https://doi.org/10.6092/issn.2531-6133/8670

Artificial Intelligence Policy: a Primer and Roadmap

RYAN CALO[†]

ABSTRACT: Talk of artificial intelligence is everywhere. People marvel at the capacity of machines to translate any language and master any game. Others condemn the use of secret algorithms to sentence criminal defendants or recoil at the prospect of machines gunning for blue, pink, and white-collar jobs. Some worry aloud that artificial intelligence will be humankind's "final invention."

This essay, prepared in connection with UC Davis Law Review's 50th anniversary symposium, explains why AI is suddenly on everyone's mind and provides a roadmap to the major policy questions AI raises. The essay is designed to help policymakers, investors, technologists, scholars, and students understand the contemporary policy environment around AI at least well enough to initiate their own exploration.

Topics covered include: justice and equity, use of force, safety and certification, privacy (including data parity) and taxation and displacement of labor. In addition to these topics, the essay will touch briefly on a selection of broader systemic questions: institutional configuration and expertise, investment and procurement, removing hurdles to accountability and correcting mental models of AI.

KEYWORDS: *Artificial Intelligence; Law; Policy; Ethics; Governance*

EDITORIAL NOTE

This essay is reprinted with author permission from Volume 51 of the UC Davis Law Review.

https://doi.org/10.6092/issn.2531-6133/8670

1. INTRODUCTION

The year is 2017 and talk of artificial intelligence (hereinafter AI) is everywhere. People marvel at the capacity of machines to translate any language and master any game.[1] Others condemn the use of secret algorithms to sentence criminal defendants[2] or recoil at the prospect of machines gunning for blue, pink, and white-collar jobs.[3] Some worry aloud that AI will be humankind's "final invention."[4]

The attention we pay to AI today is hardly new: looking back twenty, forty, or even a hundred years, one encounters similar hopes and concerns around AI systems and the robots they inhabit. Batya Friedman and Helen Nissenbaum wrote Bias in Computer Systems, a framework for evaluating and responding to machines that discriminate unfairly, in 1996.[5] The 1980 New York Times headline "A Robot Is After Your Job" could as easily appear in September 2017.[6]

The field of AI itself dates back at least to the 1950s, when John McCarthy and others coined the term one summer at Dartmouth College, and the concepts underlying AI go back generations earlier to the ideas of Charles Babbage, Ada

[†] Lane Powell and D. Wayne Gittinger Associate Professor, University of Washington School of Law (USA).

[1] *See, e.g.*, Cade Metz, *In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go*, WIRED (Jan. 27, 2016), https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go (reporting how after decades of work, Google's AI finally beat the top human player in the game of Go, a 2,500-year-old game of strategy and intuition exponentially more complex than chess).

[2] *See, e.g.*, CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 27 (2016) (comparing such algorithms to weapons of mass destruction for contributing to and sustaining toxic recidivism cycles); *see also*. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (discussing errors algorithms make when generating risk-assessment scores).

[3] *See, e.g.*, MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE xvi (2015) (predicting that machines' role will evolve from that of the worker's tool to the worker itself).

[4] JAMES BARRAT, OUR FINAL INVENTION: ARTIFICIAL INTELLIGENCE AND THE END OF THE HUMAN ERA 5 (2013) ("Our species is going to mortally struggle with this problem.").

[5] Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS, Jul. 1996, at. 330.

[6] Harley Shaiken, *A Robot Is After Your Job; New Technology Isn't a Panacea*, N.Y. TIMES, Sept. 3, 1980, at 19. For an excellent timeline of coverage of robots displacing labor, see generally Louis Anslow, *Robots Have Been About to Take All the Jobs for More than 200 Years*, TIMELINE (May 16, 2016), https://timeline.com/robots-have-been-about-to-take-all-the-jobs-for-more-than-200-years-5c9c08a2f41d.

Lovelace, and Alan Turing.[7] Although there have been significant developments and refinements, nearly every technique we use today — including the biologically-inspired neural nets at the core of the practical AI breakthroughs currently making headlines — was developed decades ago by researchers in the United States, Canada, and elsewhere.[8]

If the terminology, constituent techniques, and hopes and fears around AI are not new, what exactly is? At least two differences characterize the present climate. First, as is widely remarked, a vast increase in computational power and access to training data has led to practical breakthroughs in machine learning, a singularly important branch of AI.[9] These breakthroughs underpin recent successes across a variety of applied domains, from diagnosing precancerous moles to driving a vehicle, and dramatize the potential of AI for both good and ill.

Second, policymakers are finally paying close attention. In 1960, when John F. Kennedy was elected, there were calls for him to hold a conference around robots and labor.[10] He declined.[11] Later there were calls to form a Federal Automation Commission.[12] None was formed. A search revealed no hearings on AI in the House or Senate until, within months of one another in 2016, the House Energy and Commerce Committee held a hearing on Advanced Robotics (robots with AI) and the Senate Joint Economic Committee held the "first ever hearing focused solely on artificial intelligence."[13] That same year, the Obama White

[7] *See* Selmer Bringsjord, David Ferrucci & Paul Bello, *Creativity, the Turing Test, and the (Better) Lovelace Test*, 11 MINDS & MACHINES 3, 5 (2000); *see also* PETER STONE ET AL., STANFORD UNIV., ARTIFICIAL INTELLIGENCE AND LIFE IN 2030: REPORT OF THE 2015 STUDY PANEL 50 (2016), https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf.

[8] *See* STONE ET AL., *supra* note 7, at 50-51; Will Knight, *Facebook Heads to Canada for the Next Big AI Breakthrough*, MIT TECH. REV. (Sept. 15, 2017), https://www.technologyreview.com/s/608858/facebook-heads-to-canada-for-the-next-big-ai-breakthrough (discussing leading figures and breakthroughs with connections to Canada).

[9] *See, e.g.*, STONE ET AL., *supra* note 7, at 14; *see also* EXEC. OFF. OF THE PRESIDENT & NAT'L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 6 (2016).

[10] *See* Anslow, *supra* note 6.

[11] He did, however, give a speech on the necessity of "effective and vigorous government leadership" to help solve the "problems of automation." Senator John F. Kennedy, Remarks at the AFL-CIO Convention (June 7, 1960).

[12] *See* Anslow, *supra* note 6.

[13] Press Release, Sen. Ted Cruz, Sen. Cruz Chairs First Congressional Hearing on Artificial Intelligence (Nov. 30, 2016), https://www.cruz.senate.gov/?p=press_release&id=2902; *see also The Transformative Impact of Robots and Automation: Hearing Before the J. Econ. Comm.*, 114th Cong. (2016).

https://doi.org/10.6092/issn.2531-6133/8670

House held several workshops on AI and published three official reports detailing its findings.[14] Formal policymaking around AI abroad is, if anything, more advanced: the governments of Japan and the European Union have proposed or formed official commissions around robots and AI in recent years.[15]

This essay, prepared in connection with the UC Davis Law Review's Fiftieth Anniversary symposium, *Future-Proofing Law: From rDNA to Robots*, is my attempt at introducing the AI policy debate to recent audiences, as well as offering a conceptual organization for existing participants. The Essay is designed to help policymakers, investors, scholars, and students understand the contemporary policy environment around AI and the key challenges it presents. These include:

- justice and equity;
- use of force;
- safety and certification;
- privacy and power; and
- taxation and displacement of labor.

In addition to these topics, the Essay will touch briefly on a selection of broader systemic questions:

- institutional configuration and expertise;
- investment and procurement;
- removing hurdles to accountability; and
- correcting flawed mental models of AI.

In each instance, the Essay endeavors to give sufficient detail to describe the challenge without prejudging the policy outcome. This Essay is meant to be a roadmap, not the road itself. Its primary goal is to point the new entrant toward a

[14] *See, e.g.*, NAT'L SCI. & TECH. COUNCIL, *supra* note 9, at 12.

[15] *See* Iina Lietzen, *Robots: Legal Affairs Committee Calls for EU-Wide Rules*, EUR. PARL. NEWS (Jan. 12, 2017),http://www.europarl.europa.eu/news/en/news-room/20170110IPR57613/robots-legal-affairs-committee-calls-for-eu-wide-rules; Press Release, Japan Ministry of Econ., Trade & Indus., Robotics Policy Office Is to Be Established in METI (July 1, 2015), http://www.meti.go.jp/english/press/2015/0701_01.html.

wider debate and equip them with the context for further exploration and research.

I am a law professor with no formal training in AI. But my long-standing engagement with AI has provided me with a front row seat to many of the recent efforts to assess and channel the impact of AI on society.[16] I am familiar with the burgeoning literature and commentary on this topic and have reached out to individuals in the field to get their sense of what is important. That said, I certainly would not suggest that the inventory of policy questions I identify here is somehow a matter of consensus. I do not speak for the AI policy community as a whole. Rather, the views that follow are idiosyncratic and reflect, in the end, one scholar's interpretation of a complex landscape.[17]

The remainder of the Essay proceeds as follows. Part I offers a short background on AI and defends the terminology of policy over comparable terms such as ethics and governance. Part II lays out the key policy concerns of AI as of this writing. Part III addresses the oddly tenacious and prevalent fear that AI poses an existential threat to humanity — a concern that, if true, would seem to dwarf all other policy concerns. A final section concludes.

2. BACKGROUND

2.1. WHAT IS AI?

There is no straightforward, consensus definition of AI. AI is best understood as a set of techniques aimed at approximating some aspect of human or animal cognition using machines. Early theorists conceived of symbolic systems — the

[16] For example, I hosted the first White House workshop on AI policy, participated as an expert in the inaugural panel of the Stanford AI 100 study, organized AI workshops for the National Science Foundation, the Department of Homeland Security, and the National Academy of Sciences, advised AI Now and FAT*, and co-founded the We Robot conference.

