ARTICLE: Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence

Winter, 2017

Reporter
20 Vand. J. Ent. & Tech. L. 385 *

Length: 33832 words

Author: Michael Guihot,* Anne F. Matthew** & Nicolas P. Suzor***

*Senior Lecture: r, Commercial and Property Law Research Centre, Queensland University Technology Faculty of Law, *michael.guihot@qut.edu.au*. The Authors wish to thank participants in the We Robot 2017 conference at Yale University, especially Kate Crawford, Dustin Lewis, and Matthew Scherer for their helpful suggestions and comments.

**Lecture: r, Commercial and Property Law Research Centre, Queensland University of Technology Faculty of Law, *a.matthew@qut.edu.au*.

***Associate Professor, Queensland University of Technology Faculty of Law, *n.suzor@qut.edu.au*. Associate Professor Suzor is the recipient of an Australian Research Council DECRA Fellowship (Project Number DE160101542).

Highlight

Abstract

There is a pervading sense of unease that artificially intelligent machines will soon radically alter our lives in ways that are still unknown. Advances in artificial intelligence (AI) technology are developing at an extremely rapid rate as computational power continues to grow exponentially. Even if existential concerns about AI do not materialize, there are enough concrete examples of problems associated with current applications of AI to warrant concern about the level of control that exists over developments in this field. Some form of regulation is likely necessary to protect society from harm. However, advances in regulatory capacity have not kept pace with developments in new technologies, including AI. This is partly because regulation has become decentered; that is, the traditional role of public regulators such as governments commanding regulation has dissipated, and other participants including those from within the industry have taken the lead. Other contributing factors are dwindling government resources on one hand and the increased power of technology companies on the other. These factors have left the field of AI development relatively unregulated. Whatever the reason, it is now more difficult for traditional public regulatory bodies to control the development of AI. In the vacuum, industry participants have begun to self-regulate by promoting soft law options such as codes of practice and standards. This Article: argues that despite the reduced authority of public regulatory agencies, the risks associated with runaway AI require [*386] regulators to begin to participate in what is largely an unregulated field. In an environment where resources are scarce, governments or public regulators must develop new ways of regulating. This Article: proposes solutions to regulating the development of AI ex ante through a two-step process: first, governments can set expectations and send signals to

influence participants in AI development. The Authors adopt the term "nudging" to refer to this type of influencing. Second, public regulators must participate in and interact with the relevant industries. By doing this, they can gather information and knowledge about the industries, begin to assess risks, and then be in a position to regulate those areas that pose the most risk first. To conduct a proper risk analysis, regulators must have sufficient knowledge and understanding about the target of regulation to be able to classify various risk categories. The Authors have proposed an initial classification based on the literature that can help to direct pressing issues for further research and a deeper understanding of the various applications of AI and the relative risks they pose.

Text

[*387]

I. Introduction

When Google purchased DeepMind in 2014, its owners made it a condition of the sale that Google establish an ethics board to govern the future use of the artificial intelligence (AI) technology. [1] This insistence betrayed concerns about AI development from within the industry. Google apparently agreed to set up the ethics board, but nothing is known about the identity of the board members or the content of their discussions. On July 20, 2016, Google reported that it had deployed DeepMind's machine learning in a series of tests on one of its live data centers. [2] The tests resulted in a reported 40 percent decrease in energy consumption for the center while the AI was applied. [3] DeepMind reported that

[*388]

working at Google scale gives us the opportunity to learn how to apply our research to truly global and complex problems, to validate the impact we can have on systems that have already been highly optimised by brilliant computer scientists, and - as our data centre work shows - to achieve amazing real-world impact too. [4]

Working at "Google scale" presumably means using Google's worldwide infrastructure to test its AI systems - the opportunities for which appear to be limitless. Google has already expanded its testing using DeepMind in other areas such as to reduce global warming [5] and to improve diagnosis and treatment in healthcare. [6]

If the results of the application of AI in Google's data centers can be replicated more broadly so as to reduce the world's energy consumption, avert global warming, or enable affordable, accessible healthcare, then humanity will

---

[1] Alex Hern, Whatever Happened to the DeepMind AI Ethics Board Google Promised?, Guardian (Jan. 26, 2017, 9:50 AM), *https://www.theguardian.com/technology/2017/jan/26/google-deepmind-ai-ethics-board* [*https://perma.cc/XZ7H-XM3A*].

[2] Richard Evans & Jim Gao, DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, DeepMind (July 20, 2016), *https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/* [*https://perma.cc/E8N9-3VTF*].

[3] Id.

[4] DeepMind Collaborations with Google, DeepMind, *https://deepmind.com/applied/deepmind-for-google/* [*https://perma.cc/6EWK-8ZW9*] (last visited Mar. 13, 2017).

[5] Sam Shead, DeepMind Is Funding Climate Change Research at Cambridge as It Looks to Use AI to Slow Down Global Warming, Bus. Insider (June 21, 2017, 7:41 PM), *https://www.businessinsider.com.au/deepmind-is-funding-climate-change-research-at-cambridge-university-2017-6*.

[6] See Mustafa Suleyman, Working with the NHS to Build Lifesaving Technology, DeepMind (Nov. 22, 2016), *https://deepmind.com/blog/working-nhs-build-lifesaving-technology/* [*https://perma.cc/E9H2-SQCM*].

reap great benefits. [7] However, while the results of the tests appear laudable, some questions linger - for instance, what checks and balances were in place to govern the application of AI here? Were any risks of its application considered and ameliorated in the tests? What governance is in place to control companies testing beta versions of AI applications on a large scale? Conversely, if regulation is put in place prematurely or without proper thought and consultation, would the potential benefits that might result from the general application of these programs in other areas be retarded or lost? In short, would regulation have a chilling effect on innovation that is harmful for the long-term public interest? The Authors argue that, with these questions in mind, AI should be more actively regulated because the benefits that can be achieved through controlled or regulated application outweigh the potential negative impacts of regulating. This Article: addresses these and some of the many other issues that must be addressed by potential regulators when seeking to regulate new technologies such as AI.

Part II outlines the range of threats posed by different applications of AI and introduces the case for regulating its [*389] development. While some argue that developing AI poses an existential threat to humanity, others point to the benefits attained by relatively controlled development and application of more benign systems. The Authors contend that these arguments are at cross-purposes and distract from a more pressing need: government ought to not only play a part in guiding the development of AI for the broader benefit of humankind but also must regulate to address the very real and present problems associated with current applications of AI today. These include bias and safety concerns, the pressing effect on employment, and the inherent intrusion into our privacy caused by AI interrogating the data society generates in everyday life. Before society contemplates regulating AI, however, it is necessary to more precisely define and classify the different technologies that are often referred to as AI. This classification exercise, the Authors argue, is vital to understanding the different types of risks that regulation might seek to address. This spectrum of risks posed by different classes of AI provides the basis upon which the Authors ultimately argue for a stratified approach to regulation. This is developed further in Part V.

Part III sets out the challenges of regulating AI. The pace of innovation in AI has far outstripped the pace of innovation in regulatory tools that might be used to govern it. This is often referred to as the pacing problem of regulation. [8] In these situations, regulation lags behind or in some circumstances "decouples" from the technology it seeks to address. [9] Another core challenge regulatory agencies face lies in the difficulty in understanding the social impacts of AI on a systems level and engaging with these impacts at every (or any) stage of development. [10] A "social-systems analysis" will allow regulators to understand the operation of AI in a broad social context. [11] As the DeepMind example illustrates, the reasons for particular decisions involving the ways in which AI is developed and applied can be [*390] opaque, largely incomprehensible, [12] and sometimes even unknowable. [13] Research

---

[7] See DeepMind Collaborations with Google, supra note 4. Here, this Article: has concentrated on the work of Google, but it is only one of the major innovators in this area. Similar work on developing AI is also being carried out by Facebook, Microsoft, and Apple, to name a few. See infra notes 76-81 and accompanying text.

[8] See Kenneth W. Abbott, Introduction: The Challenges of Oversight for Emerging Technologies, in Innovative Governance Models for Emerging Technologies 1, 3 (Gary E. Marchant et al. eds., 2014); Braden R. Allenby, Governance and Technology Systems: The Challenge of Emerging Technologies, in The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem 3, 16-17 (Gary E. Marchant et al. eds., 2011).

[9] Braden R. Allenby, The Dynamics of Emerging Technology Systems, in Innovative Governance Models for Emerging Technologies, supra note 8, at 19, 43.

[10] Kate Crawford & Ryan Calo, Comment, There Is a Blind Spot in AI Research, 538 Nature 311, 311, 313 (2016), https://www.nature.com/polopoly_fs/1.20805!/menu/main /topColumns/topLeftColumn/pdf/538311a.pdf [*https://perma.cc/G2TL-NK9V*].

[11] See id. at 313.

[12] See Perri 6, Ethics, Regulation and the New Artificial Intelligence, Part II: Autonomy and Liability, 4 Info. Comm. & Soc'y 406, 410 (2001).

[13] See Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information 4 (Harvard Univ. Press 2015).

and development in AI is carried out in many different locations, at different times, and in ways that are not highly visible. The scale of research also varies and can be carried out by a single person on a home computer or at a scale that only large multinational companies such as Google can attain.

There is no shortage of advice given to regulators about how to respond to technological change. The Authors review the challenges that current and future developments in AI are likely to pose for regulators and the different and sometimes conflicting advice that commentators have urged regulators to follow. The Authors consider the urgency of developing effective mechanisms of regulation and explain how the challenges of regulating AI are different in kind to the challenges of regulating in other domains. The Authors argue that, as many public regulators now find themselves without the resources to adequately understand or intervene in the range of complex issues that rapid developments in AI present, some regulatory innovation is required. In order to meet these challenges, the Authors suggest that regulators will need to be adaptable, develop new strategies to learn about risks, and identify opportunities to influence technological developers. The Authors show that recent developments in how regulation is conceived go some way to identifying potential future strategies for public regulators but that more work is needed.

Part IV considers how public regulators such as governments face an unprecedented challenge in managing complex governance systems that include not only public regulatory agencies but also individuals, firms, market competitors, and civil society organizations that all might play some role in influencing the development of AI in different contexts. While the regulation of other emerging technologies is not directly applicable to AI, there is much that can be learned from innovations in regulation of other fields. [14] Current regulatory mechanisms, including laws governing tort, copyright, privacy, patent, and regulations that govern other emerging technologies, are either unsuitable or, for other reasons, cannot easily be applied to novel technological developments in areas such as the regulation of AI. [15] The challenge in regulating this field is magnified [*391] by fundamental uncertainty about how AI will develop and how that development may impact other challenges society will face in the future. [16]

The size and power of the multinational companies that develop most of the world's AI - such as Google, Facebook, and Microsoft - raise fundamental issues about the ability of governments to regulate in this area at all. Far fewer of the traditional tools of regulation once available to governments seeking to regulate AI remain viable or available. The Authors highlight the concerns being expressed about the rampant research and development into AI by some of the world's biggest companies, ostensibly ungoverned, [17] and propose some innovative solutions to counterbalance the power disparity. The Authors review the range of proposals and suggestions for regulating AI and consider how regulatory theory provides guidance.

Part V argues that in the context of highly constrained governance resources, some regulatory innovation is required. Some regulation theorists are experimenting with different interventions in choice architecture to set the context and environment in which choices are made so as to promote regulatory goals. [18] The Authors argue that there is a role for government to play in shaping the regulatory environment at a very broad policy level by nudging

---

[14] Roger Brownsword, So What Does the World Need Now? Reflections on Regulating Technologies, in Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes 23, 23, 30 (Roger Brownsword & Karen Yeung eds., 2008).

[15] See Allenby, supra note 9, at 20-22.

[16] See Gonenc Gurkaynak, Ilay Yilmaz & Gunes Haksever, Stifling Artificial Intelligence: Human Perils, 32 Computer L. & Security Rev. 749, 754-55 (2016).

[17] Kate Crawford and Ryan Calo raised this issue in their comment in Nature, referring to it as the "blind spot in thinking about AI." See Crawford & Calo, supra note 10, at 311.

[18] See, e.g., Frederik Zuiderveen Borgesius, Behavioural Sciences and the Regulation of Privacy on the Internet, in Nudge and the Law: A European Perspective (Alberto Alemanno & Anne-Lise Sibony eds., 2014).

or influencing beneficial development. [19] By using its influence in this way, government can seek to guide the development of AI by framing the agenda in positive ways without wholly relinquishing its traditional regulatory role. This will also allow governments to develop a fuller regulatory response over time. The multitude of different applications of AI makes it improbable that nudging would have an effect at the micro level of individual applications. At this micro level, the Authors suggest that other more concrete regulatory approaches need to be employed. For a government to influence the [*392] development of AI systems and successfully further the public interest, it must be able to understand and influence this complex and intricate web of actors that often have diverse goals, intentions, purposes, norms, and powers. [20] When the focus shifts to regulation within individual industries or of particular types of AI applications, regulatory agencies must move beyond nudging and adopt more focused, nuanced, and adaptive approaches to regulation. [21] Other theorists have proposed greater roles for regulatory agencies with specific expertise. [22] Still others have suggested that public regulators may be able to experiment with more rapid, temporary laws, [23] although the potential lack of legal certainty that results may create further problems for investors and other participants in the field. The Authors identify some of these opportunities for innovation in the work of public regulators.

A regulatory intervention in the development of AI technology must consider the spectrum of risks that different AI applications pose. Part V introduces a risk-based regulation framework to help regulators work through the different forms of AI and to identify where scarce regulatory resources should be concentrated. The Authors' initial typology presents three discrete categories: low-, medium-, and high-risk applications of AI. Of these, the Authors suggest that the most productive area for regulators to focus on at the moment is medium-to high-risk categories, but that the potential for low-risk AI to quickly develop into high risk should mean that these areas must not be completely discounted.

Part VI concludes with a suggestion for greater cooperation and information sharing between regulators and the potentially regulated. The Authors argue that, with the increase in societal concerns about the risk inherent in developing AI, regulation of AI is an inevitable and responsible approach to governance.

[*393]

II. Artificial Intelligence: What Does the Future Hold?

To be able to regulate AI, regulatory bodies must understand it and the potential risks that it poses. The Authors must look both back and into the future to see how AI has been defined and what it might become and what risks AI has posed and might pose in the future. This Part outlines some attempts that have been made to define AI and demonstrates that there is no concrete definition. [24] This has led to an informal classification system based upon the "strength" of the underlying algorithm or its ultimate effect. Traditional classifications of AI differentiate between

---

[19] Other versions of adaptive policymaking to address deep uncertainty have been proposed using various models or approaches to policymaking. See, e.g., Warren E. Walker, Vincent A.W.J. Marchau & Darren Swanson, Addressing Deep Uncertainty Using Adaptive Policies, 77 Tech. Forecasting & Soc. Change 917, 918 (2010); Warren E. Walker, S. Adnan Rahman & Jonathan Cave, Adaptive Policies, Policy Analysis, and Policy-Making, 128 Eur. J. Operational Res. 282, 283 (2001).

[20] See Julia Black, Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory' World, 54 Current Legal Probs. 103, 106-09 (2001).

[21] See Richard S. Whitt, Adaptive Policymaking: Evolving and Applying Emergent Solutions for U.S. Communications Policy, *61 Fed. Comm. L.J. 483, 487, 576 (2008)* (proposing the application of his version of "adaptive policymaking," where regulators "tinker" with "inputs, connectivity, incentives, and feedback" to encourage firms to act in ways that further the public good).

[22] See, e.g., Matthew U. Scherer, Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, *29 Harv. J.L. & Tech. 354, 357 (2016).*

[23] Wulf A. Kaal, Dynamic Regulation for Innovation 16 (Univ. of St. Thomas (Minn.) Legal Studies Research, Paper No. 16-22, 2016); see Sofia Ranchordas, Constitutional Sunsets and Experimental Legislation: A Comparative Perspective 217 (2014).

[24] See Stuart J. Russell & Peter Norvig, Artificial Intelligence: A Modern Approach 1-5 (3d ed. 2010).

"narrow" and "strong" AI. [25] This dichotomy is an unsatisfactory way of measuring AI because it rests on dissimilar considerations of breadth and strength. The Authors propose a different classification based upon the risks that each AI application poses. In this way, the Authors can begin to sort various classes of AI based on whether the AI poses a low, medium, or high risk to either society or to human safety or wellbeing. This classification is crucial to understanding how regulatory strategies can be tailored to the relevant AI risk profile. Regulatory bodies need to perform this risk analysis before they develop laws that affect a class of AI. It is important, then, to distinguish the various meanings given to the term "artificial intelligence" and the different forms AI may take. This allows the identification of a subset or range of applications of AI most suitable for governments or regulatory bodies to initially regulate.

A. Defining AI

Before defining AI, intelligence must first be defined. Intelligence in human terms has been described as a set of factors that include "consciousness, self-awareness, language use, the ability to learn, the ability to abstract, the ability to adapt, and the ability to reason." [26] Once intelligence is defined, estimations or approximations [*394] of those qualities should form the benchmark of attempts to create or simulate it - hence artificial intelligence. But for which of those characteristics of intelligence can a simulation be called AI? Must it replicate all aspects of intelligence?

John McCarthy, the computer scientist who originally coined the term AI, did not limit intelligence in AI to a replication of human intelligence but argued that machines could display other intelligences that involve "much more computing than people can do." [27] He defined AI as "the science and engineering of making intelligent machines, especially intelligent computer programs." [28] Steven Omohundro, a leading scientist in the field, adopted an external agency requirement and defined AI as a system that "has goals which it tries to accomplish by acting in the world." [29] Stuart Russell and Peter Norvig, the writers of a popular university textbook on AI, summarized eight definitions of AI differentiated by how they reflected expectations of human thinking and behavior, or rational (machine) thinking and behavior. [30] Ultimately, Russell and Norvig preferred the rational agent approach in which machine agents "operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue [the best expected outcome]." [31] To be able to display these

---

[25] See, e.g., Ray Kurzweil, The Singularity Is Near: When Humans Transcend Biology 206, 222 (Rick Kot ed., 2005).

[26] Scherer, supra note 22, at 360. Consciousness on its own has proved notoriously difficult to define, a difficulty amplified when attempting to define artificial consciousness. See Gerald M. Edelman, The Remembered Present: A Biological Theory of Consciousness 4-5 (1990); Francis Crick & J. Clark, The Astonishing Hypothesis, 1 J. Conscious Stud. 10, 10-16 (1994); Francis Crick & Christof Koch, Towards a Neurobiological Theory of Consciousness, 2 Seminars Neurosciences 263, 263-75 (1990); Stanislas Dehaene & Jean-Pierre Changeux, Experimental and Theoretical Approaches to Conscious Processing, 70 Neuron 200, 200 (2011); Christof Koch et al., Neural Correlates of Consciousness: Progress and Problems, 17 Nature Revs. Neuroscience 307, 307 (2016); Steve Torrance, Ethics and Consciousness in Artificial Agents, 22 Ai & Soc'y 495, 497-98 (2008); Paul F. M. J. Verschure, Synthetic Consciousness: The Distributed Adaptive Control Perspective, 371 Phil. Transactions Royal Soc'y B, Aug. 2016, at 1-2, http://rstb.royalsocietypublishing.org/content/royptb/371/1701/ 20150448.full.pdf [https://perma.cc/58EQ-U3ER]; Wendell Wallach, Colin Allen & Stan Franklin, Consciousness and Ethics: Artificially Conscious Moral Agents, 3 Int'l J. Machine Consciousness 177, 177 (2011).

[27] John McCarthy, What Is Artificial Intelligence?, Stan. U. (Nov. 12, 2007), http://www-formal.stanford.edu/jmc/whatisai/whatisai.html [https://perma.cc/N5YZ-QYS7].

[28] Id.

[29] Stephen M. Omohundro, The Basic AI Drives, in Artificial General Intelligence 2008: Proceedings of the First AGI Conference 483, 483 (Pei Wang et al. eds., 2008).

[30] Russell & Norvig, supra note 24, at 1-5.

[31] Id. at 4. This combination of perception, adaptability, creativity, and autonomous operation reflects what would be required of an agent to pass the Turing test. Id. at 2-3.

characteristics, AI also needs to be actuated in machinery whether that is a computer system or a robot. Typically, though, these machine behaviors have been compared against human abilities to process language, to reason, and to perceive and manipulate objects in the environment to attain predetermined goals. [32]

All of these definitions set a fairly high bar for an algorithm to attain before it meets the definition of AI. AI can therefore be differentiated from machine learning systems, and even from machine [*395] learning that learns from examining large data sets, sometimes using neural networks to make deep connections among the data. If these computations do not display the other characteristics of AI, such as operating autonomously, adapting to change, and creating and pursuing their own goals, [33] then they cannot be AI. However, while they cannot be AI based on the definitions above, machine learning systems are also, perhaps erroneously, often referred to as possessing AI.

