/dokumente/textarchiv/2018/kw26-de-enquete-kommission-kuenstliche-intelligenz/560330>, accessed Feb.22, 2019.

of instruments to foster the skills of the workforce; and

- Safeguarding the possibilities for work councils to engage in codetermination when it comes to the introduction and use of AI.

The German government also established a commission of MPs and AI experts to investigate how AI and algorithmic decision-making will affect society, with its report of recommendations due in 2020.⁹⁹ While these cross sector initiatives are underway, the German government's pioneering work in setting the world's first guidelines around autonomous vehicles is an example of a sectoral approach that seeks to strike a balance between harnessing the benefits of AI while mitigating risk.

In 2017, the German government established an *Ethics Commission on Automated and Connected Driving*. The Report recommended 20 guidelines for programming automated driving systems, which were subsequently adopted by the government. The guidelines start from an ethical imperative, that automated vehicles are safer than vehicles driven by people, and provide ethical certainty around specific scenarios, such as when an accident cannot be avoided, humans take precedence over animals and property, and no discrimination as to who should survive. The guidelines put boundaries around

situations of moral ambiguity, but stop short of fully autonomous systems, recognising that the technology is not yet capable of resolving situations in which the vehicle has to decide between the lesser of two evils. The report also outlines a role for public education in transmitting to its citizens an understanding of the principles upon which autonomous vehicles operate.

The guiding principles are as follows:

- *“Automated and connected driving is an ethical imperative if the systems cause fewer accidents than human drivers.*
- *Damage to property must take precedence over personal injury. In hazardous situations, the protection of human life must always have top priority.*
- *In the event of unavoidable accident situations, any distinction between individuals based on personal features (age, gender, physical or mental constitution) is impermissible.*
- *In every driving situation, it must be clearly regulated and apparent who is responsible for the driving task: the human or the computer.*
- *It must be documented and stored who is driving (to resolve possible issues of liability, among other things).*
- *Drivers must always be able to decide themselves whether their*

100 Government of Germany, Federal Ministry of Transport and Digital Infrastructure, 2017, Press Release: “Federal Government Adopts Action Plan on Automated Driving,” (Aug. 28), <https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/128-dobrindt-federal-government-action-plan-automated-driving.html>, accessed Feb. 9, 2019.

vehicle data are to be forwarded and used (*data sovereignty*).”¹⁰⁰

South Korea

The government of South Korea is an example of a truly comprehensive approach to building a digital society. It recognized early on the potential of AI for societies searching for new solutions to some of its oldest problems, like support for aging populations.

Although never publicly released, in 2007 the South Korean government drafted the world’s first robotics charter to provide ethical guidelines on the respective roles and functions of manufacturers, users, owners, and robots themselves. Concerned about the social and legal issues related to human-robot interaction, especially with respect to the decision-making potential of AI, the charter specified the rights and duties of users and owners, the rights and duties of robots, and standards for manufacturers.¹⁰¹

In 2016, the government released a *Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society*. The plan is all encompassing, addressing many of the areas that AI is thought to have an impact on, such as workforce preparedness, education and social welfare. The plan includes a broad range of legal and ethical reforms to

guide the development and adoption of AI, from self-regulatory practices to regulation.

More particularly, the government seeks to put in place a human-centred charter of ethics to govern data-collection and AI algorithms for developers and users (by 2018). The plan also proposes a range of regulatory measures, chief among them, protecting citizens’ personal data, establishing in law the value and ownership of data collected by private-sector companies as equal to owning real estate, clarifying manufacturers’ liabilities for accidents that result from AI-related errors, and granting rights and responsibilities to “electronic persons.”

The plan also proposes a range of administrative practices, including establishing a process for contesting and reviewing decisions made by AI, protocols for compliance around the collection of data and development of algorithms to guard against algorithmic bias, and methods for ensuring compliance around ethical standards during the development of AI.