[17] Earlier AI pioneer Herbert Simon argues that it is the duty of people who study a new technology to offer their interpretations regarding its likely effects on society. Herbert Simon, The Shape of Automation for Men and Management vii (1965). But: "Such interpretations should be, of course, the beginning and not the end of public discussion." I vehemently agree. For another interpretation, focusing on careers in AI policy, see Miles Brundage, *Guide to Working in AI Policy and Strategy*, 80,000 HOURS (2017), https://80000hours.org/articles/ai-policy-guide.

organization of abstract symbols using logical rules — as the most fruitful path toward computers that can "think."[18] However, the approach of building a reasoning machine upon which to scaffold all other cognitive tasks, as originally envisioned by Turing and others, did not deliver upon initial expectations. What seems possible in theory has yet to yield many viable applications in practice.[19]

Some blame an over-commitment to symbolic systems relative to other available techniques (e.g., reinforcement learning) for the dwindling of research funding in the late 1980s known as the "AI Winter."[20] Regardless, as limitations to the capacity of "good old fashioned AI" to deliver practical applications became apparent, researchers pursued a variety of other approaches to approximating cognition grounded in the analysis and manipulation of real world data.[21] An important consequence of the shift was that researchers began to try to solve specific problems or master particular "domains," such as converting speech to text or playing chess, instead of pursuing a holistic intelligence capable of performing every cognitive task within one system.[22]

All manner of AI techniques see study and use today. Much of the contemporary excitement around AI, however, flows from the enormous promise of a particular set of techniques known collectively as machine learning.[23] Machine learning (hereinafter ML) refers to the capacity of a system to improve its performance at a task over time.[24] Often this task involves recognizing patterns in datasets, although ML outputs can include everything from translating languages and diagnosing precancerous moles to grasping objects or helping to drive a car. As alluded to above, most every technique that underpins ML has been

[18] STONE ET AL., *supra* note 7, at 51.

[19] *Id.*

[20] *Id.*; *see also* NAT'L SCI. & TECH. COUNCIL, *supra* note 9, at 25.

[21] *See* STONE ET AL., *supra* note 7, at 51.

[22] *Id.* at 6-9. Originally the community drew a distinction between "weak" or "narrow" AI, designed to solve a single problem like chess, and "strong" AI with human-like capabilities across the boards. Today the term strong AI has given way to terms like artificial general intelligence ("AGI"), which refer to systems that can accomplish tasks in more than one domain without necessarily mastering all cognitive tasks.

[23] *See* NAT'L SCI. & TECH. COUNCIL, *supra* note 9, at 8.

[24] *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88 (2014).

around for decades. The recent explosion of efficacy comes from a combination of much faster computers and much more data.[25]

In other words, AI is an umbrella term, comprised by many different techniques. Today's cutting-edge practitioners tend to emphasize approaches such as deep learning within ML that leverage many-layered structures to extract features from enormous data sets in service of practical tasks requiring pattern recognition, or use other techniques to similar effect.[26] As we will see, these general features of contemporary AI — the shift toward practical applications, for example, and the reliance on data — also inform our policy questions.

2.2. WHERE IS AI DEVELOPED AND DEPLOYED?

Development of AI is most advanced within industry, academia, and the military.[27] Industry in particular is taking the lead on AI, with tech companies hiring away top scientists from universities and leveraging unparalleled access to enormous computational power and voluminous, timely data.[28] This was not always the case: as with many technologies, AI had its origins in academic research catalyzed by considerable military funding.[29] But industry has long held a significant role. The AI Winter gave way to the present AI Spring in part thanks to the continued efforts of researchers who once worked at Xerox Park and Bell Labs. Even today, much of the AI research occurring at firms is happening in research departments structurally insulated, to some degree, from the demands of the company's bottom line. Still, it is worth noting that as few as seven for profit institutions — Google, Facebook, IBM, Amazon, Microsoft, Apple, and Baidu in China —

[25] *See* STONE ET AL., *supra* note 7, at 51.

[26] *Id.* at 14-15; *see also* NAT'L SCI. AND TECH. COUNCIL, *supra* note 9, at 9-10.

[27] There are other private organizations and public labs with considerable acumen in AI, including the Allen Institute for AI and the Stanford Research Institute (hereinafter SRI).

[28] *See* Jordan Pearson, *Uber's AI Hub in Pittsburgh Gutted a University Lab — Now It's in Toronto*, VICE MOTHERBOARD (May 9, 2017), https://motherboard.vice.com/en_us/article/3dxkej/ubers-ai-hub-in-pittsburgh-gutted-a-university-lab-now-its-in-toronto (reporting concerns over whether Uber will become a "parasite draining brainpower (and taxpayer-funded research) from public institutions").

[29] See Joseph Weizenbaum, Computer Power and Human Reason: From Judgment to Calculation 271-272 (1976) (discussing funding sources for AI research).

seemingly hold AI capabilities that vastly outstrip all other institutions as of this writing.[30]

AI is deployed across a wide variety of devices and settings. How wide depends on whom you ask. Some would characterize spam filters that leverage ML or simple chat bots on social media — programmed to, for instance, reply to posts about climate change by denying its basis in science — as AI.[31] Others would limit the term to highly complex instantiations such as the Defense Advanced Research Project Agency's (hereinafter DARPA) Cognitive Assistant that Learns and Organizes [32] or the guidance software of a fully driverless car. We might also draw a distinction between disembodied AI, which acquires, processes, and outputs information as data, and robotics or other cyber-physical systems, which leverage AI to act physically upon the world. Indeed, there is reason to believe the law will treat these two categories differently.[33]

Regardless, many of the devices and services we access today — from iPhone autocorrect to Google Images — leverage trained pattern recognition systems or complex algorithms that a generous definition of AI might encompass.[34] The discussion that follows does not assume a minimal threshold of AI complexity but focuses instead on what is different about contemporary AI from previous or constituent technologies such as computers and the Internet.

[30] *Cf.* Vinod Iyengar, *Why AI Consolidation Will Create the Worst Monopoly in U.S. History*, TECHCRUNCH (Aug. 24, 2016), https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worst-monopoly-in-us-history (explaining how these major technology companies have made a practice of acquiring most every promising AI startup); *see also* QUORA, *What Companies Are Winning the Race for Artificial Intelligence?*, FORBES (Feb. 24, 2017), https://www.forbes.com/sites/quora/2017/02/24/what-companies-are-winning-the-race-for-artificial-intelligence/#2af852e6f5cd. There have been efforts to democratize AI, including the heavily funded but non-profit OpenAI. *See* OPENAI, https://openai.com/about (last visited Oct. 18, 2017).

[31] *See* Clay Dillow, *Tired of Repetitive Arguing About Climate Change, Scientist Makes a Bot to Argue for Him*, POPULAR SCI. (Nov. 3, 2010), http://www.popsci.com/science/article/2010-11/twitter-chatbot-trolls-web-tweeting-science-climate-change-deniers.

[32] *Cognitive Assistant that Learns and Organizes*, SRI INT'L, http://www.ai.sri.com/project/CALO (last visited Oct. 18, 2017). No relation.

[33] *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 532 (2015) [hereinafter Calo, *Robotics*].

[34] *See* Matthew Hutson, *Our Bots, Ourselves*, ATLANTIC, Mar. 2017, at 28, 28-29.

https://doi.org/10.6092/issn.2531-6133/8670

2.3. WHY AI "POLICY"?

That AI lacks a stable, consensus definition or instantiation complicates efforts to develop an appropriate policy infrastructure. We might question the very utility of the word "policy" in describing societal efforts to channel AI in the public interest. There are other terms in circulation. A new initiative anchored by MIT's Media Lab and Harvard University's Berkman Klein Center for Internet and Society, for instance, refers to itself as the "Ethics and Governance of Artificial Intelligence Fund."[35] Perhaps these are better words. Or perhaps it makes no difference, in the end, what labels we use as long as the task is to explore and channel AI's social impacts and our work is nuanced and rigorous.

This Essay uses the term policy deliberately for several reasons. First, there are issues with the alternatives. The study and practice of ethics is of vital importance, of course, and AI presents unique and important ethical questions. Several efforts are underway, within industry, academia, and other organizations, to sort out the ethics of AI.[36] But these efforts likely cannot substitute for policymaking. Ethics as a construct is notoriously malleable and contested: both Kant and Bentham get to say "should."[37] Policy — in the sense of official policy, at least — has a degree of finality once promulgated.[38] Moreover, even assuming moral consensus, ethics lacks a hard enforcement mechanism. A handful of companies dominate the emerging AI industry.[39] They are going to prefer ethical standards over binding rules for the obvious reason that no tangible penalties attach to changing or disregarding ethics should the necessity arise.

[35] *See Ethics and Governance of Artificial Intelligence*, MASS. INST. OF TECH. SCH. OF ARCHITECTURE & PLANNING, (l Oct. 15, 2017.

[36] *See e.g.*, IEEE, ETHICALLY ALIGNED DESIGN: A VISION FOR PRIORITIZING HUMAN WELLBEING WITH ARTIFICIAL INTELLIGENCE AND AUTONOMOUS SYSTEMS 2 (Dec. 13, 2016), . *Id.* at 125.

[37] *See* José de Sousa e Brito, *Derecho, Deber y Utilidad: de Bentham a Kant y de Mill a Aristóteles [Right, Duty, and Utility: From Bentham to Kant and from Mill to Aristotle]*, 17 REVISTA IBEROAMERICANA DE ESTUDIOS UTILITARISTAS [IBERO-AMERICAN MAGAZINE OF UTILITARIAN STUDIES] 91, 91-92 (2010).

[38] Law has, in H.L.A. Hart's terminology, a "rule of recognition." H.L.A. HART, THE CONCEPT OF LAW 100 (Joseph Raz et al. eds., 3rd ed. 2012).

[39] *See* Hutson, *supra* note 34.

Indeed, the unfolding development of a professional ethics of AI, while at one level welcome and even necessary, merits ongoing attention.[40] History is replete with examples of new industries forming ethical codes of conduct, only to have those codes invalidated by the federal government (the Department of Justice or Federal Trade Commission) as a restraint on trade. The National Society of Professional Engineers (hereinafter NSPE) alone has been the subject of litigation across several decades. In the 1970s, the DOJ sued the NSPE for establishing a "canon of ethics" that prohibited certain bidding practices; in the 1990s, the Federal Trade Commision sued the NSPE for restricting advertising practices.[41] The ethical codes of structural engineers have also been the subject of complaints, as have the codes of numerous other industries.[42] Will AI engineers fare differently? This is not to say that companies or groups should avoid ethical principles, only that we should pay attention to the composition and motivation of the authors of such principles, as well as their likely effects on markets and on society.