This lack of definitional clarity means that the broad label "AI" has become the vernacular term for a range of programs, algorithms, and networks that are used in a multitude of applications. For example, AI is used to refer not only to the programs underlying chess-and other game-playing programs, as well as Roomba vacuum cleaners, [34] but also to the coordinated systems controlling autonomous vehicles [35] and the personal agents developed by Microsoft, Apple, and Google, among others. [36] Some of these uses of the term AI are differentiated by descriptors such as "narrow AI" to distinguish their limited application to a single set task. [37] But when AI is developed so as to apply more broadly or with greater effectiveness, it is often referred to as becoming "stronger" [38] rather than "broader."

Complicating this definitional problem further, research by mathematicians and engineers who seek to develop self-replicating and self-aware algorithms is also said to be work "in AI." [39] There has been some attempt to distinguish this work from narrow or even [*396] stronger AI and algorithms that display these characteristics by referring to it as "strong AI." A more common reference is artificial general intelligence (AGI). As opposed to "narrow AI," AGI is said to possess "a reasonable degree of self-understanding and autonomous self-control, [has] the ability to solve a variety of complex problems in a variety of contexts, and [can] learn to solve new problems that [it] didn't know about at the time of [its] creation." [40] AGI is "subject to a variety of "drives' including self-protection, resource

---

[32] See id.

[33] See Liza Daly, AI Literacy: The Basics of Machine Learning, World Writable (Apr. 11, 2017), https://worldwritable.com/ai-literacy-the-basics-of-machine-learning-2e20f93e34b4 [https://perma.cc/P6UJ-FH4G].

[34] See Artificial General Intelligence vi (Ben Goertzel & Cassio Pennachin eds., 2007). These single-task applications are often classified as "narrow AI." See id.; see also Kurzweil, supra note 25, at 92. The bulk of AI research and development today is conducted into this narrow type of AI. See Cassio Pennachin & Ben Goertzel, Contemporary Approaches to Artificial General Intelligence, in Artificial General Intelligence, supra, at 1, 1.

[35] See Peter Stone et al., Artificial Intelligence and Life in 2030, at 18-19 (2016), https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf [https://perma.cc/ 5XZY-ATYS].

[36] See Jamie Condliffe, In 2016, AI Home Assistants Won Our Hearts, MIT Tech. Rev. (Dec. 20, 2016), https://www.technologyreview.com/s/603228/in-2016-ai-home-assistants-won-our-hearts/.

[37] See, e.g., Kai-Fu Lee, A Blueprint for Coexistence with Artificial Intelligence, Wired (July 12, 2017, 6:50 AM), https://www.wired.com/story/a-blueprint-for-coexistence-with-artificial-intelligence/.

[38] Kurzweil, supra note 25. This classification system refers to "narrow AI," as opposed to "strong AI." See id. Perhaps a clearer dichotomy would be to refer to "weak AI" and "strong AI," but this Article: retains the traditional classification.

[39] Laurent Orseau, Asymptotic Non-Learnability of Universal Agents with Computable Horizon Functions, 473 Theoretical Computer Sci. 149, 149 (2013).

[40] Artificial General Intelligence, supra note 34, at vi.

acquisition, replication, goal preservation, efficiency, and self-improvement." [41] It is generally recognized that AGI does not yet exist, but it is AGI that causes most concern to those who believe that AI creates an existential threat to humanity. The Authors discuss this further in Part II.C below.

The range of applications of AI sits on a spectrum from those applications that are not strictly AI, [42] through to narrow applications of AI (as found in chess games, etc.), to AGI. When referring to AI, then, the Authors must bear in mind this vast array of uses and misuses of the term. It is neither possible nor even desirable to govern all of these diverse uses of AI using one regulatory approach. However, the risks associated with these different applications of AI will arguably drive different regulatory responses and must therefore be treated differently. Accordingly, the Authors argue for a classification based on the risk that various AI applications pose. For public regulators that have limited resources and information, classifying AI can inform their decisions about which applications or class of AI to regulate first, and at what level.

B. Introducing Risk as a Defining Point

The Authors propose that risk should be considered as a quality that differentiates classes of AI. The Authors develop this idea further in Part III; here, however, the Authors argue that once applications of AI are classified according to the potential risk each poses to society or to the people or environment in which they are applied, then public regulators can more efficiently and effectively [*397] direct their regulatory responses. Without that knowledge, they will be grasping in the dark to even understand the regulatory problem. [43]

Even categorizing risk in relation to AI is complicated by a lack of clarity on AI's potential. On one hand is the pervasive fear that AI will develop rapidly to the point at which it will annihilate humans as a species, either through some miscalculation in replicating software or because humans are suboptimal to the machine's set goals. When talking about risk in relation to AI, it is these risks that linger just below the surface of each argument. On the other hand, others argue that the development of AI is benign and beneficial to society. However, these arguments may be at cross-purposes, the Authors argue, due to a lack of a sufficient and agreed-upon definition for AI. The Authors suggest that classifying AI based upon potential risk factors as suggested in this Article: may clarify some of these arguments so that regulation may be used where required to minimize risks, while at the same time allowing development of less risky AI with only minimal regulatory intervention. In this way, the Authors can avoid suggesting the same regulatory response to the AI in a Roomba, for example, as would be suggested to regulate autonomous weapons systems or the comparatively simple algorithms that regulate critical environmental or energy systems.

Part II.C discusses the arguments made in relation to the existential risks posed by some in relation to AI. As discussed, these arguments are often raised as reasons to regulate the development of AI, and must be addressed. Then, Part II.D outlines concrete examples of problems associated with current applications of AI in use today that the Authors argue also require a regulatory response - but for more concrete reasons.

---

[41] Steve Omohundro, Rational Artificial Intelligence for the Greater Good, in Singularity Hypotheses: A Scientific and Philosophical Assessment 161, 161 (Amnon H. Eden et al. eds., 2012).

[42] See, e.g., Adi Prakash, "Doing AI': What Legal Should Remember About Big Data, Legaltech news (July 12, 2017), http://m.legaltechnews.com/?slreturn=20170726234213/#/article/ 1202792798132/Doing-AI-What-Legal-Should-Remember-About-Big-Data?utm_ content=buffer01b0a&utm_medium=social&utm_ source=twitter.com&utm_campaign=buffer&_almReferrer=https:%2F%2Ft.co%2Fwki9DruH9A [https://perma.cc/63LL-XE2A].

[43] See Nat'l Sci. & Tech. Council, Exec. Office of the President, Preparing for the Future of Artificial Intelligence 39 (2016), https://obamawhitehouse.archives.gov/sites/default/files/ whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf [http://perma.cc/BHL5-ZKAB]. The US government has recognized this. See id. at 18 ("Agencies should draw on appropriate technical expertise at the senior level when setting regulatory policy for AI-enabled products. Effective regulation of AI-enabled products requires collaboration between agency leadership, staff knowledgeable about the existing regulatory framework and regulatory practices generally, and technical experts with knowledge of AI. Agency leadership should take steps to recruit the necessary technical talent, or identify it in existing agency staff, and should ensure that there are sufficient technical "seats at the table' in regulatory policy discussions." (emphasis omitted)).

C. Reports of the Singularity and the End of Humanity May Be Greatly Exaggerated

Perhaps the most visceral fear about the development of AI is the existential threat to humanity that is said will be caused by the [*398] rise of superintelligent machines. [44] These concerns pervade the collective consciousness in relation to AI. The Authors argue that this fear may be overstated given the current state of development in AI, but it is so ubiquitous that it informs every level of discussion. It is also addressed in every code of conduct, standard, or values statement that has been developed by those in the industry [45] and should be addressed in any regulatory intervention.

In 1965, Irving John Good proposed that society would be transformed by the invention of a machine with ultraintelligence. [46] It would surpass human intelligence and be able to design even more intelligent machines. [47] It would, he argued, be the last machine that humans would ever need to make for themselves [48] and would save humanity. [49] Good argued that this ex machina in human image would be designed from an understanding of human intellect. [50] Good's optimism was not shared by subsequent scholars, such as Vernor Vinge, who saw superintelligent machines not as saviors but as the advent of doomsday. [51] Vinge's concern was that once the machine attained human-level intelligence, it would not remain at that level for long and would reach superintelligence and beyond very quickly. [52] Vinge argued that such a machine could become aware of its own superior intelligence. [53] This event, which he described as the singularity, would spell the end of humanity. [54]

These fears are not new and are not confined to fears of AI. Age-old concerns in human mythology about humans playing "god the creator" form the basis of stories such as Frankenstein and other [*399] golem stories. [55] These stories have parallels with, and lessons for, the development of AGI. In the mythology, a golem is created, often from clay, and imbued with life through "a detailed statement of specific letter combinations that are required to bring about the "birth' of a golem" [56] - comparable to the algorithm in AI. In some golem stories, the golem obtains superhuman strength and, uncontrolled, causes destruction and mayhem. [57] The parallels to the creation of AGI with superhuman intelligence are apt. A further parallel might be drawn with the desire to regulate or control these

---

[44] See, e.g., Vernor Vinge, The Coming Technological Singularity: How to Survive in the Post-Human Era, in Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace 11, 12-14 (1993), *https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940022855.pdf* [*https://perma.cc/6UY3-C2RJ*]. Writing fifteen years after Vinge, Kurzweil appears most optimistic about the outcome of the singularity but maintains an element of caution. See Kurzweil, supra note 25, at 292-96; see also Nick Bostrom, Superintelligence: Paths, Dangers, Strategies 4-5, 279 (Keith Mansfield ed., 2014); Ray Kurzweil, Foreword: to John Von Neumann, The Computer & the Brain xi-xii (3d ed. 2012); Nick Bostrom, When Machines Outsmart Humans, 35 Futures 759, 763-64 (2003).

[45] See discussion infra Part IV.B.

[46] Irving John Good, Speculations Concerning the First Ultraintelligent Machine, 6 Advances in Computers 31, 31 (1966).

[47] Id. at 33.

[48] Id. at 31-32.

[49] Id. at 31.

[50] Id. at 78.

[51] Vinge, supra note 44, at 13.

[52] Id. at 14.

[53] Id. at 13.

[54] Id.

[55] See, e.g., Mary Shelley, Frankenstein (3d ed. 1831); Elie Wiesel, The Golem: The Story of a Legend (1983).

[56] Josepha Sherman, Storytelling: An Encyclopedia of Mythology and Folklore 204 (Josepha Sherman ed., 2008).

[57] See Wiesel, supra note 55, at 25.

fears. For example, golems were bound by Jewish law. [58] They were programmed not to kill unless necessary and could not lie. [59] This demonstrates the birth of the idea of embedding legal codes within technical code. [60]

Existential concerns have stimulated the minds of ethicists and philosophers since soon after work began on AI. [61] However, discussions about the legal ramifications of AI were typically slower to develop, and early considerations of AI and the law only appear in the early 1980s. [62] Even then, the dangers associated with the inability to understand and control AI were apparent. [63] This problem has not diminished and, if anything, has probably increased in the nearly forty years since 1981. Researchers in AI recognize that there is a potential risk that if autonomous AGI is developed, it will be difficult for a human operator to maintain control. [64]

Some of the risks seem remote or are, at this stage, only potential problems. But stories that portray the catastrophic consequences of autonomous, self-aware AI - such as those portrayed in science fiction, as well as the prophecies of researchers such as Omohundro - pervade the zeitgeist and have begun to induce a level of anxiety and fear that may well yet reach a tipping point in society's consciousness. [65] People can be particularly risk averse when they [*400] stand to lose something, [66] and governments respond to the desires and concerns of the societies they govern. One aim of the law is to predict what might go wrong and to design laws to prevent or avoid it.

It is characteristic of exponential growth that all the significant effects of the growth occur in the last short timeframe at the end of the growth set. AI has had a long gestation period. There have been many failed predictions about the imminent explosion of AI over the last sixty years, but, far from dissipating, the questions about AI's impact will only become more urgent as we draw nearer to the exponential inflection point and its growth takes a sudden and dramatic vertical trajectory. The question is whether society, after sixty years of growth, is now approaching that inflection point or is still in the slower gradual development phase. The answer must be that as society approaches the point where AI begins to develop more quickly, society should begin to prepare for and guide the development of AI in ways that will produce benefits while still avoiding existential threats as best possible. This should be the role of the law, but lawmaking processes are often criticized as being overly responsive or reactive rather than sufficiently proactive.

Those within the AI industry have already taken steps to counter concerns about AGI autonomously self-replicating out of human control. For example, Laurent Orseau and Stuart Armstrong, an engineer at DeepMind and a researcher into systemic risk, respectively, acknowledged that "reinforcement learning agents … are unlikely to

---

[58] Sherman, supra note 56, at 205.

[59] Id.

[60] See 1 Lawrence Lessig, Code: Version 2.0, at 1819 (2d ed. 2008).

[61] See Norbert Wiener, Cybernetics 16972 (2d ed. 1961).

[62] See Sam N. Lehman-Wilzig, Frankenstein Unbound: Towards a Legal Definition of Artificial Intelligence, 13 Futures 442, 443 (1981).

[63] Id. at 446; see Wiener, supra note 61, at 175.

[64] See Omohundro, supra note 41, at 172.

[65] Malcolm Gladwell identified the three characteristics that identify what he described as a "tipping point," particularly in epidemics, as "one, contagiousness; two, the fact that little causes can have big effects; and three, that change happens not gradually but at one dramatic moment[.]" Malcolm Gladwell, The Tipping Point: How Little Things Can Make a Big Difference 9 (ed. 2000).

[66] Daniel Kahneman & Amos Tversky, Prospect Theory: An Analysis of Decision Under Risk, 47 Econometrica 263, 279 (1979).

behave optimally all the time." [67] They recognized "concerns that a "superintelligent' agent may resist being shut down, because this would lead to a decrease of its expected reward," [68] and detailed how DeepMind's engineers have developed a "big red button," or an off switch for such an artificially intelligent reinforcement learning agent. Any regulation of AI might consider compulsory adoption of this program in all research and development into AGI.

However, not everyone shares these concerns. The panel that contributed to the Stanford Report on AI titled Artificial Intelligence and Life in 2030 noted that

[*401]

contrary to the more fantastic predictions for AI in the popular press, the Study Panel found no cause for concern that AI is an imminent threat to humankind. No machines with self-sustaining long-term goals and intent have been developed, nor are they likely to be developed in the near future. Instead, increasingly useful applications of AI, with potentially profound positive impacts on our society and economy are likely to emerge between now and 2030[.] [69]

While threats to humankind posed by AI may yet be some way off, it is important to listen to those in the industry who are calling for controls to be put in place now to prepare for the future. If AI ever does develop to a point where it becomes a threat to humanity, they argue, it may well be too late to do anything about it. Far from ignoring these fears and threats, any regulatory response to AI must address the risks AI poses in some manner. The more recent warnings of technology entrepreneurs like Elon Musk and scientists like Stephen Hawking about the risks of runaway AI should at least cause regulators to pause and consider whether they have appropriate risk identification and mitigation strategies in place. [70]

The call to regulate comes not only from deep human fears of the singularity but also because of more concrete problems associated with the narrow AI that currently exists, has already been implemented, and pervades our everyday lives. Part II.D analyzes some of these unforeseen problems that are occurring now in current applications of AI. These issues highlight the potential for unforeseen errors to occur. These types of demonstrable errors and unforeseen problems are the canary in the coalmine of AI development. They provide warning about how things can go wrong when society and governments allow AI systems to be developed and deployed without appropriate regulation in place. Any regulatory response needs to ensure that AI systems are designed and deployed so that they do not pose any harm (in its broadest sense) to people or society. [71]

[*402]

D. Problems Associated with Current Applications of AI

---

[67] Laurent Orseau & Stuart Armstrong, Safely Interruptible Agents, in Uncertainty in Artificial Intelligence: Proceedings of the Thirty-Second Conference 557, 557 (Alexander Ihler et al. eds., 2016), http://www.auai.org/uai2016/proceedings/uai-2016-proceedings.pdf [https://perma.cc/G3M2-MQDG].

[68] Id. at 558.

[69] Stone et al., supra note 35, at 4.

[70] See, e.g., An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence, Future of Life Inst., https://futureoflife.org/ai-open-letter/ [https://perma.cc/S549-9FV2] (last visited Oct. 30, 2017); see also Amitai Etzioni & Oren Etzioni, Keeping AI Legal, 19 Vand. J. Ent. & Tech. L. 133, 14546 (2016) (noting the concerns of "AI doomsayers" like Stephen Hawking and cautioning that "AI programs should be subject to continual oversight to ensure that their conduct does not stray from the boundaries set by human agents").

[71] See Crawford & Calo, supra note 10, at 313 (referring to this concern as the blind spot in AI research).

For the moment, the dystopian ramifications of rampant, uncontrollable AI are still the imaginings of science fiction writers. [72] The current challenge for regulating AI is the proliferation in the capabilities of relatively narrow AI systems tasked with performing specific functions. [73] Developments in AI technology have been smoldering since research on it began shortly after World War II. [74] Today, AI is at the forefront of technological development and is used in driverless vehicles, speech and facial recognition, language translation, lipreading, combatting spam and online payment fraud, detecting cancer, law enforcement, and logistics planning. Much of this AI is what can be described as narrow AI - that is, AI designed to solve a specific problem or familiar task, such as playing chess. These commercial applications of AI appear to be limitless, and the world's largest technology companies are investing heavily in its potential. For example, IBM's cognitive computing platform, Watson, has developed from its initial challenge of winning the game show Jeopardy! to being applied to provide real solutions to problems in commerce, law, and health. [75] DeepMind's AlphaGo recently defeated [*403] the human master of the complex Chinese board game Go, and Google also used DeepMind's AI to reduce the electricity consumption in Google's data centers. [76] Microsoft, meanwhile, has incorporated AI into its personal agents such as Cortana and Zo, which can perform a dizzying array of tasks and answer seemingly unlimited questions using a mellifluous (female by

---

[72] See, e.g., AI's Future Is Not So Scary, MIT Tech. Rev. (Nov. 9, 2016), *https://tribunecontentagency.com/article/ai039s-future-is-not-so-scary/* [*https://perma.cc/2FRN-SFFF*] ("The odds that artificial intelligence will enslave or eliminate humankind within the next decade or so are thankfully slim.").

[73] See Kurzweil, supra note 25, at 264, 289, 409 (explaining that there is an expectation that narrow AI will perform the task better or faster than human intelligence, given the AI's capacity to manage and consider vast arrays of data and variables); see also Ben Goertzel, Response, Human-Level Artificial General Intelligence and the Possibility of a Technological Singularity: A Reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's Critique of Kurzweil, 171 Artificial Intelligence 1161, 1162 (2007). Goertzel notes that the distinguishing features of narrow AI are that it does not understand itself, the task, or how to generalize or apply the knowledge it has learned in performing the task beyond the specific problem. For example, a narrow AI program for diagnosing one type of cancer would not itself be able to generalize its diagnostic insights to diagnose another type of cancer, though a human might be able to further develop the first AI for the subsequent purpose. Id. at 1162.

[74] McCarthy, supra note 27.

[75] AI and Cognitive Computing, IBM Res., *http://research.ibm.com/cognitive-computing/* [*https://perma.cc/BA2V-CKMC*] (last visited Oct. 30, 2017) (describing Watson as "the world's first and most-advanced AI platform"); see also Stephen Baker, Final Jeopardy: Man vs. Machine and the Quest to Know Everything (2011); Ryan Abbott, I Think, Therefore I Invent: Creative Computers and the Future of Patent Law,*57 B.C. L. Rev. 1079, 1088-91 (2016);* Betsy Cooper, Judges in Jeopardy!: Could IBM's Watson Beat Courts at Their Own Game, 121 Yale L.J. Forum 87 (2011); Jessica S. Allain, Comment, From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems, *73 La. L. Rev. 1049 (2013);* Shanna Carpenter, Video: IBM Insiders Break Down Watson's Jeopardy! Win, TED Blog (Feb. 18, 2011, 2:52 PM), *http://blog.ted.com/experts-and-ibm-insiders-break-down-watsons-jeopardy-win/* [*https://perma.cc/B7GW-MEZ2*]; IBM, IBM Watson: A System Designed for Answers, YouTube (Jan. 21, 2011), *https://www.youtube.com/watch?v=cU-AhmQ363I*; IBM, IBM Watson: How It Works, YouTube (Oct. 7, 2014), *https://www.youtube.com/watch?v=_Xcmh1LQB9I*. IBM is currently tasking Watson with learning how to help with the identification of melanoma and is seeking people's input to assist with timely, accurate detection. See Outthink Melanoma, IBM Austl., *https://www.ibm.com/cognitive/au-en/melanoma/* [*https://perma.cc/FT97-W2SS*] (last visited Oct. 30, 2017). Commercial applications of Watson include, for example, ROSS Intelligence's software marketed to lawyers as their "own personal artificially intelligent researcher … that effortlessly reads through and finds numerous answers for any legal question." ROSS Intelligence, Meet ROSS, Your Brand New Artificially Intelligent Lawyer, YouTube (Dec. 28, 2016), *https://www.youtube.com/watch?v=ZF0J_Q0AK0E*. ROSS can be asked questions in natural language, just as one might ask "any other lawyer." See Mark Gediman, Artificial Intelligence: Not Just Sci-Fi Anymore, 21 Am. Ass'n L. Libr. Spectrum, Sept.-Oct. 2016, at 34, 35-36; Paul Lippe, What We Know and Need to Know About Watson, Esq.,*67 S.C. L. Rev. 419, 420 (2016);* ROSS Intelligence, supra.

