The Inevitability of AI Law & Policy:

Preparing Government for the Era of Autonomous Machines

Ryan Clough
Public Knowledge

October 2018

This report was made possible by the generous support of the Nielsen Foundation.

Executive Summary

Artificial intelligence is emerging as a transformative technology, and is likely to grow ever more ubiquitous in the coming years. On the immediate horizon, experts in many fields foresee autonomous systems taking over activities that are central to daily life, and that previously could be done only by humans--from driving cars to diagnosing medical conditions to producing works of art and beyond.

As big data and autonomous systems become predominant, legal and political scrutiny is ramping up as well. This paper calls for a new federal authority to build the capacities necessary to govern these technologies. To be clear, this does not mean a new overarching regulator that would replace existing agencies, as sector-specific regulation of autonomous systems will almost certainly continue. Nor do we propose the immediate creation of a sweeping new rulemaking regime. Instead, the new authority would build and provide expertise and experience in AI for the rest of the government, which will be essential for at least five concrete priorities:

1. Boosting Sector-Specific Regulators and Confronting Overarching Policy Challenges Raised by AI. While sector-specific regulators will inevitably lead on the regulation of AI in many specific industries, there are drawbacks to relying solely on their narrow mandates, such as the risk of tunnel vision. An AI authority would help sector-specific regulators confront major repeatable policy challenges that cut across their domains, such as fairness, transparency, accountability, privacy, and human autonomy. This will be especially crucial given AI's well-documented and inherent potential for exploitation and disempowerment. Given the unpredictability, opacity, and power of current-generation AI, any individual can be mistreated within a larger automated system, and many historically-vulnerable groups face the most severe risks and harms. To ensure that the enormous potential benefits of AI accrue broadly across our society, government must build more capacity to protect fundamental rights at stake.

2. Protecting Public Values in Government Procurement and Implementation of AI. Governments are already implementing AI in all sorts of specific ways, but public procurement of autonomous systems raises unique challenges, such as how to effectuate public policy goals in AI design, and how to avoid unique risks of lock-in with AI vendors. An AI authority would be an indispensable advisor on public AI projects, ensuring that other parts of the government follow best practices with public AI projects (for example, to allow for sufficient transparency to the public) and avoid common pitfalls in their relationships with the private sector.

3. Attracting AI Practitioners to Civil Service, and Building Centralized and Durable AI Expertise Within Government. Other federal government agencies, as well as states and localities, do not have the resources to succeed at this on their own. And even if they did, their expertise will often be isolated, unavailable to other agencies. A new AI authority would build a cadre of AI experts and repeatable experience on many different policy issues, consulting and coordinating with other parts of the government and boosting its capacities as a whole.

4. Identifying Major Gaps in the Laws and Regulatory Frameworks that Govern AI. An AI authority would be crucial in identifying major policy issues that do not fit well (or at all) into existing laws and jurisdictions. For example, AI is likely to raise overarching questions and problems for both liability and discrimination law. An AI authority could also prepare the government for even larger, long-term policy challenges, such as the possible economic disruption caused by AI, and the safety concerns raised by general AI, if and when it develops.

5. Coordinating Strategies and Priorities for International AI Governance. Many other governments, including China, are pouring resources and attention into AI. An AI authority would assess international policy priorities in AI, and defend American interests and values as the technology and governance of AI continues to evolve around the world.

Introduction

Today, artificial intelligence is used to detect heart attacks. Hire new workers. Translate intricate texts. Drive cars. Predict crime and deploy more police to a neighborhood. Take customer service calls. Review tax returns. Decide which news, products and search results we see online. And much, much more.

Over the next decade, AI will become ever more intertwined with our economy, our society, our government, and our daily lives. Top executives at the world's biggest technology firms describe AI in reverential terms: as the "new electricity,"[1] "the most significant development in computing in my lifetime,"[2] and "one of the most important things that humanity is working on."[3] According to one forecast, seventy percent of companies will be operating an AI system by 2030, boosting economic output by $13 trillion.[4] China, the European Union, and many other countries have announced national strategies on artificial intelligence, investing billions in research and commercialization. In the United States, the Defense Department just announced another $2 billion in research funding for AI,[5] and government officials from both parties have spoken about AI's "potential to disrupt every sector of society in both anticipated and unanticipated ways."[6]

As AI becomes ever more ubiquitous, it will inevitably demand more resources and attention from government -- both in regulating AI in the private sector and in implementing AI directly in the public sector. But the current generation of artificial intelligence poses profound and novel challenges for governance. Broadly speaking, AI is different from other information technologies in at least three crucial respects:

1. **Emergence**: by its very nature, AI behaves in unpredictable ways, not fully planned or foreseen by humans. As the law professor Ryan Calo explains, this

---

[1] Ted Greenwald, *What Exactly Is Artificial Intelligence, Anyway?*, Wall Street Journal (Apr. 30, 2018).
[2] Sergey Brin, *Alphabet 2017 Founders' Letter*, https://abc.xyz/investor/founders-letters/2017/index.html.
[3] Drew Harwell, *Facebook, boosting artificial-intelligence research, says it's 'not going fast enough,'* Washington Post (July 17, 2018).
[4] Jacques Bughin, *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute Discussion Paper (September 2018).
[5] Matt McFarland, *The Pentagon is Investing $2 billion into artificial intelligence*, CNN Business (September 7, 2018), https://money.cnn.com/2018/09/07/technology/darpa-artificial-intelligence/index.html.
[6] U.S. H.R. Comm. on Oversight and Gov't Reform, Subcomm. on Info. Tech., Rise of the Machines: Artificial Intelligence and Its Growing Impact on U.S. Policy at 1 (2018).

"can lead to solutions no human would have come to on her own," with "something approaching creativity."[7]

2. **Opacity**: the emergent capabilities of today's AI systems are based primarily on machine learning algorithms, which draw dense webs of association, interpretation, and inference among massive amounts of data, building models for prediction and decision making in comparable situations. The scale and complexity of these models are typically beyond full human comprehension. And as algorithmic systems grow ever more complicated, and even begin to interact with each other, the result may be software that "no one can fully understand."[8]
3. **Power over human behavior**: while other information technologies can certainly exert powerful influence over humans, the potential capabilities of AI are on a different level. AI can make precise predictions about what a person will and will not do in a given situation, and tailor its interactions accordingly. In many instances, this could allow AI to manipulate our choices and behavior without our full awareness.

For AI systems with these core characteristics, government may have significant trouble applying existing legal and administrative regimes, which evolved to govern humans instead of unpredictable and autonomous machines.

The enormous potential of artificial intelligence is hard to dispute. If it achieves even half of what many AI optimists expect, these systems could generate enormous wealth and productivity, supercharging human capabilities and freeing us to pursue more rewarding lives and relationships. But along with this promise comes major risks, especially for the most marginalized and vulnerable in our society. It is not an exaggeration to say that AI has an inherent potential for exploitation and disempowerment. The recent history of big data and algorithmic decision-making is full of examples and evidence of systematic bias and discrimination. Even if unintended by their designers, AI systems have disfavored female applicants to a medical school, mistakenly predicted recidivism more often for black offenders than for whites, directed resources for road repair away from poorer neighborhoods, and made it less likely that

[7] Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 Calif. L. Rev. 513-63 (2015). .
[8] S. Andrew Smith, *Franken-algorithms: the deadly consequences of unpredictable code*, The Guardian, Aug. 30, 2018.

women would be shown advertisements for higher-paying jobs. Any individual can be misjudged and mistreated by artificial intelligence, but the record to date indicates that it is significantly more likely to happen to the less powerful, who also have less recourse to do anything about it.

Ultimately, these questions invoke fundamental rights and values, which are far more than just technical matters or the exclusive domain of engineers and technology firms. As the owners and operators of dominant AI systems become more and more like digital sovereigns, we risk ceding self-determination, both individually and as a democracy.