The plan likewise sets out a series of adaptive measures to prepare for the societal changes that AI promises. These include:

- *Improving the social safety net* as response to job losses and

¹⁰¹ Bensoussan, Alain, 2014, “ Le Droit des Robots La Charte Coreenne,” *Planete Robots* #25, (Jan-Fev) pp.14-15, <https://www.alain-bensoussan.com/wp-content/uploads/2014/01/24025860.pdf>, accessed Feb. 9, 2019.

transitions and income polarization;

- *Expanding flexible working hours programs and introduce a working-hour account system* to support workers in the transition to an “increasingly automated and platform-dependent industrial structure;”

- *Redefining “the concept of ‘worker’* to account for increasingly diverse forms of employment, and broaden the reach of unemployment and workers’ compensation insurance to include new forms of work; and

- *Educating toward a human-centered* technological culture that respects both humanity and technology.

City of Barcelona

The City of Barcelona is at the forefront of an emerging movement of cities rethinking the narrow technological objectives of the Smart City concept, toward a democratic, open source, and commons-based digital city built from the bottom up. At the heart of Barcelona’s smart city development is technological sovereignty, where the government and citizens determine the direction and use of technological innovations for a common purpose, and the digital rights of citizens.

Barcelona has undertaken a number

of simultaneous initiatives to get there. For instance, the City developed a digital transformation roadmap with guidelines and democratic digital standards, including:

- o A technological code of conduct, the migration to open source software, open architectures and open standards;
- o Procurement contracts that mandate transparency, open standards and open data; and
- o A data directive with data ethics, privacy and citizens’ data sovereignty at its core.

A second example, the Barcelona Digital City Plan, which was developed in consultation with citizens and stakeholders, seeks to implement projects that help solve urban issues, like access to affordable housing, and participatory urban planning for a sustainable and green city.

The City further seeks to involve citizens in decision-making using Decidim, a free open-sourced digital platform for citizen participation. Citizen involvement is encouraged at all stages, from proposing ideas for the budget to deciding bus routes.

Barcelona also forms part of DECODE (Decentralised Citizen-owned Data Ecosystems), a pilot initiative funded by the EU and launched in 2017 to address the loss of control by citizens

102 Decode, “What is Decode?” <https://decodeproject.eu/what-decode>, accessed Feb. 14, 2019.

103 Decode.

of their personal data, and the monopolization and monetization of personal data by a small group of high tech companies. At the heart of the initiative is the belief that “people should have a choice about what happens to their digital identity, who uses their data online, and for which purposes.”¹⁰² DECODE seeks to give people control over their personal data, and, in the process, build a democratic version of data ownership known as a data commons: “a data-centric digital economy where data that is generated and gathered by citizens, the Internet of Things (IoT), and sensor networks is available for broader communal use.”¹⁰³ At the heart of this vision is a strong role for cities, as custodians of the digital rights of citizens.

Barcelona is piloting two of four DECODE initiatives (Amsterdam the other two), the results of which will form the basis of policy recommendations to the European Commission on how to protect citizens’ digital sovereignty.

The first is a data commons, where citizens own and control their data, and decide what kind of data (encrypted, privacy-enhancing) they want to share, with whom, and on what terms. The second initiative is a digital platform for civic participation, which allows citizens to sign petitions anonymously, with authentication requirements in place. The platform enhances privacy and transparency, and allows for data sharing.

Comments:

These examples illustrate a range of paths to building digital societies. The German government’s pioneering work around setting, adopting and enforcing ethical guidelines around self-driving vehicles provides an example of a sectoral approach. South Korea presents a cross-sectoral approach that uses a range of measures, while Barcelona undertakes broader institutional reform, reframing the concept of smart cities and claiming technological sovereignty and digital rights for citizens. As these approaches illustrate, the key is to find the right balance in harnessing AI’s potential such that the benefits for societies outweigh the risks, and to adapt the practices that are best suited to the context and circumstances of a particular society.