[76] See, e.g., DeepMind Collaborations with Google, supra note 4.

design) computer-generated voice. [77] Microsoft's algorithm DeeperCoder is capable of writing code to solve simple problems. [78] And Facebook uses AI in its face recognition, language translation, and camera effects, as well as in its research arm - Facebook Artificial Intelligence Research (FAIR) - which is said to be "committed to advancing the field of machine intelligence." [79] Joaquin Candela, Director of Engineering for Facebook's Applied Machine Learning (AML) group, has stated that Facebook is working towards a "generalization of AI" [80] that will, it is argued, be capable of enhancing the speed at which applications can be built by "a hundred-x magnitude," expanding possibilities for impact in fields ranging from medicine to transportation. [81] Advances in AI technology are vaulting toward the exponential as computer capacity and speed [*404] double every two years. [82] The Stanford Report predicts that as driverless cars fall into common use, they will form the first public impressions of AI in a corporeal form. [83] This experience will be an important one for AI since we are on the cusp of a surge of AI with a physical embodiment. [84] The Stanford Report also predicts that, by 2030, the typical North American city will feature personal robots, driverless trucks, and flying cars. [85]

These AI systems present a spectrum of immediate issues that may require a regulatory response. Some are likely to be dealt with by developers as they come to their attention, and end users of the system may deal with others as they refine their use of the system and work with developers in overcoming issues as and when they arise. This Section outlines several of the issues that may require a regulatory response, including biases that appear in law enforcement decisions made by AI systems; safety, particularly in relation to driverless cars; the lack of a human "heart" when relying on AI in judicial decision making; privacy in relation to a vast number of applications; and the pressing problems associated with unemployment caused by increasing rates of automation supported by AI.

1. Bias

The coalescing of AI and big data opens significant possibilities for the synthesis and analysis of that data, but it also stands to compound problems that presently exist in that process. These include unintended racism, sexism,

---

[77] Microsoft's AI Vision, Rooted in Research, Conversations, Microsoft News Ctr., https://news.microsoft.com/features/microsofts-ai-vision-rooted-in-research-conversations/ [https://perma.cc/T95Y-ADBC] (last visited Mar. 13, 2017).

[78] Dave Gershgorn, Microsoft's AI Is Learning to Write Code by Itself, Not Steal It, Quartz (Mar. 1, 2017), https://qz.com/920468/artificial-intelligence-created-by-microsoft-and-university-of-cambridge-is-learning-to-write-code-by- itself-not-steal-it/ [http://perma.cc/875M-MC6U].

[79] Facebook AI Research (FAIR), Facebook Res., https://research.fb.com/category/facebook-ai-research-fair (last visited Oct. 30, 2017).

[80] Steven Levy, Inside Facebook's AI Machine, Wired (Feb. 23, 2017, 12:00 AM), https://backchannel.com/inside-facebooks-ai-machine-7a869b922ea7 [https://perma.cc/C42W-BL39].

[81] Id.

[82] This is known as "Moore's Law," after the cofounder of Intel who predicted in 1965 that computing power would double every year (later revised to every two years). See Tom Simonite, Moore's Law Is Dead. Now What?, MIT Tech. Rev. (May 13, 2016), https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/. There is some speculation that this rate of change is no longer happening. See id.; see also Pedro Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World 287 (2015) (stating that Moore's Law "is on its last legs").

[83] Stone et al., supra note 35, at 18.

[84] Id. at 15-16 (noting advancements in robotics, computer vision, natural language processing, and collaborative systems that are required to embody AI or give it functionality).

[85] See id. at 18-23 (automated vehicles); id. at 24-25 (home robots); id. at 7, 18, 20 (flying vehicles).

and discrimination in the outcomes of data analysis. [86] Ifeoma Ajunwa, Kate Crawford, and Joel [*405] S. Ford have proposed a model to regulate big data both to address privacy concerns and to allow a pathway to correct erroneous assumptions made from an assemblage of that data. [87] Bias can be difficult to detect and, if care is not taken, can "become part of the logic of everyday algorithmic systems." [88] These biases have arisen in a law enforcement context: algorithms performing predictive risk assessments of defendants committing future crimes were making errors with risk scores for black defendants, giving them high risk scores at almost double the rate of white defendants. [89] Conversely, risk scores were erroneously low for white defendants. [90] Bias also arises in the work of private platforms that filter, index, and sort online content and mediate communications. [91] Crawford sees at least some of this as a manifestation of a bias problem with data and calls for vigilance in AI system design and training to avoid built-in bias. [92] Bias issues such as these are unlikely to provoke a regulatory response if they are dealt with in AI system design. However, these issues can be ameliorated with regulation that requires either careful design or prompt troubleshooting when the issues are identified.

2. Safety

AI is being touted as a solution to a number of social problems. However, when it is implemented in a social context, it also presents a range of safety issues. [93] For example, autonomous vehicles such as [*406] cars and trucks have the potential to improve safety on roads if they succeed in reducing accidents caused by driver error such as inattention, impairment, slow reaction times, and inappropriate risk-taking. [94] Social benefits potentially

---

[86] Kate Crawford, Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics, 41 Sci. Tech. & Hum. Values 77, 82-83 (2016) [hereinafter Crawford, Can an Algorithm Be Agonistic?]; Kate Crawford, Artificial Intelligence's White Guy Problem, N.Y. Times (June 25, 2016) [hereinafter Crawford, Artificial Intelligence's White Guy Problem], https://www.nytimes.com/2016/06/26/opinion/sunday/artificial -intelligences-white-guy-problem.html [http://perma.cc/B4H2-B5G6]; Yoni Har Carmel & Tammy Harel Ben-Shahar, Reshaping Ability Grouping Through Big Data,20 Vand. J. Ent. & Tech. L. 87, 110 (2017) (discussing how historical inequalities can cause algorithms to select for similar biases in future datasets).

[87] Ifeoma Ajunwa, Kate Crawford & Joel S. Ford, Health and Big Data: An Ethical Framework for Health Information Collection by Corporate Wellness Programs, 44 J.L. Med. & Ethics 474, 476 (2016).

[88] Crawford, Artificial Intelligence's White Guy Problem, supra note 86.

[89] Julia Angwin et al., Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks., ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [http://perma.cc/3XDB-7FWX].

[90] Id.

[91] Tarleton Gillespie, The Relevance of Algorithms, in Media Technologies: Essays: on Communication, Materiality, and Society 167, 188 (Tarleton Gillespie et al. eds., 2014); Nicolas Suzor, Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms 12-13 (Sept. 2016) (unpublished symposium paper), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2909889 [https://perma.cc/JP7N-FRBL].

[92] Crawford, Artificial Intelligence's White Guy Problem, supra note 86; Dark Days: AI and the Rise of Fascism, SXSW: Schedule, http://schedule.sxsw.com/2017/events/PP93821 [https://perma.cc/999V-FZ36] (last visited Oct. 30, 2017).

[93] Patrick Lin, Keith Abney & George Bekey, Robot Ethics: Mapping the Issues for a Mechanized World, 175 Artificial Intelligence 942, 945-46 (2011); Drew Simshaw et al., Regulating Healthcare Robots: Maximizing Opportunities While Minimizing Risks, ***22 Rich. J.L. & Tech. 1, 9 (2015).*** See generally Eliezer Yudkowsky, Cognitive Biases Potentially Affecting Judgment of Global Risks, in Global Catastrophic Risks 91 (Nick Bostrom & Milan M. Cirkovic eds., 2008) (remarking that the most powerful and beneficial technologies also exhibit the greatest potential risks to society).

[94] See Stone et al., supra note 35, at 19-21.

include improving mobility for those unable to drive or those who live in heavily traffic-congested urban areas. [95] Hence, there is an urgency to deploy autonomous vehicles, and developers have already been testing autonomous vehicles on public roads. Indeed, the authors of the Stanford Report expect that "transportation is likely to be one of the first domains in which the general public will be asked to trust the reliability and safety of an AI system for a critical task." [96]

However, the safety risks present in autonomous vehicles include the risk of accidents that may not otherwise have occurred; accidents created by even minor software or hardware errors; flawed or deficient programming of software; or unethical decision-making in the face of a high-risk, multi-risk scenario. [97] Regulation is key to providing an environment that will give the technology a chance to develop to its full potential while protecting the public from unacceptable risks. [98] Public regulators are already developing regulatory frameworks for safety assurance during the development [*407] and testing phases. [99] These frameworks extend to design standards, vehicle modification, and the development of safety principles, criteria, and assurance standards that are efficient, affordable, and create a minimal "administrative burden." [100]

The success of AI in solving social problems will ultimately lie in public and regulatory confidence in its use, and much of this confidence will turn upon trust in safety assurance. [101] Safety, in this sense, ought not to be confined to physical safety but should extend to concern for nonphysical harm, [102] such as privacy, security, and the

---

[95] Id. at 18; see James Manyika et al., Disruptive Technologies: Advances That Will Transform Life, Business, and the Global Economy 78-83 (2013) (identifying government regulation as potentially both an enabler and barrier to the socioeconomic benefits of autonomous vehicles).

[96] Stone et al., supra note 35, at 18.

[97] Lin, Abney & Bekey, supra note 93, at 945. Programming issues may be highly specific and unique to certain cultures, geographical terrain, or indigenous fauna. See, for example, reports that Volvo is working on difficulties arising with the animal detection system in its autonomous vehicles when confronted with the unusual way in which kangaroos move. Jake Evans, Driverless Cars: Kangaroos Throwing off Animal Detection Software, ABC News Austl. (June 23, 2017, 5:28 PM), http://www.abc.net.au/news/2017-06-24/driverless-cars-in-australia-face-challenge-of-roo-problem/8574816 [http://perma.cc/PJ6S-ZVG4]. The system had previously been tested on moose in Sweden. Id.

[98] Upon the introduction of the Federal Automated Vehicles Policy in the United States, President Obama noted: "The quickest way to slam the brakes on innovation is for the public to lose confidence in the safety of new technologies." Barack Obama, Opinion, Barack Obama: Self-Driving, Yes, But Also Safe, Pittsburgh Post-Gazette (Sept. 19, 2016, 7:00 PM), http://www.post-gazette.com/opinion/ Op-Ed/2016/09/19/Barack-Obama-Self-driving-yes-but-also-safe/stories/ 201609200027; see U.S. Dep't of Transp., Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety 6 (2016).

[99] See, e.g., Strassenverkehrsgesetz [SVG] [Road Traffic Act], Dec. 19, 1958, SR 741.01 (Switz.); U.S. Dep't of Transp., supra note 98, at 7; Nat'l Transport Comm'n (Austl.), National Guidelines for Automated Vehicle Trials (2016); Pilot Project - Automated Vehicles, O. Reg. 306/15 (Can.); Dep't for Transport (UK), The Pathway to Driverless Cars: A Code of Practice for Testing (2015). Articles: 8 and 39 of the UN Convention on Road Traffic were amended to facilitate use of autonomous vehicles on public roads while ensuring the driver of the vehicle maintained her position in a superior role. Inland Transp. Comm., Rep. of the Sixty-Eighth Session of the Working Party on Road Traffic Safety, U.N. Doc. ECE/trans/WP.1/145, at 9-10 (2014) (amending the Vienna Convention on Road Traffic, Nov. 8, 1968, 1042 U.N.T.S. 17, 24, 43). The justifications for the amendment to the Convention on Road Traffic are included as an appendix to that document. See id. at 11.

[100] Nat'l Transport Comm'n Austl., supra note 99, at 34; Current Projects: Automated Vehicle Trial Guidelines, Nat'l Transport Commission (Austl.), http://www.ntc.gov.au/current-projects/ automated-vehicle-trial-guidelines/ [http://perma.cc/ZN4W-WZHM] (last updated May 31, 2017).

[101] Stone et al., supra note 35, at 35-36; see Simshaw et al., supra note 93, at 8-10.

[102] See David D. Luxton, Susan Leigh Anderson & Michael Anderson, Ethical Issues and Artificial Intelligence Technologies in Behavioral and Mental Health Care, in Artificial Intelligence in Behavioral and Mental Health Care 255, 255 (David D. Luxton ed., 2016); Simshaw et al., supra note 93, at 8-10; Bernd Carsten Stahl & Mark Coeckelbergh, Ethics of Healthcare Robotics: Towards Responsible Research and Innovation, 86 Robotics & Autonomous Sys. 152, 154 (2016).

dehumanization of care for people at their most vulnerable. [103] For example, the benefit of AI-enabled healthcare robots could be impeded by lack of regulation to assure public trust and confidence across a range of safety issues including these types of nonphysical harm.

These risks are most acute with personal care robots. Trust and confidence in AI-assisted robots may be hard-won in personal care situations, given that they have traditionally involved human-to-human interaction. [104] Also, to be effective and efficient, personal care robots must be able to access personal and medical information, "know[] … and possibly share the location of [*408] medication, objects, and people," [105] connect with hospital or other healthcare networks, and connect with networked technology such as personal devices including phones, wearable devices, or mobile applications. [106] The unprecedented amount of personal and medical information that could potentially be accessed, used, processed, and stored by personal care robots is vulnerable to the same privacy and security concerns raised in relation to the Internet of Things. [107] Aside from these security and privacy concerns, the healthcare context may raise unique safety concerns, for example, if an external party can hack medical devices such as pacemakers. [108] These risks escalate with unsophisticated home users. [109] As a result, society should give careful consideration to regulation that can address these concerns.

3. Legal Decision-Making

AI has been applied in highly specific legal tasks such as sentencing and judicial interpretation in an effort to improve transparency and consistency in judicial decisions. [110] However, these systems have been criticized as lacking capacity to exercise discretion and make situational value judgments. [111] Concerns have been raised about mechanistic reliance upon these applications of AI and their [*409] capacity to influence and shape the behavior of people involved in the decision-making process. [112]

---

[103] Healthcare robots include surgical, routine-task, and personal care robots. See Luxton, Anderson & Anderson, supra note 102, at 255; Simshaw et al., supra note 93, at 9-10; Stahl & Coeckelbergh, supra note 102, at 154, 157.

[104] Laurel D. Riek, Robotics Technology in Mental Health Care, in Artificial Intelligence in Behavioral and Mental Health Care, supra note 102, at 185, 194.

[105] Simshaw et al., supra note 93, at 11-12.

[106] Id. at 13-15.

[107] See Lin, Abney & Bekey, supra note 93, at 945; Simshaw et al., supra note 93, at 2. Similarly, complex safety issues arise with the noncommercial or recreational use of drones. See generally Roger Clarke & Lyria Bennett Moses, The Regulation of Civilian Drones' Impacts on Public Safety, 30 Computer L. & Security Rev. 263 (2014).

[108] See David D. Luxton et al., Intelligent Mobile, Wearable, and Ambient Technologies for Behavioral Health Care, in Artificial Intelligence in Behavioral and Mental Health Care, supra note 102, at 137, 156; Simshaw et al., supra note 93, at 22.

[109] Simshaw et al., supra note 93, at 22.

[110] See Trevor Bench-Capon & Henry Prakken, Argumentation, in Information Technology and Lawyers 61, 62 (Arno R. Lodder & Anja Oskamp eds., 2006); Maria Jean J. Hall et al., Supporting Discretionary Decision-Making with Information Technology: A Case Study in the Criminal Sentencing Jurisdiction, 2 U. Ottawa L. & Tech. J. 1, 31 (2005), *http://www.uoltj.ca/articles/vol2.1/2005.2.1.uoltj.Hall.1-36.pdf* [*https://perma.cc/6SRU-VNKD*].

[111] See Uri J. Schild, Expert Systems and Case Law 121 (1992); Hall et al., supra note 110, at 8-9; Philip Leith, The Judge and the Computer: How Best "Decision Support'?, 6 Artificial Intelligence & L. 289, 294-96 (1998); Paul Lippe, Daniel Martin Katz & Dan Jackson, Legal by Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law, *93 Or. L. Rev. 833, 849 (2015);* Brian Simpson, Algorithms or Advocacy: Does the Legal Profession Have a Future in a Digital World?, 25 Info. & Comm. Tech. L. 50, 56 (2016); John Zeleznikow, Building Decision Support Systems in Discretionary Legal Domains, 14 Int'l Rev. L. Computers & Tech. 341, 343 (2000).

[112] See Hall et al., supra note 110, at 33; Anja Oskamp & Maaike W. Tragter, Automated Legal Decision Systems in Practice: The Mirror of Reality, 5 Artificial Intelligence & L. 291, 293 (1997); Abdul Paliwala, Rediscovering Artificial Intelligence and Law:

Decision-making in the application of legal principles necessarily involves discretion. Decision-making in sentencing relies on "induction and intuition as well as the capacity to assess the social impact of decisions." [113] These have not yet proven to be among AI's greatest strengths. There is a significant body of scholarship that argues against using AI in making definitive legal decisions [114] and cautions against even a narrowly limited role for AI in informing human decisions. [115] As Brian Simpson argued, even if AI is able to approximate human discretion in sentencing decision-making, the question that remains is the extent to which "an algorithm [can] have a heart." [116] Simpson questions whether "such algorithms [can] deal with the unexpected, quirky[,] or unique individual that may require appeals to a sense of justice[.]" [117] Paul Lippe, Daniel Katz, and Dan Jackson propose that an optimal combination of AI and humans is required to provide balance. [118]

These concerns animate Article: 22 of the European Union's General Data Protection Regulation, which creates a new right for individuals "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." [119] The implication, at least in Europe, is that beginning in 2018 a human must somehow be involved in making the decisions. How effective this [*410] is likely to be remains to be seen. Public regulators in all jurisdictions similarly ought to consider the risks of allowing the involvement of AI in making automated final legal decisions.

Even where the decision is not automated but AI is used to support human decision-making, public regulators ought to be wary of undesirable risks and consequences. Reliance upon AI systems in judicial decision-making enlivens long-standing fears that reducing human processes to their most mechanistic may have an unintended regulatory effect. [120] That is, once a process is reduced to its most mechanistic, it may make the humans involved in the decision-making process more compliant or programmable to the process. [121] Even where the goal and purpose of involving AI in legal decision-making is to increase consistency, there are still risks that it will lead to standardization, [122] which in automated legal decision-making processes can have a regulatory effect on the people involved. [123] This regulatory impact may extend to an unintended chilling effect on individualization even where the legislature intended there to be some flexibility. [124] People involved in the decision-making process may

---

An Inadequate Jurisprudence?, 30 Int'l Rev. L. Computers & Tech. 107, 112-13 (2016); Steven P. R. Rose & Hilary Rose, "Do Not Adjust Your Mind, There Is a Fault in Reality' - Ideology in Neurobiology, 2 Cognition 479, 498-99 (1973); Simpson, supra note 111, at 56.

[113] Hall et al., supra note 110, at 9.

[114] See Cooper, supra note 75, at 97-99; Leith, supra note 111; Philip Leith, The Emperor's New Expert System, 50 Mod. L. Rev. 128, 128-32 (1987); Philip Leith, The Rise and Fall of the Legal Expert System, 1 Eur. J.L. & Tech. (2010), http://ejlt.org/article/view/14/1 [https://perma.cc/B4K9-ATNE]; Cass R. Sunstein, Of Artificial Intelligence and Legal Reasoning,***8 U. Chi. L. Sch. Roundtable 29, 34-35 (2001).*** See generally John O. McGinnis & Russell G. Pearce, The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services, *82 Fordham L. Rev. 3041 (2014).*

[115] See Schild, supra note 111, at 121; Lippe, Katz & Jackson, supra note 111, at 4, 13, 20; Paliwala, supra note 112, at 112-13; Zeleznikow, supra note 111, at 343.

[116] Simpson, supra note 111, at 56.

