

Policy and Society

ISSN: 1449-4035 (Print) 1839-3373 (Online) Journal homepage: www.tandfonline.com/journals/rpas20

Governance of artificial intelligence

Araz Taeihagh

To cite this article: Araz Taeihagh (2021) Governance of artificial intelligence, Policy and Society, 40:2, 137-157, DOI: 10.1080/14494035.2021.1928377

To link to this article: https://doi.org/10.1080/14494035.2021.1928377

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 04 Jun 2021.

Submit your article to this journal

Article views: 21382

View related articles

View Crossmark data

Citing articles: 16 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rpas20

POLICY AND SOCIETY
2021, VOL. 40, NO. 2, 137–157
https://doi.org/10.1080/14494035.2021.1928377

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Governance of artificial intelligence

Araz Taeihagh

Policy Systems Group, Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore

ABSTRACT

The rapid developments in Artificial Intelligence (AI) and the intensification in the adoption of AI in domains such as autonomous vehicles, lethal weapon systems, robotics and alike pose serious challenges to governments as they must manage the scale and speed of socio-technical transitions occurring. While there is considerable literature emerging on various aspects of AI, governance of AI is a significantly underdeveloped area. The new applications of AI offer opportunities for increasing economic efficiency and quality of life, but they also generate unexpected and unintended consequences and pose new forms of risks that need to be addressed. To enhance the benefits from AI while minimising the adverse risks, governments worldwide need to understand better the scope and depth of the risks posed and develop regulatory and governance processes and structures to address these challenges. This introductory article unpacks AI and describes why the Governance of AI should be gaining far more attention given the myriad of challenges it presents. It then summarises the special issue articles and highlights their key contributions. This special issue introduces the multifaceted challenges of governance of AI, including emerging governance approaches to AI, policy capacity building, exploring legal and regulatory challenges of AI and Robotics, and outstanding issues and gaps that need attention. The special issue showcases the state-of-the-art in the governance of AI, aiming to enable researchers and practitioners to appreciate the challenges and complexities of AI governance and highlight future avenues for exploration.

KEYWORDS

Governance; artificial intelligence; AI; robotics; public policy

1. Introduction

Artificial intelligence (AI) is rapidly changing how transactions and social interactions are organised in society today. AI systems and the algorithms supporting their operations play an increasingly important role in making value-laden decisions for society, ranging from clinical decision support systems that make medical diagnoses, policing systems that predict the likelihood of criminal activities and filtering algorithms that categorise and provide personalised content for users (Helbing, 2019; Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). The ability to mimic or rival human intelligence in complex problem-solving sets AI apart from other technologies, as many cognitive tasks

CONTACT Araz Taeihagh spparaz@nus.edu.sg; araz.taeihagh@new.oxon.org Lee Kuan Yew School of Public Policy, National University of Singapore, 469B Bukit Timah Road, Li Ka Shing Building, Level 2, #02-10 259771 Singapore

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

traditionally performed by humans can be replaced and outperformed by machines (Bathaee, 2018; Osoba & Welser, 2017; Sætra, 2020).

While the technology can yield positive impacts for humanity, AI applications can also generate unexpected and unintended consequences and pose new forms of risks that need to be effectively managed by governments. As AI systems learn from data in addition to programmed rules, unanticipated situations that the system has not been trained to handle and uncertainties in human-machine interactions can lead AI systems to display unexpected behaviours that pose safety hazards for its users (He et al., 2019; Helbing, 2019; Knudson & Tumer, 2011; Lim & Taeihagh, 2019). In many AI systems, biases in the data and algorithm have been shown to yield discriminatory and unethical outcomes for different individuals in various domains, such as credit scoring and criminal sentencing (Huq, 2019; Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018). The autonomous nature of AI systems presents issues around the potential loss of human autonomy and control over decision-making, which can yield ethically questionable outcomes in multiple applications such as caregiving and military combat (Firlej & Taeihagh, 2021; Leenes et al., 2017; Solovyeva & Hynek, 2018). Responsibility and liability for harms resulting from the use of AI applications remain ambiguous under many legal frameworks (Leenes et al., 2017; Xu & Borson, 2018) and the automation of routine and manual tasks in domains such as data analysis, service, manufacturing and driving enabled by machine-learning algorithms, chatbots and driverless vehicles are expected to displace millions of jobs that will not be evenly distributed within and across countries (Linkov, Trump, Poinsatte-Jones, & Florin, 2018; Taeihagh & Lim, 2019). Managing the scale and speed of AI adoption and their attendant risks is becoming an increasingly central task for governments. However, in many instances, the beneficiaries of these technologies do not bear the costs of their risks, and these risks are transferred to the society or governments (Leenes et al., 2017; Soteropoulos, Berger, & Ciari, 2018).

While there is considerable literature emerging on various aspects of AI, governance of AI is an emerging but significantly underdeveloped area. To enhance the benefits of AI while minimising the adverse risks they pose, governments worldwide need to understand better the scope and depth of the risks posed. There is a need to reassess the efficacy of traditional governance approaches such as the use of regulations, taxes, and subsidies, which may be insufficient due to the lack of information and constant changes (Guihot, Matthew, & Suzor, 2017), and the speed and scale of adoption of AI threatens to outpace the regulatory responses to address the concerns raised (Taeihagh, Ramesh, & Howlett, 2021). As such, governments face mounting pressures to design and establish new regulatory and governance structures to deal with these challenges effectively. The increasing recognition of AI governance across government, the public (Chen, Kuo, & Lee, 2020; Zhang & Dafoe, 2019, 2020) and industry is evident from the emergence of new governance frameworks in the meta-discourse on AI such as adaptive and hybrid governance (Leiser & Murray 2016; Linkov et al., 2018; Tan & Taeihagh, 2021b), and self-regulatory initiatives such standards and voluntary codes of conduct to guide AI design (Guihot et al., 2017; IEEE 2019). The first half of 2018 saw the release of new AI strategies from over a dozen countries, significant boosts in pledged financial support by governments for AI, and the heightened involvement of industry bodies in AI regulatory development (Cath, 2018), raising further questions regarding what ideas and interests

should shape AI governance to ensure inclusion and diverse representation of all members of society (Hemphill, 2016; Jobin, Ienca, & Vayena, 2019).

This special issue introduces the multifaceted challenges of governance of Artificial Intelligence, including emerging governance approaches to AI, policy capacity building, and exploring legal and regulatory challenges of AI and Robotics. This introduction unpacks AI and describes why the Governance of AI should be gaining far more attention given the myriad of challenges it presents. The introduction then summarises of the special issue articles are presented, and their key contributions are highlighted. Thanks to the diverse set of articles comprising this special issue; it highlights the state-of-the-art in the governance of AI and discusses the outstanding issues and gaps that need attention, aiming to enable researchers and practitioners to appreciate the challenges that AI brings better and understand the complexities of governance of AI and future avenues for exploration.