Five Central Challenges in the Governance of Artificial Intelligence that Call for a New AI Authority

1. Boosting Sector-Specific Regulators and Confronting Overarching Policy Challenges Raised by AI

Artificial intelligence is already prevalent in many different industries, and as the capabilities and applications of AI continue to grow, it will reach ever more broadly into our society, our economy, and in daily life. In this way, we can compare the growth and ubiquity of AI to the rise of the internet towards the end of the last century. In the latter case, what began as an obscure network for a few government departments and academics steadily evolved over time, to the point today that it is scarcely possible to think of a part of life that it has not transformed in some major way. Although the precise timeline and milestones may vary in unpredictable ways, we should expect the same transformation and ubiquity from autonomous systems.

As AI reaches new fields and activities, it will reach into the domains of an ever-greater number of existing, “sector-specific” government agencies and jurisdictions. This has long been the model for much of the administrative state. While there are many important laws of general application, the twentieth century saw enormous growth in the number and importance sector-specific regulators, which were necessary to confront

the increasingly complicated and specialized policymaking questions raised by an advanced economy.

We should expect the primacy of sector-specific regulation to continue in the age of AI. Concepts such as big data, algorithmic decision-making, and artificial intelligence encompass a multitude of technologies and applications, defying easy categorization or broad policy unity. As Andrew Burt argues, "regulating an assemblage of technology we can't clearly define is a recipe for poor laws and even worse technology."[9] Furthermore, many of the policy concerns and priorities will vary from field to field--indeed, they may not even be predictable until autonomous applications are brought to market. For example, self-driving cars may raise one set of concerns focused on traffic safety, while the use of AI by financial institutions may raise entirely different questions focused on market stability and fairness. Many of these situations would be best addressed--and in any case, will be addressed--by the existing sector-specific regulators. As the Office of Science and Technology Policy reported after soliciting public comments on AI policy, "the general consensus … was that broad regulation of AI research or practice would be inadvisable at this time," and instead supported the "adaptation" of existing regulation "to account for the effects of AI."[10]

However, even if sector-specific regulation of AI is both necessary and inevitable, it also has significant shortcomings. First, as described above, individual agencies may lack the technical capacities necessary to work with complex technologies and do their jobs. Second, many sector-specific regulators will be prone to "tunnel vision"--becoming "unduly focused on carrying out their narrow mission without attention to broader side effects of regulatory choices."[11] For example, as is evident from its name, the primary mission of the National Highway Traffic Safety Administration is to make cars safer and less crash-prone. We shouldn't expect NHTSA to give equal weight to other priorities, such as environmental impact and competition in the automotive industry--indeed, we would probably object if it did. Tunnel vision becomes especially relevant when a specialized agency takes the lead on a multifaceted problem, without other agencies to

[9] Andrew Burt, *Leave A.I. Alone*, N.Y. Times, Jan. 4, 2018.
[10] Nat'l Sci. & Tech. Council, Comm. on Tech, Exec. Office of the President, *Preparing for the Future of Artificial Intelligence* at 17 (Oct. 2016).
[11] Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83, 113 (2017)

represent competing views and priorities. Regulators will often need to strike the right balance between competing legitimate priorities--for example, protecting privacy versus enabling competition through more open data. There are strong reasons to doubt that many sector-specific regulators will excel at this on their own.

Although AI policy questions will vary by sector and application, there are also core problems that are endemic to autonomous systems, and that will pop up in many different contexts. Sector-specific regulators will often be at a disadvantage when confronting these overarching issues. They may be unable to access the relevant knowledge and experience accumulated elsewhere in the government, and may not preserve or translate the lessons to be learned from their own work.[12] As a matter of efficiency, disparate agencies may spend more time and money working out different solutions to a common problem, compared to what a single agency would formulate as a repeatable approach. Furthermore, piecemeal regulation of the same essential issue could create a "thicket" of inconsistent judgments and mandates, creating more difficulties for the developers of autonomous systems.[13]

This is where a new AI authority would come in. To be clear, this paper is not calling for an overarching regulator and/or vast new legal authority over AI to supplant existing sector-specific agencies. Instead of regulating on its own, the new body would primarily consult with and advise existing regulators when they confront AI applications within their domains. This assistance could be crucial both for industry-specific regulators (e.g., NHTSA for transportation or the FAA for Aviation) as well as for agencies that focus on particular policy concerns across multiple sectors (such as DHS for cybersecurity or USPTO for intellectual property). As an example of the latter, take the Federal Trade Commission, which enforces consumer protection laws of general applicability. There are strong reasons to expect that the FTC should continue to lead on many consumer protection issues raised by AI, both now and in the future.[14] But just

[12] *Id.* at 114.
[13] Paul Ohm & Blake Reid, *Regulating Software When Everything Has Software*, 84 Geo. Wash. L. Rev. 1672, 1696-97 (2016).
[14] *See* Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 Md. L. Rev. 785, 813-14 (2015).

as the FTC already coordinates with and relies upon the specialized expertise of many other agencies,[15] it would do the same with an AI authority.[16]

Such coordination and advice could be especially crucial for common challenges in AI governance that arise across sector-specific boundaries. We should expect these sorts of transcendent issues to predominate with AI, in part because the emergent behavior in autonomous systems often will not be easily analogous to human behavior and decision-making, adding novelty and uncertainty to policy questions.[17] Even if we put aside the future challenges that may be raised by general artificial intelligence, and focus only on the immediate issues raised by machine learning and other AI technologies already in use today, there is a wide range of common questions that will emerge across different sectors, in which the expertise and coordination of an AI authority will be paramount. Here are several of the most important.

Fairness. Foremost among public concerns about AI are fears of "bias" in its formulation, implementation and outcomes. These worries are not hypothetical. The recent history of big data and machine learning is full of discriminatory predictions and disparate impacts, often mirroring historical patterns of marginalization (based on race, gender, income, geography, and so on). For example:

- An algorithm for medical school admissions "screened out qualified female and minority applicants because it was trained on the decisions made previously by a biased admissions board."[18]
- Facial recognition software made errors in less than 1 percent of cases when identifying the gender of white men, but was wrong about dark-skinned women up to 34 percent of the time.[19]

---

[15] *Id.* at 830-31.
[16] *See also* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. Davis L. Rev. 399, 427-28 (2017).
[17] *See Ryan* Calo, *The Case for a Federal Robotics Commission*, Brookings Inst. at 5-6 (2014).
[18] Ben Buchanan and Taylor Miller, *Machine Learning for Policymakers: What It Is and Why It Matters*, Belfer Center for Science and International Affairs at 32 (June 2017).
[19] Larry Hardesty, *Study finds gender and skin-type bias in commercial artificial-intelligence systems*, MIT News Office (Feb. 11, 2018), http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212?mod=article_inline; *see also* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, Proceedings of Machine Learning Research 81:1-15 (2018).