[117] Id.

[118] See Lippe, Katz & Jackson, supra note 111, at 849.

[119] Commission Regulation 2016/679, art. 22(1), 2016 O.J. (L 119) 46 (EU), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 [http://perma.cc/L3NR-U4NU].

[120] Rose & Rose, supra note 112, at 498-99.

[121] Id.

[122] Hall et al., supra note 110, at 33.

[123] Oskamp & Tragter, supra note 112, at 293.

[124] See id.

have difficulty deviating from the standardization in order to, for example, "have a heart," [125] "introduce an element of humanity in special circumstances," [126] or consider whether the decision is in the best interests of society. [127]

The array of concerns surrounding the use of AI systems in judicial decision-making is likely to be managed by the continual refinement of how AI systems are deployed by people in the decision-making process and should ultimately be regulated.

4. Privacy

The leaps in advancement that are the promise of AI will sometimes turn on the quality and quantity of information available to it to inform AI learning. Public regulators will need to regulate to protect the privacy of individuals if large data sets are disclosed to tech companies with AI capabilities. For example, maintaining patient privacy should be paramount where data sets held by public health services are shared with technology companies. This should be [*411] so even where data is disclosed for a specific purpose and is technically compliant with current regulatory disclosure models. Even so, sensitivities surrounding well-intentioned disclosures should result in a regulatory response, even where the disclosure technically complies with existing regulatory processes. [128] Such a regulatory response may result from the existing regulatory compliance process failing to contemplate the scale of the disclosure, the use to which the data is put by AI systems, or how the data might be used and stored by private entities not previously considered an interested stakeholder in that type of data at the time the regulatory process was settled. [129] Such a regulatory response may involve the imposition of a command and control model heavily restricting future access to such data sets.

5. Unemployment

The socioeconomic and sociopolitical impact of AI is a serious risk for public regulators. The deployment of AI in workplaces via algorithms, robotics, or automation targeting increased speed, [*412] efficiency, or safety is

---

[125] Simpson, supra note 111, at 56.

[126] Hall et al., supra note 110, at 33.

[127] See Oskamp & Tragter, supra note 112, at 292; Paliwala, supra note 112, at 112-13.

[128] See, for example, the debate surrounding the disclosure of private health data of an estimated 1.3 million UK patients in a collaboration between DeepMind and the Royal Free London NHS Foundation Trust in the United Kingdom. See Suleyman, supra note 6; see also Google DeepMind: Q&A, Royal Free London NHS Found. Tr. (May 4, 2016), https://www.royalfree.nhs.uk/news-media/news/google-deepmind-qa/ [https://perma.cc/F4VS-CDB9]. DeepMind has provided information about its Independent Reviewers involved in the NHS project. DeepMind Health's Independent Review Panel, DeepMind, https://deepmind.com/applied/deepmind-health/transparency-independent-reviewers/independent-reviewers/ [https://perma.cc/S8YS-N5D5] (last visited Oct. 31, 2017). The relevant statute in the United Kingdom applicable to the disclosure of this type of data is the Data Protection Act 1998, legislation primarily regulated by the Information Commissioner's Office (ICO). See Data Protection Act 1998, c. 29, sch. 5 (Eng.). The ICO, with the assistance of the National Data Guardian, is currently reviewing these disclosures for compliance with all appropriate regulatory processes for exchange of this data. DeepMind and the Royal Free London NHS Foundation Trust have stated their belief that they satisfied all appropriate regulatory processes for these data exchanges. See DeepMind Health's Independent Review Panel, supra; Google DeepMind: Q&A, supra. The National Data Guardian has reportedly completed its report for the ICO. See Info. Comm'r's Office, ENF0605979, Data Protection Act 1998 Undertaking Follow-Up (2017), https://ico.org.uk/media/action-weve-taken/undertakings/2013395/ hscic-nhs-digital-undertaking-follow-up.pdf [https://perma.cc/ NE3N-BURL]; Jane Wakefield, Google DeepMind's NHS Deal Under Scrutiny, BBC News (Mar. 17, 2017),http://www.bbc.com/news/technology-39301901 [http://perma.cc/8CXT-PANY]. The debate surrounding this disclosure is explored in Julia Powles & Hal Hodson, Google DeepMind and Healthcare in an Age of Algorithms, Health Tech., Jan. 2016, at 1-17, http://link.springer.com/10.1007/s12553-017-0179-1 [https://perma.cc/RS8E-2RH7].

[129] See Sam Shead, The UK Data Regulator Has Ruled That Google DeepMind's First Deal with the NHS Was Illegal, Bus. Insider Austl. (July 3, 2017, 9:48 PM), https://www.businessinsider.com.au/ico-deepmind-first-nhs-deal-illegal-2017-6 [https://perma.cc/5DYH-ATAH].

expected to radically change the workforce. [130] These concerns speak to a fundamental issue beyond the economics of increased productivity. The sheer scale of the disruptive impact on wages and employment is unlikely to be matched by increased productivity and may instead "exacerbate inequality rather than promote greater opportunity and shared prosperity." [131] Regulators must consider issues such as the benefits that society can attain from AI and how regulators can support workers through the expected job displacement if the scale of that displacement is anything approaching the levels anticipated. Public regulators need to consider the socioeconomic and sociopolitical disequilibrium that might result if the AI revolution causes widespread unemployment. Ultimately, regulators must consider if society will require a living wage paid for by taxes on robots. [132]

Adverse impacts on employment will not be confined to manufacturing or blue-collar work where robots are already used. [133] While unskilled routine tasks that lend themselves to automation are at high risk, jobs that are highly skilled involving high levels of dexterity, creativity, social intelligence, collaboration, negotiation, and problem solving will also be at risk with further advances in technology. [134] Every robot introduced into the workplace is estimated [*413] to have a sizable impact on wages and employment rates. [135] As the use of robots in workplaces increases, the aggregate effect on employment and wages is expected to increase, as well. [136]

The pace of change and the sheer extent of displacement caused by the effects of automation, robots, and AI on wages and employment will be unprecedented. Workers will be marginalized and forced to "upskill" to find work. [137] The World Bank has cautioned that public regulators are "in a race between skills and technology," and for many skills, "people are losing the race." [138] At least part of the answer is to reform education and training, but as the World Bank has observed, these types of reforms have such a long lag time until they can prove effective that targeted educational reforms must begin at a young age. [139] Therefore, a regulatory response needs to consider

---

[130] See Carl Benedikt Frey & Michael A. Osborne, The Future of Employment: How Susceptible Are Jobs to Computerisation?, 114 Tech. Forecasting & Soc. Change 254, 261 (2017). McKinsey's occupational study estimates that 51 percent of US jobs ($ 2.7 trillion of wages) could be automated by 2055, or decades earlier depending on the pace of technological development. See James Manyika et al., McKinsey Glob. Inst., A Future That Works: Automation, Employment, and Productivity 6, 12 (2017) , *http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works* [*https://perma.cc/4VXX-N2Y4*]. The World Bank estimates that 57 percent of jobs in the OECD nations could be displaced. See World Bank Grp., World Development Report 2016: Digital Dividends 129 fig.2.24 (2016),*http://www.worldbank.org/en/publication/wdr2016* [*https://perma.cc/E76Q-QMAY*].

[131] World Bank Grp., supra note 130, at 249 (emphasis omitted); see Daron Acemoglu & Pascual Restrepo, Robots and Jobs: Evidence from US Labor Markets 7-10 (MIT Dep't of Econ., Working Paper No. 17-04, 2017), *https://papers.ssrn.com/abstract=2940245*.

[132] Compare Kevin J. Delaney, The Robot That Takes Your Job Should Pay Taxes, Says Bill Gates, Quartz (Feb. 17, 2017), *https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/* [*https://perma.cc/ZS2Q-EPLR*] (noting Bill Gates's belief "that governments should tax companies' use of [robots] … to fund other types of employment"), with Why Taxing Robots Is Not a Good Idea, Economist (Feb. 25, 2017), *https://www.economist.com/news/finance-and-economics/21717374-bill-gatess-proposal-revealing-about-challenge-automation-poses-why-taxing* [*https://perma.cc/5F3J-UUUB*] (criticizing Gates's proposal because, inter alia, increasing the expense of robotic labor "might further delay an already overdue productivity boom").

[133] See Acemoglu & Restrepo, supra note 131, at 1, 5.

[134] Manyika et al., supra note 130, at 49-50; see World Bank Grp., supra note 130, at 125-26; Frey & Osborne, supra note 130, at 255-58.

[135] Acemoglu & Restrepo, supra note 131, at 36 (observing that this impact will only be marginally correlated with the more usual effects of imports, other technologies, and the natural attrition of "routine jobs").

[136] Id. at 35-36.

[137] World Bank Grp., supra note 130, at 130.

[138] Id. at 131.

[139] Id.

support for education, training, and transitioning displaced workers through the process of job disruption and reemployment. [140] Public regulators ought to influence education and training agendas now to ensure the development of resilient, transferrable, and not easily automated skills that lend themselves to a lifetime of working with and adapting to technological change. [141] Longer working lifetimes and the pace of technological development may see low-skilled workers experience this type of job disruption more than once.

This Part outlined examples of problems associated with current applications of AI systems that will provoke a regulatory response. The examples provided illustrate concrete problems and the possibility of far greater, even existential, problems if the development of AI is left unattended. As set out in the next Part, regulating AI systems is an extremely difficult problem to solve. Formulating the regulatory response will be a challenging one for any regulator. As specific problems manifest, fear, anxiety, or populist concerns - whether evidence-based or not - may create an urge in the regulator to step in. However, the Authors argue for a considered, principled, and consultative approach.

[*414]

III. The Difficulty in Regulating Artificial Intelligence

Even in the simplest of industries, "regulation is extraordinarily difficult." [142] When considering the regulation of new technologies, former justice of the High Court of Australia Michael Kirby noted that "the normal organs of legal regulation often appear powerless." [143] Further along that continuum, regulating the development of AI may be the hardest task yet for regulators to tackle.

Regulation is often implemented as a means to avoid or limit risks to human health or safety, to the environment, or against some moral hazard such as gene manipulation. [144] However, the real risks of AI may yet be unknown and are perhaps unknowable. This necessarily makes them difficult to evaluate for the purposes of risk assessment, which involves balancing a range of social attitudes and will often reflect the culture and values of the society in which it is deployed. However, it is clear that the variety of applications of AI in operation today poses a range of risks.

A. The Range of Risks Associated with AI and AGI

Part II outlined a range or spectrum of classes of AI - from narrow AI through to AGI. However, the level of risk associated with the applications within each class does not directly correlate to the class. The applications of AI within a single class could pose a range of risks from low to moderate to high. Further, an application of AI in the narrow class may have the potential to become stronger as the AI learns or develops. Whether that AI could then develop into AGI and thus pose a greater risk is often unknowable. Still further, the type of risk posed by each application may not be the same within each class of AI. For example, with a particular application of AI, there might be a low risk to safety or to human life, but a high risk of a breach of privacy, or a high risk of causing unemployment. Therefore, it is too simplistic to merely take a class of AI such as narrow AI and to seek to regulate it based upon a presumed level of risk. An additional complicating factor is that similar types of application will be used differently in different industries or areas. For example, the same narrow AI application used in a product in the aviation industry may [*415] be applied to a different product in an agricultural setting. This will very likely

---

[140] Manyika et al., supra note 130, at 18.

[141] World Bank Grp., supra note 130, at 131.

[142] Bridget M. Hutter, A Risk Regulation Perspective on Regulatory Excellence, in Achieving Regulatory Excellence 101, 101 (Cary Coglianese ed., 2017).

[143] Michael Kirby, New Frontier: Regulating Technology by Law and "Code", in Regulating Technologies, supra note 14, at 367, 383.

[144] See Deryck Beyleveld & Roger Brownsword, Emerging Technologies, Extreme Uncertainty, and the Principle of Rational Precautionary Reasoning, 4 Law Innovation & Tech. 35, 35 (2012).

mean that different regulatory agencies will be required to regulate the same AI, but in different applications. Taking this complication one step further, the risk posed by the application's use in the agricultural setting may be lower than when the same AI is applied in the aviation industry. Therefore, the same AI application will have to be treated differently by two separate agencies.

Public regulators must become informed about the AI used in their field, assess the risks posed by the AI application as it is used in the industry in which they operate, and regulate appropriately. Earlier research has acknowledged that the reliability and fidelity of organizations involved ought to be evaluated based on factors including the intended use to which the technology might be put. [145] Armed with a deeper understanding of the industry and the intended use of the AI, stakeholders involved in informing the regulatory approach will be better placed to ask the right questions to assuage, or at least contextualize, their concerns about levels of risk. Iterative and cooperative involvement of all stakeholders, including public regulators, is key to avoiding the necessity to hastily adopted command and control regulatory action and its unintended consequences. [146] The Authors must therefore consider the type of risk that different classes and types of AI pose - starting with a look at the systemic nature of AI risk that exists even now.

B. Systemic Risk

Not all applications of AI will eventuate in a "singularity" scale event. [147] However, immediate systemic risk issues are present with [*416] existing AI applications. [148] Systemic risk is the embedded risk "to human health and the environment … in a larger context of social, financial and economic risks and opportunities." [149] Systemic risks exist in an atmosphere of uncertainty, and they are not restrained by sector, domain, or geography. [150] Assessed at its height, strong AI or AGI presents inherent systemic risk. [151] However, the integrated nature and

---

[145] See Phil Macnaghten & Jason Chilvers, Governing Risky Technologies, in Critical Risk Research: Practices, Politics and Ethics 99, 102 (Matthew Kearnes et al. eds., 2012); see also Jessica L. Carlo, Kalle Lyytinen & Richard J. Boland, Systemic Risk, Information Technology Artifacts, and High Reliability Organizations: A Case of Constructing a Radical Architecture, 4 Sprouts 57, 58 (2004), http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1144& context=icis2004 [https://perma.cc/HW36-87WC].

[146] See Macnaghten & Chilvers, supra note 145, at 100. Note: that Cass R. Sunstein outlined a number of paradoxes that can be brought about by inappropriate regulation: imposing stringent regulations may lead to the regulator's own administrators failing to act or refusing to enforce the regulations. Cass R. Sunstein, Paradoxes of the Regulatory State, 57 U. Chi. L. Rev. 407, 416 (1990). Further, regulations that "impose costs exclusively on new sources or entrants can be self-defeating." Id. at 417. By way of example, Sunstein noted that stringent regulation of nuclear facilities had "perpetuated the risks produced by coal, a significantly more dangerous power source." Id. at 418. Sunstein argued that these paradoxes (among others) must be borne in mind when introducing regulation. Id. at 413.

[147] Goertzel, supra note 73, at 1162. For a a more complete discussion of systemic risk, see Marjolein B.A. van Asselt & Ortwin Renn, Risk Governance, 14 J. Risk Res. 431, 436-38 (2011). Systemic risk has been studied in a technology context. See Carlo, Lyytinen & Boland, supra note 145, at 58. Numerous studies have been conducted into how the law should deal with unknown risks. See, e.g., Jaap Spier, Uncertainties and the State of the Art: A Legal Nightmare, 14 J. Risk Res. 501 (2011). Paradoxically, AI may be able to assist with the management of systemic risk. See Jerzy Balicki et al., Methods of Artificial Intelligence for Prediction and Prevention Crisis Situations in Banking Systems, in Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence 180, 181 (Jerzy Balicki ed., 2014).

[148] See Omohundro, supra note 29, at 483 (arguing that even a chess-playing robot will be "dangerous unless it is designed very carefully" because "without special precautions, it will resist being turned off, will try to break into other machines and make copies of itself, and will try to acquire resources without regard for anyone else's safety").

[149] Ortwin Renn & Andreas Klinke, Systemic Risks: A New Challenge for Risk Management, 5 Eur. Molecular Biology Org. Rep. S41, S41 (2004).

[150] van Asselt & Renn, supra note 147, at 436 (discussing and identifying these characteristics of systemic risk).

[151] See Omohundro, supra note 29 at 483. Seth D. Baum provides an example of a systemic risk as the fourteenth-century black plague in Venice, which was managed by the Venetians without knowledge or forethought of germ theory or microbiology. Seth

embeddedness of even narrow AI's deployment into complex, interdependent social, financial, and economic systems or networks amplifies the potential for risk impact, particularly where AI is deployed in a pervasive way. [152] The more complex and nonlinear these networks are, the easier it is for the impacts of an AI "incident" to proliferate rapidly throughout the network, affecting multiple stakeholders. [153] Systemic risks are problematic for regulation. While [*417] systemic risks are not unknown to public regulators, [154] the potential size and connectedness of the network that AI can access is unprecedented. For these reasons, it is unlikely that command and control models of regulation would be effective to regulate systemic risk. [155]

According to Marjolein van Asselt and Ortwin Renn, systemic risk should be managed via "a cautious and flexible strategy that enables learning from restricted errors, new knowledge, and visible effects, so that adaption, reversal, or adjustment of regulatory measures is possible." [156] Then the public regulator, business, and society can ensure that "early warning" systems are in place to detect risk if it eventuates. [157] Public regulators could initially develop agreed-upon principles that synthesize those things that need to be considered before formulating the processes necessary to govern those risks. [158]

In the regulation of AI, the mix and interplay of stakeholders will be important in the formulation of principles to regulate systemic risk, since it is nonstate stakeholders that are at an information advantage in understanding the

---

D. Baum, Risk and Resilience for Unknown, Unquantifiable, Systemic, and Unlikely/Catastrophic Threats, 35 Env't Sys. & Decisions 229, 231 (2015).

[152] van Asselt & Renn, supra note 147, at 436; see also Tero Karppi & Kate Crawford, Social Media, Financial Algorithms and the Hack Crash, 33 Theory Culture & Soc'y 73, 74, 77 (2016) (considering the connected nature of human communication and financial system algorithms and discussing how the eventual coalescence of big data and AI will compound this interconnectness of social systems). See generally Tomas Hellstrom, Systemic Innovation and Risk: Technology Assessment and the Challenge of Responsible Innovation, 25 Tech. Soc'y 369 (2003) (providing a full discussion of the systemic risks of networked technology).

[153] See Carlo, Lyytinen & Boland, supra note 145, at 59; van Asselt & Renn, supra note 147, at 436. Note: that Baum disagrees that AI (or aliens) could be considered a systemic risk since, if either risk were to eventuate and achieve world domination, humanity would have lost control of its system and be rendered incapable of managing it. Baum, supra note 151, at 234. Thus, any attempts to make systems more resilient to AI or alien invasion is misguided. Id. at 231. Baum's view of the systemic risks of AI is predicated on a vision of the systemic risk being the singularity or harbinger of doom. The Authors argue that this dismisses the systemic risk narrower AI systems might present. Notably, Baum suggests that since, in his view, AI is not a systemic threat, appropriate risk management is "not to increase resilience of affected systems but to reduce the probability of the systems being affected in the first place." Id. at 234.

[154] See Carlo, Lyytinen & Boland, supra note 145, at 59 (discussing management of risk hazards associated with nuclear facilities). Numerous studies have been conducted considering the resilience of infrastructure in the face of systemic risk from a number of eventualities. See, e.g., Jonathon Clarke et al., Resilience Evaluation and SOTA Summary Report: Realising European Resilience for Critical Infrastructure (2015); Baum, supra note 151; Seth D. Baum et al., Resilience to Global Food Supply Catastrophes, 35 Env't Sys. & Decisions 301 (2015); Daniel DiMase et al., Systems Engineering Framework for Cyber Physical Security and Resilience, 35 Env't Sys. & Decisions 291 (2015); Sabrina Larkin et al., Benchmarking Agency and Organizational Practices in Resilience Decision Making, 35 Env't Sys. & Decisions 185 (2015); Julie D. Rosati et al., Quantifying Coastal System Resilience for the US Army Corps of Engineers, 35 Env't Sys. & Decisions 196 (2015); Nicole R. Sikula et al., Risk Management Is Not Enough: A Conceptual Model for Resilience and Adaptation-Based Vulnerability Assessments, 35 Env't Sys. & Decisions 219 (2015).

[155] Neil Gunningham & Darren Sinclair, Smart Regulation, in Regulatory Theory: Foundations and applications 133, 142 (Peter Drahos ed., 2017).

[156] van Asselt & Renn, supra note 147, at 438.

[157] Id. at 438-39.