V. Conclusions

This paper explores some of the literature around the challenges of governing at a time of accelerating disruptive change, with a focus on AI technologies, which more than others raise profound ethical and social concerns, and promise far-reaching changes for the society that we know today.

Governments are following different paths to steer their societies through this unprecedented period of change, from industry self-regulation, declarations of principles, and regulatory frameworks, to more comprehensive approaches that draw on some of these as well as other measures to build a digital society.

What can be learned from these early initiatives?

There are no simple solutions; it is for governments to determine the appropriate mix of tools, practices, and approaches depending on their own legal, ethical and cultural contexts and circumstances. The key is to learn from others about how they invented solutions to build the adaptive capacity and resilience of societies in order for people to build and share a better future together.

The next working paper, the Future of Work, will examine more closely the impact of the technological revolution on work, the workplace, and the work force.

Annex I:

Google's Principles for AI Research and Product Development (2018)

1. Be socially beneficial.

... As we consider potential development and uses of AI technologies, we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides.

...We will strive to make high-quality and accurate information readily available using AI, while continuing to respect cultural, social, and legal norms in the countries where we operate. And we will continue to thoughtfully evaluate when to make our technologies available on a non-commercial basis.

2. Avoid creating or reinforcing unfair bias.

AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.

3. Be built and tested for safety.

We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm. We will design our AI systems to be appropriately cautious, and seek to develop them in accordance with best practices in AI safety research. In appropriate cases, we will test AI technologies in constrained environments and monitor their operation after deployment.

4. Be accountable to people.

We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control.

5. Incorporate privacy design principles.

We will incorporate our privacy principles in the development and use of our AI technologies. We will give opportunity for notice and consent, encourage

architectures with privacy safeguards, and provide appropriate transparency and control over the use of data.

6. Uphold high standards of scientific excellence.

Technological innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity, and collaboration. AI tools have the potential to unlock new realms of scientific research and knowledge in critical domains like biology, chemistry, medicine, and environmental sciences. We aspire to high standards of scientific excellence as we work to progress AI development.

We will work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we will responsibly share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications.

7. Be made available for uses that accord with these principles.

...We will work to limit potentially harmful or abusive applications. As we develop and deploy AI technologies, we will evaluate likely uses in light of the following factors:

- *Primary purpose and use*: the primary purpose and likely use of a technology and application, including how closely the solution is related to or adaptable to a harmful use.
- *Nature and uniqueness*: whether we are making available technology that is unique or more generally available.
- *Scale*: whether the use of this technology will have significant impact.
- *Nature of Google's involvement*: whether we are providing general-purpose tools, integrating tools for customers, or developing custom solutions.

AI applications we will not pursue

In addition to the above objectives, we will not design or deploy AI in the following application areas:

1. Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.
2. Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.
3. Technologies that gather or use information for surveillance violating internationally accepted norms.
4. Technologies whose purpose contravenes widely accepted principles of international law and human rights.

-

Source: <https://ai.google/principles/>

Bibliography:

- 2009, "An Intellectual Entente," *Harvard Magazine*, (Oct. 9), <https://harvardmagazine.com/breaking-news/james-watson-edward-o-wilson-intellectual-entente>, accessed Dec. 12, 2018.
- Benay, Alex (ed.), 2018, *Government Digital: The Quest to Regain Public Trust*, Toronto: Dundurn.
- Bensoussan, Alain, 2014, "Le Droit des Robots La Charte Coreenne," *Planete Robots* #25, (Jan-Feb) pp.14-15, <https://www.alain-bensoussan.com/wp-content/uploads/2014/01/24025860.pdf>, accessed Feb. 9, 2019.
- Borowska, Kasia, 2019, "The AI Arms Race Means We Need AI Ethics," *Forbes*, (Jan. 22), <https://www.forbes.com/sites/kasiaborowska/2019/01/22/the-ai-arms-race-means-we-need-ai-ethics/#2cab289e67a0>, accessed Feb. 6, 2019.
- Bourgon, Jocelyne, 2011, *A New Synthesis of Public Administration: Serving in the 21st Century*, Kingston: Queen's Policy Studies.
- Bourgon, Jocelyne, 2019, "NS Research Program 2019" (unpublished). Also see website: <https://www.pgionline.com/>.
- Bourgon, Jocelyne, 2017, *The New Synthesis of Public Administration Fieldbook*, Denmark: Dansk Psykologisk Forlag A/S.
- Breznitz, Dan, 2018, "Data and the Future of Growth," in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, pp.66-72, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.
- Briggs, Ray and Ed Dobner, Jennifer Dul, 2018, "Digital Reality in Government: How AR and VR can enhance Government Services," (Aug 24), Deloitte, <https://www2.deloitte.com/insights/us/en/industry/public-sector/augmented-virtual-reality-government-services.html>, accessed Nov 16, 2018.
- Cath, Corinne, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo, and Luciano Floridi, 2016, "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach" (Dec. 23), <https://ssrn.com/abstract=2906249>, accessed Jan. 25, 2019.
- Chui M., M. Harrysson, J. Manyika, R. Roberts, R. Chung, P. Nel, & A.V. Heteren, 2018, *Applying artificial intelligence for social good*. McKinsey & Co. <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good>, accessed Dec. 5, 2018.
- Ciuriak, Dan, 2018, "The Economics of Data: Implications for the Data-driven Econo

my,” in *Data Governance in the Digital Age*, edited by the Centre for International Governance Innovation, (May 2), pp. 12-19, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

Crawford, Kate, 2016, “Artificial intelligence’s white guy problem,” (June 25), http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=1, accessed Jan. 25, 2019.

Decode, “What is Decode?” <https://decodeproject.eu/what-decode>, accessed Feb. 14, 2019.

Deutscher Bundestag, 2018, “Enquete Commission on Artificial Intelligence,” <https://www.bundestag.de/dokumente/textarchiv/2018/kw26-de-enquete-kommission-kuenstliche-intelligenz/560330>, accessed Feb.22, 2019.

Down, Mina, 2019, “How Blockchain is Driving the Fourth Industrial Revolution,” *The Startup*, (Jan. 17), <https://medium.com/swlh/blockchain-ai-internet-of-things-industrial-revolution-33e99a5ab3db>, accessed Feb. 2, 2019.

Dutton, Tim, 2018, “Building an AI World: Report on National and Regional AI Strategies,” CIFAR: Ottawa, <https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies>, accessed Feb 10, 2019.

Dutton, Tim, 2018, “An Overview of National AI Strategies” <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>, accessed Jan 27, 2019.

D2L, 2018, “The Future of Work and Learning in the Age of the 4th Industrial Revolution,” <https://www.d2l.com/wp-content/uploads/2018/01/The-Future-of-Work-and-Learning-D2L.pdf>, accessed November 27, 2018.

Executive Office of the President, 2016, *Artificial intelligence, automation and the economy*, Washington, DC, <https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>, accessed Jan. 25, 2019.

Executive Office of the President National Science and Technology Council Committee on Technology, 2016, *Preparing for the future of artificial intelligence*, Washington, DC, https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf, accessed Jan. 25, 2019.

Executive Office of the President of the United States, The White House Office of Science and Technology Policy, 2018, Summary of the 2018 *White House Summit on Artificial Intelligence for American Industry*, Washington, D.C., <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>, accessed Feb. 16, 2019.