- A smartphone application in Boston identified potholes for repairs based on the accelerometers in residents' smartphones, but overlooked the likelihood that "systematic differences in smartphone ownership" would result in undercounting of road problems in poorer neighborhoods.[20]
- An employee recruiting system developed by Amazon taught itself to "downgrade resumes with the word 'women's' in them, and to assign lower scores to graduates of two women-only colleges."[21]
- An image recognition system labeled photographs of women in typical American wedding dresses as "bride" and "wedding" while labeling photographs of traditionally-dressed Indian brides as "performance art" and "costume."[22]
- A risk assessment system for criminal cases wrongly predicted that black defendants would commit another offense "at almost twice the rate as white defendants," who were "mislabeled as low risk more often than black defendants."[23]

Harmful and unfair biases in AI will vary tremendously in practice, and can arise from many different sources. For example, in some situations, bias will primarily come in the form of unintended or neglected flaws in the technology and/or inputs of AI, such as when a machine learning system relies on data is inaccurate or misleading for a minority population.[24] Such concerns have already emerged in many existing systems that rely on big data to make decisions. "Training data can be incomplete, biased or otherwise skewed, often drawing on limited and non-representative samples that are poorly defined before use."[25] For example, multiple investigations of the risk assessment systems used in both pre-trial and post-trial criminal proceedings have found that they are based on data in which defendants "with different demographic characteristics have systematically different likelihoods of apprehension and different

[20] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Cal. L. Rev. 671, 685 (2016).
[21] David Meyer, *Amazon Reportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool from Discriminating Against Women*, Fortune (Oct. 10, 2018), http://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/?utm_source=emailshare&utm_medium=email&utm_campaign=email-share-article&utm_content=20181013.
[22] James Zou and Londa Schiebinger, Design AI so that it's fair, 559 Nature 324, 325 (July 12, 2018).
[23] Julia Angwin et al., *Machine Bias*, Pro Publica (May 23, 2016).
[24] Nat'l Sci. & Tech. Council, supra note 10, at 31.
[25] Alex Campolo et al., AI Now 2017 Report 1, AI Now Inst. at 15-16.

sentencing intensities."[26] As Carol Rose of the ACLU of Massachusetts explains: "Disfavored communities and people of color who historically have been targeted for government scrutiny too often bear the brunt of dangers posed by these new technologies."[27]

In other situations, harmful bias may be more a matter of AI giving improper weight to certain factors, or ignoring larger rights and values. Even when an autonomous system myopically selects an "accurate" way to measure or maximize some narrow goal programmed into it, it may ignore or undervalue fundamental rights or ethical principles, rendering both its processes and its outcomes illegitimate. For example, in the context of criminal risk assessments, a machine learning system trying to predict recidivism might begin to weigh many attributes that would normally be out of bounds, such as "unemployment, marital status, age, education, finances, neighborhood, and family background, including family members' criminal history." Even if the system might find some of these factors as predictive and therefore useful to its programmed goal, this does not mean that it is fair or wise use them in sentencing an individual defendant, especially if they reflect and reinforce structural inequities that made that person more vulnerable to begin with.[28]

An AI authority could build crucial expertise in the various manifestations of bias, helping other regulators to clarify which specific forms of bias are most at play in their particular industries and decisions, and apply the most appropriate corrections and safeguards. Many sector-specific regulators may struggle with such questions on their own, especially because issues of AI bias have major technical dimensions. For instance, in some machine learning systems, hidden bias may primarily come from tainted training data, whereas in other instances harmful discrimination may derive from a system's logical processes of "feature selection" (determining which variables are at play in a given decision, and what weight to give them).[29] In other instances, bias may

---

[26] Osonde Osoba & William Welser IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, RAND Corp., at 15 (2017).
[27] Ina Fried, *Microsoft Tries to Write the Book on AI*, Axios, Jan. 17, 2018, https://www.axios.com/microsoft-tries-to-write-the-book-on-ai-1516219179-2a03cf29-5917-4e80-9dd5-23d12c1e7659.html.
[28] *See* Sonja B. Starr, Op-Ed., *Sentencing, by the Numbers*, N.Y. Times, Aug. 10, 2014.
[29] Joshua Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633, 680-83 (2017).

be a problem of "masking," in which the humans who design or implement a system use its complexities to both accomplish and hide some discriminatory purpose.[30]

Privacy. In recent years, digital privacy has soared in the public consciousness and on the agendas of many policymakers and regulators, both in the United States and around the world. Artificial intelligence will only bring more scrutiny. Recent breakthroughs in machine learning technologies are in large part a story of data collection.[31] As the types and quantities of training data grow exponentially, the capabilities of machine learning expand into more and more industries and walks of life.

Because current-generation AI depends upon and is often inseparable from underlying and ongoing data collection, it raises a core set of privacy risks that will cut across many different contexts and applications:

- **AI drives demand for data.** As companies, governments, and other critical institutions implement AI, they will have powerful incentives to collect more and more data, as a system's success will often depend on "ingesting as much training data as possible."[32] For example, from the many emerging applications of AI in medicine, it is easy to foresee "a world in which we're constantly under diagnostic surveillance" from sensors in our smartphones, furniture, cars, and clothes. Growing demand will also make many types of data more financially valuable to sell,[33] instigating even more collection by a wide variety of commercial entities.
- **AI boosts the supply of data.** As connected devices become more useful and widely adopted with AI, data collection also expands. For example, as the AI in digital assistants, smart speakers, and smart cameras gain more capabilities, they will inevitably be bought by more consumers and collect data from more

[30] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, Cal. L. Rev., Inc. 671, 712-13 (2016).
[31] See Microsoft, The Future Computer: Artificial Intelligence and its role in society (2018) at 22 (noting dramatic rise in data collection, to a forecasted total of 44 zettabytes (or 44 trillion gigabytes) in 2019).
[32] Alex Campolo et al., AI Now 2017 Report 1, AI Now Inst. at 29.
[33] See Buchanan and Miller, *supra* note 18, at 13 (explaining that, in many contexts, "a decent algorithm that learns from a lot of relevant data outperforms a great algorithm that learns from minimal or poor data," which is why "companies are willing to pay massive amounts of money for more data"); *see also* Jacques Bughin et al., *Artificial Intelligence: The Next Digital Frontier?*, McKinsey Global Institute (June 2017) at 33.

places.[34] Even more importantly, AI can extract and process much more useable data from raw sources of information—such as automating the recognition and categorization of human faces in surveillance videos.[35] And as more data is collected, most privacy risks will multiply.[36]

- **AI can make personal data far more revealing and insightful.** Traditionally, privacy laws and protections have distinguished more-sensitive and less-sensitive categories of personal data—such as a person's health history and tax records versus the car model she drives and what sports teams she roots for—reasoning that disclosures of the former will be more harmful and thus require tighter restrictions on collection, use, and dissemination. However, with big data and AI, these distinctions are increasingly antiquated. As Ryan Calo explains, AI "is increasingly able to derive the intimate from the available," meaning that "freely shared information of seeming innocence – where you ate lunch, for example, or what you bought at the grocery store – can lead to insights of a deeply sensitive nature."[37] For example, one AI system learned to predict household income, race, education, and voting behavior based on the car models parked at an address on Google Street View.[38] Such insights deepen further when AI can draw on multiple sources of useful data, making connections and inferences on a scale well beyond any humans capabilities.
- **AI can exercise powerful influence over human behavior.** Through empirical insights into our decisions and actions, autonomous systems can change how they interact with us to favor one outcome over another—typically without our full knowledge or understanding. In many cases, it is not an exaggeration to say that

---

[34] *E.g.*, Karl Bode, *Your Robot Vacuum Cleaner Will Soon Collect And Sell Data About You And Your Home*, Techdirt (July 25, 2017), https://www.techdirt.com/articles/20170725/06340737856/your-robot-vacuum-cleaner-will-soon-collect-sell-data-about-you-your-home.shtml.

[35] See James Vincent, Artificial intelligence is going to supercharge surveillance, The Verge (Jan 23, 2018), https://www.theverge.com/2018/1/23/16907238/artificial-intelligence-surveillance-cameras-security.

[36] Ben Buchanan, *Prepared Testimony and Statement for the Record, Game Changers: Artificial Intelligence Part II*, Hearing before the House Oversight Committee, Subcommittee on IT (https://oversight.house.gov/wp-content/uploads/2018/04/Buchanan-Harvard-Statement-AI-III-4-18.pdf).

[37] Calo, *supra* note 16, at 421; *see also Osoba and Welser*, *supra* note 25, 2017 at 12 (explaining how machine learning algorithms can "implicitly reconstruct sensitive fields" from other information).