[158] Id. at 439 (suggesting communication and inclusion, integration, and reflection as principles to consider); see Ortwin Renn, Risk Governance: Coping With Uncertainty in a Complex World 63 (2008); Bridget Hutter, Risk, Regulation and Management, in Risk in Social Science 202, 214-15 (Peter Taylor-Gooby & Jens O. Zinn eds., 2006); see also Carlo, Lyytinen & Boland, supra note 145, at 59.

underlying matrix of science and technology in this area. The necessarily diverse mix of stakeholders and heterogeneous interests may make unified agreement on principles difficult. [159] Those charged with developing principles will need to consider not only the technological and scientific concerns but [*418] also a range of societal norms and social and economic considerations. [160] Settling on a set of principles will involve an element of trust in the science and technology. Creating a culture of iterative and cooperative development could engender this trust. Progress could be smoothed by a culture of fidelity and transparency from those with technical knowledge and scientific expertise in AI. Even if fuller information is available to public regulators, it will still be difficult to know everything necessary to regulate effectively because of the opacity of AI algorithms that are not transparent on their face and are said to reside in an impenetrable black box. [161]

C. The Risks of Failing to Regulate Must Be Evaluated Against the Risks Involved in Regulating AI

Some academics have proposed that society should merely adapt existing liability regimes to avoid legal uncertainty and to avoid the difficulties associated with regulating AI. [162] The common law has long adjusted to changes in technology iteratively, and, to a large extent, this incremental approach helps to minimize the risks of incorrect decisions in regulatory policy. [163] So, for example, a judicial process that adapts tort law principles to place liability for harm on the entity that is most effectively able to mitigate the risk - the "least cost avoider" - may adequately deal with concerns about potential harm caused by autonomous cars. Proponents of an iterative, "light touch" approach favor responding to concrete problems as they arise, either through incremental adjustments to the common law or careful, limited, and predominantly sui generis legislation if and as required. [164] The attractiveness of this approach is that it avoids the necessity of evaluating prospective risks - ensuring that regulation is targeted and limited to clear harms that courts and legislatures are [*419] able to understand. Those implementing and enforcing the laws could avoid much of the uncertainty surrounding new regulatory regimes. The Authors do not subscribe to this light touch method and argue that AI requires a sui generis approach as outlined in this Article: .

Entrepreneurs and technological innovators maintain a healthy fear of regulation, which is often seen as red tape that hinders or stymies development. [165] Adam D. Thierer, for example, argues for what he terms "permissionless innovation" - that "unless a compelling case can be made that a new invention will bring serious harm to society, innovation should be allowed to continue unabated." [166] Ray Kurzweil, too, argues against regulation, preferring a free-market system to deal with problems if and when they arise - however, he does this while simultaneously urging caution. [167] Technology-rich industries have a long history of seeking to avoid the impulse to regulate that

---

[159] Carlo, Lyytinen & Boland, supra note 145, at 70.

[160] Id. at 59.

[161] Crawford has observed that the "algorithmic black box" is compounded by the fact that "algorithms do not always behave in predictable ways." Crawford, Can an Algorithm Be Agonistic?, supra note 86, at 7980. In an analysis of societal impacts of algorithms, Karppi and Crawford suggest that, instead of seeking to find transparency in algorithms, a better approach would be the development of "theories that address and analyze the broader sweep of their operations and impact[,] as well as their social, political and institutional contexts." See Karppi & Crawford, supra note 152, at 74.

[162] Chris Holder et al., Robotics and Law: Key Legal and Regulatory Implications of the Robotics Age (Part I of II), 32 Computer L. & Security Rev. 383, 386 (2016).

[163] See, e.g., Diana M. Bowman, The Hare and the Tortoise: An Australian Perspective on Regulating New Technologies and Their Products and Processes, in Innovative Governance Models for Emerging Technologies, supra note 8, at 155, 174-75; Kaal, supra note 23, at 15-16.

[164] Bowman, supra note 163, at 157; Kaal, supra note 23, at 16-18.

[165] Adam Thierer, Technopanics, Threat Inflation, and the Danger of an Information Technology Precautionary Principle, *14 Minn. J.L. Sci. & Tech. 309, 339, 375 (2012).*

[166] Adam Thierer, Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom 1 (rev. & expanded ed. 2016).

[167] Kurzweil, supra note 25, at 304.

often accompanies widespread social fears about new technologies. [168] Scholars and industry representatives have expressed important concerns about the limits of regulation in high-technology industries, and AI poses its own specific challenges for regulators. The key fear is that it may be too early to regulate AI, and that any regulation adopted today "may hinder developments that could prove essential for human existence." [169] Risk analysis too generally involves striking a balance, and the promise of AI may make taking some risk worthwhile.

However, calls for innovation without any regulation must be viewed critically. Part II of this Article: provided a number of concrete examples of potential and existing problems and risks that current applications of as-yet unregulated AI pose for society. The Authors also argued that there is at least the potential for AI development to cause harm to humanity and society. Arguing that regulation necessarily stymies innovation is a syllogistic fallacy; not all regulation stymies innovation. There are enough problems already with relatively narrow AI to persuade regulators that some regulation may indeed be necessary. While regulation may be difficult and may meet resistance from the industry, it is important that society begins to consider the regulation of this vital area. The Authors take up the challenge of contributing to AI research from a legal and regulatory [*420] perspective in this Article: . The next Section details some of the specific problems that regulators face when attempting to regulate in this area.

D. The Problems Posed by AI for Regulators

Matthew Scherer sets out four general ex ante problems with regulating research and development of AI: (1) discreetness, meaning "AI projects could be developed without the largescale integrated institutional frameworks"; (2) diffuseness, meaning AI projects could be carried out by diffuse actors in many locations around the world; (3) discreteness, meaning projects will make use of discrete components and technologies "the full potential of which will not be apparent until the components come together"; and (4) opacity, meaning the "technologies underlying AI will tend to be opaque to most potential regulators." [170]

These broad categories succinctly capture some of the major problems facing those seeking to regulate AI. Certainly, AI is being developed and deployed in many parts of the world, and it is difficult to predict what problems might arise when even two of these powerful technologies are combined. However, while there is the potential for AI development to occur without the need for large scale institutional frameworks such as government agencies, most of the investment in research and development is currently being made by large private companies such as Google, Facebook, Microsoft, Apple, and Amazon. [171] That is where major innovations and developments will be most likely to occur. This also exacerbates the opacity problem because private companies are apt to maintain secrecy, are not required to share information, and are in fact benefitting from the law of patents to protect their legitimate interests in new technology from other developers. However, these broad problems represent only the top layer of concerns; a deeper analysis reveals much more specific and fundamental problems.

Scherer also proposed a system under which an agency would be set up to certify AI systems as safe, [172] and where such certified systems would "enjoy limited tort liability" [173] while uncertified systems would be subject to full liability. This approach concentrates on the consequences of problems with AI and seeks to punish errant behavior after it has occurred. This Article: is more concerned with [*421] proposing solutions to regulating the development of AI ex ante. This Part outlines the potential difficulties associated with this process. Along with the general problems with regulating new technologies outlined above, there are a number of specific problems associated with regulating AI.

---

[168] Thierer, supra note 166, at 68-71.

[169] Gurkaynak, Yilmaz & Haksever, supra note 16, at 753.

[170] Scherer, supra note 22, at 359.

[171] See discussion supra Part IV.A.

[172] Scherer, supra note 22, at 394.

[173] Id.

1. The Pacing Problem

A particular problem that regulators face is that developments in technology outpace any attempt at regulating it. [174] In the face of the continuously increasing speed of innovation, legal and ethical oversight have lagged. [175] This pacing problem plagues the regulation of technology generally and often leads to the technology disengaging or decoupling from the regulation that seeks to regulate it. Because AI is at the forefront of scientific discovery and is developing so quickly, it is affected by this issue more than other technologies. Attempts by regulators to address the pacing problem by future-proofing legislation often result in regulatory disconnect where the laws are too general or vague to effectively serve their intended purpose or to provide meaningful guidance regarding any specific technology. [176] Regulators need to find the optimal middle ground between regulation that is ineffective because it cannot keep pace with the rate of innovation and regulation that is too general to be meaningful in specific cases.

2. Information Asymmetry and the Collingridge Dilemma

Private companies are investing heavily in AI research and development. The result is information asymmetries between those companies and public regulators seeking to understand those developments. [177] Even if lawmakers are able to obtain technical [*422] information from developers, most nontechnical individuals will still be at a loss to understand a product, let alone predict what impacts it may have on individuals, societies, and economies. [178] This is the major cause of the pacing problem, but it is also an issue for courts trying to interpret and apply any legislation that has been implemented, as well as for commentators and advocacy groups looking to hold companies accountable. It is a special problem for regulators who need to fully understand the subject of regulation.

This information problem forms the first half of the Collingridge Dilemma on the control of technology, which states that at the earliest stages of development of a new technology, regulation is difficult due to a lack of information, while in the later stages the technology is so entrenched in our daily lives that there is a resistance to regulatory change from users, developers, and investors. [179] AI has already been deployed in society in a wide variety of fields, from medical diagnostics to criminal sentencing to social media, rendering the need to address this issue even more urgent. [180]

---

[174] Gary E. Marchant, The Growing Gap Between Emerging Technologies and the Law, in The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight, supra note 8, at 19, 19.

[175] Roger Brownsword, Rights, Regulation, and the Technological Revolution 1-2 (2008); Mark Fenwick, Wulf A. Kaal & Erik P. M. Vermeulen, Regulation Tomorrow: What Happens when Technology Is Faster than the Law? 1, 5 (Univ. of Saint Thomas (Minn.) Legal Studies, Research Paper No. 16-23, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2834531; Graeme Laurie, Shawn HE Harmon & Fabiana Arzuaga, Foresighting Futures: Law, New Technologies and the Challenges of Regulating for Uncertainty, 4 Law Innovation & Tech. 1, 3 (2012); Marchant, supra note 174, at 19.

[176] See Brownsword, supra note 175, at 162; Allenby, supra note 8, at 14-15; Ray Purdy, Legal and Regulatory Anticipation and "Beaming' Presence Technologies, 6 Law Innovation & Tech. 147, 147-48 (2014); Fenwick, Kaal & Vermeulen, supra note 175, at 5.

[177] See Brownsword, supra note 175, at 162; Laurie, Harmon & Arzuaga, supra note 175, at 4.

[178] See Hasan Bakhshi, Alan Freeman & Jason Potts, State of Uncertainty: Innovation Policy Through Experimentation 4 (2011); Matthew C. Stephenson, Information Acquisition and Institutional Design, 124 Harv. L. Rev. 1422, 1457 (2011); Gregory N. Mandel, Regulating Emerging Technologies 9 (Temple Univ. Legal Studies, Research Paper No. 2009-18, 2009), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1355674.

[179] David Collingridge, The Social Control of Technology 19 (1980); Laurie, Harmon & Arzuaga, supra note 175, at 6.

[180] See Andreas Margelisch, Swiss Life, A State of the Art Report on Legal Knowledge-Based Systems 4 (1999), http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.49.6315&rep=rep1&type=pdf [https://perma.cc/B2MP-8KXS]; Jason Borenstein & Yvette Pearson, Companion Robots and the Emotional Development of Children, 5 Law Innovation & Tech. 172,

3. Little Established Ethical Guidance, Normative Agreement, or Regulatory Precedent

The ethical and social implications of introducing robots into mainstream society is a very weighty issue that remains largely unresolved, even as the consequences of this interaction are already unfolding. [181] Areas in which ethical issues arise include the use of military robots as well as human-robot relationships, such as the use [*423] of robots as sex partners, caregivers, and servants. [182] Patrick Lin et al. argue that robot ethics issues can be classified in terms of safety and errors, law and ethics, and social impact, and they consider the possibility and desirability of programming ethics into AI systems. [183] A regulatory regime for the design and deployment of robots and AI in society must consider the need to include ethics rules in the code that underpins their operation. That system of ethics must reflect a broad normative consensus on what ethical values robots and AI systems should include.

4. Regulatory Delay and Uncertainty

Regulatory delay occurs as regulators consider if and when they will approve the implementation of a new development. [184] For example, legislators may preemptively ban the commercialization of new products in response to public concerns, acting even before enough research can be conducted to ascertain whether the concerns are valid or well founded. This delay causes uncertainty for developers. [185] Investors and developers are left in the dark while legislators decide what to do, sometimes having to withdraw funding and resources from what might turn out to be a useful and lucrative innovation because they are no longer able or willing to bear the risk. [186] This effect adds to the concerns of developers about regulators seeking to regulate the development of AI.

Of course, some social benefits that may come from innovation and development of AI may well be lost or limited if regulation is implemented prematurely. [187] Cass R. Sunstein, in particular, has adverted to the problems associated with adopting what is known as the "precautionary principle" to regulate risk. [188] People, he argues, are nothing if not "predictably irrational" [189] and tend to be overly [*424] concerned with losses rather than the gains that might be made from, for example, new technology. [190] He argues that regulators should therefore avoid

---

172 (2013); B.M. Dickens & R.J. Cook, Legal and Ethical Issues in Telemedicine and Robotics, 94 Int'l J. Gynecology & Obstetrics 73, 73 (2006); Angwin et al., supra note 89.

[181] Miles Brundage, Limitations and Risks of Machine Ethics, 26 J. Experimental & Theoretical Artificial Intelligence 355, 355-72 (2014); Gordana Dodig Crnkovic & Baran Curuklu, Robots: Ethical by Design, 14 Ethics & Info. Tech. 61, 61-71 (2012); Perri 6, supra note 12, at 419-20; Bert-Jaap Koops, The Concepts, Approaches, and Applications of Responsible Innovation. An Introduction 14 (Tilburg Law Sch., Paper No. 19/2015, 2015), *https://papers.ssrn.com/abstract=2673753*.

[182] Lin, Abney & Bekey, supra note 93, at 944-46.

[183] Id. at 945-47.

[184] See Bakhshi, Freeman & Potts, supra note 178, at 4; Ronald R. Braeutigam, The Effect of Uncertainty in Regulatory Delay on the Rate of Innovation, 43 Law & Contemp. Probs. 98, 98 (1979); Stephenson, supra note 178, at 1429-30; Mandel, supra note 178, at 4.

[185] Braeutigam, supra note 184, at 98-99.

[186] Roberta Romano, Regulating in the Dark, in Regulatory Breakdown: The Crisis of Confidence in U.S. Regulation 86, 87 (Cary Coglianese ed., 2012); Mandel, supra note 178, at 1.

[187] See Kurzweil, supra note 25, at 264, 289, 409; Thierer, supra note 166, at 120.

[188] Cass R. Sunstein, Beyond the Precautionary Principle, *151 U. Pa. L. Rev. 1003, 1003-04 (2003).*

[189] See generally Dan Ariely, Predictably Irrational: The Hidden Forces That Shape Our Decisions (2008) (discussing the term "predictably irrational" and its utility for understanding behavioral economics); Daniel Kahneman, Thinking, Fast and Slow (2011) (noting that intuitive biases often prevent rational decision-making).

[190] Sunstein, supra note 188, at 1009.

regulating purely on the threat of unknown future risks. [191] However, at the same time, he warns against regulatory inaction "because a probability of harm is, under many circumstances, a sufficient reason to act." [192] Ultimately, Sunstein urges that "a wide variety of adverse effects may come from inaction, regulation, and everything in between," noting the need to "attempt to consider all of those adverse effects and not simply a subset." [193] This measured approach should find some favor. After over sixty years of developments in AI, regulation now could not be criticized as being overly reactive or precautionary. As argued in Part II, as AI development continues apace, some caution in this area is warranted. [194]

5. Coordination Across Regulatory Bodies

Coordinating the many regulatory bodies involved in a new technology is a problem that plagues every innovating industry. [195] Many groups and industry bodies have already developed codes of conduct and standards to regulate the development of AI. [196] Given the increasingly interdisciplinary nature of AI research, coordinating these industry bodies is no less a challenge for public regulators in this field. [197] An AI regulatory regime would need to account for existing laws, other governmental regulatory bodies, and self-regulatory industry bodies that develop professional codes of [*425] ethics, and it would need to do this across many different fields such as neuroscience; neurobiology; mechanical, electrical and software engineering; psychology; innovation studies; and economics and finance. [198] Soft law developments such as industry codes of practice, principles, and standards developed by groups of industry participants vary and can often be at cross-purposes. Gary E. Marchant and Wendell Wallach propose that this multiplicity of perspectives and approaches requires an "issue manager" to oversee and coordinate the various principles, codes, and other approaches. [199] Marchant and Wallach have proposed to form a Governance Coordination Committee to "provide oversight, cultivate public debate, and evaluate the ethical, legal, social, and economic ramifications of … important new technologies[.]" [200] The current efforts to attempt to govern using these industry-led soft law approaches are discussed further in Part IV.

6. Agency Capture

---

[191] See id.

[192] Id. at 1055.

[193] Id. at 1056.

[194] See supra Part II.D (discussing several issues that may require a regulatory response, including biases in law enforcement, safety, effects on judicial decision-making, reduced privacy, and unemployment caused by increasing rates of automation supported by AI).

[195] See Stuart Minor Benjamin & Arti K. Rai, Fixing Innovation Policy: A Structural Perspective, *77 Geo. Wash. L. Rev. 1, 3-5 (2008);* Lyria Bennett Moses, Agents of Change: How the Law "Copes' with Technological Change, 20 Griffith L. Rev. 763, 767 (2011); Mandel, supra note 178, at 75; Jason Potts, The National Origins of Global Innovation Policy and the Case for a World Innovation Organization 2-3 (Dec. 19, 2015) (unpublished manuscript), *https://papers.ssrn.com/abstract=2705906*.

[196] See, e.g., infra Part IV.B (discussing the codes of conduct produced by the Partnership on AI and the standards prepared by the Institute of Electrical and Electronics Engineers (IEEE)).

[197] See Moses, supra note 195, at 786-87 (describing regulatory issues associated with nanotechnology).

[198] See Peter W. B. Phillips, Governing Transformative Technological Innovation: Who's in Charge? 247-48 (2007). See generally Christopher-Paul Milne & Joyce Tait, Evolution Along the Government - Governance Continuum: FDA's Orphan Products and Fast Track Programs as Exemplars of "What Works" for Innovation and Regulation, ***64 Food & Drug L.J. 733 (2009);*** Potts, supra note 195.

[199] See Gary E. Marchant & Wendell Wallach, Governing the Governance of Emerging Technologies, in Innovative Governance Models for Emerging Technologies, supra note 8, at 136, 143.

[200] See Gary E. Marchant & Wendell Wallach, Coordinating Technology Governance, 31 Issues Sci. & Tech., Summer 2015, at 43.

Regulatory failure due to agency capture occurs where regulators become sympathetic towards the industry they are regulating. This can be the result of any number of factors, such as a high frequency of interaction between industry and regulators, industry representatives "buying off" regulators with gifts like free lunches or sponsorship to attend conferences, or a "revolving door" for employees between regulatory agencies and industry. [201] While each of these problems is relatively common throughout innovating industries, the AI industry is particularly susceptible to the revolving door issue. [202] The information asymmetry issue where AI companies hold all the relevant information about the technology makes the knowledge and expertise acquired by employees of AI developers [*426] particularly valuable to regulators, which are likely to be interested in employing former AI developers when (and if) they can.

7. Limited Enforcement Mechanisms and Jurisdiction Shopping

Added to the complexities outlined above, the major players in the development of AI - such as Google, Facebook, Microsoft, and Apple - are some of the biggest, most complex, and powerful corporations the world has seen. [203] They own and control what Marx might have described as the means of production in this field - that is, the vast array of superpowerful computers and the phalanx of the world's best and brightest mathematicians and engineers required to churn the algorithms necessary to create AI. [204] The power disparity between these players and government regulators, who often struggle to secure sufficient resources to operate, highlights the difficulties that might be faced by a regulator in trying to regulate these companies. [205]

The fact that the technology is relatively opaque [206] also makes it easier for firms to hide wrongdoing and evade regulation. Volkswagen, for example, was able to create specific code to identify the tests used by regulators to measure emissions and make its car engines appear to run more cleanly than when in normal use. [207] Similarly, recent reports suggest that Uber created a version of its app specifically designed to identify users likely to be regulators and prevent them from accessing the system to investigate concerns or collect evidence. [208]

Part III outlined various risks associated with AI and broadly grouped applications of AI into three classes based upon the risks that each pose. The Authors also highlighted the general and specific difficulties that regulators face when attempting to regulate new technologies and, particularly, AI. Part IV outlines how public [*427] regulators

---

[201] See Thomas O. McGarity, MTBE: A Precautionary Tale, *28 Harv. Envtl. L. Rev. 281, 325-26 (2004).* See generally Ben Goldacre, Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients (2013).

[202] See Anne Weismann, Silicon Valley Companies Lobby to Remain Unregulated, N.Y. Times (Oct. 24, 2016, 3:20 AM), *https://www.nytimes.com/roomfordebate/2016/10/24/silicon-valley-goes-to-washington/silicon-valley-companies-lobby-to-remain-*unregulated?referer=*https://t.co/86PwxwASfJ&nytmobile=0* [*https://perma.cc/9ER4-D6PM*].