The term "governance" has its attractions. Like policy, governance is a flexible term that can accommodate many modalities and structures. Perhaps too flexible: it is not entirely clear what is being governed and by whom. Regardless, governance carries its own intellectual baggage — baggage that, like "ethics," is complicated by the industry's dominance of AI development and application. Setting aside the specific associations with "corporate governance,"[43] much contemporary governance literature embeds the claim

[40] *See* Romain Dillet, *Apple Joins Amazon, Facebook, Google, IBM and Microsoft in AI Initiative*, TECHCRUNCH (Jan. 27, 2017), https://techcrunch.com/2017/01/27/apple-joins-amazon-facebook-google-ibm-and-microsoft-in-ai-initiative. My own interactions with the Partnership on AI, which has a diverse board of industry and civil society, suggests that participants are genuinely interested in channeling AI toward the social good.

[41] *See* Nat'l Soc'y of Prof'l Eng'rs v. United States, 435 U.S. 679 (1978); *In re* Nat'l Soc'y of Prof'l Eng'rs, 116 F.T.C. 787 (1993), 1993 WL 13009653.

[42] *See In re* Structural Eng'rs Ass'n of N. Cal., 112 F.T.C. 530 (1989), 1989 WL 1126789, at *1 (invalidating code of ethics); *see, e.g.*, *In re* Conn. Chiropractic Ass'n, 114 F.T.C. 708, 712 (1991) (invalidating the ethical code of chiropractors); *In re* Am. Med. Ass'n, 94 F.T.C. 701 (1979), 1979 WL 199033, at *6 (invalidating the ethical guidelines of doctors), amended by *In re* Am. Med. Ass'n, 114 F.T.C. 575 (1991).

[43] Brian R. Cheffins, *The History of Corporate Governance*, *in* THE OXFORD HANDBOOK OF CORPORATE GOVERNANCE 46 (Douglas Michael Wright et al. eds., 2013).

https://doi.org/10.6092/issn.2531-6133/8670

that authority will or should devolve to actors other than the state.[44] While it is true that invoking the term governance can help insulate technologies from overt government interference — as in the case of Internet governance through non-governmental bodies such as the Internet Corporation for Assigned Names and Numbers (hereinafter ICANN) and the Internet Engineering Task Force (hereinafter IETF)[45] — the governance model also resists official policy by tacitly devolving responsibility to industry from the state.[46]

Meanwhile, several aspects of policy recommend it. Policy admits of the possibility of new laws, but does not require them. It may not be wise or even feasible to pass general laws about AI at this early stage, whereas it is very likely wise and timely to plan for AI's effects on society — including through the development of expertise, the investigation of AI's current and likely social impacts, and perhaps smaller changes to appropriate doctrines and laws in response to AI's positive and negative affordances.[47] Industry may seek to influence public policy, but it is not its role ultimately to set it. Policy conveys the necessity of exploration and planning, the finality of law, and the primacy of public interest without definitely endorsing or rejecting regulatory intervention. For these reasons, I have consciously chosen it as my frame.

3. KEY QUESTIONS FOR AI POLICY

This Part turns to the main goal of the Essay: a roadmap to the various challenges that AI poses for policymakers. It starts with discrete challenges, in the sense of

[44] *See* Roderick A.W. Rhodes, *The New Governance: Governing Without Government*, 44 POL. STUD. 652, 657 (1996); *see also* WENDY BROWN, UNDOING THE DEMOS: NEOLIBERALISM'S STEALTH REVOLUTION 122-123 (2015) (noting that "almost all scholars and definitions converge on the idea that governance" involves "networked, integrated, cooperative, partnered, disseminated, and at least partly self-organized" control).

[45] The United States government stood up both ICANN and IETF, but today they run largely interdependent of state control as non-profits.

[46] *Supra* note 44 and accompanying text.

[47] *See e.g.*, Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. (2018) (arguing *inter alia* for a clarification that companies may not invoke trade secret law to avoid scrutiny of their AI or algorithmic systems by criminal defendants).

https://doi.org/10.6092/issn.2531-6133/8670

specific domains where attention is warranted, and then discusses some general questions that tend to cut across domains. For the most part, the Essay avoids getting into detail about specific laws or doctrines that require reexamination and instead emphasize questions of overall strategy and planning.

The primary purpose of this Part is to give newer entrants to the AI policy world — whether from government, industry, media, academia, or otherwise — a general sense of what kinds of questions the community is asking and why. A secondary purpose is to help bring cohesion to this multifaceted and growing field. The inventory hopes to provide a roadmap for individuals and institutions to the various policy questions that arguably require their attention. The Essay tees up questions; it does not purport to answer them.

A limitation of virtually any taxonomic approach is the need to articulate criteria for inclusion — why are some questions on this list and not others?[48] Expert's opinions may vary on the stops they would include in a roadmap of key policy issues, and I welcome critique. There are several places where I draw distinctions or parallels that are not represented elsewhere in the literature, with which others may disagree. Ultimately this represents but one informed scholar's take on a complex and dynamic area of study.

3.1. JUSTICE AND EQUITY

Perhaps the most visible and developed area of AI policy to date involves the capacity of algorithms or trained systems to reflect human values such as fairness, accountability, and transparency ("FAT").[49] This topic is the subject of considerable study, including an established but accelerating literature on

[48] *Cf.* Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1132, 1139-1142 (2011) (critiquing Daniel Solove's taxonomy of privacy). If I have an articulable criterion for inclusion, it is sustained attention by academics and policymakers. Some version of the questions in this Part appear in the social scientific literature, in the White House reports on AI, in the Stanford AI 100 report, in the latest U.S. Robotics Roadmap, in the Senate hearing on AI, in the research wish list of the Partnership on AI, and in the various important public and private workshops such as AI Now, FAT/ML, and We Robot.

[49] *See e.g.*, KATE CRAWFORD ET AL., THE AI NOW REPORT: THE SOCIAL AND ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE NEAR TERM 6-8, July 7, 2016; *Thematic Pillars*, PARTNERSHIP ON AI, (l Oct. 14, 2017).

technological due process and at least one annual conference on the design of FAT systems.[50] The topic is also potentially quite broad, encompassing both the prospect of bias in AI-enabled features or products as well as the use of AI in making material decisions regarding financial, health, and even liberty outcomes. In service of teasing out specific policy issues, the Essay separates "applied inequality" from "consequential decision-making" while acknowledging the considerable overlap.

3.1.1. INEQUALITY IN APPLICATION

By inequality in application, I mean to refer to a particular set of problems involving the design and deployment of AI that works well for everyone. The examples here include everything from a camera that cautions against taking a Taiwanese-American blogger's picture because the software believes she is blinking,[51] to an image recognition system that characterizes an African American couple as gorillas,[52] to a translation engine that associates the role of engineer with being male and the role of nurse with being female.[53] These scenarios can be policy relevant in their own right, as when African Americans fail to see opportunities on Facebook due to the platform's (now discontinued) discriminatory allowances,[54] or when Asian Americans pay more for test preparation due to a price discriminatory algorithm.[55] They can also hold downstream policy ramifications, as when a person of Taiwanese descent has

[50] FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN MACHINE LEARNING, FAT/ML, http://www.fatml.org (last visited Oct. 14, 2017). *See also infra*, note 63 (discussing the term "technological due process").

[51] *See* Adam Rose, *Are Face-Detection Cameras Racist?*, TIME (Jan. 22, 2010), http://content.time.com/time/business/article/0,8599,1954643,00.html.

[52] *See* Jessica Guynn, *Google Photos Labeled Black People "Gorillas,"* USA TODAY (July 1, 2015), https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465.

[53] Aylin Caliskan et al., Semantics Derived Automatically from Language Corpora Contain Human-Like Biases, 356 Science 183, 183-184 (2017).

[54] *See* Julia Angwin & Terry Parris JR., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

[55] *See also* Julia Angwin & Terry Parris, *The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review*, PROPUBLICA (Sept. 1, 2015), https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review.

https://doi.org/10.6092/issn.2531-6133/8670

trouble renewing a passport,[56] or a young woman in Turkey researching international opportunities in higher education finds only references to nursing.[57]

There are a variety of reasons why AI systems might not work well for certain populations. For example, the designs may be using models trained on data where a particular demographic is underrepresented and hence not well reflected. More white faces in the training set of an image recognition AI means the system performs best for Caucasians.[58] There are also systems that are selectively applied to the marginalized populations. To illustrate, police use "heat maps" that purport to predict areas of future criminal activity to determine where to patrol but in fact lead to disproportionate harassment of African Americans.[59] Yet police do not routinely turn such techniques inward to predict which officers are likely to engage in excessive force.[60] Nor do investment firms initiate transactions on the basis of machine learning that they cannot explain to wealthy, sophisticated investors.[61]

The policy questions here are at least twofold. First, what constitutes best practice in minimizing discriminatory bias and by what mechanism (antidiscrimination laws, consumer protection, industry standards) does society incentivize development and adoption of best practice?[62] And second, how do we ensure that the risks and benefits of AI are evenly distributed across society? Each

[56] *See* Seline Cheng, *An Algorithm Rejected an Asian Man's Passport Photo for Having "Closed Eyes,"* QUARTZ (Dec. 7, 2016), https://qz.com/857122/an-algorithm-rejected-an-asian-mans-passport-photo-for-having-closed-eyes.

[57] *See* Adam Hadhazy , *Biased Bots: Artificial-Intelligence Systems Echo Human Prejudices*, PRINCETON UNIV. (Apr. 18, 2017), https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices ("Turkish uses a gender-neutral, third person pronoun, 'o.' Plugged into the online translation service Google Translate, however, the Turkish sentences 'o bir doktor' and 'o bir hemşire' are translated into English as 'he is a doctor' and 'she is a nurse.'"). *See generally* Caliskan et al., *supra* note 53 (discussing gender bias within certain computer systems occupations).

[58] *See* Rose, *supra* note 51 (discussing performance and race in the context of camera software).

[59] *See* Jessica Saunders et al., *Predictions Put into Practice: A Quasi Experimental Evaluation of Chicago's Predictive Policing Pilot*, 12 J. EXPERIMENTAL CRIMINOLOGY 347, 350-51 (2016).

[60] *See* Kate Crawford & Ryan Calo, *There Is a Blind Spot in AI Research*, 538 NATURE 311, 311-312 (2016).

[61] *Id.*; *see also* Will Knight, *The Financial World Wants to Open AI's Black Boxes*, MIT TECH. REV. (Apr. 13, 2017), https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes.