2. AI – background and recent trends

Conceptions of AI date back to earlier efforts in developing artificial neural networks to replicate human intelligence, which can be referred to as the ability to interpret and learn from the information. Originally designed to understand neuron activity in the human brain, more sophisticated neural networks were developed in the late 20th century with the aid of advancements in processing power to solve problems such as image and speech recognition (Izenman 2008). These efforts led to the introduction of the concept of AI as computer programs (or machines) that can perform predefined tasks at much higher speeds and accuracy. In the most recent wave of AI developments facilitated by advancements in big data analytics, AI capabilities have expanded to include computer programs that can learn from vast amounts of data and make decisions without human guidance, commonly referred to as Machine-learning (ML) algorithms (Izenman 2008). Unlike earlier algorithms that rely on pre-programmed rules to execute repetitive tasks, ML algorithms are designed with rules about how to learn from data that involves 'inferential reasoning', 'perception', 'classification', and 'optimisation' to replicate human decision-making (Bathaee, 2018; Linkov et al., 2018). The learning process involves feeding these algorithms with large data sets, from which they seek and test complex mathematical correlations between candidate variables to maximise predictions of a specified outcome (Kleinberg et al. 2018; Brauneis & Goodman, 2018). As these algorithms adapt their decision-making rules with more experience, ML-driven decisions are primarily dependent on the data rather than on pre-programmed rules and, thus, typically cannot be predicted well in advance (Mittelstadt et al., 2016).

Among AI experts and researchers, there is a broad consensus that AI still 'falls short' of human cognitive abilities, and most AI applications that have been successful to date stem from 'narrow AI' or 'weak AI', which refer to AI applications that can perform tasks in specific and restricted domains, such as chess, image, and speech recognition (Bostrom & Ludkowsky 2014; Lele, 2019b). Narrow AI is expected to automate and replace many mid-skill professions due to their ability to execute routine, cognitive tasks at much higher speeds and accuracy than their human counterparts (Lele, 2019bb; Linkov et al., 2018). In future, it is expected that this form of AI will eventually achieve 'General AI' or 'artificial general intelligence', a level of intelligence

comparable to or surpassing humans due to the ability to generalise across different contexts that cannot be programmed in advance (Bostrom & Ludkowsky 2014; Wang & Siau, 2019). This introduction and the articles comprising this special issue focus on applications of narrow AI.

Both industry and governments worldwide have enthused over the potential societal benefits arising from AI and thus, have accelerated the technology's development and deployment across various domains. Some of the impetuses for deploying AI include increasing economic efficiency and quality of life, meeting labour shortages, tackling ageing populations and strengthening national defence, and they vary between governments according to each nation's unique strategic concerns (Lele, 2019; Taeihagh & Lim, 2019). For instance, governments in Japan and Singapore have supported the use of assistive and surgical robots in healthcare and autonomous vehicles for public transportation to meet labour shortages and tackle ageing populations (Inagaki, 2019; SNDGO 2019; Taeihagh & Lim, 2019; Tan & Taeihagh, 2021, 2021b). Cost-savings and increased productivity are the main motivations for AI adoption in various sectors, which is already transforming the manufacturing, logistic, service, and maritime industries (World Economic Forum, 2018). AI-based technologies are also a strategic military asset for countries such as China, US, and Russia, whose governments have made significant investments in robots, drones and fully autonomous weapon systems for national defence and geopolitical influence (Allen, 2019; Lele, 2019).

3. Understanding the risks of AI

Many scholars highlight the safety issues that can arise from deploying AI in various domains. A major challenge faced by most AI applications to date stems from their lack of generalizability to different contexts, in which they can face unexpected situations widely referred to as 'corner cases' that the system had not been trained to handle (Bostrom & Ludkowsky 2014; Lim & Taeihagh, 2019; Pei, Cao, Yang, & Jana, 2017). For instance, fatal crashes have already resulted from trials of Tesla's partially autonomous vehicles due to the system's misinterpretation of unique environmental conditions that it had not previously experienced during testing. While various means of detecting these corner cases in advance have been devised, such as simulating data on many possible driving situations for autonomous vehicles, not all scenarios can be covered or even envisioned by the human designers (Bolte, Bar, Lipinski, & Fingscheidt, 2019; Pei et al., 2017). Due to the complexity and adaptive nature of ML processes, it is difficult for humans to articulate or understand why and how a decision was made, which hinders the identification of corner case behaviours in advance (Mittelstadt et al., 2016). As ML decisions are highly data-driven and unpredictable, the system can exhibit vastly different behaviours in response to almost identical inputs that make it difficult to specify 'correct' behaviours and verify their safety in advance (Koopman & Wagner, 2016). In particular, scholars point out potential safety hazards that can also arise from the interaction between AI systems and their users due to the problem of automation bias, where humans afford more credibility to automated decisions due to the latter's seemingly objective nature and, thus, grow complacent and display less cautious behaviour while using AI systems (Osoba & Welser, 2017; Taeihagh & Lim, 2019). Thus, human-machine interfaces significantly shape the degree of safety, particularly in social settings that

involve frequent interactions with users such as robots for personal care, autonomous vehicles, and service providers.

The decision-making autonomy of AI significantly reduces human control over their decisions, creating new challenges for ascribing responsibility and legal liability for the harms imposed by AI on others. Existing legal frameworks for the ascribing of responsibility and liability for machine operation treat machines as tools that are controlled by their human operator based on the assumption that humans have a certain degree of control over the machine's specification (Matthias 2004; Leenes & Lucivero, 2014). However, as AI relies largely on ML processes that learn and adapt their own rules, humans are no longer in control and, thus, cannot be expected to always bear responsibility for AI's behaviour. Under strict product liability, manufacturers and software designers could be subject to liability for manufacturing defects and design defects, but the unpredictability of ML decisions implies that many erroneous decisions made by AI are beyond the control of and cannot be anticipated by these parties (Butcher & Beridze, 2019; Kim et al. 2017; Lim & Taeihagh, 2019). This raises critical questions regarding the extent to which different parties in the AI supply chain will be held liable in different accident scenarios and the degree of autonomy that is sufficient to 'limit' the responsibility of these parties for such unanticipated accidents (Osoba & Welser, 2017; Wirtz, Weyerer, & Sturm, 2020). It is also widely recognised that excessive liability risks can hinder long-run innovation and improvements to the technology, which highlights a major issue regarding how governments can structure new liability frameworks that balance the benefits of promoting innovation with the moral imperative of protecting society from the risks of emerging technologies (Leenes et al., 2017).

Given the value-laden nature of the decisions automated by algorithms in various aspects of society, AI systems can potentially exhibit behaviours that conflict with societal values and norms, prompting concerns regarding the ethical issues that can arise from AI's rapid adoption. One of the most intensively discussed issues across industry and academia is the potential for algorithmic decisions to be biased and discriminatory. As ML algorithms can learn from data gathered from society to make decisions, they could not only conflict with the original ethical rules they were programmed with but also reproduce the inequality and discriminatory patterns of society that is contained in such data (Goodman & Flaxman, 2017; Osoba & Welser, 2017; Piano, 2020). If sensitive personal characteristics such as gender or race in the data are used to classify individuals, and some characteristics are found to negatively correlate with the outcome that the algorithm is designed to optimise, the individuals categorised with these traits will be penalised over others with different group characteristics (Liu 2018). This could yield disparate outcomes in terms of risk exposure and access to social and economic benefits. Bias can also be introduced through the human designer in constructing the algorithm, and even if sensitive attributes are removed from the data, there are techniques for ML algorithms to use 'probabilistically inferred' variables as a proxy for sensitive attributes, which is much harder to regulate (Kroll et al., 2016; Osoba & Welser, 2017). The risk of bias and discrimination stemming from the optimisation process in AI algorithms reflects a dominant concern surrounding discussions of fairness in AI governance – the trade-off between equity and efficiency in algorithmic decision-making – (Sætra, 2020) and how a balance can be struck to produce socially desirable outcomes catering to the different groups' ethical preferences remains subject to debate.

