Artificial Intelligence:
An Introduction to the Legal, Policy and Ethical Issues

James X. Dempsey
Berkeley Center for Law & Technology
August 10, 2020

> "[I]ncreasingly useful applications of AI, with potentially profound positive impacts on our society and economy, are likely to emerge between now and 2030." AI 100 Study.[1]
>
> "The development of full artificial intelligence could spell the end of the human race." Stephen Hawking

The age of artificial intelligence is dawning. Already AI is widespread, appearing in multiple contexts, from medical diagnosis to driving directions to stock trading to social networking to policing. As science fiction writer William Gibson said, the future is already here, it's just not evenly distributed. It seems likely that every sector of economic activity and every aspect of social and political life will be (is already being) affected by AI. It also seems likely, however, that the full impact of AI is impossible to predict. Undoubtedly, there is hyperbole in today's predictions about AI, both positive and dystopian. In thinking about AI, we should keep in mind the observation of another visionary, Roy Amara, founder of the Institute for the Future, who said that we tend to overestimate the short term impact of a new technology, but underestimate its long term impact.

While the exact shape of the AI-influenced future is uncertain, there is widespread assumption that the impacts of AI will be profound. As the European Commission said in 2018, "The way we approach AI will define the world we live in."[2] Or, as Russia's President said in 2017, the country that masters AI will "get to rule the world."[3]

Among its many profound implications, AI poses challenges for law, corporate and government policy, and ethics. Courts are being asked to apply traditional legal doctrines to complex and purportedly unexplainable systems. Policymakers are deciding whether to modify existing regulatory structures to specifically address AI. Overarching these granular choices is the public policy challenge of promoting and shaping the development of AI in ways that will be beneficial while mitigating its negative impacts. More law, or AI-specific law, may not be the answer. The report of the AI 100 Study panel convened under the auspices of Stanford University concluded: "Rather than 'more' or 'stricter'

---

[1] PETER STONE ET AL., ARTIFICIAL INTELLIGENCE AND LIFE IN 2030, ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE 4 (2016), https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf [hereinafter AI 100 STUDY].

[2] *Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions,* COM (2018) 137 final (Apr. 4, 2018), https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF.

[3] David Meyer, *Vladimir Putin Says Whoever Leads in Artificial Intelligence Will Rule the World*, FORTUNE (Sept. 4, 2017), http://fortune.com/2017/09/04/ai-artificial-intelligence-putin-rule-world/; Russia Insight, *Whoever Leads in AI Will Rule the World* (Sept. 4, 2017), https://www.youtube.com/watch?v=2kggRND8c7Q.

regulation, policies should be designed to encourage helpful innovation, generate and transfer expertise, and foster broad corporate and civic responsibility for addressing critical societal issues raised by these technologies."[4] To decide just what policies are needed, officials in all branches and at all levels of government will need access to technical expertise in AI—to translators who can explain the technology behind AI.[5]

Although AI presents substantial legal issues, it is important to recognize that many traditional doctrines and statutes of general application could answer the issues posed by AI or at least provide the starting point for responding to those issues. As Judge Frank Easterbrook advised, rather than creating technology-specific rules, it is usually better to first develop a sound rule, then apply it to computer innovations.[6]

This paper seeks to introduce some of the types of legal, policy, and, to a lesser degree, ethical issues that AI poses. The paper focuses largely on developments and debates in the United States, with occasional reference to the law or policy frameworks of other countries. It should be viewed solely as an introduction. There are undoubtedly other issues not addressed, and for each of the issues that is mentioned there is already a rich literature that it is impossible to even summarize here.

I. What is Artificial Intelligence?

Although it represents one of the major technologies of our time, there is no common or accepted definition of artificial intelligence ("AI"). An October 2016 report issued by the Obama Administration said, "Some define AI loosely as a computerized system that exhibits behavior that is commonly thought of as requiring intelligence. Others define AI as a system capable of rationally solving complex problems or taking appropriate actions to achieve its goals in whatever real world circumstances it encounters." A 2018 book issued by Microsoft defines AI as "a set of technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways that are similar to what people do."[7] (But in key ways AI is *not* similar to human thinking.) The European Commission's Communication on AI states, "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions–with some degree of autonomy–to achieve specific goals."

---

[4] AI 100 STUDY, *supra* note 1, at 43 ("Effective governance requires more experts who understand and can analyze the interactions between AI technologies, programmatic objectives, and overall societal values.")

[5] *Id.*

[6] Frank H. Easterbrook, *Cyberspace and the Law of the Horse,* 1996 U. CHI. LEGAL F. 207, 208 (1996) ("Develop a sound law of intellectual property, then *apply* it to computer networks.").

[7] Microsoft, THE FUTURE COMPUTED: ARTIFICIAL INTELLIGENCE AND ITS ROLE IN SOCIETY 28 (2018) https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf.

AI can be divided into two basic categories: narrow (or weak) and general (or strong). Narrow AI competes with human thinking and reasoning in one domain.[8] An example of narrow AI is IBM's Deep Blue chess-playing program.[9] It could beat the best chess player in the world, but it can't play checkers. Even very robust AI may be narrow: the AI in a self-driving car could not fly an airplane or even steer a bicycle. (However, techniques learned in developing the AI for the self-driving car may make it easier to develop AI for a broad range of other functions.)

Narrow AI is already pervasive:

- AI makes trades on Wall Street,[10] determines credit scores, reads and rates resumes,[11] and interprets x-rays.[12]
- It is being integrated into policing and the criminal justice system.[13]
- AI is prominent in self-driving cars, robotic surgical equipment, and medical diagnostic systems.
- Your phone uses AI to give prompts for words when you are composing a text.
- The navigational software on the same phone uses AI to determine the fastest route home.
- Many of the "May I help you?" boxes that pop up online are chat bots, automated systems that interpret users' questions and return answers as if being provided by a human customer service representative.

---

[8] *See* Ben Dickson, *What is Narrow, General, and Super Artificial Intelligence*, TECHTALKS (May 12, 2017) https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-super-artificial-intelligence/; Peter Voss, *From Narrow to General AI*, MEDIUM (Oct. 3, 2017), https://medium.com/intuitionmachine/from-narrow-to-general-ai-e21b568155b9.

[9] *See* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U. CAL. DAVIS 399, 405 n.22 (2017).

[10] Michelle Fleury, *How Artificial Intelligence is Transforming the Financial Industry*, BBC (Sept. 16, 2015), www.bbc.com/news/business-34264380. ("About three quarters of trades on the New York Stock Exchange and Nasdaq are [now] done by algorithms . . . .")

[11] Riia O'Donnell, *AI in recruitment isn't a prediction — it's already here*, HR DIVE (Jan. 18, 2018), https://www.hrdive.com/news/ai-in-recruitment-isnt-a-prediction-its-already-here/514876/.

[12] Greg Freiherr, *Why AI By Any Name Is Sweet For Radiology*, IMAGING TECHNOLOGY NEWS (Jan. 10, 2018), https://www.itnonline.com/article/why-ai-any-name-sweet-radiology.

[13] Elizabeth E. Joh, *Artificial Intelligence and Policing: First Questions*, 41 SEATTLE UNIV. L. REV. 1139 (2018), https://ssrn.com/abstract=3168779. See also Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, National Institute of Justice (Oct. 8, 2018) (discussing NIJ support for AI research in four areas: video and image analysis, DNA analysis, gunshot detection, and crime forecasting) https://www.nij.gov/journals/280/Pages/using-artificial-intelligence-to-address-criminal-justice-needs.aspx.

- AI is behind Google search and determines what you see on Facebook; it allows Amazon to suggest to you what books to buy, Netflix to suggest what movies to watch, Spotify to compile playlists.[14]
- AI-based language translation is built into Google's search engine and is widely available in other services.[15]

Whereas narrow AI automates a single activity typically performed by a human, artificial general intelligence (AGI) can perform tasks in more than one domain. It aims to solve problems never before encountered and to learn how to perform new tasks.[16] It is often said that general AI thinks, reasons, and deduces in a manner similar to humans,[17] but, again, that is misleading in some important ways

Although there is a spectrum of developments between narrow and general AI, most commentators agree that no system yet developed can truly be designated artificial general intelligence.[18] In fact, it is debated whether artificial general intelligence will ever be attained. However, some see important steps towards AGI in systems such as Google's Deep Mind.[19]

Many definitions of AI recognize that AI is not one thing but a set of techniques. The non-profit research organization AI Now emphasizes that artificial intelligence "refers to a

---

[14] Bernard Marr, *The Amazing Ways Spotify Uses Big Data, AI and Machine Learning To Drive Business Success*, FORBES (Oct. 30, 2017), https://www.forbes.com/sites/bernardmarr/2017/10/30/the-amazing-ways-spotify-uses-big-data-ai-and-machine-learning-to-drive-business-success/#38eb7bf64bd2.

[15] Gideon Lewis-Kraus, *The Great AI Awakening*, N.Y. TIMES (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html (describing AI's role in the evolution of Google translate); Allison Linn, *Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English*, THE AI BLOG (Mar. 14, 2018), https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/. *But see* David Pring-Mill, *Why Hasn't AI Mastered Language Translation*, SINGULARITYHUB (Mar. 4, 2018), https://singularityhub.com/2018/03/04/why-hasnt-ai-mastered-language-translation/; Celia Chen, *AI-powered translation still needs work after errors mar debut at Boao Forum*, SOUTH CHINA MORNING POST (Apr. 16, 2018), http://www.scmp.com/tech/innovation/article/2141940/ai-powered-translation-still-needs-work-after-errors-mar-debut-boao.