- Finley, Klint, 2016, “Obama wants to help the government to develop AI,” (Oct.10), <https://www.wired.com/2016/10/obama-envisions-ai-new-apollo-program/>, accessed Feb 16, 2019.
- Finley, Klint, 2016, “Tech Giants Team Up to Keep AI From Getting Out of Hand,” (Sept. 28), <https://www.wired.com/2016/09/google-facebook-microsoft-tackle-ethics-ai/>, accessed February 7, 2019.
- Forbes Insights, 2015, Analytics: Don’t Forget the Human Element, (Nov), https://www.forbes.com/forbesinsights/ey_data_analytics_2015/index.html, accessed June 26, 2018.
- Francis, Diane, 2018, “The Biggest New Laws to Regulate Tech Giants—and Why They Matter,” (June 6), *Singularity Hub*, <https://singularityhub.com/2018/06/06/the-biggest-new-laws-to-regulate-tech-giants-and-why-they-matter/#sm.0000irpkce1dbrdswqo9oph18y2a0>, accessed Jan. 30, 2019.
- Gabel, Detlev and Tim Hickman, 2018, “The Rapid Evolution of Data Protection Laws,” in *The International Comparative Legal Guide to Data Protection 2018*, London: Global Legal Group, pp.1-5, https://iapp.org/media/pdf/resource_center/Legal_Guide_To_Data_Protection_2018.pdf, accessed Feb 8, 2019.
- Gibbs, Samuel, 2017, “Elon Musk leads 116 experts calling for outright ban of killer robots,” *The Guardian*, Aug. 20, <https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war>, accessed Feb. 28, 2019.
- Government of Canada, 2017, Speaking Notes for The Honourable Scott Brison, President of the Treasury Board of Canada, at FWD50, (Nov 2, 2017), https://www.canada.ca/en/treasury-board-secretariat/news/2017/11/speaking_notes_forth_ehonourablecottbrisonpresidentofthetreasury.html, accessed Dec.5, 2018.
- Government of Germany, Federal Ministry for Economic Affairs and Energy, 2018, Press Release: “Federal Government adopts Artificial Intelligence Strategy,” (Nov. 16) <https://www.de.digital/DIGITAL/Redaktion/EN/Meldungen/2018/2018-11-16-federal-government-adopts-artificial-intelligence-strategy.html>, accessed Feb 9, 2018.
- Government of Germany, Federal Ministry of Transport and Digital Infrastructure, 2017, Press Release: “Federal Government Adopts Action Plan on Automated Driving,” (Aug. 28), <https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/128-do-brindt-federal-government-action-plan-automated-driving.html>, accessed Feb. 9, 2019.
- Government of Singapore, Personal Data Protection Commission, 2018, “Discussion Paper on Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI,” (June 5), p.2, <https://www.pdpc.gov>.

sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/
Discussion-Paper-on-AI-and-PD---050618.pdf, accessed Feb. 8, 2019.

Gov.UK, Cabinet Office, Government Digital Service, 2017, *Policy Paper: Government Transformation Strategy: Background*, (Feb 9), <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-background>, accessed Nov. 16, 2018.

Greenfield, Adam, 2017, *Radical Technologies: The Design of Everyday Life*. Verso: N.Y.

Haggart, Blayne, 2018, "The Government's Role in Constructing the Data-driven Economy," in *Data Governance in the Digital Age*, edited by the Centre for International Governance Innovation, (May 2), pp.20-24, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

Harris, Tristan, 2018, "DreamForce 2018 Keynote: How to Stop Technology from Destabilizing the World," (Sept. 25), <https://www.youtube.com/watch?v=UfMzyBkzuEc>, accessed Dec 12, 2018.

Hebenstreit, Sonja and Isabel Funk-Leisch, 2018, "Austria," in *The International Comparative Legal Guide to Data Protection 2018*, London: Global Legal Group, pp.20-29, https://iapp.org/media/pdf/resource_center/Legal_Guide_To_Data_Protection_2018.pdf, accessed Feb 8, 2019.

Helbing, Dirk, 2015, "Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies," <https://arxiv.org/abs/1504.03751>, accessed Jan.24, 2019.

Hume, Kathryn, Daniel Moore and Michael Zerbs, 2018, "The AI revolution needs a rule book. Here's a beginning" in the *Globe and Mail*, (Dec. 6), accessed Dec.7, 2018.