A vast body of literature and government reports have highlighted issues of data privacy and surveillance that can arise from AI applications. As algorithms in AI systems utilise sensors to collect data and big data technologies to store, process and transmit data through external communication networks, there have been concerns regarding the potential misuse of personal data by third parties and increasing calls for more holistic data governance frameworks to ensure reliable sharing of data within and between organisations (Gasser & Almeida, 2017; Janssen, Brous, Estevez, Barbosa, & Janowski, 2020). AI systems store extensive personal information about their users that can be transmitted to third parties to profile individuals' preferences, such as using past travel data collected in autonomous vehicles to tailor advertisements to passengers (Chen et al., 2020; Lim & Taeihagh, 2018), using personal and medical information collected by personal care robots and networked medical devices for the surveillance of individuals (Guihot et al., 2017; Leenes et al., 2017; Tan, Taeihagh, & Tripathi, 2021). The ownership of such data and how AI system developers should design these robots to adhere to privacy laws are key concerns that remain to be addressed (Chen et al., 2020; Leenes et al., 2017). Surveillance is also a key concern over the use of AI in many domains, such as surveillance robots in the workplace that monitor employee performance and government agencies potentially using autonomous vehicles to track passenger movements with negative implications for democratic freedoms and personal autonomy (Leenes et al., 2017; Lim & Taeihagh, 2018).

The autonomy assumed by AI systems to make decisions in place of humans can introduce ethical concerns in their application across various sectors. Studies have underlined the potential for personalisation algorithms used by digital platforms to undermine the decision-making autonomy of data subjects by filtering information presented to users based on their preferences and influencing their choices. By exerting control over an individual's decision and reducing the 'diversity of information' provided, personalisation algorithms can reduce personal autonomy and, thus, be construed as unethical (Mittelstadt et al., 2016). In healthcare, the use of robots to provide personal care services has prompted concerns over the potential loss of autonomy and dignity of care recipients if robots excessively restrict patients' mobility to avoid dangerous situations (Leenes et al., 2017; Tan et al., 2021). Studies have yet to examine how these risks can be balanced against their potential benefits for autonomy in other scenarios, such as autonomous vehicles increasing mobility for the disabled and elderly (Lim & Taeihagh, 2018), and personal care robots offering patients greater freedom of movement with the assurance of being monitored (Leenes et al., 2017). In the military, autonomous weapon systems such as drones and unmanned aerial vehicles have been developed to improve the precision and reliability of military combat, planning and strategy, but there has been increasing momentum across industry and academia, including prominent figures, highlighting their ethical and legal unacceptability (Lele, 2019; Roff, 2014). Central to these concerns is the delegation of authority to a machine to exert lethal force 'independently of human determinations of its moral and legal legitimacy' and the lack of controllability over these adaptive systems that could amplify the consequences of failure, prompting fears of a dystopian future where such weapons inflict casualties and escalate crises at a much larger scale (Firlej & Taeihagh, 2021; Scharre, 2016; Solovyeva & Hynek, 2018).

Unemployment and social instability resulting from the automation of routine cognitive tasks remains one of the most publicly debated issues concerning AI adoption

(Frey & Osborne, 2017; Linkov et al., 2018). The effects of automation are already felt in industries such as the manufacturing, entertainment, healthcare, finance, and transport sectors as companies increasingly invest in AI to reduce labour costs and boost efficiency (Linkov et al., 2018). While technological advancements have historically created new jobs as well, there are concerns that the distribution of employment opportunities is uneven across sectors and skill levels. Studies show that highly routine and cognitive tasks that characterise many middle-skilled jobs are at a high risk of automation. In contrast, tasks with relatively lower risks of automation are those that machines cannot easily replicate – this includes manual tasks in low-skilled, service occupations that require flexibility and 'physical adaptability', as well as high-skilled occupations in engineering and science that require creative intelligence (Frey & Osborne, 2017; World Economic; Forum, 2018). As high- and low-skilled occupations benefit from increased wage premiums and middle-skilled jobs are being phased out, automation could exacerbate income and social inequalities (Alonso et al. 2018).

4. Governing AI

4.1 Why AI governance is important

Understanding and managing the risks posed by AI is crucial to realise the benefits of the technology. Increased efficiency and quality in the delivery of goods and services, greater autonomy and mobility for the elderly and disabled, and improved safety from using AI in safety-critical operations such as in healthcare, transport and emergency response are the many socio-economic benefits arising from AI that can propel smart and sustainable development (Agarwal, Gurjar, Agarwal, & Birla, 2015; Lim & Taeihagh, 2018; Yigitcanlar et al., 2018). Thus, as AI systems develop and increase in complexity, their risks and interconnectivity with other smart devices and systems will also increase, necessitating the creation of both specific governance mechanisms, such as for healthcare, transport and autonomous weapons, as well as a broader global governance framework for AI (Butcher & Beridze, 2019).

4.2 Challenges to AI governance

The high degree of uncertainty and complexity of the AI landscape imposes many challenges for governments in designing and implementing effective policies to govern AI. Many challenges posed by AI stem from the nature of the problem, which are highly unpredictable, intractable and nonlinear, making it difficult for governments to formulate concrete objectives in their policies (Gasser & Almeida, 2017; Perry & Uuk, 2019).

ML systems' inherent opacity and unpredictability pose technical challenges for governments in ensuring the accountability of AI. Firstly, the opacity of complex ML algorithms remains a major barrier to AI governance as it limits the extent of transparency, explainability and accountability that can be achieved in AI systems (Lim & Taeihagh, 2019; Mittelstadt et al., 2016). Even with mandated levels of transparency and explainability of algorithms, it is impossible for the experts themselves to interpret how certain algorithmic outputs are derived from its inputs and designing the algorithm to be more explainable reduces their complexity, which has

been shown to undermine accuracy and performance (Felzmann, Villaronga, Lutz, & Tamò-Larrieux, 2019; Piano, 2020). This issue is highlighted as a severe limitation of the EU General Data Protection Regulation in increasing algorithmic transparency to tackle discrimination (Goodman & Flaxman, 2017). Algorithms are often kept intentionally opaque by their developers to prevent cyber-attacks and to safeguard trade secrets, which is legally justified by intellectual property rights, and the complexity of extensive datasets used by ML algorithms makes it nearly impossible to identify and remove all variables that are correlated with sensitive categories of personal data (Carabantes, 2020; Goodman & Flaxman 2016; Kroll et al., 2016). As most individuals lack sufficient technical literacy or the willingness to pay for accessing such expertise to help them to interpret these explanations, requirements by the likes of GDPR for mandated explanations are unlikely to inform and/or empower them (Carabantes, 2020; Felzmann et al., 2019). Secondly, ML decisions are unpredictable as they are derived from the data and can differ significantly with slight changes in inputs. The lack of human controllability over the behaviour of AI systems suggests the difficulty of assigning liability and accountability for harms resulting from software defects, as manufacturers and programmers often cannot predict the inputs and design rules that could yield unsafe or discriminatory outcomes (Kim et al. 2017; Kroll et al., 2016).