[38] Glyn Moody, Using AI To Identify Car Models In 50 Million Google Street Views Reveals A Wide Range Of Demographic Information, Techdirt (Jan. 17, 2018), https://www.techdirt.com/articles/20180104/03261938926/using-ai-to-identify-car-models-50-million google-street-views-reveals-wide-range-demographic-information.shtml 1/17/18.

AI may come to know us better than we know ourselves by predicting our behavior more accurately, and finding ways to influence it that we may neither know nor understand. The proliferation of machine learning could add up to a world of "exquisite and hyper-targeted control"[39] – for example, in workplace systems that monitor and manage employees[40] – reshaping many of the power dynamics that govern our daily lives. This power to influence and control may be especially harmful and intrusive if it accrues mostly to large firms and bureaucracies, who may prioritize their own interests over the rights and agency of those individuals who are the objects of an AI system's predictions and actions. Ultimately, these concerns go beyond conventional notions of privacy, reaching fundamental questions about whether and how individual autonomy will endure within centralized, AI-equipped systems.

An AI authority would help sector-specific regulators and other parts of government confront these profound privacy challenges. In some instances, this will be a matter of sharing experience with the common problems raised by particular technologies and/or the best practices for addressing them. For example, what are the most pressing privacy risks from natural language processing, across the many settings in which it may be deployed (education, law enforcement, medicine, at home, etc.)? What computational methods for minimizing privacy risks (such as techniques for "differential privacy"[41]) are effective in which situations?

Beyond the technical realm, sector-specific agencies may also struggle with the delicate balances between privacy and other competing objectives, such as:

- Making more government data available for AI. As industry and other experts call for the government to help advance AI development by making more and better public data available, more agencies will have to decide whether and how to

[39] Calo, *supra* note 16, at 423.
[40] AI Now, *supra* note 31, at 11-12 ("By selectively exploiting workers' behavior, often without workers' consent or even knowledge, these technologies have the potential to make workers complicit in their own exploitation."); *see also* Imani Moise, *What's on Your Mind? Bosses Are Using Artificial Intelligence to Find Out*, Wall Street Journal (Mar. 28, 2018).
[41] Buchanan and Miller, *supra* note 18, at 30.

share the data under their custody and control, and perhaps whether to collect even more of it.[42]

- Requiring sufficient and useful transparency into the data inputs and/or outputs of AI systems, and deciding who (the public, outside researchers, confidential investigators, etc.) should get access to what information.[43]
- Promoting competition among AI providers in specific fields, such as through regulations on data sharing and interoperability between different proprietary systems — which may be necessary to prevent a dominant firm from locking in its customers, but may also come with significant privacy and security risks.[44]

An AI authority could offer vital expertise and experience on these overarching privacy questions, as they continue to emerge in different sectors and agency jurisdictions. This will be especially important for the most vulnerable communities, which sector-specific regulators may often misunderstand or neglect even when their privacy interests are the most endangered.[45] An AI authority should be best positioned to ring alarm bells about major potential harms that are going unaddressed — whether from AI privacy risks, excessive control over individual choices and opportunities, or discriminatory impacts on discrete populations.

Overall, given that much of the near-term development of AI is focused on how it interacts and collaborates with people, problems of privacy and self-determination will be unavoidable.[46] Rather than leaving privacy regulation in AI solely to fragmented agencies and reactive policymaking, an AI authority could build capacity and consistency throughout government — to identify and protect the most essential privacy and autonomy interests at stake in AI, while also fostering its development and successful adoption.

---

[42] *See, e.g.*, Organization for Economic Cooperation and Development ("OECD"), *AI: Intelligent Machines, Smart Policies Conference Summary*, OECD Digital Economy Papers No. 270 (Aug. 2018) at 6.
[43] *See, e.g.*, Darrell West and John Allen, How artificial intelligence is transforming the world, Brookings Institution, https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/ ("In general, the research community needs better access to government and business data, although with appropriate safeguards to make sure researchers do not misuse data in the way Cambridge Analytica did with Facebook information.").
[44] *See, e.g.*, Hal Varian, Artificial Intelligence, Economics, and Industrial Organization (June 2018), https://www.nber.org/chapters/c14017.pdf.
[45] AI Now, *supra* note 31, at 4 ("Without contextual knowledge, informed consent, and due processes mechanisms, these systems can create risks that threaten and expose already vulnerable populations."
[46] See AI100 2016 at 17.

Explainability. One of the unique and fundamental challenges raised by AI involves how to understand what it "decides" and why. In other words, how can humans explain and usefully understand what autonomous systems do? This is already a major question for existing deep learning technologies, which often cannot provide meaningful causal explanations beyond mere correlation--in other words, telling us only "what will happen but not why."[47] Some experts have gone as far as saying that it might be impossible to make machine learning truly explainable to humans,[48] because the ways in which they learn are "almost entirely alien."[49] Even if these predictions are overstated, at a minimum there is significant technical difficulty in making AI explainable, with research ongoing.[50]

Most significant regulations of AI will depend on making a system explainable and understandable--in other words, accountable--in at least some way. For example, if an AI system is used in employment decisions, to decide hiring, firing, and promotions, then any regulation of such decisions--for example, to prevent unlawful employment discrimination--requires some understanding of why the system acted the way it did towards a particular employee. Given its technical complexities and uncertainties, most sector-specific regulators will not be in a good position to confront AI explainability. The shared expertise and experience of an AI authority would be crucial here. It would identify the different practical types and meanings of AI explainability, and help other agencies to determine which ones were appropriate in different contexts. In some cases, explainability may be primarily a matter of making a system's outcomes consistent and predictable, whereas in others it might be necessary to show that AI satisfied substantive and/or legal criteria for its decision-making processes.[51] Crucially, an AI authority could also help other parts of the government to formulate the correct technical demands for AI design and/or implementation that are necessary to make a system understandable in a specific policy context.

---

[47] Calo, *supra* note 17, at 414.
[48] *See, e.g.*, Will Knight, *The Dark Secret at the Heart of AI*, Mass. Inst. of Tech. Tech. Rev., Apr. 11, 2017.
[49] Tutt, *supra* note 11, at 87-89.
[50] Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, Research Paper No. 17-8, Univ. of Penn. Law School Inst. for Law and Econ.,1147, 1154 (2017).
[51] Andrew D. Selbst, *A Mild Defense of Our New Machine Overlords*, 70 Vand. L. Rev. 87, 103 (2017).

Transparency and oversight. Regardless of which industry AI may be applied in and how it is regulated, there will always be ongoing questions of how to ensure that an autonomous system is working properly, complying with applicable laws and regulations, and not causing hidden harms or unintended consequences. There are many different specific methods and forms of oversight for autonomous systems, from reviewing datasets to verifying and certifying system components and processes to ongoing outcome testing.[52] The oversight methods that are appropriate will depend upon the technical details of a particular autonomous system along with the policy goals at play. For example, as medical applications of AI continue to expand, health care regulators will need specific and ongoing measurements of the safety and efficacy of AI-influenced treatments, as well as their cost effectiveness.[53] Instead of each sector-specific regulator reinventing the wheel in coming up with their separate oversight standards for AI, an AI authority would identify various needs and best practices, and then assist other agencies in selecting the right tools for a given context.