[203] See, e.g., Rana Foroohar, Release Big Tech's Grip on Power, Fin. Times (June 19, 2017), *https://www.ft.com/content/173a9ed8-52b0-11e7-a1f2-db19572361bb*.

[204] See generally Jonathan Taplin, Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy (2017) (remarking that Google, Facebook, Amazon, and Apple shaped the Internet into its current form without any regulatory oversight, rendering these companies "more powerful than most people realize").

[205] See, e.g., infra Parts IV.A, IV.B (discussing the relative lack of power held by governments vis-a-vis the corporate players in the field). This discussion will form the basis of a further paper on power relations in regulating AI.

[206] See generally Pasquale, supra note 13, at 10.

[207] See Russell Hotten, Volkswagen: The Scandal Explained, BBC News (Dec. 10, 2015), *http://www.bbc.com/news/business-34324772* [*https://perma.cc/JLT8-XWF7*].

[208] See Associated Press, Uber Deploys Secret Weapon Against Undercover Regulators, Bus. Insider (Mar. 3, 2017, 4:39 PM), *http://www.businessinsider.com/ap-uber-deploys-secret-weapon-against-undercover-regulators-2017-3* [*https://perma.cc/CA2Z-MUSH*].

will need to adopt new strategies to begin to regulate AI as the old strategies lose effectiveness. One strategy will be to regulate based upon the relative risks associated with particular applications of AI.

IV. The Need for Regulatory Innovation

Regulators face an extremely difficult challenge in responding to AI. As discussed above, regulators find it difficult to keep up with the pace of change; do not have all the information they require; must avoid overregulation and uncertainty; require, but cannot rely too heavily on, specialist knowledge obtained from industry; have to make do with enforcement mechanisms that are only partially effective; and need to make clear and justifiable policy decisions in a field that is highly contested. Traditionally, governments had the information and resources that put them in the best position to regulate in most instances. Society has seen a long period where government held legislative control of the state. In the field of new technology, at least, the machinery of control is drifting away from government and is becoming decentered. This Part reviews this decentering of regulation and outlines some examples of peer or self-regulation that has begun to proliferate in the vacuum of government control.

The Authors review both the regulatory theory literature and the legal literature on the regulation of technology. As will be shown, these theories have clear limitations when asked to respond to the development of new technologies but may still provide some guidance to regulators seeking to approach regulating AI. Regulatory theories that have developed over the last two decades, such as responsive regulation and really responsive regulation, are normative and propose what good regulation should include. They are also, by definition, "responsive" and hence presuppose the existence of a regulatory framework. As such, they are best used to guide interactions between regulators and the regulated when regulatory systems are already in place. Further, responsive regulation is limited in its ability to regulate new technologies that exhibit the kinds of characteristics set out in Part III above - it lacks the flexibility required to react quickly enough in such a dynamic field. Further, while much can be learned from regulation of other emerging technologies, the regulation of AI must be sui generis. While in its nascent stages, it will require a more nuanced set of regulatory approaches.

[*428]

A. Regulating with Limited Resources in a Decentralized Environment

For a long time, regulation was thought of mainly in terms of legal commands and sanctions. The state, in the classical model of regulation, is a powerful entity that can command obedience through a monopoly on the legitimate use of force. [209] It is now widely recognized that there are far more techniques in the regulation toolbox than "command and control" style rules backed by sanctions. [210] Ian Ayres and John Braithwaite's concept of "responsive regulation," for example, sets out a graduated pyramid of interventions by the state in policing behavior in order to encourage and direct an optimal mix of regulatory work by private and public entities. [211] The responsive element of "responsive regulation" is that as the regulatory response moves up the pyramid, "escalating forms of government intervention will reinforce and help constitute less intrusive and delegated forms of market regulation." [212] That is, responsive regulation still requires government to assert a "willingness to regulate more intrusively" and by so doing can guide the regulation where it is most effective, mostly through "less intrusive and less centralized forms of government intervention." [213] Ayres and Braithwaite proposed a pyramid of enforcement measures by government with the most intrusive command and control regulation at the apex and less intrusive measures such as self-regulation at the base. Government still maintains the ability and responsibility to ultimately

---

[209] See generally John Austin, The Province of Jurisprudence Determined (ed. 1832); Thomas Hobbes, Leviathan (Edward White & David Widger eds., Project Gutenberg 2009) (1651).

[210] See Julia Black, Critical Reflections on Regulation, 27 Austl. J. Legal Phil. 1, 4 (2002).

[211] See Ian Ayres & John Braithwaite, Responsive Regulation: Transcending the Deregulation Debate 6 (Donald R. Harris et al. eds., 1992).

[212] Id. at 4.

[213] Id.

regulate if required. [214] The threat relies on government's ability to inflict varying degrees of discretionary punishment or other forms of persuasion within the pyramidal structure if the regulated entity fails to comply with initial regulatory attempts - this is referred to as the tit-for-tat approach. [215] The critical effect of responsive regulation is to highlight developments in alternative means of regulating other than command and control - and therefore avoid some of the more problematic effects of blunt regulatory tools. It appears, however, to be a tool that is still too blunt to hone new [*429] technologies such as AI. Part of the reason for this is that the traditional role of government has diminished over time.

1. The Traditional Role of Government in Regulation

When considering the regulatory role of the contemporary state in 2007, Christopher Hood and Helen Margetts listed four resources of regulation that governments have in differing degrees in differing contexts: nodality, authority, treasure, and organization. [216] Nodality refers to the government's central position as a receiver and distributor of information that allows it access to and control of the full range of information. [217] Governments hold a strategic position with nearly full information about the area and topic of regulation. [218] Authority refers to the authority of the government to determine what is legal [219] and to "demand, forbid, guarantee, [and] adjudicate." [220] Treasure refers to the government's assets, both in money and other tangible assets such as buildings and equipment, which give it the power to control development at the time and place of its choosing. [221] Organization refers to the government's human resources with the knowledge and skills to be able to carry out any required task. [222] This includes arrangements of "[people] (soldiers, workers, bureaucrats), land, buildings, materials, computers and equipment." [223] The interaction of these roles traditionally held by government simplifies analysis of the role of government in regulation. [224] When these theories are applied to the difficult task of regulating AI, the challenges that regulators face are clearly visible. In these contexts, the government's nodality, authority, treasure, and organization have been depleted or usurped and are not always sufficient to match that of the major technology companies such as Google, Facebook, Microsoft, and Apple. Part of the challenge of effectively regulating AI is to identify opportunities for regulatory agencies to influence other actors when these four resources are limited.

Similarly, the responsive regulation model depends on a strong regulatory state that is ultimately able to use sanctions to direct [*430] behavior. However, as stated, that no longer fully reflects practical realities. Responsive regulation may still be an effective means of governing more traditional industries, such as the production of wool in Australia, [225] but the Authors argue that responsive regulation is not sufficiently flexible and nuanced to apply to a dynamic environment such as the development of AI. Further, it relies on the power of the state to impose the ultimate sanction at the apex of the pyramid; that is, the command and control regulation of an industry. The notion of government as the apex of power structures is arguably no longer applicable, if it truly ever was solely the case.

---

[214] See id. at 4-5.

[215] See id. at 37-38.

[216] Christopher C. Hood & Helen Z. Margetts, The Tools of Government in the Digital Age 5-6 (2007).

[217] Id. at 5.

[218] Id. at 6.

[219] Id.

[220] Id. at 5.

[221] See id. at 6.

[222] See id.

[223] Id.

[224] Id. at 12.

[225] For examples of potential uses for responsive regulation, see John Braithwaite, The Essence of Responsive Regulation, 44 U.B.C. L. Rev. 475, 480-83 (2011).

This is especially so when considering the power, global reach, and diffuse company structures of companies operating at "Google scale."

2. Decentered Regulation

Over the last three decades, regulatory scholars and regulatory agencies have been grappling with the "decentering" of regulation, which recognizes that regulation is not the exclusive work of states and that the power of the state to command obedience is reduced. [226] As Julia Black contends, a "decentered understanding" of regulation involves "complexity, fragmentation of knowledge and of the exercise of power and control, autonomy, interactions and interdependencies, and the collapse of the public/private distinction." [227] The hallmarks of "decentered" regulation, she argues, are that it is "hybrid (combining governmental and non-governmental actors), multi-faceted (using a number of different strategies simultaneously or sequentially), and indirect." [228] The current environment surrounding the development of AI shows that if regulation of AI is to succeed, that regulation must evolve in an environment that displays these characteristics. Those regulating in this field need to understand and work within these parameters. Black argues that decentered regulation

should be indirect, focusing on interactions between the system and its environment. It should be a process of co-ordinating, steering, influencing, and balancing interactions between actors/systems, to organise themselves, using such techniques as proceduralization, collibration, feedback loops, redundancy, and above all, countering variety with a variety. [229]

[*431] Regulators must address the challenges of regulating with limited resources. These resource constraints have curbed the impact of regulatory bodies in general. However, they are particularly debilitating in the context of new technologies that involve a steep learning curve and require regulatory bodies to engage deeply in the industry. Regulatory agencies that seek to regulate AI in this environment should first seek to engage with and work with the relevant actors to learn about and grapple with the complexities in the field. By doing this, they can begin to understand the motivations of the relevant players so that they might start to influence the direction AI development will take. This process, as Black recommends, will involve recurring loops of discussion and feedback where effective ideas are fostered and redundant notions are jettisoned. [230] Public regulators faced with resource constraints must do this while also managing a shifting regulatory environment where they are subject to pressure from interest groups and citizens to pursue conflicting agendas and must also consider how regulation of AI might affect regulatory work in other fields and industries. Regulators must also be able to reflect on the effectiveness of their strategies, often in an information vacuum, and be able to change strategies when one approach does not work, is ineffective, or even is retrograde. [231]

B. Self-Regulation and Peer Regulation

One result of decentered regulation is that governments that once held a central position of power and influence have ceded some of that influence and power to a dissipated group of regulatory participants. Where political influence and power exist in those industries, self-regulation evolves and becomes the default position. In recent years, prominent figures from within the AI industry have begun to warn about the need to ensure that the

---

[226] See, e.g., Black, supra note 20, at 105 (emphasizing the difference between the "decentred" analysis of regulation and the regulatory state).

[227] Black, supra note 210, at 8.

[228] Black, supra note 20, at 111.

[229] Id.

[230] Id.

[231] See Robert Baldwin & Julia Black, Really Responsive Regulation, 71 Mod. L. Rev. 59, 61 (2008).

development and deployment of AI technology is effectively regulated. [232] In the absence of government-led intervention, those within the industry are regulating themselves. This is not typical self-regulation under the auspices of a formal government agency but is self-regulation in a vacuum of government input.

[*432] Some academics have described a kind of self-regulation where the influence of corporate peers guides the behavior of industry participants. [233] Bob Jessop describes a system of governance that limits the role of regulatory bodies and emphasizes "the reflexive self-organization of independent actors involved in complex relations of reciprocal interdependence, with such self-organization being based on continuing dialogue and resource-sharing to develop mutually beneficial joint projects and to manage the contradictions and dilemmas inevitably involved in such situations." [234] This appears to describe what is happening in practice in relation to AI. Jessop emphasizes the role of self-organization of stakeholders to include

(1) the more or less spontaneous, bottom-up development by networks of rules, values, norms and principles that they then acknowledge and follow; [and] (2) increased deliberation and participation by civil society groups through stakeholder democracy, putting external pressure on the state managers and/or other elites involved in governance. [235]

This bottom-up development is happening now in the development of AI. Prominent industry participants have developed several codes of conduct and practice, and the next phase of coordinating these strategies has begun - all outside the auspices of government control.

The challenges of regulating fast-moving technology are so great that industry self-regulatory approaches are often presented as the most effective mechanism to manage risk. Industry bodies are already forming to respond to fears about the ongoing deployment of AI systems in ways that could be interpreted as staving off what they might describe as clumsy and heavy-handed public regulation. One of the most prominent efforts is the Partnership on AI between Google, DeepMind, Facebook, Microsoft, Apple, Amazon, and IBM, together with the American Civil Liberties Union and the Association for the Advancement of Artificial Intelligence (AAAI). [236] The Partnership on AI's purpose statement is to "benefit people and society," [237] and it is said to have been "established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI [*433] and its influences on people and society." [238] It has developed a series of tenets for the development of AI that commit its members to ongoing engagement with stakeholders to protect the privacy, security, and other human rights of individuals. [239] In doing so, the Partnership is taking on the role of a self-regulatory association and potentially warding off more enforceable state-imposed regulatory obligations.

---

[232] See Omohundro, supra note 29, at 483-93; Stuart Russell, Daniel Dewey & Max Tegmark, Research Priorities for Robust and Beneficial Artificial Intelligence, AI Mag., Winter 2015, at 105, https://futureoflife.org/data/documents/research_priorities.pdf [https://perma.cc/5C2G-2H6X].

[233] Bob Jessop, Governance and Metagovernance: On Reflexivity, Requisite Variety and Requisite Irony, in Governance as Social and Political Communication 101, 101 (Henrik P. Bang ed., 2003).

[234] Id.

[235] Bob Jessop, State Theory, in Handbook on Theories of Governance 71, 82 (Christopher Ansell & Jacob Torfing eds., 2016).

[236] Founding Partners, Partnership on AI to Benefit People & Soc'y, https://www.partnershiponai.org/#s-founding-partners [https://perma.cc/KT3D-CSSY] (last visited Nov. 2, 2017).

[237] Id.

[238] Partnership on AI to Benefit People & Soc'y, https://www.partnershiponai.org/ [https://perma.cc/3FD4-R6G6] (last visited Nov. 2, 2017).

[239] Tenets, Partnership on AI to Benefit People & Soc'y, https://www.partnershiponai.org/tenets/ [https://perma.cc/H6ZB-CV3Z] (last visited Nov. 2, 2017).

Another industry-led attempt to regulate AI was developed at the Future of Life Institute's Asilomar conference in January 2017. [240] The twenty-three Asilomar principles, as they are known, are grouped under three headings: research issues, ethics and values, and longer-term issues. [241] Principles falling within the longer-term issues include Principle 22, titled "Importance." It states that "advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources." [242] Principle 23, titled "Risks," notes that "risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact." [243] Further, Principle 24, titled "Recursive Self-Improvement," notes that "AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures." [244] These principles reflect concerns that even those within the industry hold about the development of AGI. The Asilomar principles contain a similar basket of issues that are reflected in other industry codes or values statements in relation to AI. [245] While they express well-meaning principles of behavior, it is uncertain who will enforce these control measures and what sanctions may be levied for their breach.

[*434] Another industry body, the Institute of Electrical and Electronics Engineers (IEEE), recently produced a discussion paper titled Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Artificial Intelligence and Autonomous Systems. [246] The Ethically Aligned Design project aimed to "bring together multiple voices in the Artificial Intelligence and Autonomous Systems communities to identify and find consensus on timely issues." [247] Those issues address concerns about how to ensure that AI does not infringe human rights, that the decisions of autonomous systems are accountable and transparent, and that there are checks in place to minimize risks through enhanced education. [248]

Proponents of AI have sought to counter the fears long expressed by science fiction authors by highlighting the positive and benign applications of AI already in place today. [249] Developers suggest that technical contingency plans like DeepMind's "big red button" are in place in case AI gets out of hand. [250] The implication is that up to this limit - the "nuclear option" of shutting down rogue AI completely - the developers of AI are already effectively regulating its development through initiatives like the Partnership on AI and the principles set out by IEEE. In this regard, the Partnership on AI has endorsed the US government's report Preparing for the Future of Artificial

---

[240] Asilomar AI Principles, Future of Life Inst., https://futureoflife.org/ai-principles/ [https://perma.cc/KAT2-DBMX] (last visited Nov. 2, 2017).

[241] Id.

[242] Id.

[243] Id.

[244] Id.

[245] See, e.g., Inst. Elec. & Elecs. Eng'rs, Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems 5 (2016), http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf; Tenets, supra note 239.

[246] See Inst. Elec. & Elecs. Eng'rs, supra note 245.

[247] Id. at 3.

[248] Id. at 5.

[249] See Evans & Gao, supra note 2; Glossary, Stottler Henke, https://www.stottlerhenke.com/artificial-intelligence/glossary/ [https://perma.cc/F4R3-65NX] (last visited Nov. 2, 2017).

[250] See generally Etzioni & Etzioni, supra note 70, at 142; Orseau & Armstrong, supra note 67, at 557.

Intelligence. [251] It is in the interests of industry participants such as the Partnership on AI not to disavow the government's position. It shows that the industry is very capable of self-regulating and that it is in lock-step with the government and its public regulators. In a classic statement of self-regulation that usurps the [*435] traditional role of public regulators, the Partnership on AI has stated that it will continue to pursue "ongoing engagement … to bring stakeholders together, create best practices, share findings and insights, and to contribute to charting a path forward." [252] Perhaps, given the government's retreat from regulating in this area, the Partnership on AI may be best placed to continue this work for the time being.

It may well be that self-regulation will be effective in mitigating the most important risks of the development and deployment of AI systems. However, there is also a risk that self-regulation may not be sufficient. [253] First, industry codes or [*436] principles are not obligatory. The principles or codes are often drafted broadly as vision or values statements that do not contain any mandatory requirements but are, rather, guides to practice that may be ignored. [254] Second, they lack effective enforcement regimes. Even if they do contain some element of obligation,

---

[251] Partnership on AI Expresses Support for White House Report on Artificial Intelligence, Partnership on AI to Benefit People & Soc'y (Oct. 12, 2016), https://www.partnershiponai.org/2016/10/partnership-ai-expresses-support-white-house-report-artificial-intelligence/ [https://perma.cc/ 4BSY-GJP5]. For the original White House report, see Nat'l Sci. & Tech. Council, supra note 43. For that report's accompanying strategic report, see Nat'l Sci. & Tech. Council, Exec. Office of the President, The National Artificial Intelligence Research and Development Strategic Plan (2016), https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf [https://perma.cc/AX3Q-KACQ]. See generally Ed Felten & Terah Lyons, The Administration's Report on the Future of Artificial Intelligence, White House Blog (Oct. 12, 2016, 6:02 AM), https://obamawhitehouse.archives.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence [https://perma.cc/MRP9-UTBU] (announcing the release of the White House report).

[252] Partnership on AI Expresses Support for White House Report on Artificial Intelligence, supra note 251.

[253] See Jodi L. Short, Self-Regulation in the Regulatory Void: "Blue Moon" or "Bad Moon"?, 649 Annals Am. Acad. Pol. & Soc. Sci. 22, 23 (2013). Self-regulation relies on the strong and "credible threat" of state-based regulation. See Michael P. Vandenbergh, The Private Life of Public Law, 105 Colum. L. Rev. 2029, 2037-39 (2005). It has been argued that self-regulation fails, or at least is unreliable, without the ever-present threat of state-based sanctions. See, e.g., Ian Ayres & John Braithwaite, Tripartism: Regulatory Capture and Empowerment, 16 Law & Soc. Inquiry 435, 489-90 (1991); Christopher Kevin Walker, Neoliberalism and the Reform of Regulation Policy in the Australian Trucking Sector: Policy Innovation or a Repeat of Known Pitfalls?, 37 Pol'y Stud. 72, 75-76 (2016); see also David P. McCaffrey & David W. Hart, Wall Street Polices Itself: How Securities Firms Manage the Legal Hazards of Competitive Pressures 176 (1998); Christine Parker, The Open Corporation: Effective Self-Regulation and Democracy 72 (2010); Joseph V. Rees, Reforming the Workplace: A Study of Self-Regulation in Occupational Safety 72 (1988); Jay A. Sigler & Joseph E. Murphy, Interactive Corporate Compliance: An Alternative to Regulatory Compulsion ix (1988); Neil A. Gunningham, Dorothy Thornton & Robert A. Kagan, Motivating Management: Corporate Compliance in Environmental Protection, 27 Law & Pol'y 289, 290-91 (2005); Andrew A. King & Michael J. Lenox, Industry Self-Regulation Without Sanctions: The Chemical Industry's Responsible Care Program, 43 Acad. Mgmt. J. 698, 698 (2000); Jay P. Shimshack & Michael B. Ward, Regulator Reputation, Enforcement, and Environmental Compliance, 50 J. Envtl. Econ. & Mgmt. 519 (2005). Gregory Jackson et al., however, reject that public regulation and self-regulation are diametrically opposed choices, instead arguing that their relationship is symbiotic. Gregory Jackson et al., Regulating Self-Regulation? The Politics and Effects of Mandatory CSR Disclosure in Comparison 1 (Mar. 1, 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2925055. The literature acknowledges that self-motivation and reputation figure into the motivational mix. See Roland Benabou & Jean Tirole, Incentives and Prosocial Behavior,96 Am. Econ. Rev. 1652, 1654 (2006). The deterrent effect of public regulation has somewhat of a paradoxical effect on self-regulation and can dampen other intrinsic and external motivating factors, such as earnest goodwill and concern for reputation. See Ayres & Braithwaite, supra note 211, at 19; Robert Baldwin, Martin Cave & Martin Lodge, Understanding Regulation: Theory, Strategy, and Practice 261-62 (2d ed. 2012); Jodi L. Short & Michael W. Toffel, Making Self-Regulation More Than Merely Symbolic: The Critical Role of the Legal Environment, 55 Admin. Sci. Q. 361, 386 (2010). See generally Eugene Bardach & Robert A. Kagan, Going by the Book: The Problem of Regulatory Unreasonableness (2002) (arguing for a flexible method of regulatory implementation in order to alleviate problems arising from competing visions of regulation); Fiona Haines, Corporate Regulation: Beyond "Punish or Persuade' (1998) (discussing the interaction between evolving capitalism and regulatory policymaking).