Data governance is a key issue surrounding the debate on AI governance, as multiple organisational and technological challenges exist that impede effective control over data and attribution of responsibility for data-driven decisions made by AI systems (Janssen et al., 2020). Data fragmentation and lack of interoperability between systems limits an organisation's control over data flows throughout its entire life cycle and shared roles between different parties in data sharing clouds the chain of accountability and causation between AI-driven decisions/events and the parties involved in facilitating that decision (Janssen et al., 2020).

Existing governance and regulatory frameworks are also ill-equipped to manage the societal problems introduced by AI due to the insufficient information required for understanding the technology and regulatory lags behind AI developments. Major technology companies and AI developers such as Google, Facebook, Microsoft, and Apple possess huge informational and resource advantages over governments in regulating AI, which significantly supersedes governments' traditional role in distributing and controlling resources in society (Guihot et al., 2017). The information asymmetries between technology companies and regulators compound the difficulty for the latter in understanding and applying new or existing legislation to AI applications (Taeihagh et al., 2021). Regulators are struggling to meet these informational gaps and are lagging behind due to rapid developments in the technology, which in turn leads to the formulation of laws that are 'too general' or 'vague' and lack specificity to effectively regulate the technology (Guihot et al., 2017; Larsson, 2020). In particular, lawmakers can be deterred from outlining specific rules and duties for algorithm programmers to allow for future experimentation and modifications to code to improve the software, but doing so provides room for programmers to evade responsibility and accountability for the system's resulting behaviour in society (Kroll et al., 2016). These challenges demonstrate how the four governing resources traditionally used by governments for regulation are insufficient to manage the risks arising from AI and the need for governments to find new

ways of acquiring information and devising effective policies that can adapt to the evolving AI landscape (Guihot et al., 2017; Wirtz et al., 2020).

Amidst the issues with 'hard' regulatory frameworks, industry bodies and governments have increasingly adopted self-regulatory or 'soft law' approaches to govern AI design, but they remain limited in their effectiveness. Soft law approaches refer to 'nonbinding norms and techniques' that create 'substantive expectations that are not directly enforceable'. Industry bodies have released voluntary standards, guidelines, and codes of conduct (IEEE 2019), and governments alike have formed expert committees to devise AI strategies and released multiple AI ethics guidelines and governance frameworks (PDPC et al. 2020; AI; Hleg, 2019; Jobin et al., 2019). Guidelines can be amended and adapt more rapidly than traditional regulation and thus are advantageous in keeping pace with technological developments (Hemphill, 2020; Larsson, 2020). This approach has been adopted in response to previous emerging technologies and can promote ethical, fair, and non-discriminatory practices in AI design, but many issues exist regarding their efficacy. Firstly, the voluntary nature of self-regulatory initiatives cannot assure that the outlined principles will always be adhered to, particularly as they are often not subject to uniform enforcement standards (Butcher & Beridze, 2019; Hemphill, 2016). Secondly, governments will face the challenge of ensuring consistent application of these guidelines in designing the same AI technology across different sectors if the principles differ across multiple guidelines and are not well-coordinated with regulations (Cath et al. 2018; Guihot et al., 2017).

In addition, self-regulation alone could be insufficient and even undesirable for AI governance due to their inability to ensure inclusivity and representation of diverse stakeholders. The deep involvement of industry stakeholders in developing ethical principles and regulations for AI raises concerns that corporate interests dominate AI regulations, which has been an ongoing critique of self-regulatory initiatives to govern emerging technologies in general (Hemphill, 2016). For instance, major technology companies have exerted significant influence over the framing of AI issues and the formulation of AI policy, such as through lobbying efforts and their inclusion in AI expert groups formed by governments (Cath et al. 2018; Jobin et al., 2019). Studies have highlighted the risks of regulatory capture by AI industries due to the significant informational advantages of AI developers that make their latter's technological expertise 'particularly valuable' for regulators (Guihot et al., 2017). Furthermore, the 'inscrutable' and highly opaque nature of ML algorithms could be used by corporations to legitimise deep industry involvement in AI regulations and the choices made behind these regulations are often made away from public scrutiny. The unbridled influence of corporations, as well as key political figures in the AI landscape, could exacerbate power imbalances and social inequalities, as the ideologies and interests of an elite few could manifest themselves through the design of AI and the decisions that they make in society (Jobin et al., 2019). To ensure greater inclusivity and diversity in AI governance, more research is required to examine the key actors, their roles, the dominant ideas, and values promoted in AI policies, whether there is a global convergence in these values across different countries and the degree to which these values reflect society's interests or are politically motivated (Jobin et al., 2019; Mulligan & Bamberger, 2018).

4.3 Steps forward for AI governance

The conceptual framing of AI is a crucial determinant of how the problems introduced by AI are understood, whether they are included in policies and the degree of priority afforded to the problem in public policy formulation (Perry & Uuk, 2019), but the issue of framing has yet to be extensively discussed as a key component of AI governance. As AI is still developing with the potential to grow more salient and diverse, the complexity of its challenges suggests that decision-making in AI systems needs to be carefully conceptualised according to their context of application, and these framing processes should be subject to public debate (Cunneen, Mullins, & Murphy, 2019). For instance, how policymakers frame the relative importance of different ethical principles arising from a particular AI application, conflicts between ethical principles, and the justification of their importance have critical implications on the resulting trade-offs from designing AI systems and the public's compliance with different ethical guidelines (Piano, 2020). The initial framing of AI is also critical to avoid amplifying perceived risks and fears surrounding new technologies and instead promote a more balanced discourse on what AI is, the goals and norms that AI should reinforce in society, and the design requirements to maximise its benefits while minimising its risks (Cath et al. 2018; Cunneen et al., 2019)

To address the governance challenges posed by the uncertainty and complexity of AI developments, there are increasing calls for the adoption of innovative governance approaches such as adaptive governance and hybrid or 'de-centred' governance (Dafoe, 2018; Linkov et al., 2018b; Pagallo, Casanovas, & Madelin, 2019; Tan & Taeihagh, 2021b). Characteristic of adaptive and hybrid governance is the diminished role of the government in controlling the distribution of resources in society. This is defined in hybrid governance as a combination of state and non-state actors, or blurring the public/private distinction by different degrees where regulation exists as a combination of industry standards and 'public regulatory oversight' (Guihot et al., 2017; Hemphill, 2016). Hybrid governance can exist in the forms of co-regulation, enforced self-regulation, and meta-regulation (Hemphill, 2016), all of which emphasise the increasing role played by non-state actors and the need for 'ongoing assessment of the balance of power' between private and public actors (Leiser & Murray, 2016). Similarly, adaptive governance emphasises the need to shift away from 'command and control' measures (Gasser & Almeida, 2017) towards more flexible approaches characterised by the iterative adjustment and improvement of regulations and policies as new information is gathered (Li, Taeihagh, De Jong, & Klinke, 2021; Linkov et al., 2018; Tan & Taeihagh, 2021b). Adaptive approaches are purported to be advantageous for proactively identifying and addressing the risks introduced from ML systems that are expected to change over time, as well as raising the public's awareness of AI and engaging with the public to identify new issues that have not yet entered the government's agenda (Cihon, Maas, & Kemp, 2020; Linkov et al., 2018; Pagallo et al., 2019). Flexibility is critical to enable diverse groups of stakeholders to build consensus around the norms and trade-offs in designing AI systems, as well as for global AI governance to be applicable across different geographical, cultural, and legal contexts and aligned with existing standards of democracy and human rights (Gasser & Almeida, 2017; Wirtz, Weyerer, & Geyer, 2019). Examples of adaptive governance include laws that require regular risk assessments of the regulated activity,

soft law approaches that involve collaboration with the affected stakeholders to develop guidelines, and legal experimentation and regulatory sandboxes to test innovative frameworks for liability and accountability for AI that will be adapted in iterative phases (Cath et al. 2018; Hemphill, 2020; Linkov et al., 2018; Philipsen, Stamhuis, & De Jong, 2021).