[16] *See* Jonathan Howard, *A Big Data Cheat Sheet: From Narrow AI to General AI*, MEDIUM (May 23, 2017), https://blog.statsbot.co/3-types-of-artificial-intelligence-4fb7df20fdd8.

[17] *See* Peter Voss, *From Narrow to General AI*, MEDIUM (Oct. 3, 2017), https://medium.com/intuitionmachine/from-narrow-to-general-ai-e21b568155b9.

[18] Ben Dickson, *What is Narrow, General, and Super Artificial Intelligence*, TECHTALKS (May 12, 2017) https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-super-artificial-intelligence/ ("Narrow AI is the only form of Artificial Intelligence that humanity has achieved so far.")

[19] *See id.*

constellation of technologies, including machine learning, perception, reasoning, and natural language processing."[20] Recent developments in AI combine a number of technologies:

- Algorithms. Many AI systems involve algorithms, which can be defined as recipes for processing data or performing some other task. Much of the concern that was expressed several years ago with the fairness and transparency of algorithmic decision-making now is being cast in terms of AI.
- Machine learning (ML). A machine learning algorithm can process data and make predictions without relying solely on pre-programmed rules. For example, an ML system can use data about some known (often human-classified) objects or events of a particular category ("training data") to identify correlations that can be used in order to make assessments about other objects or events of the same kind.[21] The algorithm can "learn" by tuning the weightings of features it relies on in the data–essentially testing multiple different weightings—to optimize its predictions, so the quality of its predictions improves over time and with more data.
- Deep learning is a sub-field of machine learning, where algorithms perform two important tasks that human programmers had previously performed: defining what features in a dataset to analyze and deciding how to weight those factors to deliver an accurate prediction.[22]
- Neural networks. Deep learning uses neural networks, which are programs that, by their interconnections, roughly approximate the neurons in a brain.[23] A neural network analyzes inputs and makes a prediction; if the prediction is wrong, the deep learning algorithm adjusts the connections among the neurons until prediction accuracy improves.
- Natural language processing. AI system have gotten much better at interpreting human language, both written and spoken.

---

[20] AI NOW INSTITUTE, THE AI NOW REPORT: THE SOCIAL AND ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE NEAR-TERM (July 7, 2016), https://ainowinstitute.org/AI_Now_2016_Report.pdf.

[21] David Kelnar, *The Fourth Industrial Revolution: A Primer on Artificial Intelligence*, MEDIUM (Dec. 2, 2016), https://medium.com/mmc-writes/the-fourth-industrial-revolution-a-primer-on-artificial-intelligence-ai-ff5e7fffcae1 ("All machine learning is AI, but not all AI is machine learning."). There are more than 15 approaches to machine learning, each of which uses a different algorithmic structure to optimize predictions based on the data received. Id. For a more nuanced description of machine learning, see Ben Buchanan and Taylor Miller, *Machine Learning for Policymakers*, Belfer Center (June 2017) https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf.

[22] *See* David Kelnar, *The Fourth Industrial Revolution: A Primer on Artificial Intelligence*, MEDIUM (Dec. 2, 2016), https://medium.com/mmc-writes/the-fourth-industrial-revolution-a-primer-on-artificial-intelligence-ai-ff5e7fffcae1 ("All deep learning is machine learning, but not all machine learning is deep learning.").

[23] *See id.*

Recent advances in machine learning and deep learning techniques have drawn on two key resources: (1) huge increases in computational power and (2) the availability of massive and ever growing amounts of data.[24] Indeed, some of the attention currently devoted to AI is a continuation of the attention that four or five years ago was lavished on big data. (The role of big data in AI research has policy implications discussed below.)

AI offers the potential to solve problems that humans cannot solve on their own, especially those involving large amounts of data and large numbers of options. AI could correct for human error and bias. For example, an AI-based automobile may avoid drunk driving accidents[25] and AI-based risk assessment programs can avoid racial bias in credit and criminal sentencing decisions.[26]

However, AI is not magic. All AI programs involve human decisions and trade-offs. Algorithms are not value-free. AI may replicate human error or bias or introduce new types of errors or bias.[27] Judges, regulators, and policymakers need to understand these biases and how they may arise in seemingly objective, data-driven processes. A self-driving car may struggle with ethical choices that humans easily process, such as choosing between hitting a shopping cart and a baby stroller.[28] An AI system intended to allocate police resources where crime is highest may replicate past bias in patterns of policing.[29] AI

---

[24] *See* Calo, *supra* note 9, at 402.

[25] *See* Dorothy Glancy et al., *A Look at the Legal Environment for Driverless Vehicles* (National Academies Press 2016) [hereinafter NAS Driverless Cars Study].

[26] *See* AI 100 STUDY, *supra* note 1, at 43.

[27] A large and growing body of research identifies racial and gender bias in various AI-based systems. See, for example, Patrick Grother, Mei Ngan and Kayee Honaoka, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, NIST (Dec. 2019) https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf (study of 189 facial recognition algorithms from 99 developers found that most had higher error rates with Black, Native American, and Asian faces); Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, Machine Bias (Pro Publica 2016)[hereinafter Pro Publica Study] (finding that software used to measure recidivism was twice as likely to mistakenly flag Black defendants as being at a higher risk of committing future crimes and twice as likely to incorrectly flag white defendants as low risk): Marcelo Prates, Pedro Avelar and Luis Lamb, Assessing Gender Bias in Machine Translation – A Case Study with Google Translate, Neural Computing and Applications (March 2019) https://www.researchgate.net/publication/332030363_Assessing_gender_bias_in_machine_translation_a_case_study_with_Google_Translate; Latanya Sweeney, Discrimination in Online Ad Delivery (January 28, 2013) https://ssrn.com/abstract=2208240.

[28] *See, e.g.*, Alex Hern, *Self-Driving Cars Don't Care About Your Moral Dilemmas*, THE GUARDIAN (Aug. 22, 2016, 10.08 AM), https://www.theguardian.com/technology/2016/aug/22/self-driving-cars-moral-dilemmas.

[29] Rashida Richardson, Jason Schultz, and Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice* (February 13, 2019), New York University Law Review Online https://ssrn.com/abstract=3333423

trained on data that reflects biases that infected past decisions could incorporate those biases into future decision-making, yet give such decisions the appearance of objectivity.[30] The complexity of AI poses challenges to accountability. Human programmers may not be able to explain how a neural network made its predictions.[31] Accountability may be stymied by proprietary claims that developers of AI-based products use to shield their underlying algorithms.[32] A growing body of literature questions the reliability of AI for certain applications,[33] while another body of research is uncovering the vulnerability of AI systems to adversarial attack.

A new field of AI research addresses the black box problem: explainable AI.[34] Companies have emerged that claim to offer explainable AI.[35]

II. A Sampling of the Legal and Ethical Issues Posed by AI

Countries around the world already have laws that address the apportionment of liability for injuries resulting from unreasonable behaviors or defective products, that define intellectual property rights, that seek to ensure fairness in credit and employment decisions, that protect privacy, and so on. By and large, "[t]here are no exceptions to these laws for AI systems."[36] Nor need there be.

However, as they have in the face of other technological changes, courts will encounter challenges in applying traditional rules to AI, and regulatory agencies and legislatures must determine whether special rules are needed.

---

[30] *See* AI 100 STUDY, *supra* note 1, at 43.

[31] *See id.*

[32] Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. Rev. 54 (2019).

[33] See, for example, Matthew Salganik, et al., Measuring the predictability of life outcomes with a scientific mass collaboration, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (2020) https://www.pnas.org/content/117/15/8398 (160 teams, some using complex machine-learning methods and working with thousands of predictor variable on thousands of families, were not able to make accurate predictions for basic life outcomes).

[34] Tathagata Chakraborti et al., *The Emerging Landscape of Explainable AI Planning and Decision Making* (02/26/2020), https://deepai.org/publication/the-emerging-landscape-of-explainable-ai-planning-and-decision-making.

[35] See, for example, Sebastian Moss, Lockheed Martin partners with DarwinAI for Explainable AI, AI BUSINESS (May 26, 2020) https://aibusiness.com/document.asp?doc_id=761328&site=aibusiness.

[36] MICROSOFT, THE FUTURE COMPUTED: ARTIFICIAL INTELLIGENCE AND ITS ROLE IN SOCIETY (2018), https://news.microsoft.com/cloudforgood/_media/downloads/the-future-computed-english.pdf.

A. Product Liability

Most countries have laws establishing civil liability for negligent or unreasonable behavior that causes damage and for addressing the harms caused by defective products. For AI, as they have in other areas, legislators may find it desirable to adopt statutes to clarify or modify these rules, or they may delegate rulemaking authority to regulatory bodies. Meanwhile, courts will fit AI into the existing legal frameworks.

"Robots cannot be sued,"[37] but their manufacturers and operators can. Already, there has been extensive litigation against manufacturers by workers injured on the job by robots, over the safety of surgical robots, over autopilot systems in airplanes, and over the software already embedded in automobiles.[38] By and large, the courts have applied traditional concepts to conceptualize liability and apportion it among machines, their makers, and users.

Globally, legal rules defining liability for products vary, but there has been a distinct movement away from reliance on negligence and warranty towards the concept of strict liability for defective products.[39] This approach holds product manufacturers liable for "defects" in the design or manufacture of the products they make or for failure to provide sufficient warning of the risks of such products.