[62] *See e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 730-732 (2016) (discussing the strengths and weaknesses of employing antidiscrimination laws in the context of data mining).

set of questions is already occupying considerable resources and attention, including within the industries that build AI into their products, and yet few would dispute we have a long way to go before resolving them.

3.1.2. CONSEQUENTIAL DECISION-MAKING

Closely related, but distinct in my view, is the question of how to design systems that make or help make consequential decisions about people. The question is distinct from unequal application in general in that consequential decision-making, especially by government, often takes place against a backdrop of procedural rules or other guarantees of process.[63] For example, in the United States, the Constitution guarantees due process and equal protection by the government,[64] and European Union citizens have the right to request that consequential decisions by private firms involve a human (current) as well as a right of explanation for adverse decisions by a machine (pending).[65] Despite these representations, participants in the criminal justice system are already using algorithms to determine whom to police, whom to parole, and how long a defendant should stay in prison.[66]

There are three distinct facets to a thorough exploration of the role of AI in consequential decision-making. The first involves cataloguing the objectives and values that procedures and processes are trying to advance in a particular context. Without a thorough understanding of what it is that laws, norms, and other safeguards are trying to achieve, we cannot assess whether existing systems

[63] *See generally* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2007-2008) (arguing that AI decision-making jeopardizes constitutional procedural due process guarantees and advocating instead for a new "technological due process").

[64] *See* Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 110 (2014); *see also* Barocas & Selbst, *supra* note 62.

[65] *See* Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, 38 AI MAGAZINE, at 1 (2017), https://arxiv.org/pdf/1606.08813.pdf.

[66] *See* Saunders et al., *supra* note 59 (discussing heat zones in predictive policing); Angwin et al., *supra* note 2 (discussing the use of algorithmically-generated risk scores in criminal sentencing); *see also* JOSEPH WALKER, *State Parole Boards Use Software to Decide Which Inmates to Release*, WALL STREET JOURNAL (Oct. 11, 2013), https://www.wsj.com/articles/state-parole-boards-use-software-to-decide-which-inmates-to-release-1381542427.

are adequate let alone design new systems that are.[67] This task is further complicated by the tradeoffs and tensions inherent in such safeguards, as when the Federal Rules of Civil Procedure call simultaneously for a "just, speedy, and inexpensive" proceeding[68] or where the Sixth Amendment lays out labor-intensive conditions for a fair criminal trial that also has to occur quickly.[69]

The second facet involves determining which of these objectives and values can and should be imported into the context of machines. Deep learning, as a technique, may be effective in establishing correlation but unable to yield or articulate a causal mechanism.[70] AI here can say what will happen but not why. If so, the outputs of multi-layer neural nets may be inappropriate affiants for warrants, bad witnesses in court, or poor bases for judicial determinations of fact.[71] Notions such as prosecutorial discretion, the rule of lenity,[72] and executive pardon may not admit of mechanization at all. Certain decisions, such as the decision to take an individual off life support, raise fundamental concerns over human dignity and thus perhaps cannot be made even by objectively well-designed machines.[73]

A third facet involves the design and vetting of consequential decision-making systems in practice. There is widespread consensus that such systems

[67] *See generally* Citron, *supra* note 63 (discussing the goals of technological due process); Crawford & Schultz, *supra* note 64 (discussing due process and Big Data); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017) (arguing that current decision-making processes have not kept up with technology).

[68] FED. R. CIV. P. 1. I owe this point to my colleague Elizabeth Porter.

[69] U.S. CONST. amend. VI (requiring that a defendant be allowed to be presented with nature and cause of the accusations, to be confronted with the witnesses against him, to compel favorable witnesses, and to have the assistance of counsel, all as part of a speedy and public trial).

[70] *See* Jason Millar & Ian Kerr, *Delegation, Relinquishment, and Responsibility: The Prospect of Expert Robots*, *in* ROBOT LAW 102, 126 (Ryan Calo et al. eds., 2015).

[71] *Id.*; Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. Pa. L. REV. 871, 877-879 (2016) (discussing emerging technologies' interactions with current Fourth Amendment jurisprudence). *See generally* Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2017) (discussing machines as witnesses).

[72] The rule of lenity requires courts to construe criminal statutes narrowly, even where legislative intent appears to militate toward a broader reading. *E.g.*, McBoyle v. United States, 283 U.S. 25, 26-27 (1931) (declining to extend a stolen vehicle statute to stolen airplanes). For an example of a discussion of the limits of translating laws into machine code, *see* Harry Surden & Mary-Anne Williams, *Technological Opacity, Predictability, and Self-Driving Cars*, 38 CARDOZO L. REV. 121, 162-163 (2016).

[73] *See* James H. Moor, *Are There Decisions Computers Should Never Make?*, 1 NATURE & SYSTEM 217, 226 (1979).. This concern is also reflected *infra* in Part II.B concerning the use of force.

should be fair, accountable, and transparent.[74] However, other values — such as efficiency — are less well developed. The overall efficiency of an AI-enabled justice system, as distinct from its fairness or accuracy in the individual case, constitutes an important omission. As the saying goes, "justice delayed is justice denied": we should not aim as a society to hold a perfectly fair, accountable, and transparent process for only a handful of people a year.

Interestingly, the value tensions inherent in processual guarantees seem to find analogs, if imperfect ones, in the machine learning literature around performance tradeoffs.[75] Several researchers have measured how making a system more transparent or less biased can decrease its accuracy overall.[76] More obviously than efficiency, accuracy is an important dimension of fairness: we would not think of rolling a die to determine sentence length as fair, even if it is transparent to participants and unbiased as to demographics. The policy challenge involves how to manage these tradeoffs, either by designing techno-social systems that somehow maximize for all values, or by embracing a particular tradeoff in a way society is prepared to recognize as valid. The end game of designing systems that reflect justice and equity will involve very considerable, interdisciplinary efforts and is likely to prove a defining policy issue of our time.

3.2. USE OF FORCE

A special case of AI-enabled decision-making involves the decision to use force. As alluded to above, there are decisions — particularly involving the deliberate taking of life — that policymakers may decide never to commit exclusively to machines. Such is the gist of many debates regarding the development and deployment of autonomous weapons.[77] International consensus holds that people

[74] *Supra* notes 49–50 and accompanying text.

[75] *See Jon Kleinberg et al., Inherent Trade-Offs in the Fair Determination of Risk Scores,* 2017 *Proc. Innovations Theoretical Computer Sci.* 2, available at https://arxiv.org/abs/1609.05807.

[76] *Id.* at 1.

[77] Note that force is deployed in more contexts than military conflict. We might also ask after the propriety of the domestic use of force by border patrols, police, or even private security guards. For a discussion of these issues, see Elizabeth E. Joh, *Policing Police Robots*, 64 UCLA L. REV. DISCOURSE 516 (2016).

should never give up "meaningful human control" over a kill decision.[78] Yet debate lingers as to the meaning and scope of meaningful human control. Is monitoring enough? Target selection? And does the prescription extend to defensive systems as well, or only to offensive tactics and weapons? None of these important questions appear settled.[79]

There is also the question of who bears responsibility for the choices of machines. The automation of weapons may seem desirable in some circumstances or even inevitable.[80] It seems unlikely, for example, that the United States military would permit its military rivals to have faster or more flexible response capabilities than its own whatever their control mechanism.[81] Regardless, establishing a consensus around meaningful human control would not obviate all inquiry into responsibility in the event of a mistake or war crime. Some uses of AI presuppose human decision but nevertheless implicate deep questions of policy and ethics — as when the intelligence community leverages algorithms to select targets for remotely operated drone strikes.[82] And there are concerns that soldiers will be placed into the loop for the sole purpose of absorbing liability for wrongdoing, as anthropologist Madeline Clare Elish argues.[83] Thus, policymakers must work toward a framework for responsibility around AI and force that is fair and satisfactory to all stakeholders.

[78] *See* HEATHER M. ROFF & RICHARD MOYES, MEANINGFUL HUMAN CONTROL, ARTIFICIAL INTELLIGENCE AND AUTONOMOUS WEAPONS, Apr. 16, 2016, *available at* http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf.

[79] *See, e.g.*, Rebecca Crootof, *A Meaningful Floor for "Meaningful Human Control*,*"* 30 TEMP. INT'L & COMP. L.J. 53, 54 (2016) ("[T]here is no consensus as to what 'meaningful human control' actually requires.").

[80] Kenneth Anderson and Matthew Waxman in particular have made important contributions to the realpolitik of AI weapons. *See, e.g.*, Kenneth Andersn & Matthew Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, HOOVER INST. (Apr. 9, 2013), available at http://www.hoover.org/research/law-and-ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can (arguing that automated weapons are both desirable and inevitable).

[81] *Id.*

[82] *See generally* John Naughton, *Death by Drone Strike, Dished Out by Algorithm*, THE GUARDIAN (Feb. 21, 2016), ("General Michael Hayden, a former director of both the CIA and the NSA, said this: 'We kill people based on metadata.'").

[83] Madeleine Claire Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction* 1 (Mar. 20, 2016) (Columbia Univ. & Data & Soc'y Inst., We Robot 2016 Working Paper), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757236; *see also* Madeleine Claire Elish & Tim Hwang, *When Your Self-Driving Car Crashes, You Could Still Be the One Who Gets Sued*, QUARTZ (July 25, 2015), (applying this same reasoning to drivers of automatic cars).

https://doi.org/10.6092/issn.2531-6133/8670

3.3. SAFETY AND CERTIFICATION

As the preceding section demonstrates, AI systems do more than process information and assist officials to make decisions of consequence. Many systems — such as the software that controls an airplane on autopilot or a fully driverless car — exert direct and physical control over objects in the human environment. Others provide sensitive services that, when performed by people, require training and certification. These applications raise additional questions concerning the standards to which AI systems are held and the procedures and techniques available to ensure those standards are being met.[84]

3.3.1. SETTING AND VALIDATING SAFETY THRESHOLDS

Robots and other cyber-physical systems have to be safe. The question is how safe, and how do we know. In a wide variety of contexts, from commercial aviation to food safety, regulatory agencies set specific safety standards and lay out requirements for how those standards must be met. Such requirements do not exist for many robots.