New governance frameworks can also be adapted from the approaches taken to regulate previous emerging technologies (Gasser & Almeida, 2017). Studies have analysed hybrid governance to regulate the Internet and emerging digital technologies. For instance, Leiser and Murray (2016) highlight the need to account for the increasing role of non-state actors, particularly private actors such as technology companies that control the exchange of information across the globe, transnational private actors that are developing design principles, as well as the role of civil society groups in ensuring accountability of the former two. Lessons could be drawn from the experiences of governing previous emerging technologies such as the Internet, nanotechnology, aviation safety and space law (Butcher & Beridze, 2019; Snir, 2014) to inform the governance of AI and other new emerging technologies. In addition, a key research agenda for future studies on AI governance would be to analyse the distinctive features of AI technology that warrants different approaches from previous technologies.

An emerging body of literature has proposed governing AI systems through their design, where social, legal, and ethical rules can be enforced through code to regulate the behaviour of AI systems (Leenes & Lucivero, 2014). For instance, provisions in data protection laws can be translated into technical specifications for robots equipped with cameras to automatically blur faces to prevent categorisation of individuals based on sensitive characteristics such as ethnicity, as well as to remove data after it has exceeded a specified storage timeframe or to restrict third party access to particular categories of data (Leenes et al., 2017). However, several implementation challenges must also be tackled to govern AI systems through code effectively. Many legal and ethical rules cannot be translated into explicit code, which includes 'subtle exceptions' that often require value judgements and consideration of contextual factors for interpretation, and the resulting machine interpretation also depends on how it was initially designed (Leenes & Lucivero, 2014; Mulligan & Bamberger, 2018). Other risks that governments need to manage from such an approach are potential manipulations of encoded rules to 'subvert regulatory aims', which is easily masked by the opacity of ML processes (Mulligan & Bamberger, 2018).

Common to recent studies in their proposed frameworks for AI governance is the emphasis on building broad societal consensus around AI ethical principles and ensuring accountability, but there is a need for studies examining how these frameworks can be implemented in practice. Gasser and Almeida's (2017) three-tiered framework comprises a technical layer involving the AI system processes and data structures, a layer for the ethical design of AI, and the third layer encompassing AI's societal implications and the role of regulation and legislation. Rahwan (2018) proposes extending the 'human-in-the-loop' approach to a broader 'society-in-the-loop' approach where society is first responsible for finding consensus on the values that should shape AI and the distribution of benefits and costs among different stakeholders. Other proposals centre on the need for greater centralisation and cross-cultural cooperation to improve coordination among national approaches (Cihon et al., 2020; Óhéigeartaigh, Whittlestone, Liu, Zeng, & Liu, 2020). Cihon et al. (2020) propose a framework to centralise the current fragmented state

of international AI governance, outlining methods to monitor and coordinate national approaches and examining the disadvantages of centralisation relative to decentralisation, such as regulatory lags and limited adaptability to rapidly evolving risks. Among these approaches, there are increasing calls to produce more concrete specifications on implementing these governance frameworks in practice and identifying the parties in government that are responsible for leading different aspects of AI governance (Wirtz et al., 2020).

5. Overview of the special issue articles

The articles in the special issue tackle the issues of governing AI and robotics through case studies and comparative analyses addressing gaps in the literature pertaining to examining the risks and benefits of deploying AI in different applications and sectors, identifying dominant ideals envisioned in the meta-discourse on AI governance and their implications. Furthermore, exploration of new legal/regulatory/governance approaches taken by various countries/governments to govern AI and examination of approaches which can enhance implementation of AI and cross-country comparisons of AI governance approaches is conducted. Authors have explored the limitations/effectiveness of different AI governance approaches through case studies and derived key lessons to facilitate policy learning. Below a brief summary of the articles in the special issue on the Governance of AI and Robotics is presented.

5.1 Framing governance for a contested emerging technology: insights from AI policy (Ulnicane, Knight, Leach, Stahl, & Wanjiku, 2021*)*

Whether AI develops in socially beneficial or problematic ways largely depends on public policies, governance arrangements and regulation. In recent years many national governments, international organisations, think tanks, consultancies, and civil society organisations have launched their AI strategies outlining benefits and risks as well as initial governance and regulatory frameworks. There are many questions regarding the regulation and governance of emerging disruptive technologies (Taeihagh et al., 2021). This article address the following question regarding the governance of AI: What governance and regulatory models are emerging to facilitate benefits and avoid risks of AI? What role do different actors – governments, international organisations, business, academia, and civil society – play in the emerging governance of AI? Do radical technological innovations in AI require radical or rather incremental innovations in governance? Are emerging governance and technology frameworks in different countries, regions and organisations converging or diverging and why?

To study these questions, Ulnicane et al. (2021) draw on a dataset of more than 60 AI strategies from 'national governments, international organisations, consultancies, think tanks and civil society organisations' worldwide. The interdisciplinary research framework of the article draws on concepts and approaches from Science, Technology and Innovation Studies, Policy Analysis and Political Science. It consists of two main pillars. First, to study risks, uncertainties, and unintended consequences of AI in policy documents, the concepts of policy framing and positive and negative expectations are used. Second, to study emerging governance and regulation frameworks outlined in the policy

documents, concepts of governance and governance models are used. AI policy documents are analysed to establish how they frame AI. The emerging governance of AI is analysed according to diverse governance models such as market, participatory, flexible, and deregulated (Peters, 2001). This article contributes to the studies of emerging disruptive technologies by analysing how the framing of risks and uncertainties of AI leads to the development of specific governance and regulatory arrangements mapping similarities and differences across countries, regions, and organisations.

5.2 Steering the governance of artificial intelligence: national strategies in *perspective (Radu, 2021)*

The latest wave of AI developments is built around deep learning, speech and image recognition, and data analytics tools deployed daily, such as banking, e-learning, medical diagnosis – and more recently, smart vehicles (Radu, 2021). The article empirically investigates the governance responses to AI developments worldwide. The article examines recent AI national strategies worldwide to uncover the modalities through which regulation is articulated at the state level.

The article highlights the key elements of the dominant regulatory discourses and compares AI national projects. Using content analysis and comparative methods, Radu (2021) examines the strategies of numerous countries to determine the articulation of AI governance and the co-production of political and socio-technological constructs. Building on the empirical evidence, the article highlights how the collective representation of a technical project is predefined politically and is encapsulated into a vision that is integrated into regulating the relevant sectors of society (Flichy, 2008). As such, the article contributes to the policy discourse around AI by conceptualisation the debates and critically examining the governance of AI and its complex effects in the future.