In the US, questions of liability are largely a matter of common law, augmented by statute and varying somewhat state to state, but the principle of strict liability for defective products is dominant. In Europe, movement towards a strict liability regime began in 1977 with the Council of Europe Convention on Products Liability in regard to Personal Injury and Death. In 1985, the European Union adopted a Product Liability Directive that created a regime of strict liability for defective products.[40] To take one other example, in Japan, under the Product Liability Act of 1994, manufacturers face liability for injuries and

---

[37] *United States v. Athlone Indus., Inc.*, 746 F.2d 977, 979 (3d Cir. 1984).

[38] *See* Quinn Emanuel Trial Lawyers, *Artificial Intelligence Litigation: Can the Law Keep Pace with the Rise of the Machines* (2018), https://www.quinnemanuel.com/the-firm/publications/article-december-2016-artificial-intelligence-litigation-can-the-law-keep-pace-with-the-rise-of-the-machines/.

[39] *See generally* Baker McKenzie, *Asia Pacific Product Liability Guide* 207 (July 2017), https://www.bakermckenzie.com/en/insight/publications/2017/03/ap-product-liability-guide; GETTING THE DEAL THROUGH, LIABILITY IN 29 JURISDICTIONS WORLDWIDE (Harvey Kaplan, Gregory Fowler & Simon Castley, eds., 2014), http://www.acc.com/_cs_upload/vl/membersonly/Article/1394895_1.pdf.

[40] Council Directive 85/374/EEC, 1995 O.J. (L 210), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31985L0374:en:HTML. The preamble to the Directive specifically notes the role of technology: "liability without fault on the part of the producer is the sole means of adequately solving the problem, peculiar to our age of increasing technicality, of a fair apportionment of the risks inherent in modern technological production[.]"

losses caused by products found to be defective.[41] In cases where it is unclear whether the accident was caused by the human operator or defects in the equipment, evidentiary rules have been established for allocating blame.[42]

Strict liability is not absolute liability. The definition of design defect turns on the reasonableness of the choices made by the manufacturer, and the process of showing what is reasonable or not often involves competing experts.[43]

1. Case-study: autonomous vehicles

Consider the likely litigation that will arise around self-driving cars. A well-established body of law already defines the legal liability of the operators and manufacturers of traditional automobiles. Lawsuits against the drivers of automobiles typically rely on a negligence theory. Suits against manufacturers more often proceed under the theory of strict products liability.

Unless legislatures act to adopt special rules for autonomous vehicles, courts will apply these doctrines. Negligence concepts are likely to still apply to the operators of autonomous vehicles (posing the question, perhaps, of who should be classified as the operator) and strict products liability doctrine will apply to manufacturers. Courts may encounter evidentiary problems in cases where it is difficult to tell whether an AI robot or human operator caused an accident, but this may be not unlike the issues posed by a traditional auto accident.

An in-depth 2016 study[44] of autonomous vehicles predicted that, overall, cases involving auto accidents will decrease as driving becomes safer with the diffusion of AI in automobiles. Where accidents do occur, the type of claims will evolve over time. As driverless vehicles become more common and their users grow more competent, claims against users will be replaced by claims that allege defects in driverless vehicles, shifting liability "upwards" from drivers to manufacturers. These cases will rely on products liability law, with design defect and warning defect claims expected to be more common than manufacturing defect claims.[45]

---

[41] Iwata Godo, *Product Liability in Japan*, LEXOLOGY (May 10, 2018,) https://www.lexology.com/library/detail.aspx?g=e013ff26-a955-4faa-bba0-5b7be84baf4f.

[42] For example, courts have allowed plaintiffs to use circumstantial evidence to establish "manufacturing defects . . . where the facts reveal that the (presumed) defect destroys the evidence necessary to prove that defect or where the evidence is otherwise unavailable through no fault of the plaintiff." *See In re Toyota Motor Corp. Unintended Acceleration Mktg., Sales Practices, & Prod. Liab. Litig.*, 978 F. Supp. 2d 1053, 1097 (C.D. Cal. 2013).

[43] *See generally* David G. Own, *Design Defects*, 73 MISSOURI L. REV. 291 (2008).

[44] NAS Driverless Cars Study, *supra* note 25.

Salient will be the question of what is a defect in design? The advent of driverless cars will likely pose questions about how such a car should be designed. For example, should driverless cars be designed to always obey the speed limits? How should they deal with the "trolley problem?"[46] Is it a defect not to equip them with sensors that block their operation by a driver who is intoxicated? Many similar questions may be posed. "Complications may arise when product liability claims are directed to failures in software, as computer code has not generally been considered a 'product' but instead is thought of as a 'service,' with cases seeking compensation caused by alleged defective software more often proceeding as breach of warranty cases rather than product liability cases."[47]

Policymakers may choose to divert from traditional tort law in developing liability doctrines for AI. One approach would be to create an AI certification process, limiting tort liability for those who obtain certification, but imposing strict liability on uncertified systems.[48] Another approach would be to adopt a regulatory system based on testing similar to that for drugs and medical devices.[49] A third approach, suggested in Europe, would be an obligatory insurance scheme.[50] In the United States, given the federal system of government, the present gridlock in Congress, and the significant ability of industry to block or neuter new regulatory legislation, any comprehensive solution seems unlikely.

---

[45] NAS Driverless Cars Study, *supra* note 25.

[46] *See, e.g.*, Jay Donde, *Self-Driving Cars Will Kill People. Who Decides Who Dies?,* WIRED (Sept. 21, 2017), https://www.wired.com/story/self-driving-cars-will-kill-people-who-decides-who-dies/ ("To understand the trolley problem, first consider this scenario: You are standing on a bridge. Underneath you, a railroad track divides into a main route and an alternative. On the main route, 50 people are tied to the rails. A trolley rushes under the bridge on the main route, hurtling towards the captives. Fortunately, there's a lever on the bridge that, when pulled, will divert the trolley onto the alternative route. Unfortunately, the alternative route is not clear of captives, either — but only one person is tied to it, rather than 50. Do you pull the lever?").

[47] Quinn Emanuel Trial Lawyers, *supra* note 37.

[48] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J. LAW & TECH. 353 (2016).

[49] See David Danks and Alex John London, *Regulating Autonomous Systems: Beyond Standards,* IEEE Intelligent Systems (2017).

[50] See *Report to the European Parliament, with Recommendations to the Commission on Civil Law Rules on Robotics*, from the Committee on Legal Affairs, Mady Delvaux, Rapporteur (Jan. 1, 2017) http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN. For more on the current EU landscape around liability and ethics of AI, see Nathalie Nevejans, *European Civil Law Rules in Robotics* (Oct. 2016), published by the Directorate-General for Internal Policies, http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf.

The history of safety regulation with respect to traditional automobiles suggests that an incremental approach will be followed, combining a hybrid of tort doctrines and specific regulatory mandates (such as the decision, after many years of debate, to mandate air bags).[51]

In the United States, as of September 2019, 36 states had adopted laws on autonomous vehicles,[52] but these laws typically do not alter products liability law to address autonomous vehicles.[53] Some merely establish task forces to study the issues. Others have sought to promote the development and use of autonomous vehicles by authorizing their operation and by permitting "platooning" of vehicles by creating exemptions to the normal following-too-closely rules. Some, such as a Florida law signed in June 2019, allow the operation of fully autonomous vehicles regardless of whether a human operator is physically present in the vehicle. The Florida law is unusual in specifying that the automated driving system, when engaged, shall be deemed to be the operator of an autonomous vehicle, regardless of whether a the person is physically present in the vehicle.[54]

Many of the state laws have called for studies on the safety of autonomous vehicles. A few have directed regulatory bodies to adopt safety standards. For example, a California statute requires the California Department of Motor Vehicles to adopt safety standards and performance requirements for autonomous vehicles.[55] In February 2018, the California Department of Motor Vehicles promulgated new rules which allow manufacturers to obtain a "driverless testing and/or deployment permit" if they meet certain relatively high-level certification requirements.[56] At the federal level, the

---

[51] See Nidhi Kalra, *Challenges and Approaches to Realizing Autonomous Vehicle Safety*, Testimony submitted to the House Energy and Commerce Committee, Subcommittee on Digital Commerce and Consumer Protection on February 14, 2017, https://democrats-energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/Testimony-Kalra-DCCP-Hrg-Self-Driving-Cars-2017-02-14.pdf.

[52] Some of the 36 have adopted multiple laws. Links to these laws can be found at the website of the National Conference of State Legislatures, which maintains databases of enacted state laws, executive orders and proposed laws on self-driving cars. See NAT'L CONFERENCE OF STATE LEGISLATURES, SELF-DRIVING VEHICLES ENACTED LEGISLATION (Sept. 9, 2019), http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx.

[53] One exception is Michigan, which amended its products liability law specifically to protect manufacturers from liability when an operator or repairman modifies an autonomous vehicle and to specify that a "motor vehicle mechanic or a motor vehicle repair facility that repairs an automated motor vehicle according to specifications from the manufacturer of the automated motor vehicle is not liable in a product liability action for damages resulting from the repairs." Mich. Comp. Laws Serv. § 600.2949b.

[54] HB 311, signed June 13, 2019.

[55] Cal. Veh. Code § 38750(d) (instructing the California Department of Motor Vehicles to adopt regulations governing "testing, equipment, and performance standards.").