Furthermore, an AI authority could also help decide the appropriate level of public and/or stakeholder transparency for a given AI system. There are many different dimensions of possible transparency for AI, ranging from who receives information (public vs. some limited monitors) to the timing of disclosure (before or after AI deployment, how frequent) to the depth and content of what information is actually disclosed. While maximizing transparency will often be crucial to protecting the public interest in AI behavior--for example, by allowing the humans affected by an AI system to know that it is safe and legitimate, and how they may need to adjust their own behavior in response--there will almost always be countervailing costs and considerations that will weigh against full and unlimited transparency. For example, disclosure of source code may be appropriate in some limited cases but useless and/or harmful in many others.[54] As with fairness and explainability, similar transparency and oversight problems are likely to arise repeatedly across different sectors. An AI authority would provide other regulators and officials with expert guidance about the extent and nature

[52] Tutt, *supra* note 11, at 108.
[53] Christopher Mims, *The AI Doctor Will See You Now*, Wall St. J. May 20, 2018.
[54] Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 Yale L.J. & Tech. 103, 130 (2018).

of transparency that is appropriate for a given situation, versus what may not be. As part of this, the authority could help other agencies correctly assess the merits of specific objections to transparency measures from regulated entities, which may reflect legitimate interests but may also be a smokescreen to hobble regulators.

2. Protecting Public Values in Government Procurement and Implementation of AI

As in the private sector, applications of artificial intelligence are multiplying at all levels of government. Today, federal, state, and local government already use machine learning algorithms in a variety of different ways:

Types of Public Applications	Examples
Resource Allocation	● Predictive policing systems that allocate patrol officers to different areas within a jurisdiction based on expected rates, types, and timing of crimes (used in Chicago, Los Angeles, and many other cities)[55] ● Flint, MI system that prioritized certain city water pipes for replacement based on risk of lead contamination[56]
Investigation	● IRS "risk-based collection model" to predict most likely cases of tax evasion that warrant further investigation[57] ● EPA ToXCast program predicts toxicity of chemicals, identifying priorities for further testing[58] ● Chicago system that identifies restaurants as

[55] *See* Brauneis & Goodman, *supra* note 52, at 146-150.
[56] *See* Coglianese & Lehr, *supra* note 28, at 1152.
[57] *Id.*
[58] *Id.*

	priorities for food safety inspections, based on wide range of data including online reviews
Adjudication	• Recidivism and public safety assessment tools used in pre-trial and sentencing decisions in criminal cases • Medical billing systems used to approve and deny Medicaid claims
Infrastructure Management	• Systems that adjust stoplight times based on predicted traffic flows[59]

As autonomous systems handle a wider range of routine bureaucratic and administrative tasks, public executives and budget committees will likely find the productivity gains and cost-cutting possibilities irresistible. For example, automated bots may soon take over much of the public's direct interactions with government offices, answering phone calls and processing applications for permits and licenses. But cost and efficiency are not the only potential benefits of government AI. If implemented correctly, automated analysis may improve the quality and rigor of government decisions, unearthing antiquated assumptions, ineffective habits, and entrenched human biases.[60] And other emerging applications of AI may improve public services and make them more widely available--for example, if autonomous driving technologies allows buses and trains to run more frequently and reliably, or if robot tutors help students learn more outside the classroom.[61]

However, the benefits of public-sector AI are far from assured, as it also brings significant risks. As AI technologies advance, we should expect the emergent behavior of autonomous systems to further evolve, likely in both sophistication and unpredictability. As a result, it may be increasingly untenable to understand these

[59] *Id.*
[60] *See* Brauneis & Goodman, *supra* note 52, at 116.
[61] *See* Nat'l Sci. & Tech. Council, *supra* note 10, at 15.

systems in terms of the conventional human processes they are replacing or supplementing. Much of the previous managerial experience of government officials may be irrelevant or even misleading when trying to implement AI. And similar to the common questions that cut across sector-specific regulators in the private sector, most public implementations of AI will confront a core set of challenges that will repeat across different contexts. The experience and expertise of an AI authority could be indispensable to the rest of the public sector in confronting each of the following.

Writing Public Policies and Values into Autonomous Systems. Executing public policy and government services through autonomous systems is rarely straightforward. To begin with, there is the threshold question of whether AI is appropriate for a particular public function. Just because the technology may allow for AI to be used does not necessarily mean that a government should actually use it in a given context. For example, inherent limitations in explainability and accountability may rule out significant reliance on AI to perform high-stakes adjudications.[62] In other contexts, it may be impossible to build enough trust in the fairness of its outcomes to achieve legitimacy with stakeholders and the public.[63]

The involvement of private vendors in algorithmic design leads to another set of dangers, including opacity, public disempowerment, and loss of accountability. Public officials who have ceded the development of predictive algorithms to private vendors may not participate in and may be unaware of the policy decisions that are incorporated into those algorithms. Public employees who use the output of a predictive algorithm to inform their decisions may not understand the design and limitations. Private participation in public administration through algorithmic governance raises concerns that data will be used to hurt citizens and weaken public authority. The risk is that the corporation controlling the data and analytics occupies the command center of day-to-day governance while the democratically accountable officials, unable to control the data, move to the periphery.

An AI authority would serve as an expert guardian of public prerogatives as governments implement their own autonomous systems. One way to do this would be to

[62] *See* Calo, *supra* note 16, at 414.
[63] *See* Dillon Reisman et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI Now Inst. at 13 (2018).

formulate and administer rigorous review procedures that would precede major public implementations of AI, such as the "algorithmic impact assessments" proposed by researchers at AI Now.[64]

Managing Private-Sector AI Providers. A related concern is that private vendors come to lock governments into proprietary systems. Some commentators warn that "smart" projects are simply vehicles to sell municipalities comprehensive data management systems owned and managed by the vendor. Service contracts can make governments dependent on the technology provider for upgrades and ongoing development, locking the government into proprietary technologies whose costs and pace of innovation they cannot control. Lock-in may extend beyond technical systems to the physical infrastructure in which they are embedded.

An AI authority could counsel other government customers about how to minimize the risks of lock-in and other risks in their relationships with private AI vendors. For example, if a contractor is providing a custom algorithm for a jurisdiction, then it could be appropriate for that jurisdiction to insist on ownership, or at least on a license for its own use and use by other jurisdictions. In all cases, government agencies should assert ownership over reports that assess risks in that jurisdiction based on data provided by that jurisdiction.

An AI authority should also review proposed contracts and other business arrangements to ensure that they do not create undue barriers to a government customer safeguarding important public values, such as transparency and due process. Instead, governments should use their contracting powers to insist on appropriate record creation, provision, and disclosure.

Public Accountability and Legitimacy. Accountability is a crucial issue with any autonomous system, but it takes on particular significance in the public sector. An algorithmic process is accountable when its stakeholders, empowered by meaningful transparency, can intervene to effect change in the algorithm, or in its use or implementation. In the public sphere, this entails that government actually be held accountable by the voting public for the algorithms it deploys.[65] Such accountability

[64] *Id.* at 4.
[65] Calo, *supra* note 16, at 430; Brauneis, *supra* note 52, at 132.

requires not perfect transparency—complete knowledge of an algorithm's rules of operation and process of creation and validation—but the lower standard of meaningful transparency—knowledge sufficient to approve or disapprove of the algorithm's performance. Records short of the underlying computer code may suffice to provide the necessary input. Of course, accountability in practice could well require public education and political processes that are beyond what we can address here. But meaningful transparency will be the necessary first step. An AI authority could prioritize accountability in public implementations of AI, setting standards to guarantee it and specific practices and procedures to accomplish it.