5.3 The Governance of Artificial Agency *(Gahnberg, 2021)*

Gahnberg argues for the need for new regulatory strategies to conceptualise the challenges of governing artificial agency. In the article, he argues that the notion of 'intelligence' is a vague metric of the success of the system and, the key characteristic of an AI system is its overall ability to act as an autonomous agent within a specific environment (Gahnberg, 2021). In the article, he posits that while different AI applications range from use in filtering Spam to killer robots, they share a common characteristic of being delegated authority to perform tasks regarding a specific objective(s).

The article underlines the social and legal constraints for delegating this authority to an artificial agent, including legal and ethical provisions that may prohibit tasking the agent to pursue certain objectives (e.g. using lethal force). Gahnberg (2021) further highlights the role of non-state actors in the development of governance mechanisms, such as in the case of development of standards and best practices developed by the technical community for fail-safe mechanisms or initiatives such as the Algorithmic Justice League that aims to mitigate the risks of algorithmic bias through creating inclusive training data. By examining the challenges of governing artificial agency, the article provides insights into the debates about AI governance.

5.4 Governing the adoption of robotics and autonomous systems in long-term care in Singapore (Tan & Taeihagh, 2021*)*

Autonomous systems and robotics have been dubbed as a viable technological solution to address the ever-increasing demand for long-term care globally, which is exacerbated by ageing populations (Robinson, MacDonald, & Broadbent, 2014; Tan et al., 2021). However, adopting new technologies in long-term care, like all other emerging technologies, involves unintended consequences and poses risks (Taeihagh et al., 2021). For instance, there are issues about safety, privacy, cybersecurity, liability, influence to the existing health workforce, patient's autonomy, patient's dignity, moral agency, social justice, and trust that need to be addressed by the government in the process of adoption of autonomous systems (Sharkey & Sharkey, 2012; Stahl & Coeckelbergh, 2016; Tan et al., 2021). Addressing these risks and ethical issues warrant the assessment of the government's policy capacity (Wu, Ramesh, & Howlett, 2015) and the various governance strategies that government chooses to deploy (no response, prevention-oriented, precaution-oriented, control-oriented, toleration-oriented and adaptation oriented) in the implementation of these autonomous systems (Li, Taeihagh, & De Jong, 2018, Li et al., 2021; Taeihagh & Lim, 2019).

The article is a theoretically-informed empirical study that examines the adoption of autonomous systems in long-term care as a policy measure to arrest the issue of rising social care demand due to the ageing population by examining Singapore as a case study. The article reviews various existing and potential applications of autonomous systems in social care and long-term care for older people in Singapore. It then discusses the technological risks and ethical implications of deploying these systems to users, society, and the economy at large. Finally, it assesses the extent to which governance strategies influence the policy response adopted by the Singapore government, specifically in balancing the need for innovation and addressing uncertainties arising from the deployment of emerging technologies in healthcare. Insights gathered are important for policy learning, especially in understanding the governance process of novel and emerging technologies as possible solutions to address the demand-supply mismatch in the health sector. The analysis will showcase the interactive dynamics of policy capacity, governance strategies and policy response as three major ingredients in the governance of emerging technologies and calibrate their respective proportions to facilitate effective implementations of novel technologies in healthcare (Tan & Taeihagh, 2021).

5.5. Exploring governance dilemmas of disruptive technologies: the case of care robots in Australia and New Zealand (Dickinson, Smith, Carey, & Carey, 2021*)*

While there is a range of experiments ongoing worldwide applying AI and robotics in various care settings (Tan et al., 2021), with high expectations for positive outcomes, Dickinson et al. (2021) are concerned with the unexpected consequences and risks, and their article explores the use of robots in care services in the Australian and New Zealand. They point out that while there is a burgeoning literature on robots, much of it is around technical efficacy, acceptability, or the legal ramifications. They point out the lack of

attention to the implementation of robots in care settings and their implications for policymaking.

Informed by semi-structured interviews with policymakers, care providers, suppliers of robots and technology experts, Dickinson et al. (2021) elicit a series of 'dilemmas' of governance faced in this space and identify three dilemmas relating to independence and surveillance, the re-shaping of human interactions, the dynamics of caregiving and receiving care illustrating some of the tensions involved in the governance of robotics in care services and draw attention to the issues that governments need to address for the effective adoption of these technologies.

5.6 Law and tech collide: foreseeability, reasonableness, and advanced driver assistance systems *(Leiman, 2021)*

There have been a number of fatalities involving automated vehicles since May 2016, which have been highly publicised. Simultaneously, millions of serious injuries and fatalities have occurred due to collisions by human-operated vehicles. In jurisdictions where compensations depend on the establishment of fault, many of the injured or their dependents will not recover any compensations (Leiman, 2021). In many Australian jurisdictions, the standard of care required presents significant challenges when applied to partly, highly, and fully automated vehicles (ibid).

Leiman (2021) explores how the existing regulatory framework in Australia considers perceptions of risk in establishing legal liability in the case of motor vehicles and whether this approach can be applied to automated vehicles. The article examines whether the law itself may affect the perceptions of risk in view of the data generated by the automated vehicles and considers the efficacy of no-fault or hybrid schemes for legal liability in existence in Australia and compares them with the alternative legislation passed by the Parliament in the UK that imposes responsibility on the insurers. Leiman also discusses proposals concerning the role of government in assuring safety and fault in the case of Automated vehicles.

5.7 Co-regulating algorithmic disclosure for digital platforms. *Di Porto and Zuppetta (2021)*

Di Porto and Zuppetta (2021) explore the increasing ability of IT-mediated platforms to adopt algorithmic decision making and whether disclosure self-regulation, as it currently stands at the EU level, is an appropriate strategy to address the risks posed by the increasing adoption of algorithmic decision making taken by private entities. While general data protection regulation (GDPR) requires platforms to provide information about the treatment of personal data by AI systems and self-assess the risks of data breaches, there is no reference made to the different abilities of the receiving entities to understand the meaning and consequences of algorithmic decisions.

The article argues that the disclosure self-regulation should be rethought considering the new AI developments and points out that since final consumers and SMEs are unaware of the technical underpinnings of these platforms and the value of personal data, the regulators should step in and address this issue. Furthermore, as platforms have increasingly deployed AI and Big data, they have gained the ability to manipulate the information they produce, thus weakening the validity of consumers and small businesses choices (Di

Porto & Zuppetta, 2021). The authors advocate that the regulators should tackle information asymmetry, low bargaining power and wrong information and endorse an enforced co-regulatory approach that allows participation of platforms and consumers, and development of personalised discloses based on the needs of the consumers to empower them.

Acknowledgments

Araz Taeihagh is grateful for the funding support provided by the Lee Kuan Yew School of Public Policy, National University of Singapore. The special issue editor would like to thank the editors of Policy and Society Journal and all the anonymous reviewers for their constructive feedback and support for this and other articles in the special issue on the Governance of AI and Robotics. A special thanks to Hazel Lim, Devyani Pande, and Siying Tan of the Policy Systems Group and the events team at the Lee Kuan Yew School of Public Policy for their support in various aspects of organising this special issue and the accompanying workshop held on August 30-31, 2019.

Funding

This work was supported by the National University of Singapore.