Department of Transportation, through the National Highway Traffic Safety Administration (NHTSA), has taken a largely non-regulatory approach, issuing voluntary guidelines intended to promote innovation and support state-level policy development.[57] NHTSA has taken regulatory action on a case-by-case basis, for example, issuing a cease and desist letter to a company selling a product that disabled a safety feature on Tesla vehicles that monitors the driver's hands on the steering wheel.[58]

In March 2020, NHTSA proposed changing its rules to modernize its crashworthiness standards to remove what it considered to be unnecessary barriers to vehicles equipped with Automated Driving Systems (ADS) and the unconventional interior designs that are expected to accompany these vehicles, including the lack of driving controls.[59] The rulemaking was limited to the crashworthiness standards, proposing a unified set of regulatory text applicable to vehicles with and without ADS functionality. Comments were due May 29, 2020.

B. Health and Safety Regulation

Autonomous vehicles are just one example of the wide array of consumer and commercial products and processes where traditional health and safety standards will have to be applied to—and adapted for—AI-based developments. In the US, the use of AI in devices that deliver medical diagnostics and treatments is subject to regulation by the Food and Drug Administration (FDA), posing questions of how to categorize certain

---

[56] *See* Cal. Code Regs. tit. 13, §§ 227.00—227.52; *see also* California Department of Motor Vehicles, *Driverless Testing and Public Use Rules for Autonomous Vehicles Approved* (Feb. 26, 2018), https://www.dmv.ca.gov/portal/dmv/detail/pubs/newsrel/2018/2018_17. The regulations, for example, require that manufacturers "certify the autonomous test vehicle complies with requirements that include a communication link between the vehicle and remote operator, a process to communicate between the vehicle and law enforcement, and an explanation of how the manufacturer will monitor test vehicles." They also require certifications relating to training, cyberattack detection/defense/response, and that the autonomous technology is "designed to detect and respond to roadway situations in compliance with California Vehicle Code."

[57] See U.S. Dept. of Transportation, Automated Driving Systems 2.0: A Vision for Safety (2017) https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf. A follow-on report, referred to as AV 4.0, complied Administration efforts to supporting AV technology growth and described 10 principles guiding federal efforts to protect users and communities, promote efficient markets, and facilitate coordinated efforts to ensure a standardized Federal approach to. Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0 (2020) https://www.transportation.gov/sites/dot.gov/files/docs/policy-initiatives/automated-vehicles/360956/ensuringamericanleadershipav4.pdf.

[58] NHTSA, Consumer Advisory: NHTSA Deems 'Autopilot Buddy" Product Unsafe (June 19, 2018) https://www.nhtsa.gov/press-releases/consumer-advisory-nhtsa-deems-autopilot-buddy-product-unsafe.

[59] NHTSA, Notice of Proposed Rulemaking, Occupant Protection for Automated Driving Systems (March 30, 2020) https://www.govinfo.gov/content/pkg/FR-2020-03-30/pdf/2020-05886.pdf.

products and how to assess their safety and efficacy. The use of drones falls under the authority of the Federal Aviation Administration (FAA).[60] For consumer-facing AI systems, regulation by the Federal Trade Commission (FTC) may come into play. Financial markets using AI technologies, such as in high-frequency trading, come under regulation by the Security Exchange Commission (SEC). [61]

This sector-by-sector approach, in which existing regulatory agencies covering specific industries address the implications of AI as it is deployed in situations within their specific areas of competency and jurisdiction, may be the most effective way to respond to AI's wide reach. The AI 100 Study panel concluded that attempts to regulate AI in general would be misguided, since there is no clear definition of AI and the risks and considerations associated with AI are very different in different domains. "Instead, policymakers should recognize that to varying degrees and over time, various industries will need distinct, appropriate, regulations that touch on software built using AI or incorporating AI in some way."[62]

Even within a specific domain, part of the challenge with regulating AI results from the difficulty in defining it. If the definition of AI is overinclusive, it may unintentionally sweep in existing technologies that do not raise the same concerns as advanced AI.[63] For example, Nevada was compelled to rewrite its autonomous vehicles statue when it realized it had defined "autonomous vehicle" as any substitution of AI for a human operator, thus sweeping many existing vehicles into the ambit of the regulation because standard cars already use AI—for example, automatic braking systems when the car detects a nearby object.[64] On the other hand, underinclusive definitions may allow deployment of risky technologies without adequate consideration.

1. Case Study: Machine Learning in Software as a Medical Device

The Food and Drug Administration (FDA) has for some years recognized that software can be a medical device (SaMD) subject to regulation for safety and effectiveness. Under existing FDA rules, manufacturers of SaMD, like manufacturers of other devices, submit a

---

[60] The FAA rules impose registration and operating requirements on drone use. For example, the rules prohibit flying drones higher than an altitude of 400 feet or faster than a speed of 100 mph. *See* Fact Sheet — Small Unmanned Aircraft Regulations (Part 107), FEDERAL AVIATION ADMINISTRATION (Mar. 23, 2018), https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=22615.

[61] *See* AI 100 STUDY, *supra* note 1, at 44.

[62] AI 100 STUDY, *supra* note 1, at 48.

[63] *See* Calo, *supra* note 9.

[64] *See id.*

marketing application to FDA prior to initial distribution of their medical device, with the submission type and data requirements based on the risk of the SaMD. And, as with hardware devices, when the manufacturer upgrades or otherwise modifies its software, it may be required to make a new submission to the FDA before marketing the modified version.

But machine learning software may be constantly modifying itself, using new data to refine its algorithm. This raises the critical question of when a continuously learning AI/ML SaMD may require a premarket submission for an algorithm change. Requiring premarket submission for every change would essentially deny the benefits of ML.

In April 2019, the FDA issued a discussion paper and request for feedback, based on the recognition that "[t]he traditional paradigm of medical device regulation was not designed for adaptive AI/ML technologies, which have the potential to adapt and optimize device performance in real-time to continuously improve healthcare for patients."[65] A new regulatory framework was needed for AI/ML-based SaMD, the agency concluded. Such an approach, the FDA said, "would need to maintain reasonable assurance of safety and effectiveness of AI/ML-based SaMD, while allowing the software to continue to learn and evolve over time to improve patient care." The new approach that the FDA proposed is based on a concept known as "total product lifecycle," an approach that the FDA said would "facilitate[] a rapid cycle of product improvement and allow[] these devices to continually improve while providing effective safeguards." Under the approach, manufacturers would be expected to embrace the general principles of culture of quality and organizational excellence and the more specific "good ML practices." The FDA said that it would expect manufacturers to monitor their AI/ML devices and incorporate a risk management approach in the development, validation, and execution of algorithm changes. The proposed framework also called for increased transparency. The FDA noted that its proposed new framework "may require additional statutory authority to implement fully."

C. Fraud

Should consumers be told when they are dealing with a robot instead of a human? The deceptive use of chatbots has come up in litigation, under traditional fraud doctrine. *In re Ashley Madison Customer Data Sec. Breach Litig.*, the district court declined to dismiss a claim that defendant acted fraudulently when it used AI "bots" to impersonate women

[65] FDA, Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device: Discussion Paper and Request for Feedback https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/SoftwareasaMedicalDevice/UCM635052.pdf. See *Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices* (April 2, 2019) https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm635083.htm.

and communicate with human users to induce them into making purchases on the company's website.[66]

The launch of a Google digital assistant product that could call restaurants and "speak" to human employees, apparently fooling them into thinking they were talking to real people, sparked calls for a norm ensuring that human should be informed that they were talking to a robot. Google voluntarily promised that it would provide such notice,[67] but that did not end the controversy. In 2018, California adopted a bill that, with certain exceptions, makes it unlawful for any person to use a bot to communicate or interact with another person in California online with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election.[68] The provisions became operative on July 1, 2019.

D. Intellectual Property

As investment capital pours into AI technology and companies apply for and seek to enforce AI-related patents, agencies and courts are beginning to consider how to apply principles of intellectual property (IP) law to AI.

1. Patents

Patent applications for AI-related inventions are on the rise worldwide.[69] And a lot of AI patents are being granted. In the US, the Patent Office has issued thousands of patents on AI technologies.[70] The issuance by the USPTO of new patent eligibility guidelines for 2019

---

[66] *See In re Ashley Madison Customer Data Sec. Breach Litig.*, 148 F. Supp. 3d 1378, 1379 (JPML 2015).

[67] Richard Nieva, Google Says It's Designing Duplex with 'Disclosure Built-in,' CNET (May 11, 2018), https://www.cnet.com/news/google-says-its-designing-duplex-with-disclosure-built-in/ ("Duplex stirred up plenty of debate about whether or how a such a realistic-sounding virtual assistant should identify itself to humans. Google had previously said it wanted to make it so people would know when they're talking to a bot. On Thursday, Google said explicitly that it will design disclosures into the feature.").

[68] SB 1001, amending the Business and Professions Code to add new sections 17940-43. Madeline Lamo and Ryan Calo have suggested that a legal requirement that all bots identify themselves might run into First Amendment problems. Madeline Lamo and Ryan Calo, Regulating Bot Speech, 66 UCLA L. REV. 988 (2019) https://www.uclalawreview.org/regulating-bot-speech/.

[69] WIPO, TECHNOLOGY TRENDS 2019: ARTIFICIAL INTELLIGENCE (2019) https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf [hereinafter WIPO 2019].