3. Building Durable and Centralized AI Expertise Within Government

Regardless of whatever policy challenges AI may bring in either the private or public sectors, the government will need significant expertise and deliberative capacities to handle them. This is true whether the government is implementing its own AI systems, regulating the private sector, or negotiating with other governments on international matters. Both technical proficiency and practical policymaking experience will be vital. As the Office of Science and Technology Policy explained in 2016, "[e]ffective regulation of technologies such as AI requires agencies to have in-house technical expertise to help guide regulatory decision-making."[66] This is a matter of broad consensus: the government needs more expertise on AI. As the IEEE argues, "the U.S. Government does not yet have sufficient technical expertise to effectively regulate AI," creating a serious danger that policymaking will "fail to support innovation, adhere to American principles, and protect public safety."[67]

Unfortunately, there are many reasons to doubt that the government's existing structure will build enough of this capacity:

- **Difficulty attracting and retaining technical experts in government service:** This is primarily a result of real-world constraints. As Ryan Calo notes,

[66] Nat'l Sci. & Tech. Council, Comm. on Tech, Exec. Office of the President, *Preparing for the Future of Artificial Intelligence* at 17 (Oct. 2016).
[67] IEEE-USA, *Artificial Intelligence Research Development and Regulation* (Feb. 10, 2017) at 1; *see also* Buchanan and Miller, *supra* note 18, at 41 ("It is not an exaggeration to say that before long, for example, the majority of Congressional committees will find machine learning impacting the areas they oversee.").

"[g]overnment bodies are slow to hire up and face steep competition from industry."[68] This is a particular problem in the fields of artificial intelligence and big data, where experts are in high demand, with many large companies complaining of a shortage of talent.[69] Nor should we expect that sector-specific agencies with many other mandates will have the foresight or resources to hire dedicated experts in autonomous systems.

- **The inefficiencies of fragmented expertise:** In the face of increasingly complex AI systems, individual agencies may lack the technical capacities necessary to understand the technology and do their jobs. But even if a sector-specific agency can survive a learning curve on some particular issue, the expertise and experience it gains will often be unavailable in practice to the rest of the government.
- **The risks of private-sector advice and influence:** while AI policymaking will necessarily rely on the input of experts and stakeholders from industry, academia, and civil society, there are inherent risks if the government grows too reliant on outside advice. For-profit AI firms, especially those with entrenched connections in Washington, will often lobby the government primarily to advance their own narrow interests, which may skew presentation of the relevant facts and have little-to-no regard for the long-term.[70] Without substantial expertise of its own, government agencies and legislators will rarely be able to sort through the cacophony of arguments and advice.
- **The limitations of state and local governments**: all of the above problems are compounded at the state and local levels. States and localities will face many of the same complexities and common issues as their federal counterparts in dealing with an economy that increasingly relies on AI, while also implementing systems of their own. But compared to the federal government, states and

[68] Calo, *supra* note 2, at 428.
[69] Cade Metz, *A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit*, N.Y. Times, Apr. 19, 2018.; *see also* Michael C. Horowitz et al., *Strategic Competition in an Era of Artificial Intelligence*, Ctr. for a New American Sec. (2018).
[70] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J. of Law and Tech. 2 at 379-80 (2016).

localities have even fewer resources and smaller budgets, worse prospects for hiring experts of their own, and even higher risks of fragmented capacities.[71]

A new AI authority could address all of these shortcomings. As Ryan Calo proposed for a Federal Robotics Commission, its staff should consist of "engineers and others with backgrounds in mechanical and electrical engineering, computer science, and human-computer interaction, right alongside experts in law and policy."[72] Such an "interdisciplinary" agency could focus on attracting the "best and brightest" in several different ways. It could identify the compensation levels necessary to be competitive in recruiting qualified AI experts, and make full use of existing mechanisms for the hiring of "specialized talent outside of the traditional government pay scale."[73] An AI authority should also explore other ways to attract the right staff, such as pipeline training programs similar to what DARPA has attempted within the military,[74] or limited-time fellowships within the government for academic and private-sector experts.[75] The authority's mission and prestige may be just as important. Whereas AI specialists may be skeptical about joining a sector-specific agency and being limited to its narrow field, the prospects of serving in an office dedicated to AI and helping to chart the future of AI policymaking across all levels of government could be far more attractive.[76]

One of the AI authority's primary mandates should be to consult with, coordinate, and make its expertise available to the rest of the federal government--Congress, other agencies, and even the courts where appropriate. Here, the authority would build a repository of knowledge, experience, and relationships, helping other policymakers and regulators to be far more efficient and effective than if they tried to confront AI issues in isolation. It is also important this expertise is available not only to other federal agencies but also to state and local governments.

4. Identifying Major Gaps and Neglected Problems in the Laws and Regulatory Frameworks that Will Govern AI

---

[71] Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83, 113 (2017).; Darrell M. West & John R. Allen, How Artificial Intelligence is Transforming the World, Brookings Inst. at 21-22 (2018).
[72] Ryan Calo, The Case for a Federal Robotics Commission, Brookings Institution (Sept. 2014) at 11.
[73] Aaron Mannes, Institutional Options for Robot Governance (2015) at 18, http://robots.law.miami.edu/2016/wp-content/uploads/2015/07/Mannes_RobotGovernanceFinal.pdf.
[74] Nat'l Sci. & Tech. Council, supra note 10, at 25.
[75] IEEE-USA, supra note 66, at 5.
[76] See Mannes, supra note 72, at 18; Calo, supra note 71, 11-12.

AI is likely to require significant adjustments to a variety of existing legal and regulatory frameworks, especially those built on assumptions about governing human behavior that many not work for autonomous machines. An AI authority could identify the most urgent gaps as they emerge, and help advise legislators and regulators at all levels of government about how to address them. The following discusses possible examples, including legal liability, discrimination law, AI's economic impact, and general AI safety.

Liability for AI decisions. The common law system in the United States for apportioning fault and compensating injury evolved over several centuries. Even in situations where it has been supplanted by statutes, core doctrines such as negligence and strict liability regulate individuals and businesses of all stripes, determining who is responsible for what risks, what they must do to mitigate them, and how much they must pay when they fail to act responsibly. However, autonomous systems may unsettle many of these liability rules, which are premised on assumptions about human behavior and judgment that often do not apply to artificial intelligence. For example:

- Where an AI system makes an error and causes an injury, it will often be complicated, if not functionally impossible, to determine what person or entity was responsible for the wrongdoing, especially when the AI acts in unpredictable ways or when multiple people and entities have supplied the code, data, and/or physical computers on which it runs.[77]
- The concept of foreseeability--which is at the heart of many liability doctrines--may be either less useful or radically altered by AI, as we should often expect the behavior of such systems to "legitimately surprise all involved," particularly when one autonomous system interacts with another.[78]
- While strict liability may be more useful and necessary with software-driven autonomous systems, current law typically applies strict liability only in narrow situations, and generally not to the sorts of "intangible" products like software that are most comparable to AI.

[77] *See* Calo, *supra* note 2, at 534-4; *see also* Calo, *supra* note 17, at 416; AI100 2016 at [46].
[78] Calo, *supra* note 2, at 554-45 ("Should these systems prove deeply useful to society, as many envision, some other formulation than foreseeability may be necessary to assess liability. … As a consequence, we may see a broader role for risk mitigation within the law. The combination of data promiscuity mixed with the capacity to do physical harm can make unpacking liability impractical.").

- In practice, judges and juries may struggle with fact-finding in cases involving wrongdoing by autonomous systems.[79]

The fragmented nature of liability laws--many of which operate at the state level--may make it even more complicated to apply them to autonomous systems. Determining the right liability frameworks for AI is beyond the scope of this paper. But it seems likely that an expansion of strict liability and/or other collective insurance frameworks will be necessary. For example, some experts have suggested that AI providers should be held liable for any of the decisions or actions of their systems, even when not foreseeable or understandable.[80] In part this is because of a "structural information asymmetry" between AI processes and many consumers and users.[81]

More is at stake than just risk mitigation and injury compensation. Excessive or poorly-calibrated liability laws could threaten other policy goals in AI, such as openness, competition, and fostering innovation. An AI authority could be essential adapting existing liability frameworks to AI in a coherent way, using its expertise to examine the competing objectives at play, identifying the particular doctrines and situations that are most unsettled, and helping Congress, state legislatures, and the courts to adjust the law as necessary.