Members of Congress and others have argued that we should embrace, for instance, driverless cars, to the extent that robots are or become safer drivers than humans.[85] However, "safer than humans" seems like an inadequate standard by which to vet any given autonomous system. Must the system be safer than humans unaided or humans assisted by cutting-edge safety features? Must the system be safer than humans overall or across all driving conditions? And just *how much* safer must driverless cars be than people before we tolerate or incentivize them? These are ultimately difficult questions not of technology but of policy.[86]

[84] *See* Henrik I. Christensen & al., FROM INTERNET TO ROBOTICS: A ROADMAP FOR US ROBOTICS 105-109, Nov. 7, 2016, http://jacobsschool.ucsd.edu/uploads/docs/2016/roadmap3-final-2b.pdf; *see also* STONE ET AL., *supra* note 7, at 42.

[85] *See, e.g.*, *Self-Driving Vehicle Legislation: Hearing Before the Subcomm. on Digital Commerce & Consumer Prot. of the H. Comm. on Energy & Commerce*, 115th Cong. (2017) (providing the opening statement of Rep. Greg Walden, Chairman, Subcomm. on Digital Commerce and Consumer Protection).

[86] *See generally* GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS (1970) (discussing different policies of adjudicating accident law).

Even assuming policymakers set satisfactory safety thresholds for driverless cars, drone delivery, and other instantiations of AI, we need to determine a proper and acceptable means of verifying that these standards are met. This process has an institutional or "who" component, as in, who does the testing (e.g., government testing, third-party independent certification, and self-certification by industry). It also has a technical or "how" component, as in, what are the testing methods (e.g., unit testing, fault-injection, virtualization, and supervision).[87] Local and international standards can be a starting point, but considerable work remains — especially as new potential applications and settings arise. For example, we might resolve safety thresholds for drone delivery or warehouse retrieval only to revisit the question anew for sidewalk delivery and fast food preparation.

There are further complications still. Some systems, such as high speed trading algorithms that can destabilize the stock market or cognitive radio systems that can interfere with emergency communications, may hold the potential, alone or in combination, to cause serious indirect harm.[88] Others may engage in harmful acts such as disinformation that simultaneously implicate free speech concerns.[89] Policymakers must determine what kinds of non-physical or indirect harms rise to the level that regulatory standards are required. Courts have a role in setting safety policy in the United States though the imposition of liability. It turns out that AI — especially AI that displays emergent properties — may pose challenges for civil liability.[90] Courts or regulators must address this misalignment. And markets also have a role, for instance, through the availability and conditions of insurance.[91]

[87] *Cf.* Bryant Walker Smith, *How Governments Can Promote Automated Driving*, 47 N.M. L. REV. 99, 101 (2017) (discussing different avenues through which government can promote automated driving and prepare community conditions to facilitate seamless integration of driverless cars once they become road-worthy).

[88] *See* Ryan Calo THE CASE FOR A FEDERAL ROBOTICS COMMISSION 9-10 (2014), https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/ [hereinafter CALO, COMMISSION].

[89] *See e.g.,* Bence Kollanyi et al. BENCE KOLLANYI ET AL., *Bots and Automation over Twitter during the Second U.S Presidential Debate* (Oct. 19, 2016), http://comprop.oii.ox.ac.uk/wpcontent/uploads/sites/89/2016/10/Data-Memo-Second-Presidential-Debate.pdf.

[90] *See* Calo, *Robotics*, *supra* note 33, at 538-45.

[91] For an overview, see Andrea Bertolini et al., *On Robots and Insurance*, 8 INT'L J. SOC. ROBOTICS 381,

https://doi.org/10.6092/issn.2531-6133/8670

3.3.2. CERTIFICATION

A closely related policy question arises where AI performs a task that, when done by a human, requires evidence of specialized skill or training.[92] In some contexts, society has seemed comfortable thus far dispensing with the formal requirement of certification when technology can be shown to be capable through supervised use. This is true of the autopilot modes of airplanes, which do not have to attend flight school. The question is open with respect to vehicles.[93] But what of technology under development today, such as autonomous surgical robots, the very value of which turns on bringing skills into an environment where no one has them? And how do we think about systems that purport to dispense legal, health, or financial advice, which requires adherence to complex fiduciary and other duties pegged to human judgment? Surgeons and lawyers must complete medical or law school and pass boards or bars. This approach may or may not serve an environment rich in AI, a dynamic that is already unfolding as the Food and Drug Administration works to classify downloadable mobile apps as medical devices[94] and other apps to dispute parking tickets.[95]

3.3.3. CYBERSECURITY

Finally, it is becoming increasingly clear that AI complicates an already intractable cybersecurity landscape.[96] First, as alluded to above, AI increasingly acts directly and even physically on the World.[97] When a malicious party gains

381 (2016) (discussing the need for adaptations in the insurance industry to respond to robotics).

[92] CHRISTENSEN ET AL., *supra* note 84, at 105.

[93] *See* MARK HARRIS, *Will You Need a New License to Operate a Self-Driving Car?*, IEEE SPECTRUM (Mar. 2, 2015), (discussing the current unsettled state of licensing schemes for "passengers" of driverless cars).

[94] *See* MEGAN MOLTENI, *Wellness Apps Evade the FDA, Only to Land in Court*, WIRED (Apr. 3, 2017).

[95] *See* AREZOU REZVANi, *'Robot Lawyer' Makes the Case Against Parking Tickets*, NPR (Jan. 16, 2017).

[96] *See generally* GREG ALLEN & TANIEL CHAN, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY (2017) (discussing ways of advancing policy on AI and national security).

[97] *See supra* Part II.B.

access to a cyber-physical system, suddenly bones instead of bits are on the line.[98] Second, ML and other AI techniques have the potential to alter both the offensive and defensive capabilities around cybersecurity, as dramatized by a recent competition held by DARPA where AI agents attacked and defended a network autonomously.[99] AI itself creates a new attack surface in the sense that ML and other techniques can be coopted purposefully to trick the system — an area known as adversarial machine learning. New threat models, standards, and techniques must be developed to address the new challenges of securing information and physical infrastructures.

3.4. PRIVACY AND POWER

Over the past decade, the discourse around privacy has shifted perceptibly.[100] What started out as a conversation about individual control over personal information has evolved into a conversation around the power of information more generally (i.e., the control institutions have over consumers and citizens by virtue of possessing so much information about them).[101] The acceleration of AI , which is intimately tied to the availability of data, will play a significant role in this evolving conversation in at least two ways: (1) the problem of pattern recognition and (2) the problem of data parity. Note that unlike some of the policy questions discussed above, which envision the consequential deployment of imperfect AI, the privacy questions that follow assume AI that is performing its assigned tasks only too well.

[98] *See* Ryan Calo, *Open Robotics*, 70 MD. L. REV. 571, 593-601 (2011) (discussing how robots have the ability to cause physical damage and injury).

[99] *See Cyber Grand Challenge*, DEF CON 24, https://www.defcon.org/html/defcon-24/dc-24-cgc.html (last visited Sept. 18, 2017); *see also* DARPA, *"Mayhem" Declared Preliminary Winner of Historic Cyber Grand Challenge*, DEF. ADVANCED RES. PROJECTS AGENCY (Aug. 4, 2016), https://www.darpa.mil/news-events/2016-08-04.

[100] The flagship privacy law workshop — Privacy Law Scholars Conference — recently celebrated its tenth anniversary, although of course privacy discourse goes back much further.

[101] *See, e.g.*, Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1952-58 (2013) (providing examples of how institutions have used surveillance to blackmail, persuade, and sort people into categories).

https://doi.org/10.6092/issn.2531-6133/8670

3.4.1. THE PROBLEM OF PATTERN RECOGNITION

The capacity of AI to recognize patterns people cannot themselves detect threatens to eviscerate the already unstable boundary between what is public and what is private.[102] AI is increasingly able to derive the intimate from the available. This means that freely shared information of apparent innocence — where you ate lunch, for example, or what you bought at the grocery store — can lead to insights of a deeply sensitive nature. With enough data about you and the population at large, firms, governments, and other institutions with access to AI will one day make guesses about you that you cannot imagine —what you like, whom you love, what you have done.[103]

Several serious policy challenges follow. The first set of challenges involves the acceleration of an existing trend around information extraction. Consumers will have next to no ability to appreciate the consequences of sharing information. This is a well-understood problem in privacy scholarship.[104] The community has addressed these challenges to privacy management under several labels, from databases to big data.[105] In that the *entire purpose* of AI is to spot patterns people cannot, however, the issue is rapidly coming to a head. Perhaps the mainstreaming of AI technology will increase the pressure on policymakers to step in and protect consumers. Perhaps not. Researchers are, at any rate, already exploring various alternatives to the status quo: fighting fire with fire by putting AI in the hands of consumers, for example, or abandoning notice and choice altogether in favor of rules and standards.[106] Whatever path we take should bear in

[102] *Cf.* Margot E. Kaminski et al., *Averting Robot Eyes,* 76 MD. L. REV. 983 (2017) (explaining the sensory capabilities of robots with limited AI).

[103] *See e.g.*, KASHMIR HILL, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012).
Tal Z. Zarsky has been a particularly close student of this phenomenon. *See generally* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503 (2013) (describing the types of trends and behaviors governments strive to predict with collected data).

[104] *See* Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1889-93 (2013).

[105] *See* Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1424-28 (2001); Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1003-09 (2017).

[106] For example, Decide.com was an artificially intelligent tool to help consumers decide when to purchase products and services. Decide.com was eventually acquired by eBay. *See generally* JOHN COOK, *eBay Acquires Decide.com, Shopping Research Site Will Shut Down Sept. 30*, GEEKWIRE (Sept. 6, 2013).

mind the many ways powerful firms can subvert and end run consumer interventions and the unlikelihood that consumers will keep up in a technological arms race.

Consumer privacy is under siege. Citizens, meanwhile, will have next to no ability to resist or reform surveillance.[107] Two doctrines in particular interact poorly with the new affordances of AI, both related to the reasonable expectation of privacy standard embedded in American constitutional law. Firstly, the interpretation of the Fourth Amendment by the courts that citizens enjoy no reasonable expectation of privacy in public or from a public vantage does not seem long for this world.[108] If everyone in public can be identified through facial recognition, and if the "public" habits of individuals or groups permit AI to derive private facts, then citizens will have little choice but to convey information to a government bent on public surveillance. Secondly, and related, the interpretation by the courts that individuals have no reasonable expectation of privacy in (non-content) information they convey to a third party, such as the telephone company, will continue to come under strain.[109]

Here is an area where grappling with legal doctrine seems inevitable. Courts are policymakers of a kind and the judiciary is already responding to these new realities by requiring warrants or probable cause in contexts involving public movements or third-party information. For example, in *United States v. Jones*, the Supreme Court required a warrant for officers to affix a GPS to a defendant's vehicle for the purpose of continuous monitoring. Five Justices in *Jones* articulated a concern over law enforcement's ability to derive intimate information from public travel over time.[110] There is a case before the Court at the time of writing

[107] *See generally* Ryan Calo, *Can Americans Resist Surveillance?*, 83 U. CHI. L. REV. 23 (2016) (analyzing the different methods American citizens can take to reform government surveillance and the associated challenges).