[70] Deputy Under Secretary of Commerce for Intellectual Property and Deputy Director of the USPTO Laura Peter, *USPTO announces Federal Register Notice on artificial intelligence patent issues* (Aug. 6, 2019) https://www.uspto.gov/blog/director/entry/uspto_announces_federal_register_notice.

may have led to an increase in the rate of issuance of AI-related patents.[71]

Under US law, how many of these AI patents are valid is an open question.[72] The courts have held that abstract ideas are not eligible for patent protection, unless there is something additional that transforms the nature of the claim into a patent-eligible application, some "inventive concept—i.e., an element or combination of elements that is sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the [ineligible concept] itself."[73] An inventive concept can be found, for example, "in the non-conventional and non-generic arrangement of known, conventional pieces."[74] However, the use of a computer to execute an abstract is not sufficient to transform it into a patent-eligible invention.[75] In *Mayo Collaborative Servs. v. Prometheus*

---

[71] See Mark Nowotarski, *AI Patents Make a Comeback at USPTO, Finance Patents Are Still Struggling*, IP WATCHDOG (Mar. 26, 2019) https://www.ipwatchdog.com/2019/03/26/ai-patents-make-comeback-finance-patents-still-struggling/id=107680/. Previously, it had been observed that that artificial intelligence technologies were likely to be heavily scrutinized and less likely to be allowed. Kate Gaudry and Samuel Hayim, *Artificial Intelligence Technologies Facing Heavy Scrutiny at the USPTO*, IP WATCHDOG (Nov. 28, 2018) https://www.ipwatchdog.com/2018/11/28/artificial-intelligence-technologies-facing-heavy-scrutiny-uspto/id=103762/. The 2019 patent subject matter eligibility guidance is available at https://www.ipwatchdog.com/wp-content/uploads/2019/01/revised-101-guidance-2018-28282.pdf.

[72] See generally Susan Y. Tull and Paula E. Miller, *Patenting Artificial Intelligence: Issues of Obviousness, Inventorship and Patent Eligibility,* Journal of Robotics, Artificial Intelligence and Law (2018) https://www.finnegan.com/images/content/1/9/v2/197825/PUBLISHED-The-Journal-of-Robotics-Artificial-Intelligence-L.pdf; World Economic Forum, ARTIFICIAL INTELLIGENCE COLLIDES WITH PATENT LAW (April 2018) http://www3.weforum.org/docs/WEF_48540_WP_End_of_Innovation_Protecting_Patent_Law.pdf.

[73] *Alice Corp. v. CLS Bank Int'l*, 134 S. Ct. 2347 (2014). See also *Electric Power Group, LLC v. Alstom S.A.*, 830 F.3d 1350 (Fed. Cir. 2016). Europe has a similar rule: Article 52 of the European Patent Convention specifies that "schemes, rules and methods for performing mental acts, playing games or doing business, and programs for computers" are not to be regarded as inventions eligible for patent protection. Convention on the Grant of European Patents, art. 52, Oct. 5, 1973, 1065 U.N.T.S. 199, http://documents.epo.org/projects/babylon/eponet.nsf/0/029F2DA107DD667FC125825F005311DA/$File/EPC_16th_edition_2016_en.pdf. On November 1, 2018, the European Patent Office issued updated Guidelines for Examination with subsection specifically on artificial intelligence and machine learning. https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_3_1.htm. *See* generally https://www.epo.org/ai and Mizuki Hasiguchi, *The Global Artificial Intelligence Revolution Challenges Patent Eligibility Laws*, 13 J. BUS. & TECH. L. 1 (2017), http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=1280&context=jbtl.

[74] *BASCOM Glob. Internet Servs., Inc. v. AT&T Mobility LLC*, 827 F.3d 1341, 1350 (Fed. Cir. 2016). Different opinions of the Federal Circuit express slightly different formulations of the *Alice* rule, pointing in different directions. See, for example, *McRO, Inc. v. Bandai Namco Games Am. Inc*., 837 F.3d 1299, 1314 (Fed. Cir. 2016) ("a specific means or method that improves [a] technology" may be patentable, while "a result or effect that itself is the abstract idea and merely invoke[s] generic processes and machinery" would not be); *Enfish v. Microsoft,* 822 F.3d 1327, 1335 (Fed. Cir. 2016) ("[s]oftware can make non-abstract improvements to computer technology just as hardware improvements can").

[75] *Content Extraction & Transmission LLC v. Wells Fargo Bank, N.A.,* 776 F.3d 1343, 1347-48 (Fed. Cir. 2014).

Labs., Inc., the Supreme Court re-emphasized that laws of nature or physical phenomena are also not patentable, even if their discovery yields revolutionary improvements in diagnosis or care.[76]

Under the *Alice* test, it seems that some, possibly many, AI systems or concepts will be found to be too abstract to be patentable. "Of particular relevance … are the courts' invalidations of patent claims for covering subject matter that could be performed through an 'ordinary mental process,' 'in the human mind' or by 'a human using a pen and paper' under the *Alice Corporation v. CLS Bank International* test. This creates a tension with patenting AI inventions because the goal of AI is often to automate or better perform human tasks and activities."[77] Likewise, the *Mayo* principle may doom many uses of AI to discover correlations in the world of bio-chemistry and other naturally-occurring phenomena.

In 2015, in a case involving the use of an expert system to test equipment operators for intoxication, the Federal Circuit held that the system was not patent-eligible because it constituted an abstract idea in that it was directed at something performed by humans absent automation.[78] The court also held that the claims failed the test of being "sufficiently inventive" because they did not specify how the system would work or if it would provide advantages over existing technology. The court left room for AI claims that involve a "specific implementation," rather than an abstract idea.[79] In another case, a district court case invalidated a patent concerning the "automated resolution of IT incidents" as being directed to an abstract idea. In dicta, the judge stated that the idea of a self-driving car could not be patented in the abstract.[80] In yet another case on an AI-based invention, the court stated, "To the extent artificial intelligence inventions … involve an inventive concept, they could be patentable even if they have, at their core, an abstract concept," but that court went on to hold the particular claim before it invalid as embodying an unpatentable "abstract idea."[81]

---

[76] *Mayo Collaborative Servs. v. Prometheus Labs., Inc.,* 566 U.S. 66 (2012); *Ariosa Diagnostics, Inc. v. Sequenom*, 788 F.3d 1371 (Fed. Cir. 2015).

[77] John G. Flaim and Yoon Chase, *"Abstract idea" in U.S. jurisprudence*, in WIPO report, *supra* note 67. See, for example, *Neochloris, Inc. v. Emerson Process Mgmt. LLLP*, 140 F. Supp. 3d 763, 774 (N.D. Ill. 2015) ("[t]here is no inventive concept when a computer [using a neural network] just replicates what a person can do, only more quickly and accurately").

[78] *See Vehicle Intelligence & Safety LLC v. Mercedes-Benz USA, LLC*, 635 F. App'x 917 (Fed. Cir. 2015), cert. denied, 136 S. Ct. 2390 (2016).

[79] *See* Quinn Emanuel Trial Lawyers, *supra* note 37.

[80] *Hewlett Packard Co. v. ServiceNow, Inc.,* 2015 U.S. Dist. LEXIS 29384 at *30 (N.D. Cal. Mar. 10, 2015).

[81] *Blue Spike, LLC v. Google Inc*., No. 14-CV01650-YGR, 2015 U.S. Dist. LEXIS 119382, at *13-16 (N.D. Cal. 8 September 2015), *aff'd per curiam*, *Blue Spike v. Google Inc.*, No. 2016-1054, 2016 U.S. App. LEXIS 20371 (Fed. Cir. 2016).

In one case specifically involving machine learning, the patent was invalidated because its claims were directed to the abstract concept of testing and refining mathematical algorithms.[82] The first step in the process generated learned functions or regressions from data. That, the court held, was not a patentable concept. Likewise, the step in which the invention took the learned functions, evaluated their effectiveness, and selected those most effective to create a rule set also involved mathematical processes that not only could be performed by humans but also went to the general abstract concept of predictive analytics rather than any specific application. "While they may invoke computers as a tool for this process, the claims do not make a specific improvement on an existing computer-related technology." In sum, the court said, the "abstract concept of testing and refining mathematical algorithms" was patent-ineligible (a conclusion that would condemn many ML systems to patent ineligibility.) Searching for an inventive concept sufficient to transform these abstract ideas into a patent-eligible application, the court found none. The invention did not solve a sufficiently specific problem. Instead, the court found, the patent claims addressed only the universal problem in any analytical framework of choosing between a more generally applicable or more specific and customized model.[83]

Despite these barriers, some AI-based or algorithm-based patents have been upheld by the US courts and by the Patent Trial and Appeal Board (PTAB). In 2018, the Federal Circuit pointed the way when it held, in *Vanda Pharmaceuticals Inc. v. West-Ward Pharmaceuticals,* that a method of treatment that utilizes an underlying natural law was still patentable because the natural law was sufficiently incorporated into a practical application, that is, the method of treatment.[84] The PTAB "has held that claims of improving the creation of three-dimensional structures of biomolecules are patent eligible even where that method is based on computer-based, algorithmic modeling."[85] Likewise, claims directed to improving medical analysis using predictive analytics have been found patent eligible.[86] A key, it seems, is to draft claims that focus not on the

---

[82] *PurePredictive, Inc. v. H2O.AI, Inc.,* (N.D. Cal. Aug. 29, 2017). See also *Hyper Search LLC v. Facebook Inc.,* No. CV 17-1387, 2018 WL 6617143, at *10 (D. Del. Dec. 17, 2018).

[83] See also *Blue Spike, LLC v. Google Inc*., No. 14-CV01650-YGR, 2015 U.S. Dist. LEXIS 119382, at *13-16 (N.D. Cal. 8 September 2015), *aff'd per curiam*, *Blue Spike v. Google Inc.*, No. 2016-1054, 2016 U.S. App. LEXIS 20371 (Fed. Cir. 2016).