Unlawful discrimination by autonomous systems. AI also poses challenges for many of the legal doctrines that have evolved to restrict discrimination on the bases of race, gender, age, and other suspect classifications. To begin with, it is unlikely that laws focusing on intentional discrimination--such the concept of disparate treatment[82]--will be of much use with AI. In limited cases, it may be possible that a plaintiff might show that the designers or operators of an AI system acted with a discriminatory animus--for example, in providing a certain set of training data to a machine learning system. But this will be the exception. Otherwise, alleged discrimination by an

[79] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J. of Law & Tech. 354, 389-90 (2016).
[80] Oren Etzioni, *How to Regulate Artificial Intelligence*, New York Times (Sept. 1, 2017).
[81] European Consumer Consultative Group, Policy Recommendations for a Safe and Secure Use of Artificial Intelligence, https://ec.europa.eu/info/sites/info/files/eccg-recommendation-on-ai_may2018_en.pdf.
[82] *See* Barocas and Selbst, *supra* note 20, at 694-96.

autonomous system could be effectively immune from disparate treatment analysis, because it lacks any intent comparable to a human decisionmaker.[83]

Instead, it will often be more realistic to show unlawful discrimination through disparate impact, which does not rely on a determination of intent. But this concept also has its problems and complexities when applied to AI. In most disparate impact cases, distinguishing and striking the right balance between "legitimate" decisional factors and illegitimate and unnecessary discrimination will be extraordinarily complicated and error-prone.[84] For example, we should expect that AI systems in the employment context "will often both predict future job performance and have some disparate impact."[85] Furthermore, if a system's decision-making processes are largely a black box to human observers, disparate impact may often be impossible to apply in practice.[86]

These challenges will likely cut across multiple specific fields of discrimination law. An AI authority would be in the best position to consider these issues broadly, and make recommendations to help many different lawmakers, enforcement agencies, and adjudicators adapt anti-discrimination laws.

Adapting to AI's Economic Impact. As they take on more and more human activities and decisions, AI may upend broad swaths of our current economy. Already, many economists attribute slower wage growth and other macroeconomic trends to technological automation.[87] Even if the technology does not quickly progress into realm of general AI, the expansion of machine learning and associated systems could displace large numbers of workers and transform major industries. For example, the McKinsey Global Institute estimates that "60 percent of all occupations have the potential for about a third of their activities to be automated."[88] Another forecast predicted that 14% of workers in OECD countries would be "at a high risk of losing their jobs to automation in the coming years … with another 32% of the workforce seeing substantial change in how their jobs are carried out." Predictions about the economic

---

[83] *See id.* at 698-701.
[84] *See* Coglianese & Lehr, *supra* note 48, at 1200; *see also* Barocas & Selbst, *supra* note 20, at 711-12.
[85] *Id.* at 712.
[86] *See* Coglianese & Lehr, *supra* note 20, at 1198-99.
[87] *See* Nat'l Sci. & Tech. Council, Comm. on Tech, Exec. Office of the President, *Artificial Intelligence, Automation, and the Economy* at 8.
[88] Campolo, *supra* note 31.

and labor effects of AI are widely debated, with optimists arguing that it will ultimately create more jobs to replace those that are lost.[89]

While the precise timing and details of such disruption may be uncertain, there is a broad consensus among economists and other experts that AI will have a systemic economic impact, with the highest risk of harm to lower-skilled workers.[90] As the Office of Science and Technology Policy found in 2016, if the productivity gains from AI "accrue to a select few … [i]nstead of broadly shared prosperity for workers and consumers, this might push towards reduced competition and increased wealth inequality."[91] For example, Spencer Overton of the Joint Center for Political and Economic Studies argues that communities of color will be disproportionately at risk, estimating that "27% of black workers are concentrated in just 30 jobs at high risk of automation."[92] An AI authority could prepare government for the economic disruptions unleashed by AI, using its expertise to identify the most pressing trends and problems, and coordinating policy responses and long-term strategies with many other parts of the government, "to help Americans who are disadvantaged by these changes and to ensure that the enormous benefits of AI and automation are developed by and available to all."[93]

AI Safety. In contrast to the many of the near-term policy issues, the greatest fears about AI safety are more contingent on the technology's long-term progress. To be sure, the narrow AI applications of today still give rise to plenty of safety concerns--such as whether and when autonomous vehicles will be ready for wide use. But these safety questions tend to vary by context and application, and are typically inseparable from whether an AI system is "doing its job" correctly. As a result, sector-specific regulators will necessarily take the lead on the specific safety issues that come before them.

---

[89] *E.g.*, Castro 2017.
[90] *See* Nat'l Sci. & Tech. Council, supra note 10.
[91] *Id.*; *see also* Bughin, supra note 4, at 21 (discussing "negative externalities" from AI disruption "such as loss of domestic consumption during unemployment" and predicting costs totaling trillions of dollars).
[92] Ina Fried, AI is the future of discrimination—and fairness, Axios (Aug. 1, 2018), https://www.axios.com/ai-as-the-future-of-discrimination-fairness-415a69aa-3bf7-476b-88fd-cb95698a6001.html?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axioslogin&stream=top-stories.
[93] Nat'l Sci. & Tech. Council, supra note 4.

However, many AI experts have focused on more profound questions: whether and how AI can be made subservient to human interests, and whether AI may ultimately come to dominate us or even threaten our existence. These concerns are mostly about general AI, and thus are bound up in the predictive debate over whether and when it will emerge. Some researchers are very worried about the deeper safety risks of AI, and expect them to become real in the coming decades. Controversially, Elon Musk has described AI as a "fundamental existential risk for human civilization," saying "we should be really concerned."[94] Nick Bostrom of the University of Oxford has claimed that "we're like small children playing with a bomb."[95] But many other experts are far more optimistic about, or even dismissive of, such fears. For example, the researcher Raymond Perrault has pointed out that current AI is still "extremely limited" in terms of general intelligence,[96] and the robotics developer Rodney Brooks has predicted that "malevolent AI … is nothing to worry about for a few hundred years at least."[97] Ryan Calo argues that existential fears about superintelligence can distract from efforts to address the far more real and imminent policy challenges from present-day AI.

Regardless of how well-grounded and imminent the safety risks of AI may or may not be, an AI authority could play a vital role -- in monitoring technological progress towards more general AI, in consulting with and learning from the many different experts in the field, and in developing major policies on AI safety if and when they become necessary. Like the major economic impacts of AI, the government should not delay in building an institutional foundation to confront major safety, even if their future is uncertain.

---

[94] Aatif Sulleyman, *Elon Musk: AI Is a 'Fundamental Existential Risk for Human Civilisation' and Creators Must Slow Down*, The Independent, Jul. 17, 2017, https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-ai-human-civilisation-existential-risk-artificial-intelligence-creator-slow-down-tesla-a7845491.html.

[95] Tim Adams, *Artificial intelligence: 'We're like children playing with a bomb'*, The Guardian, June 12, 2016.

[96] James Vincent, *Artificial Intelligence Isn't as Clever as We Think, but That Doesn't Stop It Being a Threat*, The Verge, Dec. 1, 2017, https://www.theverge.com/2017/12/1/16723238/ai-artificial-intelligence-progress-index.

[97] Paul Ford, *Our Fear of Artificial Intelligence*, MIT Tech. Rev., Feb 11, 2015; see also AI100 2016 at [6] ("The frightening, futurist portrayals of Artificial Intelligence that dominate films and novels, and shape the popular imagination, are fictional. Unlike in the movies, there is no race of superhuman robots on the horizon or probably even possible.").