Invest in Estonia, 2018, "AI and Kratt momentum," (Oct.18), <https://e-estonia.com/ai-and-the-kratt-momentum/>, accessed Feb 6, 2019.

Jones, HESSIE, "Geoff Hinton Dismissed the Need for Explainable AI: 8 Experts Explain why he's Wrong," *Forbes*, (Dec. 20), <https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/#ccc0b8d756d3>, accessed Dec. 20, 2018.

Klugman, Ian, 2018, "Shift Happens: Governments and the Fourth Industrial Revolution," in *Government Digital: The Quest to Regain Public Trust*, edited by Alex Benay, Toronto: Dundurn.

Lawrence, Matthew, 2017, "Emerging Technology: an interview with Adam Greenfield. 'God forbid that anyone stopped to ask what harm this might do us,'" *IPPR*, (Oct 13), <https://www.ippr.org/juncture-item/emerging-technology-an->

interview-with-adam-greenfield-god-forbid-that-anyone-stopped-to-ask-what-harm-this-might-do-us, accessed Dec. 7, 2018.

Lucas, Louise and Emily Feng, 2018, “Inside China’s surveillance state,” *Financial Times*, (July 20), <https://www.ft.com/content/2182eebe-8a17-11e8-bf9e-8771d5404543>, accessed Feb. 21, 2019.

Lung, Nicky, 2018, “Singapore announces initiatives on AI governance and ethic,” (June 5), <https://www.opengovasia.com/singapore-announces-initiatives-on-ai-governance-and-ethics/>, accessed Feb 8, 2019.

Martineau, Paris, 2018, “The Co-opting of French Unrest to Spread Disinformation,” *Wired*, (Dec. 11), <https://www.wired.com/story/co-opting-french-unrest-spread-disinformation/>, accessed Dec. 12, 2018.

McKinsey Analytics, 2018, “Notes from the AI frontier: AI Adoption advances, but foundational barriers remain,” (Nov), <https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain>, accessed Feb. 21, 2019.

Medhora, Rohinton, 2018, “Data Governance in the Digital Age,” in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, pp.2-5, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

Merchant, Brian, 2018, “Here’s Who Stands to Lose Most as AI Sweeps the Globe,” (Dec. 12), <https://gizmodo.com/heres-who-stands-to-lose-in-the-booming-male-dominated-1831059055>, accessed Dec. 13, 2018.

Mialhe, Nicolas and Cyrus Hodes, 2017, “Making the AI Revolution Work for Everyone: A Report to the OECD” (March), The Future Society, AI Initiative, <https://www.tuftsgloballeadership.org/sites/default/files/images/resources/Mialhe%20Reading.pdf>, accessed Dec.7, 2018.

Millar, Dr. Jason, Brent Barron, Dr. Koichi Hori, Rebecca Finlay, Kentaro Kotsuki, and Dr. Ian Kerr, 2018, “Discussion Paper for G7 Multistakeholder Conference on Artificial Intelligence, “Theme 3: Accountability in AI Promoting Greater Societal Trust,” Dec. 6, Montreal, Canada, p. 3, [https://www.ic.gc.ca/eic/site/133.nsf/vwapj/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf/\\$FILE/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf](https://www.ic.gc.ca/eic/site/133.nsf/vwapj/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf/$FILE/3_Discussion_Paper_-_Accountability_in_AI_EN.pdf), accessed Feb. 6, 2018.

Obar, Jonathan and Brenda McPhail, 2018, “Preventing Big Data Discrimination in Canada,” in *Data Governance in the Digital Age Special Report*, Centre for International Governance Innovation, pp.56-64, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.

Orcutt, Mike, 2018, “Ethereum thinks it can change the world. It’s running out of

time to prove it,” *MIT Technology Review*, (Dec. 13), <https://www.technologyreview.com/s/612507/ethereum-thinks-it-can-change-the-world-its-running-out-of-time-to-prove-it/>, accessed Dec 13, 2018.