[84] *Vanda Pharms Inc. v. West-Ward Pharms Int'l Ltd.,* 887 F.3d 1117 (Fed. Cir. 2018). The patentability of a claim that focused on using an equation in a practical way was already articulated in *Diamond v. Diehr*, 450 U.S. 175, 179 & n.5 (1981).

[85] Joshua Reisberg and Ross Blau**,** *Navigating Uncertainty In Digital Health Patent Eligibility, Law360 (March 2, 2020),* citing *Ex part Zhou*, No.2018-004159 (PTAB, April 12, 2019).

[86] *Id.,* citing *Ex parte Attanapola*, No. 2018-005031 (PTAB July 1, 2019).

algorithm but on the application of the algorithm to produce a result, to solve a problem.

In August 2019, the US Patent and Trademark Office issued a request for comments on patenting AI inventions.[87] The RFC set out a series of questions, including whether there are any eligibility considerations unique to AI inventions and what is or should be the law where AI contributes to the conception of an invention. In November, the PTO extended the deadline for public comments to January 10, 2020.

Other countries have more generous standards for issuing AI-related patents. China, for example, has outpaced the U.S. in granting AI-related patents.[88]

In December, 2019, the World Intellectual Property Organization (WIPO) launched a consultation process on AI and IP, inviting feedback on an issues paper designed to help define the most pressing questions likely to face IP policymakers as AI increases.[89]

2. Copyright

Under copyright law, it has long been clear that computer programs are copyrightable, but a program's algorithms are not.[90]

A separate question is whether to extend intellectual property protections to works that AI creates. (Machines can write news stories, fiction and music; they can paint pictures. Algorithms can produce other algorithms and AI systems can discover new things.) In the US, the Copyright Office has stated that it "will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author."[91] How much creative input or

---

[87] USPTO, Request for Comments on Patenting Artificial Intelligence Inventions, 84 Fed. Reg. 44889 (Aug. 27, 2019) https://www.federalregister.gov/documents/2019/08/27/2019-18443/request-for-comments-on-patenting-artificial-intelligence-inventions. The PTO has created a webpage assembling its AI-related notices and other resources. https://www.uspto.gov/initiatives/artificial-intelligence

[88] *See* Echo Huang, *China Has Shot Far Ahead of the U.S. On Deep-Learning Patents*, QUARTZ (Mar. 2, 2018), https://qz.com/1217798/china-has-shot-far-ahead-of-the-us-on-ai-patents/; Ralph Jennings, *China Leads the U.S. in Patent Applications for Blockchain and Artificial Intelligence*, FORBES (May 17, 2018).

[89] https://www.wipo.int/pressroom/en/articles/2019/article_0017.html.

[90] Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1187 (1986). *See* UNITED STATES COPYRIGHT OFFICE, COPYRIGHT REGISTRATION OF COMPUTER PROGRAMS, CIRCULAR 61 (Sept. 2017), https://www.copyright.gov/circs/circ61.pdf ("A computer program is a set of statements or instructions to be used directly or indirectly in a computer to bring about a certain result. Copyright protection for a computer program extends to all of the copyrightable expression embodied in the program. The copyright law does not protect the functional aspects of a computer program, such as the program's algorithms, formatting, functions, logic, or system design.").

intervention by the human is enough to render the output of a machine copyrightable is a question of case-by-case line-drawing. A number of other countries take a broader approach in extending copyright protection to the output of machines.[92]

A further separate question is who owns the IP in a work created by a machine.[93] The copyright statutes of several countries expressly make it clear that the author of a computer-generated work is "the person by whom the arrangements necessary for the creation of the work are undertaken."[94]

E. Professional Ethics and Liability in Law and Healthcare

As AI becomes more useful for highly skilled professions, such as law and medicine, courts and policymakers must determine how to impose and apportion liability for malpractice.[95] The failure to take advantage of AI could be malpractice, but so could the unquestioning reliance on AI. Across professions, the reliance on technology can induce "skill fade," as professionals cease to exercise their skills and end up less proficient when the technology fails.[96]

In the practice of law, and considering just the professional responsibility framework in the United States, AI may implicate many provisions of the model code of professional conduct maintained by the American Bar Association. Under Model Rule 1.1, for example, lawyers are under a duty of competence, which includes understanding the benefits and risks associated with using relevant technology in the course of their practice.[97] The duty

---

[91] U.S. Copyright Office, The Compendium of U.S. Copyright Office Practices, Chapter 300, § 313.2 (revised Sept. 29, 2017).

[92] Robert C. Denicola, *Ex Machina: Copyright Protections for Computer-Generated Works*, 69 RUTGERS UNIV. L. REV. 251, 281-82 (2016) http://www.rutgerslawreview.com/wp-content/uploads/2017/07/Robert-Denicola-Ex-Machina-69-Rutgers-UL-Rev-251-2016.pdf.

[93] Andres Guadamuz, *Artificial intelligence and copyright*, WIPO MAGAZINE (Oct. 2017) http://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.

[94] *See* Denicola, *supra* note 91, at 281-82 (citing laws of Ireland, UK, New Zealand, South Africa, and India). Prof. Samuelson has argued under U.S. law that, if computer generated works were copyrightable, in general the user of the program that generated the work should be considered its author. Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1226 (1986).

[95] Examples of use of AI in legal practice include RavelLaw, a program that uses natural language processing to automate case law analysis and legal research, as well as Ross Intelligence, which uses Watson to process natural language to help with case law review. Healthcare providers use AI in robotic surgical instruments and cancer treatment devices, and also use products like Watson to recommend medical treatment. *See* Quinn Emanuel Trial Lawyers, *supra* note 37.

[96] Brian Sheppard, Skill Fade: The Ethics of Lawyer Dependence on Algorithms and Technology, Legal Solutions, Vol. 19, No. 2 (March 2018).

of communication, Model Rule 1.4, may obligate law firms to let their clients know if they use AI tools. The duty of confidentiality obligates lawyers to ensure third-party AI providers are using best practices to keep client data safe. AI systems may implicate Model Rule 5.3, which requires a lawyer to take responsibility for non-lawyer assistance. "Despite the widespread adoption of AI tools to conduct contract reviews and legal research, among a host of other tasks, there has been no corresponding uptick in guidance from regulatory bodies on how lawyers can ethically use these increasingly sophisticated tools."[98] Meanwhile, consumer-facing tools that offer legal advice or assistance to individuals may run afoul of prohibitions on the unauthorized practice of law.[99]

In August 2019, the American Bar Association adopted a resolution on AI and the legal profession: RESOLVED, That the American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ("AI") in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI."[100]

The medical profession has also begun to address the professional implications of AI. At its annual meeting in June 2018, the American Medical Association adopted broad policy recommendations on AI, which the AMA refers to as "augmented intelligence."[101] Most substantively, the policy states that the AMA will promote the development of thoughtfully designed, high-quality, clinically validated health care AI that is designed and evaluated in keeping with best practices in user-centered design; is transparent; conforms to leading standards for reproducibility; identifies and takes steps to address bias and avoids introducing or exacerbating health care disparities; and safeguards patients' and

---

[97] Comment 8 to Model Rule 1.1 states: "To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology … ." As of February 2019, 35 states had formally adopted this comment into the commentary on their own rules. https://www.lawsitesblog.com/tech-competence.

[98] Sam Reisman, *Ethics Rules Have Not Evolved with AI, GCs Say*, LAW360 (Apr. 23, 2018).

[99] Thomas Spahn, *Is Your Artificial Intelligence Guilty of the Unauthorized Practice of Law?*, 24 Rich J.L. & Tech., no. 4 (2018).

[100] Nicolas Economou, Artificial Intelligence and the Law: the ABA's important and timely contribution (Aug. 26, 2019) https://www.americanbar.org/groups/business_law/publications/committee_newsletters/legal_analytics/2019/201908/ai_law/

[101] AMA, *AMA passes first policy recommendation on augmented intelligence* (June 14, 2018) https://www.ama-assn.org/press-center/press-releases/ama-passes-first-policy-recommendations-augmented-intelligence. The policy is discussed at length in AMA, *Augmented intelligence in health care* (2018) https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf.

other individuals' privacy interests and preserves the security and integrity of personal information.[102]

When it comes to apportioning liability between a professional and the maker of the AI tools she uses, courts may apply a number of liability theories. Under a vicarious liability theory, courts may decide that an autonomous machine can be analogized to an employee and impose vicarious liability on a hospital that fails to adequately supervise the quality of medical care provisioned in its facility.[103] On the other hand, if a court analogizes an autonomous machine to a typical medical device (rather than employee), then products liability claims may apply to defective equipment.[104] In products liability cases, the "learned intermediary doctrine" may preclude plaintiffs from suing medical AI manufacturers directly because the manufacturer owes no duty to the patient.[105]

F. Contracts

Legal experts in the United States, Canada, and elsewhere have long since outlined through model laws the conditions under which software can enter into a binding contract.[106] In 2000, Congress enacted legislation recognizing that computers can make binding contracts.[107]

G. Substantive Criminal Law

AI may pose questions under the criminal law. "As AI applications engage in behavior that were it done by a human, would constitute a crime, courts and other legal actors will have to puzzle though whom to hold accountable on what theory."[108]

---

[102] In February 2019,the AMA Journal of Ethics devoted its entire issue to AI. https://journalofethics.ama-assn.org/issue/artificial-intelligence-health-care.

[103] *See* Quinn Emanuel Trial Lawyers, *supra* note 37.