[254] Inst. Elec. & Elecs. Eng'rs, supra note 245, at 5.

participants may lack the will to enforce those obligations. [255] Third, many different suggested approaches, such as the IEEE standards or the principles proposed by the Partnership on AI or the Asilomar principles, vary in their content and focus and lack a central governing body that will coordinate direction and compliance. [256]

Certainly, it will be important to avoid regulation that is ineffective or unduly stymies research and development. The Authors suggest that governments need to consider and engage with the concerns and risks associated with AI now in order to protect public interests that industry-led regulation is unsuited to addressing. [257]

C. The Evolving Nature of Regulation

Despite the efforts of those within the industry to self-regulate, the task of regulating the development and deployment of AI is increasingly pressing. The AI Now Report, prepared after the AI Now public symposium hosted by the White House and New York University's Information Law Institute in July 2016, [258] set out several key recommendations for future work in AI development. One of those recommendations was to

increase efforts to improve diversity among AI developers and researchers, and broaden and incorporate the full range of perspectives, contexts, and disciplinary backgrounds into the development of AI systems. The field of AI should also support and promote interdisciplinary AI research initiatives that look at AI systems' impact from multiple perspectives, combining the computational, social scientific, and humanistic. [259]

The ongoing pace of change and the notoriously slow response of lawyers and regulators create real challenges for this type of multidisciplinary collaboration. So much so that, in a cri de coeur, the [*437] Ethically Aligned Design report noted that "there is much to do for lawyers in this field that thus far has attracted very few practitioners and academics despite being an area of pressing need." [260] The report calls on lawyers to be "part of [the] discussions on regulation, governance, [and] domestic and international legislation in these areas." [261]

Stuart Russell, Daniel Dewey, and Max Tegmark set out two policy questions they argue need to be addressed by regulators, academics, and those in the industry: "(1) What is the space of policies worth studying, and how might they be enacted? (2) Which criteria should be used to determine the merits of a policy?" [262] They proposed that the qualities of these policies should include "verifiability of compliance, enforceability, ability to reduce risk, ability to avoid stifling desirable technology development, likelihood of being adopted, and ability to adapt over time to changing circumstances." [263] It appears inevitable that there will eventually be some form of regulation of AI. The

---

[255] Id. at 18.

[256] See Marchant & Wallach, supra note 199, at 43-44.

[257] See Baldwin, Cave & Lodge, supra note 253, at 259-80; Rob Baggott, Regulatory Reform in Britain: The Changing Face of Self-Regulation, 67 Pub. Admin. 435, 444 (1989); Black, supra note 20, at 115; Short, supra note 253, at 23.

[258] Kate Crawford et al., The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term (2016), https://artificialintelligencenow.com/media/documents/AINowSummary Report_3_RpmwKHu.pdf [https://perma.cc/Q9T6-HJJU].

[259] Id. at 5.

[260] Inst. Elec. & Elecs. Eng'rs, supra note 245, at 89.

[261] Id.

[262] Russell, Dewey & Tegmark, supra note 232, at 107.

[263] Id. Other principles of good governance that might be added to this list include that policies should be "participatory, consensus oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive and [should] follow[] the rule of law." See U.N. Econ. & Soc. Comm'n for Asia & the Pacific, What is Good Governance?, http://www.unescap.org/sites/default/files/good-governance.pdf [https://perma.cc/8768-NHCN] (last visited Nov. 2, 2017).

European Union has begun to develop Civil Law Rules on Robotics [264] that will ultimately govern the development of robotics and AI in Europe, which is further discussed in Part V below.

Table 1 below sets out some of the major theories of regulation that have evolved over the last two decades. Regulatory theory has developed from the prominent but increasingly less influential command and control style to more and more nuanced and adaptive approaches as increasingly complex situations have demanded. As the Authors suggest, many of these theories are either inappropriate or would be ineffective when regulating AI.

[*438]

Table 1 - Theories of Regulation

Theory	Guiding principles
"Responsive Regulation" [265] : a "tit-for-tat" approach to enforce compliance by persuasion and education before escalating up a "pyramid" of more punitive sanctions	Regulators should: . Think in context; . Listen actively (build commitment with stakeholders); . Engage with fairness; . Praise those who show commitment; . Signal a preference for support and education . Signal a range of escalating sanctions that may be used if necessary; . Engage a wider network of partners as regulatory responses increase in severity; . Elicit active responsibility from stakeholders where possible; and . Evaluate regulations and improve practices. [266]
"Smart Regulation"	Regulators should: . Prefer a mix of regulatory instruments while avoiding "smorgasboardism"; . Prefer less interventionist measures; . Escalate up a pyramid of sanctions when required (responsive regulation);

[264] See Comm. on Legal Affairs, Rep. with Recommendations to the Comm. on Civil Law Rules on Robotics, Eur. Parl. Doc. A8-0005/2017 (Jan. 27, 2017) [hereinafter EU Robotics Report], http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+ A8-2017-0005+0+DOC+PDF+V0//EN [https://perma.cc/HQ8U-VXVC].

[265] See generally Ayres & Braithwaite, supra note 211 (introducing and explaining the concept of "Responsive Regulation").

[266] Braithwaite, supra note 225, at 476.

	. Empower third parties to act as surrogate regulators; and . Maximize opportunities for win-win outcomes by encouraging businesses to go "beyond compliance." [267]
"Risk-Based Regulation"	Hampton Review: . "Regulators, and the regulatory system as a whole, should use comprehensive risk assessment to concentrate resources on the areas that need them most"; . "Regulators should be accountable for the efficiency and effectiveness of their activities, while remaining independent in the decisions they take"; . "All regulations should be written so that they are easily understood, easily implemented, and easily enforced, and all interested parties should be consulted when they are being drafted"; . "No inspection should take place without a reason"; . "Businesses should not have to give unnecessary information, nor give the same piece of information twice"; . "The few businesses that persistently break regulations should be identified quickly, and face proportionate and meaningful sanctions"; . "Regulators should provide authoritative, accessible advice easily and cheaply"; . "When new policies are being developed, explicit consideration should be given to how

[267] Neil Gunningham & Darren Sinclair, Designing Smart Regulation 2, http://www.oecd.org/env/outreach/33947759.pdf [https://perma.cc/FKS5-ZEXT] (last visited Nov. 2, 2017).

	they can be enforced using existing systems and data to minimise the administrative burden imposed"; . "Regulators should be of the right size and scope, and no new regulator should be created where an existing one can do the work"; and . "Regulators should recognise that a key element of their activity will be to allow, or even encourage, economic progress and only to intervene when there is a clear case for protection." [268]
"Regulatory Craft" (focusing on problem solving)	Regulators should: . "Nominate potential problems for attention"; . "Define the problem precisely"; . "Determine how to measure impact"; . "Develop solutions or interventions"; . "Implement the plan, with . . . periodic monitoring, review, and adjustment"; and . "Close project, allowing for long-term monitoring and maintenance." [269]
"Really Responsive Regulation"	Regulators should be responsive to: . Firms' compliance responses (Responsive Regulation); but also . The "attitudinal settings" (operating and cognitive framework of the target of regulation); . The institutional environment;

[268] Philip Hampton, Reducing Administrative Burdens: Effective Inspection and Enforcement 7 (2005), http://news.bbc.co.uk/nol/shared/bsp/hi/pdfs/bud05hampton_ 150305_640.pdf [https://perma.cc/6BWC-A5T9].

[269] Malcolm K. Sparrow, The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance 142 (2000).

	. The "logics of different regulatory tools and strategies"; . The regulatory regime's own performance and effects; and . Changes in priorities, circumstances, and objectives. [270]
"Really Responsive Risk-Based Regulation"	In applying risk-based regulation, regulators should: . Be "responsive to regulated firms' behavior, attitude, and culture; institutional environments; interactions of controls; regulatory performance; and change"; . "Take attitudinal matters on board"; . Identify how attitudes vary across regulatory tasks; and . "Be clear about the degree to which any particular regulatory task can and should be guided by a risk-scoring system." [271] Risk-based regulation must focus on: . "Detecting undesirable or non-compliant behavior, . Responding to that behavior by developing tools and strategies, . Enforcing those tools and strategies on the ground, [and] . Assessing their success or failure, and modifying them accordingly." [272]

[*441] Table 1 outlines a number of theories that describe traditional methods of regulation. While no one theory would apply as a whole to the regulation of AI, a risk-based approach in combination with several of the elements of Really Responsive Regulation and Smart Regulation may ultimately prove effective. Julia Black and Robert

---

[270] Baldwin & Black, supra note 231, at 61, 73.

[271] Julia Black & Robert Baldwin, Really Responsive Risk-Based Regulation, 32 Law & Pol'y 181, 193, 210 (2010).

[272] Id. at 187 (emphasis in original).

Baldwin's Really Responsive Risk-Based Regulation[273] is perhaps the closest to this approach. It requires regulators to be "responsive to regulated firms' behavior, attitude, and culture; institutional environments; interactions of controls; regulatory performance; and change."[274] However, Black and Baldwin could not have foreseen the changes in the AI environment that would occur subsequent to their article's publication in 2010. The speed of change in the regulatory environment surrounding AI makes it difficult for regulators to react in good time. The proliferation of AI into daily life has been fast and furtive. The consequence is that public regulators will need to be even more responsive in the forms Black and Baldwin suggest. Perhaps a Really Really Responsive Risk-Based Regulation would be required. However, the Authors suggest a more nuanced approach is required in Part V below.

The regulation of AI requires a theory of regulation that is not bound by the normative straits in which the theories above have evolved. Those theories detail a normative approach to regulation. [275] They presuppose a regulatory environment already being in place and the ability of the government to impose its control if ultimately required to do so. As this Article: has argued, this is no longer the case. However, the main problem that each theory faces when it comes to new technologies such as AI is that the mechanisms to respond to change are too slow. They require the machinery of the state to respond to changes in the regulatory environment, but that machinery is not easily engaged and, when engaged, responds too slowly.

Meanwhile, others have offered different and sometimes more concrete suggestions for how regulatory agencies can deal with the particular difficulties of regulating fast-moving technological change:

[*442]

Table 2 - Applications of Strategies

Theory	Guiding principles
"Adaptive Policymaking"	Regulation should be: . Cautious; . Macroscopic; . Incremental; . Experimental; . Contextual; . Flexible; . Provisional; . Accountable; and . Sustainable. [276]
One Hundred Year Study on AI	Government should: . Accrue greater technical expertise in AI; . Remove impediments to research on the social impacts of AI; . Increase public and private funding for research on the social impacts of AI;

[273] See generally id.

[274] Id. at 210.

[275] See Ayres & Braithwaite, supra note 211, at 17; Baldwin & Black, supra note 231, at 1718; Black & Baldwin, supra note 271, at 194. See generally Neil Gunningham, Peter Grabosky & Darren Sinclair, Smart Regulation: Designing Environmental Policy (Keith Hawkins ed., 1998).

[276] Whitt, supra note 21, at 50004.

	. Resist pressure for "more" and "tougher" regulation that stifles innovation or forces innovators to leave the jurisdiction; . Encourage a "virtuous cycle" of accountability, transparency, and professionalization among AI developers; and . Continually re-evaluate policies in the context of research on social impacts. [277]
White House Report: Preparing for the Future of Artificial Intelligence	Regulatory agencies should: . Recruit and develop technical expertise in AI; . Develop a workforce with "more diverse perspectives on the current state of technological development"; . Use risk-assessment to identify regulatory needs; . Avoid increasing compliance costs or slowing development or adoption of beneficial innovations where possible; and . Avoid premature regulation that could stifle innovation and growth. [278]
Experimental Innovation Policy (OECD Report: Making Innovation Work)	The quality and efficiency of public expenditure on regulation targeted at innovation can be improved by an experimental approach to policymaking. Regulators should accordingly: . Embed diagnostic monitoring and evaluation into regulatory programs at the outset; . Collaborate closely with

[277] Stone et al., supra note 35, at 1011.

[278] Nat'l Sci. & Tech. Council, supra note 43, at 1-2, 17-18.

	private firms and non-governmental actors; and . Share and compare results of policy experimentation with other jurisdictions. [279]

Table 2 lists a set of strategies rather than broad theories. They are more practically applicable than theoretical. In that vein, many scholars have suggested specific regulatory tools that may be useful in regulating new technologies such as AI, which include the following:

. Enhancing flexibility through temporary regulation by using experimental legislation [280] and through sunset clauses to "define adaptable goals and enable the adjustment of laws and regulations according to the evolution of circumstances." [281]

. Creating "regulatory sandboxes" to allow firms to "roll out and test new ideas … without being forced to comply with the applicable set of rules and regulations." [282]

. Developing "anticipatory rulemaking" [283] techniques that leverage feedback processes to enable "rulemakers to adapt to regulatory contingencies if and when they arise because a [*444] feedback effect provides relevant, timely, decentralized, and institution-specific information ex-ante." [284]

. Making increased use of data analysis to identify what, when, and how to regulate. [285]

. Utilizing the iterative development of the common law to adapt rules to new technological contexts where possible, and developing new specialist regulatory agencies where they are particularly needed. [286]

. Using "legal foresighting" to identify and explore possible future legal developments, in order to discover shared values, develop shared lexicons, forge a common vision of the future, and take steps to realize that vision. [287]

. Creating new multi-stakeholder fora to help overcome information and uncertainty issues that stifle innovation or inhibit effective regulation. [288]

---

[279] OECD & World Bank, Making Innovation Policy Work: Learning From Experimentation 4 (Mark A. Dutz et al. eds., 2014), *http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/making-innovation-policy-work_9789264185739-en#.WeQWXkzMzUo* [*https://perma.cc/DDU5-SWH5*].

[280] Fenwick, Kaal & Vermeulen, supra note 175, at 24 (recommending to engage in policy and regulatory experiments by comparing different regulatory regimes and embracing "contingency, flexibility and an openness to the new").

[281] Ranchordas, supra note 23, at 212. See generally Romano, supra note 186 (discussing regulation of financial markets through sunset provisions).

[282] Fenwick, Kaal & Vermeulen, supra note 175, at 25.

[283] Kaal, supra note 23, at 19-20.

[284] Wulf A. Kaal & Erik P. M. Vermeulen, How to Regulate Disruptive Innovation - From Facts to Data, 57 Jurimetrics (forthcoming 2017) (manuscript at 25), *https://papers.ssrn.com/abstract=2808044*.

[285] Id. at 2; see Jason Potts, Brett Henderson & Gerard Roe, Detecting New Industry Emergence Using Government Data: A New Analytic Approach to Regional Innovation Policy 3 (Apr. 12, 2016) (unpublished manuscript), *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2763978*.

[286] Scherer, supra note 22, at 395.

[287] Laurie, Harmon & Arzuaga, supra note 175, at 3.

[288] Mandel, supra note 178, at 1, 4.

While there is no shortage of suggested regulatory responses, it is hard to distil a clear set of concrete recommendations from the wide and varied literature. This may partly be due to the disparate nature of AI, including the definitional problems outlined in Part II. Ultimately, one of the key problems is that while there are common regulatory challenges across different areas of innovation and technology policy, there are also highly context-specific challenges. [289] Ensuring regulatory approaches are closely connected with their contexts requires individual responses to different technologies in different locations at different times. As Roger Brownsword points out, this means that inevitably, "the details of the regulatory regime will always reflect a tension between the need for flexibility (if regulation is to move with the technology) and the demand for predictability and consistency (if regulatees are to know where they stand)." [290] Brownsword concluded that "while we should try to develop stock (tried and trusted) responses to challenges that we know [*445] to be generic, simple transplantation of a particular regulatory response from one technology to another is not always appropriate." [291]

The spectrum of regulatory approaches from command and control to self-regulation or peer regulation presents a quandary for those trying to regulate in this area. There is no quick fix that can be implemented to resolve the problems which the Authors have outlined. The next Part considers some practical and innovative means to begin the process of regulating the development of AI that includes considering a number of tools from within self-regulation and risk regulation theories. The Authors conclude that while these theories may eventually influence the regulation of AI, there is currently a moment in time where all of the stakeholders may be able to influence the development and regulation of AI through cooperation and collaboration in the nascent stages of development. In this way, all stakeholders can have a role and a stake in the way that regulation develops. This may take the form of overt self-imposed industry codes of practice or conduct from the participants [292] and involve less intrusive and direct guidance from public regulators - what might be termed a "nudge."

V. Strategies to Regulate Artificial Intelligence

Part IV outlined a number of theories of regulation and detailed some of their deficiencies when it comes to regulating AI. The Authors also outlined some theories of regulation that may not be applicable to regulating AI. This Part argues that in the lag time it takes to properly devise an appropriate regulatory structure to address AI, public regulatory bodies should begin to exert their influence on the nascent development of AI so as to broadly guide its development in beneficial ways. The Authors then suggest that public regulators should begin to develop risk-based strategies to most effectively target their limited regulatory resources. Regulating the risk profile for AI outlined in Part III requires a staggered approach where the highest risks, as assessed by public regulators, are addressed first. At the very least, regulators should be taking steps now to establish what risks pertain to the various classes and types of AI and should be in a position to regulate if eventually required. However, as governments have so far shown an inability to engage with regulation in this area, the Authors suggest that there is a broader and more immediate role for the state in influencing the development of AI systems but that doing so well will require some [*446] innovation in regulatory practices. The Authors further suggest that this can be done immediately while the harder and more onerous task of preparing risk profiles can happen over a longer term. The recent Stanford Report recommended a "vigorous and informed debate" to "steer AI in ways that enrich our lives and our society." [293] How government regulators may actually be able to steer AI development, however, is a crucial and, as yet, unanswered question. This Part considers how public regulatory agencies may be able to adopt

---

[289] Roger Brownsword & Karen Yeung, Regulating Technologies: Tools, Targets and Thematics, in Regulating Technologies, supra note 14, at 3, 6.

[290] Brownsword, supra note 14, at 27.

[291] Brownsword & Yeung, supra note 289, at 6; see Brownsword, supra note 14, at 32.

[292] See discussion supra Part II.

[293] Stone et al., supra note 35, at 49.

strategies to "nudge" [294] the development of AI. In this way, regulators may be able to influence those responsible for designing and deploying AI systems to do so in a way that furthers the public interest.

A. The Influence or "Nudging" of Regulators

Much has been made of nudge theory in recent years. [295] Psychological observations as applied in behavioral economics reveal that normative human behavior can be skewed or distorted by inherent human biases. [296] Nudge theory proposes that by exploiting these biases, humans can be nudged to behave so as to achieve an outcome desired by the nudger. The theory has tended to focus on nudging individual behaviors. However, recent work has examined how behavioral economics approaches might influence a broader spectrum of decision-makers. [297] In an example used in a study of environmental policymaking, Elke Weber argued that "decisions could be reframed in ways that might affect choices [by] changing the focus of such decisions from individuals to groups." [298] She argued that "cultures that emphasize the importance of affiliation and social goals over autonomy and individual goals have been shown to influence the way in which decisions under risk and uncertainty get made." [299] Weber argued further that "the goal of environmental policy is to change behavior of companies, governing boards and committees, and members of the general public in the direction of [*447] more sustainable, long-term, and socially and environmentally responsible actions." [300] Weber concluded that "conventional policy interventions are not using the full range of goals that motivate behavior and changes in behavior … [and] do not utilize the full range of processes that people use to decide on a course of action." [301] These regulatory interventions apply the idea of nudging in its broadest sense. It is not only the behavior of the individual that can be the target of behavioral policymaking. The theory can be used to influence those who govern companies, such as boards of directors. In this way, regulatory policy can shape the behaviors of companies and, even more broadly, groups of companies within industries.