5. Coordinating Strategies and Priorities for International AI Governance

Going back decades, the United States has dominated much of the computer and information technology industries, from semiconductors starting in the 1950s to microprocessors in the 1970s and 1980s to the search engines, social media networks, and other internet platforms of today. Many American firms and researchers are at the forefront of artificial intelligence development, globally. However, as AI continues to improve and proliferate, American dominance is far from assured. Many countries have identified AI as a central priority for their economic and technological futures, and begun investing significant public resources to foster AI industries and capabilities within their borders. For example, Japan, France, the United Kingdom, and Germany have all published national strategies for the development of AI.[98]

More than any other, China stands out as the most significant national competitor to the United States for hegemony over AI. Of course, China has been increasingly successful in building other information technology industries. More consumer electronics and computing hardware are manufactured there than in anywhere else in the world.[99] And in part due to the exclusion of foreign competitors, China has grown major internet platform and e-commerce companies of its own, such as Baidu and tencent. In AI, China is determined to build on this success. Going back to at least 2014, China has poured billions of dollars and government attention into both the technological development and concrete commercialization of AI.[100] According to some metrics, Chinese researchers now publish more papers than U.S.-based researchers on deep learning.[101] This year, China published a national plan to "become the world leader in artificial intelligence and create an industry worth $150 billion to its economy."[102] In part as a result of China's sustained investment, several observers are

[98] Nicholas Thompson, *Emmanuel Macron Talks to Wired About France's AI Strategy*, Wired, Mar. 31, 2018; Jiji, AI Researchers to be Focus of Government's 'Integrated Innovation Strategy,' The Japan Times, Jun. 3, 2018; Zoe Webster, *AI and Data at the Heart of UK Industrial Strategy*, Innovate UK, Jun. 28, 2018, https://innovateuk.blog.gov.uk/2018/06/28/ai-and-data-at-the-heart-of-uk-industrial-strategy/.
[99] Will Knight, *China's AI Awakening*, Mass. Inst. of Tech. Tech. Rev., Oct. 10, 2017; Matt Schiavenza, *China's Dominance in Manufacturing--in One Chart*, The Atlantic, Aug. 5, 2013.
[100] Michael C. Horowitz et al., *supra* note 68.
[101] Brian Fung, *China Has Now Eclipsed Us in AI Research*, The Wash. Post, Oct. 13, 2016.
[102] Cade Metz, *As China Marches Forward on A.I., the White House Is Silent*, N.Y. Times, Feb. 12, 2018.

predicting that China will be, at the very least, a serious competitor in AI, even if the United States has a lead right now. According to one researcher, "it is indisputable that Chinese authors are a significant force in A.I., and their position has been increasing drastically in the past five years."[103] China may have a particular advantage in data collection to fuel AI development, from both the size of its population and the absence of constraints on surveillance and data sharing, relative to the United States and other industrialized democracies.[104]

From an American perspective, China's rise in AI may raise particular concerns, different and more pressing than those raised by any other national competitor. First, China has a history of building domestic industries through aggressive and often heavy-handed measures against foreign firms, such as demanding the transfer of technical know-how and other intellectual property as a condition of market access. If China continues these tactics in AI, it could gain substantial advantages over other countries. Second, AI may become significant in the larger geopolitical and economic competition between China and the United States.[105] For example, artificial intelligence will likely have major military applications, where a Chinese lead in the technology would raise national security concerns.

Third, and most important, if the Chinese government (and/or firms closely connected to the government) gains significant influence over how AI develops, the technology may come to reflect starkly different values and priorities than what would prevail in the United States. For example, it is doubtful that China will foster democratic values in the design and governance of AI, such as transparency and accountability mechanisms that are useful to civil society and the public at large. Furthermore, just like China has prioritized strict government control of online communication and expression, it is already pursuing AI applications that may enable deeper authoritarian control.[106]

Despite the intensifying competition over AI development, to date there have only nascent coordinated efforts within the U.S. government to consider its strategic position, or to identify what the government should do to defend American interests as AI

---

[103] John Markoff & Matthew Rosenberg, *China's Intelligent Weaponry Gets Smarter*, N.Y. Times, Feb. 3, 2017.
[104] See Bergen and Ramli 8/14/17.
[105] *See generally* Horowitz, supra note 68.
[106] *E.g.*, Christina Larson, *Who needs democracy when you have data?*, MIT Technology Review (Aug. 20, 2018).

spreads internationally.[107] As James Lewis of the Center for Strategic and International Studies puts it, "[i]t's a race in the new generation of computing. The difference is that China seems to think it's a race and America doesn't."[108] An AI authority could help fill this gap in a number of different ways. It could formally and regularly assess the state of American competitiveness in AI, major technological developments around the world, and the actions of other governments to foster their own AI industries. In consultation with other parts of the government, an AI authority could recommend policies and other measures with this international perspective at the forefront. The authority could advise the U.S. Trade Representative and the State Department on when and how to respond to foreign governments that act against American AI firms and/or favor their own industries. At the same time, the authority could also advise Congress and the rest of the executive branch on the wisdom, necessity, and implementation of U.S. public support for AI development--for example, by coordinating priorities with the various agencies that award and oversee government research grants.

An AI authority could also be indispensable as international policymaking on AI ramps up. As autonomous systems become even more capable and ubiquitous, a variety of important technical and policy questions will likely emerge in international and/or multi-stakeholder forums, such as:

- Setting technical standards for the design and operation of AI systems in many different contexts, including standards that are directly tied to major policy issues, like accountability and transparency.[109]
- Cross-border flows of data and other AI communications, including how much of the infrastructure in an AI system must be localized within particular countries.
- Harmonizing sector-specific AI regulations--for example, to enable genuine international competition.[110]

---

[107] In August 2018, Congress passed the John S. McCain National Defense Authorization Act for Fiscal Year 2019, which includes a provision establishing a "National Security Commission on Artificial Intelligence" within the executive branch, comprised of representatives from various executive departments. This Commission's mandate includes a review of U.S. competitiveness in "artificial intelligence, machine learning, and other associated technologies," along with other policy concerns relevant to AI's impact on national security and defense.

[108] Paul Mozur & John Markoff, *Is China Outsmarting America in A.I.?*, N.Y. Times, May 27, 2017.

[109] *See* IEEE-USA, *supra* note 66.

[110] See IEEE-USA 2017 at __; OECD Sept 2018 at __.

The sustained expertise of an AI authority would likely be essential on such questions. Similar to the international bureaus of existing agencies like the Federal Communications Commission and the U.S. Patent and Trademark Office, the government's AI experts could participate directly in international deliberations over AI policy. This could be especially crucial in areas when the American views of AI diverges sharply from that of other countries. An AI authority could help coordinate with allies and civil society, to defend American interests in international policymaking while promoting core values that will shape the development and governance of AI over the long term.

CONCLUSION

Overall, many skeptics of AI regulation emphasize the prospect of overregulation--arguing that it will threaten the "permissionless innovation" necessary for beneficial AI to emerge.[111] But this risk may be higher with the tunnel vision inherent in sector-specific regulation, as specialized regulators can be prone to focusing heavily on their established mandates--such as ensuring safety and minimizing the risk of harm--at the expense of innovation in a particular field. In contrast, an agency with broader authority over autonomous systems, cutting across multiple sectors and applications, will be more likely to prioritize innovation and technological development, particularly if they are written into its founding mandate. Thus, somewhat counterintuitively, creating a new authority may lessen this risk of over-regulation. Furthermore, even most skeptics of AI regulation recognize that some government oversight necessary, and that the ubiquity and transformative potential of AI will inevitably push AI regulation to the forefront of policymakers' agendas. A new AI authority would confront this reality head on, helping the rest of the government adapt to autonomous technologies in the most efficient and forward-thinking ways.

[111] Andrea O'Sullivan, Don't Let Regulators Ruin AI, MIT Technology Review (Oct. 24, 2017), https://www.technologyreview.com/s/609132/dont-let-regulators-ruin-ai/.