[108] *See* Joel Reidenberg, *Privacy in Public*, 69 U. MIAMI L. REV. 141, 143-47 (2014).

[109] Courts and statutes tend to recognize that the content of a message such as an email deserves greater protection than the non-content that accompanies the message, that is, where it is going, whether it is encrypted, whether it contains attachments, and so on. *Cf.* Riley v. California, 134 S. Ct. 2473 (2014) (invalidating the warrantless search and seizure of a mobile phone incident to arrest).

[110] *See* United States v. Jones, 565 U.S. 400, 415-17, 428-31 (2012).

https://doi.org/10.6092/issn.2531-6133/8670

concerning the ability of police to demand historic location data about citizens from their mobile phone provider.[111]

On the other hand, in the dog-sniffing case *Florida v. Jardines*, the Court also reaffirmed the principle that individuals have no reasonable expectation of privacy in contraband such as illegal drugs.[112] Thus, in theory, even if the courts resolve to recognize a reasonable expectation of privacy in public and in information conveyed to a third party, courts might still permit the government to leverage AI to search exclusively for illegal activity. Indeed, some argue that AI is not a search at all given that no human need to access the data unless or until the AI identifies something unlawful.[113] Even discounting the likely false positives, a reasonable question for law and policy is whether we want to live in a society with perfect enforcement.[114]

The second set of policy challenges involves not what information states and firms collect, but the way highly granular information gets deployed. Again, the privacy conversation has evolved to focus not on the capacity of the individual to protect their data, but on the power over an individual or group that comes from knowing so much about them. For example, firms can manipulate other market participants through a fine-tuned understanding of the individual and collective cognitive limitations of consumers.[115] Bots can gain our confidences to extract personal information.[116] Politicians and political operatives can micro-target messages, including misleading ones, in an effort to sway aggregate public attention.[117] All of these capacities are dramatically enhanced by the ability of AI

[111] Carpenter v. United States, 819 F.3d 880, 886 (6th Cir. 2016), *cert. granted*, 137 S. Ct. 2211 (2017).

[112] *See* Florida v. Jardines, 569 U.S. 1, 8-9 (2013).

[113] *See, e.g.*, Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 551 (2005) (arguing that a search does not occur until information is presented on a screen for a human to see, as opposed to simply being processed by the computer or transferred to a hard drive).

[114] *See* Christina M. Mulligan, *Perfect Enforcement of Law: When to Limit and When to Use Technology*, 14 RICH. J. L. & TECH., no. 4, 2008, at 78-102.

[115] *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1001-02 (2014) [hereinafter Calo, *Digital Market Manipulation*].

[116] *See* Ian R. Kerr, *Bots, Babes, and the Californication of Commerce*, 1 U. OTTAWA L. & TECH. J. 284, 312–17 (2004) (presciently describing the role of chat bots in online commerce).

[117] Ira S. Rubenstein, *Voter Privacy in the Age of Big Data*, 2014(5) WIS. L. REV. 861, 866–67 (2014).

https://doi.org/10.6092/issn.2531-6133/8670

to detect patterns in a complex world. Thus, a distinct area of study is the best law and policy infrastructure for a world of such exquisite and hyper-targeted control.

3.4.2. THE DATA PARITY PROBLEM

The data-intensive nature of machine learning, the technique yielding the most powerful applications of AI at the moment, has ramifications that are distinct from the pattern recognition problem. Simply put, the greater access to data a firm has, the better positioned it is to solve difficult problems with ML. As Amanda Levendowski explores, ML practitioners have essentially three options in securing sufficient data.[118] They can build the databases themselves, they can buy the data, or they can use "low friction" alternatives such as content in the public domain.[119] The last option carries perils for bias discussed above. The first two are avenues largely available to big firms or institutions such as Facebook or the military.

The reality that a handful of large entities (literally, fewer than a human has fingers) possess orders of magnitude more data than anyone else leads to a policy question around data parity. Smaller firms will have trouble entering and competing in the marketplace.[120] Industry research labs will come to outstrip public labs or universities, to the extent they do not already. Accordingly, cutting-edge AI practitioners will face even greater incentives to enter the private sphere, and ML applications will bend systematically toward the goals of profit-driven companies and not society at large. Companies will possess not only more and better information but a monopoly on its serious analysis.

Why label the question of asymmetric access to data a "privacy" question? I do so because privacy ultimately governs the set of responsible policy outcomes that arise in response to the data parity problem. Firms will, and already do, invoke consumer privacy as a rationale for not permitting access to their data.

[118] *See* Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. (forthcoming 2018) (manuscript at 23, 27-32), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3024938.

[119] *Id.*

[120] *Id.* at 26 (attributing this in part to the fact that larger firms have access to much more data).

This is partly why the AI policy community must maintain a healthy dose of skepticism toward "ethical codes of conduct" developed by industry.[121] Such codes are likely to contain a principle of privacy that, unless carefully crafted, operates to help shield the company from an obligation to share training data with other stakeholders.

A related question involves access to citizen data held by the government. Governments possess an immense amount of information; data that citizens are obligated to provide to the state forms the backbone of the contemporary data broker industry.[122] Firms big and small, as well as university and other researchers, may be able to access government data on comparable terms. But there are policy challenges here as well. Governments can and sometimes should place limits and conditions around sharing data.[123] In the United States at least, this means carefully crafting policies to avoid constitutional scrutiny as infringements on speech. The government cannot pick and choose with impunity the sorts of uses to which private actors place data released by the state.[124] At the same time, governments may be able to put sensible restrictions in place before compelling citizens to release private data.

To be clear: I do not think society should run roughshod over privacy in its pursuit of data parity. Indeed, I present this issue as a key policy challenge precisely because I believe we need mechanisms by which to achieve a greater measure of data parity without sacrificing personal or collective privacy. Some within academia and industry are already working on methods — including differential privacy and federated training — that seek to minimize the privacy impact of granting broader access to data-intensive systems.[125] The hard policy

[121] *See supra* Part I.

[122] *See* Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899, 1904 (2015).

[123] *Cf.* Julia Powles & Hal Hodson, *Google DeepMind and Healthcare in An Age of Algorithms*, HEALTH TECH. (Mar. 16, 2017), https://link.springer.com/article/10.1007%2Fs12553-017-0179-1 (outlining an incident where Google Deepmind accessed sensitive patient information, and what the British government could do to minimize that access).

[124] *See* Sorrell v. IMS Health Inc., 564 U.S. 552, 579-80 (2011).

[125] *See* James Vincent, *Google Is Testing a New Way of Training its AI Algorithms Directly on Your Phone*, VERGE (Apr. 10, 2017), https://www.theverge.com/2017/4/10/15241492/google-ai-user-data-federated-learning; *see also* Cynthia Dwork, *Differential Privacy*, *in* AUTOMATA, LANGUAGES AND PROGRAMMING 1, 2-3 (Michele Bugliesi et al. eds., 2007),

question is how to incentivize technical, legal, social, and other interventions that safeguard privacy even as AI is democratized.

3.5. TAXATION AND DISPLACEMENT OF LABOR

A common concern, especially in public discourse, is that AI will displace jobs by mastering tasks currently performed by people.[126] The classic example is the truck driver: many have observed that self-driving vehicles could obviate, or at least radically transform, this very common role. Machines have been replacing people since the Industrial Revolution (which posed its own challenges for society). The difference, many suppose, is twofold: first, the process of automation will be much faster, and second, very few sectors will remain untouched by AI's contemporary and anticipated capabilities.[127] This would widen the populations that could feel AI's impact and limit the efficacy of temporary unemployment benefits or retraining.

In its exploration of AI's impact on America, the Obama White House specifically inquired into the impact of AI on the job force and issued a report recommending a thicker social safety net to manage the upcoming disruption.[128] Some predict that new jobs will arise even as old ones fall away, or that AI will often improve the day to day of workers by permitting them to focus on more rewarding tasks involving judgment and creativity with which AI struggles.[129] Others explore the eventual need for a universal basic income, presumably underwritten by gains in productivity for automation, so that even those displaced entirely by AI have access to resources.[130] Still others wisely call for more

https://link.springer.com/content/pdf/10.1007%2F11787006.pdf [https://doi.org/10.1007/11787006_1].

[126] *See, e.g.*, FORD, *supra* note 3 ("[M]achines themselves are turning into workers . . . .").

[127] *See* ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 126–28 (2014).

[128] *See* EXEC. OFFICE OF THE PRESIDENT, ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY 35-42 (2016), https://obamawhitehouse.archives.gov/sites/whitehouse. gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF.

[129] *See* BRYNJOLFSSON & MCAFEE, *supra* note 127, at 134-38.

[130] Queena Kim, *As Our Jobs Are Automated, Some Say We'll Need a Guaranteed Basic Income*, NPR WEEKEND EDITION (Sept. 24, 2016), http://www.npr.org/2016/09/24/495186758/as-our-jobs-are-automated-some-say-well-need-a-guaranteed-basic-income.

https://doi.org/10.6092/issn.2531-6133/8670

and better information specific to automation so as to be able to better predict and scope the effects of AI.[131]

In addition to assessing impact and addressing displacement, policymakers will have to think through the effects of AI on the public fisc. Taxation is a highly complex policy domain that touches upon virtually all aspects of society; AI is no exception. Robots do not pay taxes, as the IRS once remarked in letter.[132] Bill Gates, Jr. thinks they should.[133] Others warn that a tax on automation amounts to a tax on innovation and progress.[134] Ultimately, federal and state policymakers will have to figure out how to keep the lights on in the absence of, for instance, the bulk of today's income taxes.