In Weber's example, policies are directed to influence the environmental responsibility of companies. [302] The Authors argue that similarly broad policies directed at companies developing AI would begin to influence or guide beneficial behaviors by those companies. If governments are unable to fully participate yet in gathering information because of resource constraints or because of the diffuse nature of AI development, they can begin to shape the behavioral environment by proposing policy statements that foster the beneficial and benign development of AI. This approach has several immediate benefits for public regulators. Foremost, it is relatively inexpensive; it does not require a great deal of investment to be able to set broad policy indicators that outline the regulators' attitude to AI development. It also would buy the regulator time to take on the task of fully engaging with the regulatory environment as outlined in this Article: .

B. Examples of Influence as Regulation

---

[294] See generally Richard H. Thaler & Cass R. Sunstein, Nudge: Improving Decisions About Health, Wealth, and Happiness 5 (2008) (explaining the concept of libertarian paternalism).

[295] See, e.g., id. at 6-8.

[296] See generally Kahneman, supra note 189 (explaining how biases present themselves).

[297] Saurabh Bhargava & George Loewenstein, Behavioral Economics and Public Policy 102: Beyond Nudging, 105 Am. Econ. Rev. 396, 396-401 (2015); Brigitte C. Madrian, Applying Insights from Behavioral Economics to Policy Design, 6 Ann. Rev. Econ. 663, 663-88 (2014).

[298] Elke U. Weber, Doing the Right Thing Willingly: Using the Insights of Behavioral Decision Research for Better Environmental Decisions, in The Behavioral Foundations of Public Policy 380, 388 (Eldar Shafir ed., 2013).

[299] Id.

[300] Id. at 391.

[301] Id.

[302] Id.

The approach of the US government in attempting to shape behaviors of those developing AI is in its infancy. As an early indicator, the government has shown that it was, until recently, prepared to consult with groups of stakeholders. In 2016, the Office of Science and Technology Policy (OSTP) conducted a series of public workshops held at the University of Washington, Stanford University, Carnegie Mellon University, and New York University. [303] The OSTP also participated in various industry conferences and sought public comment in the form of a Request for Information. [304] As a further signal of its policy intention to shape behaviors, the US government [*448] also published two documents: Preparing for the Future of Artificial Intelligence, [305] and The National Artificial Intelligence Research and Development Strategic Plan. [306] The latter states:

AI presents some risks in several areas, from jobs and the economy to safety, ethical, and legal questions. Thus, as AI science and technology develop, the Federal government must also invest in research to better understand what the implications are for AI for all these realms, and to address these implications by developing AI systems that align with ethical, legal, and societal goals. [307]

The Preparing for the Future of Artificial Intelligence report made twenty-three recommendations on what government agencies, schools and universities, and AI professionals could do to prepare for the future of AI. [308] The Authors argue that this document on its own had the effect of engaging with and shaping or influencing the development of AI. Evidence for the immediate impact that the report had includes that it was adopted by Partnership on AI. [309]

These strategies might be seen in a number of different ways. First, the government is seen to be consultative, and is attempting to engage with stakeholders in the area. Kenneth Abbott noted that "modern regulatory policy, including risk regulation policy, views public communication, input and participation as essential." [310] He cited the 2012 OECD recommendations on regulatory policy that "call for "open government,' including transparency and communication, stakeholder engagement throughout the regulatory process, and open and balanced public consultations." [311] Second, it could be seen as an information-gathering exercise - a necessary first step in the risk regulation literature as well as in behavioral economics theories. Third, the government could be seen to be signposting its intention to regulate if necessary.

The US government, by engaging with AI and those responsible for developing it and publishing its stated intentions, sent a clear signal to all those involved in the developing field of AI. It showed that the government was engaged in the conversations and was prepared to stake a claim in the game. This also may be seen as the government seeking to influence or nudge decision makers in the AI industry and to shape behaviors within that field. The government's emphasis on beneficial development clearly articulates its intentions and focus and sends a clear signal to the entire industry [*449] in the United States and, more broadly, in the Western world. Because many of the companies that develop AI are based in the United States, such a clear policy signal from the US government would obviously have an influential effect on the behaviors of the major AI companies and the people who work within them.

---

[303] Nat'l Sci. & Tech. Council, supra note 43, at 12.

[304] Id.

[305] Id.

[306] Nat'l Sci. & Tech. Council, supra note 251, at v.

[307] Id. at 15.

[308] Nat'l Sci. & Tech. Council, supra note 43, at 40-42.

[309] Partnership on AI Expresses Support for White House Report on Artificial Intelligence, supra note 251.

[310] Abbott, supra note 75, at 10.

[311] Id.

However, in a worrying development, the US government appears to have retreated from its laudable participation in the development of AI. The Preparing for the Future of Artificial Intelligence report has been removed from the government's website and archived. [312] Similarly, the government's National Artificial Intelligence Research and Development Strategic Plan [313] is no longer available on the White House website and is presumably no longer government policy. Retreating from its former position sends an altogether different and equally strong message: that regulation is not a priority, regulation is not going to happen in the near future, and the government is, at least for now, uninterested in the development of AI. Therefore, if AI is to be regulated in any meaningful way in the absence of government direction, it may well be up to those developing the AI to control its development. However, this is hardly the ideal solution. It is unfortunate that the planned policy no longer can have the influential effect that it once had.

While the US government has retreated from its role as influencing the development of AI, the European Parliament has taken positive steps. In February 2017, it passed a resolution to recommend to the EU Commission (EC) to develop Civil Law Rules on Robotics (which included AI). [314] The resolution recommends that the European Union adopt rules on liability for issues arising from robots and AI [315] and also recommends that the EC designate a European Agency for Robotics and Artificial Intelligence to govern robotics and AI. The Agency would

provide the technical, ethical and regulatory expertise needed to support the relevant public actors, at both Union and Member State level, in their efforts to ensure a timely, ethical and well-informed response to the new opportunities and challenges, in particular those of a cross-border nature, arising from technological developments in robotics, such as in the transport sector. [316]

[*450] The resolution recommends a system of registration of so-called "smart robots," the definition of which is wide enough to capture AI. The registration would apply across the European Union. [317] The resolution also recommends developing a Code of Ethical Conduct for researchers and designers in robotics and AI to "act responsibly and with absolute consideration for the need to respect the dignity, privacy and safety of humans." [318] This move by the European Parliament and the EC sends a clear signal to the industry intended to influence the research, development, and design of robots and AI, at least in Europe. Once set up, the Agency for Robotics and Artificial Intelligence will begin to gather the much-needed technical, ethical, and regulatory expertise to begin the regulatory process. This initiative represents the most advanced work towards regulation of AI today and should be lauded as a model for the rest of the world.

There is therefore a place for government policy to shape the behavior of those in the AI industry. At the same time, though, more needs to be done to begin the process of developing regulation. Part IV proposed that a Really Really Responsive Risk-Based Regulatory framework will be most effective. The risk-based regulatory approach will allow regulatory bodies to target their intervention to the most pressing elements of AI development based upon a risk analysis.

C. Risk-Based Regulation of AI

---

[312] A search of the White House website returns no results for Preparing for the Future of Artificial Intelligence. The document can be found at Nat'l Sci. & Tech. Council, supra note 43.

[313] Nat'l Sci. & Tech. Council, supra note 251.

[314] See Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics, Eur. Parl. Doc. TA 51 (2017), http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP// NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN [https://perma.cc/HB6N-JP7V].

[315] See EU Robotics Report, supra note 264, at 20.

[316] Id. at 10.

[317] Id. at 20.

[318] Id. at 21.

Parts III and IV outlined a number of risk profiles for various classes of AI. Given the risks posed by AI, it is appropriate that regulation responds to those risks. Risk-based frameworks usually entail the following sequence: first, the regulator sets the level and type of risks it will tolerate; second, the regulator conducts some form of risk assessment and assesses the likelihood of the risk eventuating; third, the regulator will evaluate the risk and rank the regulated entities on their level of risk - high, medium or low; and fourth, the regulator will allocate resources according to the level of risk that it has assessed. [319] These tasks are usually carried out by a regulatory agency after consultation with those within the industry.

In the regulation of AI, then, public regulators must undertake a risk analysis of current applications of AI. After the regulator has assessed and set the level of risk that it might tolerate, it must gather as much information about the state of affairs as possible. This can be done by consulting with those already in the industry and [*451] participating in or organizing information sessions such as roundtables that involve all relevant stakeholders. The risks can only be properly assessed with all relevant information. Only when public regulatory agencies or governments are aware of the issues will they be in a position to properly rank the risks that meet or exceed their tolerance levels and to allocate the necessary resources to regulate the risks involved. The Authors' initial triage of risks posed by various applications of AI in Part III could then be refined and further developed in a feedback loop after multiple consultation processes.

Van Asselt and Renn emphasized the need for communication and inclusion when assessing risk. They argued that "various actors are included, [and] play a key role in framing [the] risk." [320] This inclusion includes "roundtables, open forums, negotiated rule-making exercises, mediation, or mixed advisory committees, including scientists and stakeholders." [321] They emphasized that "it is important to know what the various actors label as risk problems. In that view, inclusion is a means to an end: integration of all relevant knowledge and inclusion of all relevant concerns." [322] The participants, they argue, should include "a range of actors which have complementary roles and diverging interests." [323] Bridget Hutter also noted that to achieve regulatory excellence, "regulators must have access to accurate information so that they have a clear idea of the risks they are regulating." [324] As outlined in Part IV, relevant industry parties are forming industry-level associations and groups to share information and agree on principles and shared values. As discussed, this has already resulted in a range of principles and proposed standards by which many in the industry have agreed to be bound. However, government and regulatory bodies must now engage in the process. The US government in particular, up until recently, had shown that it was willing to take the lead in this information gathering and sharing phase of the regulatory process. [325] It is essential for the government to continue this level of involvement if it is to put itself in a position to be able to regulate effectively. Without such involvement, it will continue to have little influence on the direction that AI development takes. At the same time, regulatory bodies need to begin to assess and rank the various risks associated with AI applications.

[*452]

D. Classifying the Risks

The high costs of, and challenges to, effective regulatory intervention require that the attention of regulators should be carefully focused on the areas posing the greatest risk. The Authors argue in Part III that different claims to AI can be refined into, at the very least, three broad subcategories based upon whether the AI (a) is a narrow and single-use AI, (b) displays some characteristics of operating autonomously or may pursue its own goals, or (c) is or

---

[319] See Black & Baldwin, supra note 271, at 184-85.

[320] van Asselt & Renn, supra note 147, at 440.

[321] Id.

[322] Id. at 441.

[323] Id.

[324] Hutter, supra note 142, at 104.

[325] See Nat'l Sci. & Tech. Council, supra note 43, at 12; Nat'l Sci. & Tech. Council, supra note 251, at 34.

displays some of the characteristics of AGI. Each of these classes poses different risks, and those risks vary within classes depending on the AI application. Within each category, there are many subcategories of application. Relatively benign applications of AI such as in Roomba, Pandora, or simple game applications can be placed within the low-risk category. On the next level, we include more robust applications such as the AI in AlphaGo, the more experimental aspects of AI work carried out by Google, Facebook, Microsoft, Apple, Amazon, and any other large player experimenting with AI, as referred to in Part II above. [326] The third class includes the more concerning aspects associated with experiments seeking to attain AGI. This group would include research conducted by mathematicians and engineers who seek to create either a self-replicating AI or AGI without concern or knowledge aforethought for its ultimate capabilities. [327] The problem with regulating AI identified in Part II is that each of these applications could plausibly lay claim to being, applying, or using AI. However, each category does not and cannot justify or require the same regulatory response, and some applications may not even require a regulatory response at this stage. [328] It is only when the risk profile of an AI application increases that a regulatory response may be required. For example, more and less risky applications of AI will exist within a single class of AI (e.g., narrow AI). However, without a risk analysis, the level of risk of each application within a class is as yet unascertained.

[*453] The class of AI that poses the greatest risk to humanity as a systemic risk is AGI. The Authors discuss AGI, even though it does not currently exist, because it requires an immediate regulatory response if it is indeed not already too late to regulate its research and development. AI professionals are already experimenting with self-replication and AI autonomy. While these experiments do not yet reach the level of AGI, they remain a very high potential, and perhaps imminent, risk. If one of these experiments, through accident or serendipity, creates a form of AGI, then the concerns expressed by many in the industry become reality, and the chance to control its behavior may well be lost. Lethal autonomous weapons also pose an extremely high risk to human well-being, but this subcategory of AI application is subject to its own unique regulatory environment and is outside the scope of this Article: . [329]

A further complication in regulating AI using a risk-based strategy arises because none of the risks or classes of AI are static. The level of risk posed by applications within each class may increase or decrease. Various push and pull factors will move the applications in each class up and down depending on features that either ameliorate or accentuate the risks associated with its use. The risks posed by narrow applications may become stronger and hence may ultimately become AGI. The question for regulators is at what point they should intervene. Should they begin to regulate as soon as AI poses some risk, or should they wait until an imminent risk is apparent? A further complication is that, at this stage, relevant regulators are not even in a position to discern which application of AI fits within which class. No clear system of classification currently exists. The Authors' suggestion is to begin classifying based on the level of risk each application currently poses.

Yet a further complication arises because the same public regulator or regulatory agency will not regulate all (or even more than one) of the applications within each class. The identified classes are separated broadly by risk factors and not by application type. So, even though they may be in the same class and on the same level of risk for the purposes of our classification, the regulators who might respond to concerns raised by the use of AI in

---

[326] See supra text accompanying notes 76-81.

[327] See James Babcock, Janos Kramar & Roman Yampolskiy, The AGI Containment Problem, in Artificial General Intelligence: 9th International Conference 2 (2016); Orseau & Armstrong, supra note 67, at 558; see also Eliezer Yudkowsky & Marcello Herreshoff, Tiling Agents for Self-Modifying AI, and the Lobian Obstacle 2 (Oct. 7, 2013) (unpublished manuscript), https://intelligence.org/files/TilingAgentsDraft.pdf [https://perma.cc/3DE5-WUY3]; Paul Christiano, Cryptographic Boxes for Unfriendly AI, LessWrong (Dec. 18, 2010, 8:28 AM),http://lesswrong.com/lw/3cz/cryptographic_boxes_for_unfriendly_ai/ [http://perma.cc/Q8VY-2RJ2].

[328] See Omohundro, supra note 41, at 162; Omohundro, supra note 29, at 483. There are some advantages for businesses that publicly advertise their product's use of AI. As a marketing ploy, therefore, businesses sometimes claim that their product uses AI when it actually does not. See, e.g., Prakash, supra note 42.

[329] For a thorough investigation of this topic, including suggestions on possible regulation in the area, see Dustin A. Lewis, Gabriella Blum & Naz K. Modirzadeh, War-Algorithm Accountability 98 (2016).

autonomous vehicles will not be required to consider, for example, Google's use of AI to reduce electricity consumption in its data centers.

The Authors suggest that it is the role of governments and regulatory bodies to begin to influence the direction that AI is to take [*454] at a broad level and to attempt to intercede now in its development. In Parts V.A and V.B, the Authors provided suggestions on the role of public regulators as to how this might be done. In the meantime, this Article: 's initial proposal is for governments or public regulators to take steps towards regulating AI by obtaining information, joining and commencing conversations with stakeholders in the industries that use AI, and influencing the development of AI in ways beneficial to society. The Authors contend that the most dangerous (and as-yet unattained) class of AI - AGI - should be regulated now, and serious questions about its development should be considered and discussed among AI professionals now.

VI. Conclusion

On May 21, 1946, as scientists were still experimenting with the new power of nuclear energy, Louis Slotin, a Canadian physicist who had worked on the Manhattan project to develop nuclear weapons during World War II, was preparing to conduct an experiment in a lab in the New Mexico desert. [330] Slotin was slowly lowering a hemispherical beryllium tamper over a piece of plutonium to excite the neutrons that were emitting from the plutonium core. This process would create a small nuclear reaction so that the scientists could measure the results. [331] The process was aptly referred to as "tickling the dragon's tail." [332] Slotin slipped and dropped the beryllium tamper directly onto the core, causing a momentary but powerful reaction that irradiated the whole room. Slotin bore the brunt of the reaction. [333] He died a painful death nine days later from radiation poisoning. [334]

Seventy years later, AI scientists, engineers, and technicians are experimenting with a new scientific development with potentially destructive capabilities. If we are to heed the allegory in the golem stories or the metaphor of the dragon's tail, society must come to the conclusion that any such danger, no matter its potential, should be carefully handled. The Authors do not suggest a draconian, command and control type of regulation, and do not even think it would work. However, the Authors do suggest a new and more nuanced, responsive, and adaptive regulation developed to foster innovation and [*455] to minimize the risks of AI. This approach, as with the approach in relation to the treatment of nuclear weapons, needs a global solution and will not be easy.

In the last two decades, the face of technology, the institutions involved, and therefore the AI regulatory space has changed dramatically. This period has seen the rise of some of the biggest technology companies - including Microsoft, Apple, Facebook, and Google - as major leaders in AI. It is arguable that in terms of new technology development, including AI, these companies hold the lion's share of regulatory resources. [335] Public regulators, by contrast, appear to be increasingly in the difficult position of needing to find mechanisms to regulate technology they have only limited capabilities to understand, by influencing firms that are very well-resourced and connected and that can exercise substantial choice about the jurisdictions in which they operate.

There are encouraging signs from recent publications - certainly the emphasis on more research into the social impacts of AI by both the US government and private coalitions is encouraging. Still, the rhetoric of avoiding overregulation is worrying - even the biggest and most well-resourced government regulators are hesitant and probably will not be particularly well equipped to deal with this issue any time soon. For smaller regulators -

---

[330] Martin Zeilig, Dr. Louis Slotin and "The Invisible Killer", Beaver, Aug.-Sept. 1995, at 20-26.

[331] Id.

[332] See Alex Wellerstein, The Demon Core and the Strange Death of Louis Slotin, New Yorker (May 21, 2016), *http://www.newyorker.com/tech/elements/demon-core-the-strange-death-of-louis-slotin* [*http://perma.cc/SF8E-VMRU*].

[333] Zeilig, supra note 330, at 20-26.

[334] Id.

[335] See Hood & Margetts, supra note 216, at 5-6 (discussing resources of nodality, authority, treasure, and organization).

including those outside of the United States, there is almost no chance of successfully intervening in current technological development. Governments are left to try to influence or nudge the development of AI at the broad policy level. This remains one of the only roles that might remain available to government, given the changing power dynamics between government and these large companies.

There are benefits to self-regulation, particularly where public regulators lack the requisite knowledge to understand the problem that needs regulating. Self-regulation has in its favor that it involves iterative and cooperative development of standards with input from various stakeholders at the coalface of the problem. The downside to self-regulation is that it works best where there is some imminent threat of state-based penalty for noncompliance. As discussed, governments are at a disadvantage, probably for the first time in history at this scale, against the major corporate stakeholders in AI.

The US government is perhaps best able to shape the development of AI because many of the major AI companies are based in the United States. Recent studies in behavioral policymaking [*456] suggest that the attitudinal settings of people within groups shape the development of the group. The government recently set about the task of informing itself about AI, and it has set out both a strategic and a research and development policy that seeks to influence beneficial development of AI. By setting out its agenda as it has and by investing in collaboration with industry participants, the US government had set a positive benchmark that sought to sway participants in the field. Whether this can be called nudging or not is moot, but the intention was clear. However, the government has more recently retreated from this stance; this is regrettable. The latest retrograde steps send an equal and opposite message to AI developers. In a positive sign, though, the European Union has taken proactive steps toward regulation of robots and AI, and other countries might do well to replicate its example.

Because regulators do not yet have the expertise or even enough information to create expertise, if we are ever to ensure AI is developed in a way that is beneficial for humanity, developers must acknowledge both their social obligation to share information (be transparent and accountable) with others, and the critical importance of collaborations with thinkers from other disciplines. The ethics board set up by DeepMind and Google, and the Partnership on AI, are great examples of this. However, the problems that face potential regulators attempting to regulate such a dynamic field illustrate that more collaboration and information sharing between all relevant parties is required if society is to safely reap the benefits of AI.

The risks that different classes of AI pose lie along a spectrum. Similarly, the different applications of AI pose different and variable risks within the field in which they are applied. Public regulators must begin to engage with researchers and professionals in the area to gain the necessary information required to be able to identify and regulate in relation to the greatest risks that AI poses. By adopting a risk-based approach, public regulators will be able to target their approaches to achieve the most efficient and effective regulatory outcomes.

Vanderbilt Journal of Entertainment and Technology Law

Vanderbilt Journal of Entertainment and Technology Law