Price Waterhouse Coopers, 2017, “Sizing the prize: What’s the real value of AI for your business and how can you capitalize?” p.6, <https://www.pwc.com.au/government/pwc-ai-analysis-sizing-the-prize-report.pdf>, accessed Feb. 21, 2019.

Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker, 2018, “Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,” AI Now Institute, <https://ainowinstitute.org/aiareport2018.pdf>, accessed Feb, 6, 2018.

Robbins, Mark, 2018, “Open data: an open-ended question,” *Policy Options*, (Jan 17), <http://policyoptions.irpp.org/magazines/january-2018/open-data-an-open-ended-question/>, accessed Nov 16, 2018.

Roose, Kevin, 2018, “Is Tech too Easy to Use?” *The New York Times*, (Dec.12), <https://www.nytimes.com/2018/12/12/technology/tech-friction-frictionless.html>, accessed Dec. 12, 2018.

Sample, Ian, 2018, “Thousands of leading AI Researchers sign pledge against killer robots,” (July 18), *The Guardian* <https://www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots>, accessed Feb. 27, 2019.

Shoham, Yoav, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, December 2018, “The AI Index 2018 Annual Report,” AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>, accessed Dec. 12, 2018.

Simonite, Tom, 2018, “Tech Firms Move to Put Ethical Guard Rails Around AI,” (May 16) <https://www.wired.com/story/tech-firms-move-to-put-ethical-guard-rails-around-ai/>, accessed Feb. 8, 2019.

-, 2019, “Smart Dubai launches guidelines on ethical use of Artificial Intelligence,” (Jan 8), *Gulf News Tech*, <https://gulfnews.com/technology/smart-dubai-launches-guidelines-on-ethical-use-of-artificial-intelligence-1.61317752>, accessed Feb. 7, 2019.

The White House, 2019, “Executive Order on Maintaining American Leadership in Artificial Intelligence,” (Feb 11), <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>, accessed Feb. 17, 2019.

Turner, Jacob, 2017, “Letting Facebook control AI regulation is like letting the NRA

control gun laws,” (Dec 6), <https://qz.com/1145668/letting-facebook-control-ai-regulation-is-like-letting-the-nra-control-gun-laws/>, accessed Jan. 27, 2018.

University of Guelph, 2018, “Press Release: U of G Launches Ethical Artificial Intelligence Centre,” (Dec 14), <https://news.uoguelph.ca/2018/12/u-of-g-launches-ethical-artificial-intelligence-centre/>, accessed Dec. 14, 2018.

Viechnicki, Dr. Peter and William Eggers, 2017, “How much time and money can AI save government?” (April 26), Deloitte, <https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/artificial-intelligence-government-analysis.html>, accessed Nov 16, 2018.

West, Darrell and John Allen, 2018, “How artificial intelligence is transforming the world,” (April 24), <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/>, accessed Feb. 21, 2019.

West, Darrell, 2018, “The role of corporations in addressing AI’s ethical dilemmas,” <https://www.brookings.edu/research/how-to-address-ai-ethical-dilemmas/>, accessed Feb. 8, 2019.

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kazianas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Shulz, Oscar Swartz, 2018, “AI Now Report 2018,” AI Now Institute, https://ainowinstitute.org/AI_Now_2018_Report.pdf, accessed Dec.20, 2018.

World Economic Forum, 2016, *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*, World Economic Forum, http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf, accessed Feb.2, 2019.

World Economic Forum, 2018, *The Future of Jobs Report 2018*, World Economic Forum, (January), http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf, accessed Feb.2, 2019.

Wylie, Bianca, 2018, “Governance Vacuums and How Code is Becoming Law,” in, *Data Governance in the Digital Age*, edited by the Centre for International Governance Innovation, (May 2), pp.86-90, <https://www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf>, accessed Nov. 8, 2018.



NS is an international co-operation initiative led by The Honourable Jocelyne Bourgon P.C., O.C.