3.6. CROSS-CUTTING QUESTIONS (SELECTED)

The preceding list of questions is scarcely exhaustive as to consequences of AI for law and policy. Notably missing is any systemic review of the ways AI challenges existing legal doctrines. For example, that AI is capable of generating spontaneous speech or content raises doctrinal questions around the limits of the First Amendment as well as the contours of intellectual property.[135] Below, this Essay discusses the prospect that AI will wake up and kill us, which, if true, would seem to render every other policy context moot.[136] But the preceding inventory does cover most of the common big picture policy questions that tend to dominant serious discourse around AI.

[131] I am thinking particularly of the ongoing work of Robert Seamans at NYU Stern. *E.g.*, Robert Seamans, *We Won't Even Know If a Robot Takes Your Job*, FORBES (Jan. 11, 2017), https://www.forbes.com/sites/washingtonbytes/2017/01/11/we-wont-even-know-if-a-robot-takes-your-job/#36c2a0894bc5.

[132] *Treasury Responds to Suggestion that Robots Pay Income Tax*, 25 TAX NOTES 20 (1984) ("[I]nanimate objects are not required to file income tax returns.").

[133] *See* 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.

[134] Steve Cousins, *Is a "Robot Tax" Really an "Innovation Penalty"?*, TECHCRUNCH (Apr. 22, 2017), https://techcrunch.com/2017/04/22/save-the-robots-from-taxes.

[135] RONALD COLLINS & DAVID SKOVER, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE (2018); *see* Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 5 STAN. TECH. L. REV. 5, 21–27 (2012); James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 670 (2016).

[136] *Infra* Part III.

https://doi.org/10.6092/issn.2531-6133/8670

In addition to these specific policy contexts such as privacy, labor, or the use of force, recurrent issues arise that cut across domains. I have selected a few here that deserve greater attention: determining the best institutional configuration for governing AI, investing collective resources in AI that benefit individuals and society, addressing hurdles to AI accountability, and addressing our tendency to anthropomorphize technologies such as AI. I will discuss each of these systemic questions briefly in turn.

3.6.1. INSTITUTIONAL CONFIGURATION AND EXPERTISE

The prospect that AI presents individual or systemic risk, while simultaneously promising enormous potential benefits to people and society if responsibly deployed, presents policymakers with an acute challenge around the best institutional configuration for channeling AI. Today AI policy is done, if at all, by piecemeal approach; federal agencies, states, cities, and other government units tackle issues that most relate to them in isolation. There are advantages to this approach similar to the advantages of experimentation inherent in federalism — the approach is sensitive to differences across contexts and preserves room for experimentation.[137] But some see the piecemeal approach as problematic, calling, for instance, for a kind of FDA for algorithms to vet every system with a serious potential to cause harm.[138]

AI prefigures into a common, but I think misguided, observation about the relationship between law and technology. The public sees law as too slow to catch up to technologic innovation. Sometimes it is true that particular laws or regulations become long outdated as technology moves beyond where it was when the law was passed. For example, the Electronic Communications Privacy Act (“ECPA”), passed in 1986, interacts poorly with a post Internet environment in part because of ECPA’s assumptions about how electronic communications would

[137] New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting) (articulating the classic concept that states serve as laboratories of democracy).

[138] *E.g.*, Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 91, 104-06 (2017).

https://doi.org/10.6092/issn.2531-6133/8670

work.[139] But this is hardly inevitable, and often political. The Federal Trade Commission has continued in its mission of protecting markets and consumers unabated, in part because it enforces a standard — that of unfair and deceptive practice — that is largely neutral as to technology.[140] In other contexts, agencies have passed new rules or interpreted rules differently to address new techniques and practices.

The better-grounded observation is that government lacks the requisite expertise to manage society in such a deeply technically-mediated world.[141] Government bodies are slow to hire up and face steep competition from industry. When the State does not have its own experts, it must either rely on the self-interested word of private firms (or their proxies) or experience a paralysis of decision and action that ill-serves innovation.[142] Thus, one overarching policy challenge is how best to introduce expertise about AI and robotics into all branches and levels of government so they can make better decisions with greater confidence.

The solution could involve new advisory bodies, such as an official Federal Advisory Committee on Artificial Intelligence with an existing department or even a standalone Federal Robotics Commission.[143] Or it could involve resuscitating the Office of Technology Assessment, building out the Congressional Research Service, or growing the Office of Science and Technology Policy. Yet another approach involves each branch hiring its own technical staff at every level. The technical knowledge and affordances of the government — from the ability to test claims in a laboratory to a working understanding of AI in lawmakers and the judiciary — will ultimately affect the government's capacity to generate wise AI policy.

[139] *See* Orin S. Kerr, *The Next Generation Communications Privacy Act*, 162 U. PA. L. REV. 373, 375, 390 (2014).

[140] *See* Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MD. L. REV. 785 (2015).

[141] *See* CALO, COMMISSION, *supra* note 88, at 4.

[142] *Id.* at 2, 6-10 (listing examples of scenarios where a state or federal government had difficulty with new technologies when it lacked expertise).

[143] *Id.* at 3; Tom Kranzit, *Updated: Washington's Sen. Cantwell Prepping Bill Calling for AI Committee*, GEEKWIRE (July 10, 2017), https://www.geekwire.com/2017/washingtons-sen-cantwell-reportedly-prepping-bill-calling-ai-committee.

3.6.2. INVESTMENT AND PROCUREMENT

The government possesses a wide variety of means by which to channel AI in the public good. As recognized by the Obama White House, which published a separate report on the topic, one way to shape AI is by investing in it.[144] Investment opportunities include not only basic AI research, which advance the state of computer science and help ensure the United States remains globally competitive, but also support of social scientific research into AI's impacts on society. Policymakers can be strategic about where funds are committed and emphasize, for example, projects with an interdisciplinary research agenda and a vision for the public good.

In addition, and sometimes less well-recognized, the government can influence policy through what it decides to purchase.[145] States are capable of exerting considerable market pressures. Thus, policymakers at all levels ought to be thinking about the qualities and characteristics of the AI-enabled products government will purchase and the companies that create them. Policymakers can also use contract to help ensure best practice around privacy, security, and other values. This can in turn move the entire market toward more responsible practice and benefit society overall.

3.6.3. REMOVING HURDLES TO ACCOUNTABILITY

Many AI systems in use or development today are proprietary, and owners of AI systems have inadequate incentives to open them up to scrutiny. In many contexts, outside analysis is necessary for accountability. For example, in the context of justice and equity, defendants may seek to challenge adverse risk

[144] NETWORKING & INFO. TECH. RES. & DEV. SUBCOMM., NAT'L SCI. & TECH. COUNCIL, THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN 15-22 (Oct. 2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf.

[145] *See, e.g.*, Smith, *supra* note 87, at 118-19 (discussing procurement in connection with driverless cars); Whittington et al., *supra* note 122, at 1908-09 (discussing procurement in connection with open municipal data).

scores.[146] In the context of safety and certification, third parties seek to verify claims of safety or to evidence a lack of compliance. Several reports, briefs, and research papers have called upon policymakers to remove actual or perceived barriers to accountability, including: (1) trade secret law;[147] (2) the Computer Fraud and Abuse Act;[148] and (3) the anti-circumvention provision of the Digital Millennium Copyright Act.[149] This has led a number of experts to recommend the formal policy step of planning to remove such barriers in order to foster greater accountability for AI.

3.6.4. MENTAL MODELS OF AI

The next and final Part is devoted to a discussion of whether AI is likely to end humanity, itself partly a reflection of the special set of fears that tend to accompany anthropomorphic technology such as AI.[150] Policymakers arguably owe it to their constituents to hold a clear and accurate mental model of AI themselves and may have a role in educating citizens about the technology and its potential effects. Here they face an uphill battle, at least in the United States, due to decades of books, films, television shows, and even plays that depict AI as a threatening substitute for people.[151] That the task is difficult, however, does not discharge policymakers from their responsibilities.

[146] *See, e.g.*, Loomis v. State, 881 N.W.2d 749, 759 (Wis. 2016) (explaining that although a defendant may not challenge the algorithms themselves, he or she may still review and challenge the resulting scores).

[147] *E.g.*, Rebecca Wexler REBECCA WEXLER, *When a Computer Program Keeps You in Jail*, N.Y. TIMES (June 13, 2017), https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html.

[148] *E.g.*, CRAWFORD ET AL., *supra* note 49; STONE ET AL., *supra* note 7.

[149] *Id.*

[150] *Infra* Part III.

[151] There are examples dating back to the origin of the word robot. *See* Danny *78 Years Ago Today, BBC Aired the First Science Fiction Television Program*, SMITHSONIAN (Feb. 11, 2016), https://www.smithsonianmag.com/smart-news/78-years-ago-today-bbc-aired-first-science-fiction-television-program-180958126. There are also examples from the heyday of German silent film, METROPOLIS (Universum Film 1927), and contemporary American cinema, EX MACHINA (Universal Pictures International 2014). But the robot-as-villain narrative is not ubiquitous. Adults in Japan, for instance, grew up reading Astro Boy, a Manga or comic in which the robot is a hero. *Astro Boy [Mighty Atom] (Manga)*, TEZUKA IN ENGLISH, http://tezukainenglish.com/wp/?page_id=138 (last visited Oct. 18, 2017).

https://doi.org/10.6092/issn.2531-6133/8670

At a more granular level, the fact that instantiations of AI — such as Alexa (Echo), Siri, and Cortana, not to mention countless chat bots on a variety of social media platforms — take the form of social agents presents special challenges for policy driven by our hardwired responses to social technology as though it were human.[152] These challenges include the potential to influence children and other vulnerable groups in commercial settings and the prospect of disrupting civic or political discourse[153] or the further diminution of possibilities for solitude through a constant sense of being in the presence of another.[154] Others are concerned about the prospect of intimacy, in all its forms, between people and machines.[155] Whatever the particulars, that even the simplest AI can trigger social and emotional responses in people requires much more study and thought.

4. ON THE AI APOCALYPSE

Some set of readers may feel I have left out a key question: does AI present an existential threat to humanity? If so, perhaps all other discussions constitute the policy equivalent of rearranging deck chairs on the Titanic. Why fix the human world if AI is going to end it?

My own view is that AI does not present an existential threat to humanity, at least not in anything like the foreseeable future. Further, devoting disproportionate attention and resources to the AI apocalypse has the potential to distract policymakers from addressing AI's more immediate harms and challenges, and could discourage investment in research on AI's present social

[152] *See generally* Kate Darling, *"Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy*, *in* ROBOT ETHICS 2.0 FROM AUTONOMOUS CARS TO ARTIFICIAL INTELLIGENCE (Patrick Lin et al. eds., 2017).