[104] *See id.*

[105] *See id.*; *cf. Banker v. Hoehn*, 278 A.D.2d 720, 721, 718 N.Y.S.2d 438, 440 (2000).

[106] Ian R. Kerr, *Ensuring the Success of Contract Formation in Agent-Mediated Electronic Commerce*, 1 ELEC. COMMERCE RESEARCH 183 (2001).

[107] 15. U.S.C. 7001(h) ("A contract or other record relating to a transaction in or affecting interstate or foreign commerce may not be denied legal effect, validity, or enforceability solely because its formation, creation, or delivery involved the action of one or more electronic agents so long as the action of any such electronic agent is legally attributable to the person to be bound. *See* UNIF. ELEC. TRANSACTIONS ACT § 14(1) (NAT'L CONFERENCE OF COMM'RS ON UNIF. STATE LAWS 1999) ("A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.").

[108] AI 100 STUDY, *supra* note 1, at 46.

Consider, for example, vehicular homicide. Traditionally, the driver of a motor vehicle would face criminal liability if death resulted from the reckless, dangerous or negligent operation of the vehicle. If the accident occurred while the vehicle was driving autonomously, could the driver still face criminal liability? Could the manufacturer ever be criminally liable?."[109] If the autonomous vehicle was programmed to minimize harm to persons, would this constitute sufficient care to protect the manufacturer from being found criminally liable?[110] As with other criminal cases involving corporations, it might be very hard to determine *who* at the manufacturing company should face punishment.[111]

H. Criminal Procedure and Due Process - Admissibility of AI-Based Evidence

When confronted with evidence and expert opinion based on AI systems, courts will apply long-standing rules of evidence, especially those regarding scientific evidence and expert opinion. For example, in April 2019, a district court, applying Rule 702 of the Federal Rules of Evidence and the Supreme Court's guidance in *Daubert v. Merrell Dow Pharm., Inc.,* 509 U.S. 579 (1993), excluded an expert's opinion regarding the results of a DNA analysis using a probabilistic genotyping program because, the court found, a foundational input to the program's calculation could not be reliably estimated.[112]

Professor Andrea Roth at UC Berkeley has proposed a comprehensive set of safeguards for the use of machine testimony, consisting of credibility testing in the form of front-end design, input, and operation protocols; pretrial disclosure and access rules; authentication and reliability rules; impeachment and courtroom testing mechanisms; jury instructions; and corroboration rules. She argues that the Sixth Amendment right to confront witnesses can be applied to machine testimony through a "right of meaningful impeachment."[113]

For many years, judges and others in the criminal justice system have been trying to eliminate racial bias in sentencing. One tool that has been turned to is algorithmic scoring

---

[109] *See* Jeffrey K. Gurney, *Crashing Into The Unknown: An Examination Of Crash-Optimization Algorithms Through The Two Lanes Of Ethics And Law*, 79 ALB. L. REV. 183, 242 (2016).

[110] *See id.* at 244.

[111] *See id.*

[112] *United States v. Williams*, Case No. 3:13-cr-00764-WHO-1, N.D. Cal. April 29, 2019). See Katherine Kwong, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 Harv. J.L. & Tech. 275 (2017).

[113] Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2016).

systems, which take a range of information about an individual and produce a risk score. However, some of these systems have been accused of bias. A study of one, called COMPAS, found that it overestimated the risk posed by Black defendants and underestimated the risk posed by whites.[114]

In 2016, the Wisconsin Supreme Court confronted such criticisms.[115] Without trying to decide if the COMPAS system was in fact biased or not, the Court permitted its use with limitations. The court held any presentence report containing a COMPAS risk assessment must inform the sentencing court about concerns regarding the risk assessment's accuracy. Moreover, the Court specified that risk scores could not be used to determine the threshold question of whether to incarcerate a person or the severity of the sentence. Instead, they could be used as one factor in probation and supervision. The Court also held that the fact that the system's weighting of factors is proprietary did not amount to a due process violation if the sentencing report included limitations and cautions regarding the assessment's accuracy.[116]

In April 2019, the Partnership on AI, a non-profit organization comprising civil society groups, corporate developers and users of AI, and academic AI research labs, issued a report on algorithmic risk assessment tools in the US criminal justice system.[117] The report outlined ten requirements that jurisdictions should weigh and address before further use of risk assessment tools in the criminal justice system. It concluded that none of the existing tools met all ten requirements. The report focused on the use of risk assessment tools in the context of pretrial detentions, but noted "many of our observations also apply to their uses for other purposes such as probation and sentencing."

A February 2019 report[118] came to a somewhat more optimistic conclusion: "Taken together, the current body of research on pretrial risk assessment tools supports their ability to identify defendants at different rates of failure to appear and pretrial arrest, and

---

[114] *See* Pro Publica Study, *supra* note 27.

[115] *State v. Loomis*, 881 N.W.2d 749, 371 Wis. 2d 235 (2016).

[116] See also Malenchik v. State, 928 N.E.2d 564, 575 (Ind. 2010) (allowing algorithmic risk assessment score to be "considered as a supplemental source of information to assist a trial court in formulating the manner a sentence is to be served"); State v. Gordon, No. 17-0395, 2018 WL 2084847, at *9 (Iowa Ct. App. May 2, 2018) (vacating a defendant's prison term because the district court considered the defendant's risk level scores as an aggravating factor when imposing the sentence without statutory authority to do so).

[117] Partnership on AI, Report on Algorithmic Assessment Tools in the U.S. Criminal Justice System (2019) https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/.

[118] Sarah Desmarais and Evan Lowder, Pretrial Risk Assessment Tools: A Primer for Judges, Prosecutors and Defense Attorneys (Safety and Justice Challenge 2019) http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf.

leaves open the possibility that they could have a positive impact on pretrial decisions and outcomes." Even this was caveated: "However, the research methods and statistics used in these [validity] studies often fail to meet the standards of practice in the field of risk assessment and the standards for educational and psychological testing more generally. Further, there has been no independent evaluation or synthesis of this research, limiting more definitive conclusions regarding the predictive validity of pretrial risk assessment tools overall and with respect to specific tools and pretrial outcomes." The report emphasized that even valid (predictively accurate) tool would not have positive effects unless properly implemented.

I. Policing

Law enforcement agencies are increasingly using algorithmic predictive policing systems to forecast criminal activity and allocate police resources. However, a 2019 study found that, in numerous jurisdictions, these systems are built on data collected within the context of flawed, racially fraught, and sometimes unlawful practices, including systemic data manipulation, falsified police reports, unlawful use of force, planted evidence, and unconstitutional searches.[119] The best AI system available, if fed falsified or otherwise flawed data, will produce biased or otherwise flawed predictions, "which in turn risk perpetuating additional harm via feedback loops."[120] Transparency may enable citizens to discover what policy judgments these algorithms embody and to evaluate their utility and fairness, but it is not easy.[121] See section II.K.1 below for a discussion of facial recognition.

J. Anti-discrimination Laws

As algorithms and AI-based systems make decisions affecting individuals, a growing chorus of concern has been raised about whether such decisions are fair and reliable.[122] The problem is that many systems being deployed are so complex that it is hard to even explain why they reach the decisions they do. The Princeton computer scientist Arvind

---

[119] Rashida Richardson, Jason Schultz and Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice* (February 13, 2019), New York University Law Review Online (forthcoming) https://ssrn.com/abstract=3333423.

[120] Id. See also Danielle Ensign et al., Runaway Feedback Loops in Predictive Policing, 81 PROC. OF MACHINE LEARNING RES. (2018) https://arxiv.org/pdf/1706.09847.pdf.

[121] Robert Brauneis and Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J. L & TECH. 103 (2018).

[122] Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence,* 66 UCLA L. Rev. 54 (2019) (discussing the relationship between AI and civil rights and arguing that society must focus on the role of private corporations in addressing algorithmic accountability through codes of conduct, impact statements, and whistleblower protection).

Narayanan summarized the concerns in a tweet: "Today's AI/ML is uninterpretable, biased, and fragile. When it works, we don't understand why."

To some extent, current law addresses questions of causality in decision-making by requiring employers, for example, to show a business justification for a system that produces discriminatory impact.[123] Likewise, the Equal Credit Protection Act ("ECOA") prohibits credit discrimination on the basis of certain protected characteristics, such as race.[124] A lender that uses AI to make credit decisions could violate ECOA—even if the AI algorithm does not explicitly consider race—if its lending practices result in a disparate impact on a racial group.[125]

In some cases, however, there may be no need to rely on disparate impact: the process of categorizing individuals in a context subject to anti-discrimination laws may be expressly based on protected categories. In March, 2019, the Department of Housing and Urban Development (HUD) brought a complaint against Facebook for violation of the fair housing laws, based on Facebooks use of algorithms and machine learning to deliver advertisements for housing. The case involved Lookalike Audiences, a Facebook targeting tool that allows an advertiser to provide Facebook with a list of current employees or customers and then creates a new target audience of users who share common qualities with those in the source audience. The complaint alleged:

> To group users by shared attributes, to create a Lookalike Audience, to determine an ad's "actual audience" during the ad delivery phase, and to price each ad for each user, Respondent [Facebook] combines the data it has about user attributes and behavior on its platforms with data it obtains about user behavior on other websites and in the non-digital world. Respondent then uses machine learning and other prediction techniques to classify and group users so as to project each user's likely response to a given ad. In doing so, Respondent inevitably recreates groupings defined by their protected class.[126]

---

[123] See, however, Solon Barocas and Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016) (arguing that existing law, especially the Title VII anti-discrimination law as interpreted by the courts, largely fails to address the discrimination that can result from data mining).