Notes on contributor

Araz Taeihagh (DPhil, Oxon) is the Head of Policy Systems Group at the Lee Kuan Yew School of Public Policy, Principal Investigator at the Centre for Trusted Internet and Community at the National University of Singapore, and Visiting Associate Professor at Dynamics of Inclusive Prosperity Initiative, at the Rotterdam School of Management, Erasmus University. He has lived and conducted research on four continents, and since 2007, his research interest has been on the interface of technology and society. Taeihagh is interested in socio-technical systems and focuses on the unique challenges that arise from introducing new technologies to society (e.g., crowdsourcing, sharing economy, autonomous vehicles, AI, MOOCs, ridesharing). Taeihagh focuses on: a) how to shape policies to accommodate new technologies and facilitate sustainable transitions; b) the effects of these new technologies on the policy process; and c) changing the way we design and analyse policies by developing innovative practical approaches to address the growth in the interdependence and complexity of socio-technical systems. Araz has more than two decades of experience working with firms on PMC and Design issues relating to chemical, petroleum and construction projects and provides technical and policy consultations on environmental, transportation, energy, and technology related issues.

ORCID

Araz Taeihagh http://orcid.org/0000-0002-4812-4745

COI-statement

No potential conflict of interest was reported by the author.

References

Agarwal, P. K., Gurjar, J., Agarwal, A. K., & Birla, R. (2015). Application of artificial intelligence for development of intelligent transport system in smart cities. *Journal of Traffic and Transportation Engineering*, *1*(1), 20–30.

Allen, G. C. (2019). *Understanding China's AI Strategy: Clues to Chinese strategic thinking on artificial intelligence and national security*. Washington, DC: Center for a New American Security.

Alonso Raposo, M., Grosso, M., Després, J., Fernándezmacías, E., Galassi, C., Krasenbrink, A., Krause, J., Levati, L., Mourtzouchou, A., Saveyn, B., Thiel, C. and Ciuffo, B. (2018). *An analysis of possible socio-economic effects of a Cooperative, Connected and Automated Mobility (CCAM) in Europe*. European Union. Luxembourg: Publications Office of the European Union,

Bathaee, Y. (2018). The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*, *31*(2), 889.

Bolte, J. A., Bar, A., Lipinski, D., & Fingscheidt, T. (2019). Towards corner case detection for autonomous driving. In *2019 IEEE Intelligent Vehicles Symposium*, pp.438–445. Paris, France. doi: 10.1109/IVS.2019.8813817

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge Handbook of Artificial Intelligence*, *316*, 334.

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale JL & Tech*, *20*, 103.

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? *The RUSI Journal*, *164*(5–6), 88–96.

Carabantes, M. Black-box artificial intelligence: an epistemological and critical analysis. AI & Soc 35, 309–317 (2020). https://doi.org/10.1007/s00146-019-00888–w

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges.

Chen, S. Y., Kuo, H. Y., & Lee, C. (2020). Preparing society for automated vehicles: Perceptions of the importance and urgency of emerging issues of governance, regulations, and wider impacts. *Sustainability*, *12*(19), 7844.

Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the Future: Investigating Architectures for International AI Governance. *Global Policy*, *11*(5), 545–556.

Cunneen, M., Mullins, M., & Murphy, F. (2019). Autonomous vehicles and embedded artificial intelligence: The challenges of framing machine driving decisions. *Applied Artificial Intelligence*, *33*(8), 706–731.

Dafoe, A. (2018). *AI governance: A research agenda; future of humanity institute*. Oxford, UK: University of Oxford.

Di Porto, F., & Zuppetta, M. (2021). Co-regulating algorithmic disclosure for digital platforms. In *Policy and society* (pp. 1–22).

Dickinson, H., Smith, C., Carey, N., & Carey, G. (2021). Exploring governance dilemmas of disruptive technologies: The case of care robots in Australia and New Zealand. *Policy & Society, This Issue*.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 2053951719860542.

Firlej, M., & Taeihagh, A. (2021). *Regulating human control over autonomous systems*. Regulation & Governance. https://doi.org/10.1111/rego.12344

Flichy, P. (2008). *Understanding technological innovation: A socio-technical approach*. Cheltenham, UK: Edward Elgar Publishing.

Forum, W. E. (2018). The future of jobs report 2018. World Economic Forum, Geneva, Switzerland.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

Gahnberg, C. (2021). The governance of artificial agency. In *Policy and society*.

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58–62.
Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*(3), 50–57.
Guihot, M., Matthew, A. F., & Suzor, N. P. (2017). Nudging robots: Innovative solutions to regulate artificial intelligence. *Vand. J. Ent. & Tech. L*, *20*, 385.
He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, *25*(1), 30–36.
Helbing, D. (2019). Societal, economic, ethical and legal challenges of the digital revolution: From big data to deep learning, artificial intelligence, and manipulative technologies. In Helbing D. (ed) *Towards digital enlightenment* (pp. 47–72). Cham: Springer. https://doi.org/10.1007/978-3-319-90869-4_6
Hemphill, T. A. (2016). Regulating nanomaterials: A case for hybrid governance. *Bulletin of Science, Technology & Society*, *36*(4), 219–228.
Hemphill, T. A. (2020). The innovation governance dilemma: Alternatives to the precautionary principle. *Technology in Society*, *63*, 101381.
Hleg, A. I. (2019). High level expert group on artificial intelligence. ethics guidelines for trustworthy AI. In *European commission*. Brussels, Belgium. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
Huq, A. Z. (2019). Racial equity in algorithmic criminal justice. *Duke Law Journal*, *68*(6), 1043–1134.
IEEE 2019. (2019). *The IEEE global initiative on ethics of autonomous and intelligent systems. ethically aligned design: A vision for prioritising human well-being with autonomous and intelligent systems* (First Edition ed.). IEEE. Springer, Cham. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html
Inagaki, K. (2019). Japan's demographics make good case for self-driving cars. *Financial Times*. https://www.ft.com/content/382aef5c-a7d7-11e9-984c-fac8325aaa04
Izenman, A. J. (2008). Modern multivariate statistical techniques. *Regression, classification and manifold learning, 10*, 978–0. Springer, New York, NY.
Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organising data for trustworthy artificial intelligence. *Government Information Quarterly*, *37*(3), 101493.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of ethics guidelines.*Nature Machine Intelligence.*1(9), 389–399.
Kim, S. (2017). Crashed software: Assessing product liability for software defects in automated vehicles. *Duke L. & Tech. Rev*, *16*, 300.
Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10, 113–174.
Knudson, M., & Tumer, K. (2011). Adaptive navigation for autonomous robots. *Robotics and Autonomous Systems*, *59*(6), 410–420.
Koopman, P., & Wagner, M. (2016). Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*, *4*(1), 15–24.
Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *U. Pa. L. Rev*, *165*, 633.
Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Journal of Law and Society*, *1*, 23.
Leenes, R., & Lucivero, F. (2014). Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design. *Law, Innovation and Technology*, *6*(2), 193–220.
Leenes, R., Palmerini, E., Koops, B. J., Bertolini, A., Salvini, P., & Lucivero, F. (2017). Regulatory challenges of robotics: Some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology*, *9*(1), 1–44.
Leiman, T. (2021). Law and tech collide: Foreseeability, reasonableness and advanced driver assistance systems. In *Policy and society* (pp. 1–22).