[153] *See* Calo, *Digital Market Manipulation*, *supra* note 115; Kerr, *supra* note 116; Mulligan, *supra* note 114, at P101.

[154] *See* Ryan Calo, *People Can Be So Fake: A New Dimension to Privacy and Technology Scholarship*, 114 PENN ST. L. REV. 809, 843-46 (2010).

[155] *E.g.*, Noel Sharkey et al., OUR SEXUAL FUTURE WITH ROBOTS: A FOUNDATION FOR RESPONSIBLE ROBOTICS CONSULTATION REPORT 1 (2017), http://responsiblerobotics.org/wp-content/uploads/2017/07/FRR-Consultation-Report-Our-Sexual-Future-with-robots_
Final.pdf.

impacts.[156] How much attention to pay to a remote but dire threat is itself a difficult question of policy. If there is *any risk* to humanity then it follows that some thought and debate is worthwhile. But too much attention has real-world consequences.

Entrepreneur Elon Musk, physicist Stephen Hawking, and other famous individuals apparently believe AI represents civilization's greatest threat to date.[157] The most common citation for this proposition is the work of a British speculative philosopher named Nick Bostrom. In *Superintelligence*, Bostrom purports to demonstrate that we are on a path toward developing AI that is both enormously superior to human intelligence and presents a significant danger of turning on its creators.[158] Bostrom, it should be said, does not see a malignant superintelligence as *inevitable*. But he presents the danger as acute enough to merit serious consideration.

A number of prominent voices in AI have convincingly challenged *Superintelligence*'s thesis along several lines.[159] Firstly, they argue that there is simply no path toward machine intelligence that rivals our own across all contexts or domains. Yes, a machine specifically designed to do so can beat any human at chess. But nothing in the current literature around ML, search, reinforcement learning, or any other aspect of AI points the way toward modeling even the intelligence of a lower mammal in full, let alone human intelligence.[160] Some say

[156] *See generally* Crawford & Calo, *supra* note 60 ("Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems . . . .").

[157] *Cf.* Sonali Kohli, *Bill Gates Joins Elon Musk and Stephen Hawking in Saying Artificial Intelligence Is Scary*, QUARTZ (Jan. 29, 2015), https://qz.com/335768/bill-gates-joins-elon-musk-and-stephen-hawking-in-saying-artificial-intelligence-is-scary (discussing how many industry juggernauts believe AI poses a threat to mankind).

[158] *See generally* NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES (2014) (exploring the "most daunting challenge humanity has ever faced" and assessing how we might best respond).

[159] *See* Raffi Khatchadourian, *The Doomsday Invention*, THE NEW YORKER (Nov. 23, 2015), https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom. In other work, Bostrom argues that we are likely all living in a computer simulation created by our distant descendants. Nick Bostrom, *Are We Living in a Computer Simulation?*, 53 PHIL. Q. 243 (2003). This prior claim raises an interesting paradox: if AI kills everyone in the future, then we cannot be living in a computer simulation created by our decedents. And if we are living in a computer simulation created by our decedents, then AI did not kill everyone. I think it a fair deduction that Professor Bostrom is wrong about something.

[160] *See* Erik Sofge, *Why Artificial Intelligence Will Not Obliterate Humanity*, POPULAR SCI. (Mar. 19, 2015),http://www.popsci.com/why-artificial-intelligence-will-not-obliterate-humanity.
Australian computer scientist Mary Anne Williams once remarked to me, "We have been doing

https://doi.org/10.6092/issn.2531-6133/8670

this explains why claims of a pending AI apocalypse come almost exclusively from the ranks of individuals such as Musk, Hawking, and Bostrom who lack work experience in the field.[161] Secondly, critics of the AI apocalypse argue that *even if* we were able eventually to create a superintelligence, there is no reason to believe it would be bent on world domination, unless this were for some reason programmed into the system. As Yann LeCun, deep learning pioneer and head of AI at Facebook colorfully puts it: "computers do not have testosterone".[162]

Note that the threat to humanity could come in several forms. The first is that AI wakes up and purposefully kills everyone out of animus or to make more room for itself. This is the stuff of Hollywood movies and books by Daniel Wilson and finds next to no support in the computer science literature (which is why we call it science *fiction*).[163] The second is that AI accidentally kills everyone in the blind pursuit of some arbitrary goal — for example, an irresistibly powerful AI charged with making paperclips destroys the Earth in the process of mining for materials.[164] Fantasy is replete with examples of this scenario as well, from The Sorcerer's Apprentice in Disney's Fantasia to the ill-fated King Midas who demands the wrong blessing.[165] A third is that a very bad individual or group uses AI as part of an attempt to end human life.

Even if you believe the mainstream AI community that we are hundreds of years away from understanding how to create machines capable of formulating

artificial intelligence since that term was coined in the 1950s, and today robots are about as smart as insects.".

[161] *See* Connie Loizos, *This Famous Roboticist Doesn't Think Elon Musk Understands AI*, TECHCRUNCH (July 19, 2017), https://techcrunch.com/2017/07/19/this-famous-roboticist-doesnt-think-elon-musk-understands-ai (quoting Rodney Brooks as noting that AI alarmists "share a common thread, in that: they don't work in AI themselves").

[162] Dave Blanchard, *Musk's Warning Sparks Call for Regulating Artificial Intelligence*, NPR (July 19, 2017), http://www.npr.org/sections/alltechconsidered/2017/07/19/537961841/musks-warning-sparks-call-for-regulating-artificial-intelligence (citing an observation by Yan LeCun that the desire to dominate is not necessarily correlated with intelligence).

[163] *See* DANIEL H. WILSON, ROBOPOCALYPSE: A NOVEL (2011). Wilson's book is thrilling in part because Wilson has training in robotics and selectively adds accurate details to lend verisimilitude.

[164] *E.g.*, BOSTROM, *supra* note 158, at 123.

[165] ARISTOTLE, POLITICS 17 (B. Jowett trans., Oxford, Clarendon Press 1885) (describing Midas' uncontrollable power to turn everything he touched into gold); FANTASIA (Walt Disney Productions 1940) (where an army of magically enchanted brooms ceaselessly fill a cauldron with water and almost drown Mickey Mouse). I owe the analogy to King Midas to Stuart Russell, a prominent computer scientist at UC Berkeley who is among the handful of AI experts to join Musk and others in worrying aloud about AI's capacity to threaten humanity.

an intent to harm, and would not do so anyway, you might be worried about the second and third scenarios. The second argument has its attractions: people can set goals for AI that lead to unintended consequences. Computers do what you tell them to do, as the saying goes, not what you want them to do. But it is also important to consider the characteristics of the system AI doomsayers envision. This system is simultaneously *so primitive* as to perceive a singular goal, such as making paperclips, arbitrarily assigned by a person, and yet *so advanced* as to be capable of outwitting and overpowering the sum total of humanity in pursuit of this goal. I find this combination of qualities unlikely, perhaps on par with the likelihood of a malicious AI bent on purposive world domination.

Perhaps more worrying is the potential that a person or group might use AI in some way to threaten all of society. This is the vision of, for example, Daniel Suarez in his book *Daemon*[166] and has been explored by workshops such as *Bad Actors in AI* at Oxford University.[167] We can imagine, for instance, a malicious actor leveraging AI to compromise nuclear security, using trading algorithms to destabilize the market, or spreading misinformation through AI-enabled micro-targeting to incite violence. The path from malicious activity to existential threat, however, is narrow, and for now the stuff of graphic novels.[168]

Only time can tell us for certain who is wrong and who is right. Although it may not be the mainstream view among AI researchers and practitioners, I have attended several events where established computer scientists and other smart people reflected some version of the doomsday scenario.[169] If there is even a remote chance that AI will wake up and kill us (i.e., if the AI apocalypse is a low

[166] Daniel Suarez, Daemon (2009).

[167] *See Bad Actors and Artificial Intelligence Workshop*, The Future of Humanity Inst. (Feb. 24, 2017), https://www.fhi.ox.ac.uk/bad-actors-and-artificial-intelligence-workshop.

[168] *See e.g.*, Alan Moore, Dave Gibbons & John Higgins, Watchmen 382-90 (1995) (graphically portraying the chaos that ensues after a villain engineers a giant monster cloned from a human brain to destroy New York).

[169] *See, e.g.*, *Past Events*, The Future of Life Inst., https://futureoflife.org/past_events (last visited Oct. 18, 2017) (cataloguing past events hosted by the Future of Life Institute, an organization that is devoted to "safeguarding life and developing optimistic visions of the future, including positive ways for humanity to steer its own course considering new technologies and challenges").

probability, high loss problem), then perhaps we should pay some attention to the issue.

The strongest argument against focusing overly on Skynet or HAL in 2017 is the opportunity cost. AI presents numerous pressing challenges to individuals and society in the very short term. The problem is not that AI "will get too smart and take over the world," computer scientist Pedro Domingos writes, "the real problem is that [it's] too stupid and [has] already."[170] By focusing so much energy on a quixotic existential threat, we risk, in information scientist Solon Barocas' words, an AI Policy Winter.

4.1. CONCLUSION

This Essay had two goals. First, it sought to provide a brief primer on AI by defining AI in relation to previous and constituent technologies and by noting the ways the contemporary conversation around AI may be unique. One of the most obvious breaks with the past is the extent and sophistication of the policy response to AI in the United States and around the world. Thus the Essay sought, second, to provide an inventory or roadmap of the serious policy questions that have arisen to date. The purpose of this inventory is to inform AI policymaking, broadly understood, by identifying the issues and developing the questions to the point that readers can initiate their own investigation. The roadmap is idiosyncratic to the author but informed by longstanding participation in AI policy.

AI is remaking aspects of society today and likely to shepherd in much greater changes in the coming years. As this Essay emphasized, the process of societal transformation carries with it many distinct and difficult questions of policy. Even so, there is reason for hope. We have certain advantages over our predecessors. The previous industrial revolutions had their lessons and we have access today to many more policymaking bodies and tools. We have also made

[170] PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 286 (2015).

interdisciplinary collaboration much more of a standard practice. But perhaps the greatest advantage is timing: AI has managed to capture policymakers' imaginations early enough in its life-cycle that there is hope we can yet channel it toward the public interest. I hope this Essay contributes in some small way to this process.