Leiser, M., & Murray, A. (2016). The role of non-state actors and institutions in the governance of new and emerging digital technologies. In The oxford handbook of law, regulation and technology, Eds. R. Brownsword, E. Scotford, K. Yeung, & O. U. Press. (pp. 670–704)

Lele, A. (2019). Disarmament, arms control and arms race. In *Disruptive technologies for the militaries and security* (pp. 217–229). Singapore: Springer.

Lele, A. (2019b). Artificial intelligence. In *Disruptive technologies for the militaries and security* (pp. 139–154). Singapore: Springer.

Li, Y., Taeihagh, A., & De Jong, M. (2018). The governance of risks in ridesharing: A revelatory case from Singapore. *Energies*, *11*(5), 1277.

Li, Y., Taeihagh, A., De Jong, M., & Klinke, A. (2021). Toward a commonly shared public policy perspective for analysing risk coping strategies. *Risk analysis, 41(3), 519–532*. https://doi.org/10.1111/risa.13505

Lim, H. S. M., & Taeihagh, A. (2018). Autonomous vehicles for smart and sustainable cities: An in-depth exploration of privacy and cybersecurity implications. *Energies*, *11*(5), 1062.

Lim, H. S. M., & Taeihagh, A. (2019). Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. *Sustainability*, *11*(20), 5791.

Linkov, I., Trump, B., Poinsatte-Jones, K., & Florin, M. V. (2018). Governance strategies for a sustainable digital world. *Sustainability*, *10*(2), 440.

Linkov, I., Trump, B. D., Anklam, E., Berube, D., Boisseasu, P., Cummings, C., Ferson, S., Florin, M., Goldstein, B., Hristozov, D., Jensen, K.A., Katalagarianakis, G., Kuzma, J., Lambert, J.H., Malloy, T., Malsch, I., Marcomini, A., Merad, M., Palma-Oliveira, J., Perkins, E., Renn, O., Seager, T., Stone, V., Vallero, D., Vermeire, T. (2018b). Comparative, collaborative, and integrative risk governance for emerging technologies. *Environment Systems and Decisions*, 38 (2), 170–176. https://doi.org/10.1007/s10669-018-9686–5

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*,*6*(3), 175–183.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679.

Mulligan, D. K., & Bamberger, K. A. (2018). Saving governance-by-design. *Calif.L.Rev*, *106*, 697.

Óhéigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. In *Philosophy & technology*, *33*(4), 571–593.

Osoba, O. A., & Welser, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation. Santa Monica, California.

Pagallo, U., Casanovas, P., & Madelin, R. (2019). The middle-out approach: Assessing models of legal governance in data protection, artificial intelligence, and the Web of Data. *The Theory and Practice of Legislation*, *7*(1), 1–25.

PDPC (2020). (2020). *Personal data protection commission of Singapore, Infocommunications Media Development Authority (IMDA) and Singapore Digital (SG:D). empowering possibilities model artificial intelligence governance framework* (Second Edition ed.). Singapore. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

Pei, K., Cao, Y., Yang, J., & Jana, S. (2017). Towards practical verification of machine learning: The case of computer vision systems. *arXiv Preprint*, *arXiv:1712.01785*.

Perry, B., & Uuk, R. (2019). AI governance and the policymaking process: Key considerations for reducing AI risk. *Big Data and Cognitive Computing*, *3*(2), 26.

Peters, G. (2001). *The Future of Governing: Four Emerging Models*. Kansas: University Press of Kansas.

Philipsen, S., Stamhuis, E. F., & De Jong, M. (2021). Legal enclaves as a test environment for innovative products: Toward legally resilient experimentation policies 1. In *Regulation & governance*. https://doi.org/10.1111/rego.12375

Piano, S. L. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 1–7.

Radu, R. (2021). Steering the governance of artificial intelligence: National strategies in perspective. In *Policy and society*.
Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14.
Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics*, *6*(4), 575–591.
Roff, H. M. (2014). The strategic robot problem: Lethal autonomous weapons in war. *Journal of Military Ethics*, *13*(3), 211–227.
Sætra, H. S. (2020). A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government. *Technology in Society, 62*. 101283.
Scharre, P. (2016). *Autonomous weapons and operational risk*. Washington, DC: Center for a New American Security.
Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, *14*(1), 27–40.
SNDGO (Smart Nation and Digital Government Office) (2019). Assistive technology and robotics in healthcare. *Smart Nation Singapore*. https://www.smartnation.sg/what-is-smart-nation/initiatives/Health/assistive-technology-and-robotics-in-healthcare
Snir, R. (2014). Trends in global nanotechnology regulation: The public-private interplay. *Vand. J. Ent. & Tech. L*, *17*, 107.
Solovyeva, A., & Hynek, N. (2018). *Going beyond the «Killer robots» debate: Six dilemmas autonomous weapon systems raise. Central European Journal of International & Security Studies, 12*(3).
Soteropoulos, A., Berger, M., & Ciari, F. (2018). Impacts of automated vehicles on travel behaviour and land use: An international review of modelling studies. *Transport reviews, 39*(1), 29–49.
Stahl, B. C., & Coeckelbergh, M. (2016). Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems*, *86*, 152–161.
Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, *39*(1), 103–128.
Taeihagh, A., Ramesh, M., & Howlett, M. (2021). Assessing the regulatory challenges of emerging disruptive technologies. In *Regulation & Governance*. https://doi.org/10.1111/rego.12392
Tan, S. Y., & Taeihagh, A. (2021). Governing the adoption of robotics and autonomous systems in long-term care in Singapore. In *Policy and society* (pp. 1–21).
Tan, S. Y., & Taeihagh, A. (2021b). Adaptive governance of autonomous vehicles: Accelerating the adoption of disruptive technologies in Singapore. *Government Information Quarterly*, *38*(2), 101546.
Tan, S. Y., Taeihagh, A., & Tripathi, A. (2021). Tensions and antagonistic interactions of risks and ethics of using robotics and autonomous systems in long-term care. *Technological Forecasting and Social Change*, *167*, 120686.
Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2021). Framing governance for a contested emerging technology: Insights from AI policy. In *Policy and Society* (pp. 1–20).
Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)*, *30*(1), 61–79.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615.
Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated ai governance framework for public administration. *International Journal of Public Administration*, *43*(9), 818–829.
Wu, X., Ramesh, M., & Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. *Policy and Society*, *34*(3–4), 165–171.
Xu, H., & Borson, J. E. (2018). The future of legal and ethical regulations for autonomous robotics. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.2362–-2366. IEEE. Madrid, Spain, Madrid Municipal Conference Centre

Yigitcanlar, T., Kamruzzaman, M., Foth, M., Sabatini, J., Da Costa, E., & Ioppolo, G. (2018). Can cities become smart without being sustainable? A systematic review of the literature. In *Sustainable cities and society*. *45*, 348–365.

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Oxford, UK: University of Oxford.

Zhang, B., & Dafoe, A. (2020). US public opinion on the governance of artificial intelligence. In *Proceedings of the AAAI/ACM Conference on AI, Ethics and Society*, pp.187–193.Zhang, B., & Dafoe, A. (2020). US public opinion on the governance of artificial intelligence. In *Proceedings of the AAAI/ACM Conference on AI, Ethics and Society*, pp.187–193. New York, NY, USA.