[124] *See* 12 C.F.R. § 202.4.

[125] *See generally* 1-8 Federal Fair Lending and Credit Practices Manual § 8.01 (2018); *see also* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions,* 89 WASH. L. REV. 1, 10—16 (2014) (discussing how individuals may be scored by artificially intelligent algorithms in areas like credit scoring without "technological due process" and may result in inaccurate, arbitrary, or discriminatory scores).

[126] https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf, para. 20.

In March 2019, Facebook settled five cases regarding the discriminatory impact of the Lookalike Audiences.[127] It created a separate tool for use in placing ads regarding housing, employment and credit, called Special Ad Audiences, with an algorithm that does not use certain categories, "including age, gender, ZIP code or similar categories."[128]

However, in August 2019, HUD issued a proposed rule to immunize automated decisions from disparate impact challenges. Under the proposal, a defendant relying on an algorithmic model could defeat a claim at the prima facie stage by showing that the model's inputs do not include close proxies for protected classes; a neutral third-party determined that the model has predictive value; or a third party created the model.[129]

More examples of racial or gender bias are coming to light. According to a 2018 report in Reuters, in 2014 Amazon created an internal system to automate hiring decisions but within a year concluded that the algorithm discriminated against women. After trying and failing to fix the problem, Amazon abandoned the effort entirely. "In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word 'women's,' as in 'women's chess club captain.' And it downgraded graduates of two all-women's colleges, according to people familiar with the matter."[130] Such algorithmically-based discrimination may be seen as distinct from situations where an online platform expressly allows targeting on the basis of gender, age or other protected characteristics.[131]

---

[127] Emily Dreyfuss, *Facebook Changes Its Ad Tech to Stop Discrimination*, WIRED (March 19, 2019) https://www.wired.com/story/facebook-advertising-discrimination-settlement.

[128] Facebook, Help: Choosing a Special Ad Category, https://www.facebook.com/business/help/298000447747885.

[129] https://www.federalregister.gov/documents/2019/08/19/2019-17542/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard. See John Villasenor and Virginia Foggo, *Why a proposed HUD rule could worsen algorithm-driven housing discrimination*, BROOKINGS (April 16, 2020) https://www.brookings.edu/blog/techtank/2020/04/16/why-a-proposed-hud-rule-could-worsen-algorithm-driven-housing-discrimination/.

[130] Jeffrey Dastin, Amazon scraps secret AI recruiting tool that showed bias against women, Reuters (Oct. 9, 2018) https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[131] For example, a Carnegie Mellon study found that simulated users selecting a gender in Google's Ad Settings received employment-related advertisements at differing rates along gender lines despite identical web browsing patterns. See Amit Datta et al., *Discrimination in Online Advertising: A Multidisciplinary Approach*, 81 PROC. OF MACHINE LEARNING RES. 1 (2018) http://proceedings.mlr.press/v81/datta18a/datta18a.pdf. In 2016, ProPublica reported that Facebook's platform allowed advertisers to exclude Black, Hispanic, and other "ethnic affinities" from seeing ads, including housing and employment ads. There was no AI involved; Facebook simply allowed advertisers to choose which races saw their ads. Julia Angwin and 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. See also, Ariana Tobin and Jeremy Merrill, *Facebook is Letting Job Advertisers Target Only*

Discrimination can creep into AI-based processes in subtle ways. A decisionmaker that would never use race as a factor should be wary of using Zip Codes, since Zip Code can be a proxy for race. But many proxies for race can be more subtle. For example, researchers in 2019 reported finding bias in a commercial algorithm widely used in the US to guide health care decisions.[132] The goal behind the algorithm was to identify the patients most at risk, to ensure that they got priority treatment. The researchers found that Black patients assigned the same level of risk by the algorithm were actually sicker than white patients. The authors estimated that this racial bias reduced the number of Black patients identified for extra care by more than half. The bias arose because the algorithm used health costs as a proxy for health needs. Less money is spent on Black patients who have the same level of need, and the algorithm thus falsely concluded that Black patients were healthier, and thus less deserving of priority treatment, than equally sick white patients.

Provisions in the EU's General Data Protection Regulation, which took effect May 25, 2018, specify that every individual has a right (1) to be informed of the existence of automated decision-making, including profiling, and to be given meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject, and (2) to opt-out of any automated processing, including profiling, "which produces legal effects concerning him or her or similarly significantly affects him or her."[133] The exact scope of the new provisions is yet unclear. However, some industry critics have argued that the GDPR will have a negative impact on the development and use of AI in Europe, putting firms there at a disadvantage compared with their competitors in North America and Asia.[134]

1. Debiasing AI

Men, PROPUBLICA (Sept. 18, 2018) https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men. Facebook promised to modify its ad targeting tool to eliminate the ability to exclude individuals of certain races from seeking housing, employment, and credit ads. Ariana Tobin, *Facebook Promises to Bar Advertisers From Targeting Ads by Race or Ethnicity. Again*, PROPUBLICA (July 25, 2018). However, in December 2019, researchers found that Special Ad Audiences, Facebook's answer to biased ad placement, did little to reduce demographic biases in target audiences. Piotr Sapiezynski, *Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences* (Dec. 17, 2019), https://arxiv.org/pdf/1912.07579.pdf. See also Miranda Bogen and Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, UPTURN (Dec. 2018).

[132] Ziad Obermeyer at al, *Dissecting racial bias in an algorithm used to manage the health of populations*, SCIENCE (25 Oct 2019), https://science.sciencemag.org/content/366/6464/447.

[133] Commission Regulation, Arts. 13, 15 and 22 and Recital 71, General Data Protection Regulation, 2016 O.J. (L 119) (EC) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

[134] Nick Wallace & Dan Castro, *The Impact of the EU's New Data Protection Regulation on AI*, INFORMATION TECH. & INNOVATION FOUNDATION (Mar. 26, 2018), https://itif.org/publications/2018/03/26/impact-eu-new-data-protection-regulation-ai.

In 2019, Sendhil Mullainathan, a co-author of the study of algorithms used to guide health care decisions, stated, "Discrimination by algorithm can be more readily discovered and more easily fixed" than bias by humans.[135] However, others have concluded that avoiding bias is not simple. In December 2019, researchers found that Special Ad Audiences, Facebook's answer to biased ad placement, did little to reduce demographic biases in target audiences.[136] The removal of bias often requires a deep understanding of the political choices inherent in training data.[137]

K. Surveillance and Privacy (including Consumer Rights or Data Protection)

Privacy experts have been warning for some time that we live in a Golden Age of surveillance.[138] Information and communications technologies woven into our personal and professional lives generate enormous quantities of information: Internet browsing records, cell phone location data, emails and texts, credit card transactions, banking records, travel history and plans, health and activity measurements, and much more. This data, largely the by-product of services voluntarily adopted by individual consumers, is held by private sector entities and is readily available to governments.[139] The Internet of Things will dramatically increase the amount of information generated and disclosed to third parties. Meanwhile, governments have been deploying their own networks of video cameras, automated license plate reader, cell site simulators, and other sensors.[140]

The near-exponential growth in the variety and volume of personally identifiable data that characterizes our digital age may be dismissed as old news. AI will give the trendlines

---

[135] Sendhil Mullainathan, *Biased Algorithms Are Easier to Fix Than Biased People*, NEW YORK TIMES (Dec. 6, 2019) https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

[136] Piotr Sapiezynski, *Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences* (Dec. 17, 2019), https://arxiv.org/pdf/1912.07579.pdf. See Ava Kofman and Ariana Tobin, *Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement,* PROPUBLICA (Dec. 13, 2019), https://www.propublica.org/article/Facebook-Ads-Can-Still-Discriminate-Against-Women-and-Older-Workers-Despite-a-Civil-Rights-Settlement.

[137] Kate Crawford and Trevor Paglen, *Excavating AI: The Politics of Training Sets for Machine Learning* (Sept. 19, 2019) https://excavating.ai. See Will Knight, *AI Is Biased: Here's How Scientists Are Trying to Fix It*, WIRED (Dec. 19, 2019), https://www.wired.com/story/ai-biased-how-scientists-trying-fix.

[138] *See, e.g.*, Peter Swire, *The Golden Age of Surveillance*, SLATE (July 15, 2015), http://www.slate.com/articles/technology/future_tense/2015/07/encryption_back_doors_aren_t_necessary_we_re_already_in_a_golden_age_of.html

[139] *See* BULK COLLECTION: SYSTEMATIC GOVERNMENT ACCESS TO PRIVATE SECTOR DATA (Fred H. Cate & James X. Dempsey, eds., Oxford 2017).

[140] See Catherine Crump, *Surveillance Policy Making By Procurement*, 90 WASH. L. REV. 1595 (2016).

some upward boost. For example, AI-enabled digital assistants such as Alex and Siri are additional data collection devices that consumers willingly bring into their homes.

However, the greatest significance of AI for the surveillance potential of corporations and governments is that AI will vastly increase their ability to analyze this information and use it to make decisions about individuals. To take just one example of the intersection of sensors and AI: In Shenzhen, China, the AI firm Intellifusion installed cameras to capture photos of pedestrians crossing the street against a red light. Using the national database of photographs of citizens, the system's AI-based facial recognition technology can identify a jaywalker and display his photo, his family name, and part of his government ID number on large LED screens at the intersection – before the individual gets to the other side of the street. Shenzhen traffic police began using the system at major intersections in April 2017. In the 10 months thereafter, as many as 13,930 jaywalking offenders were recorded and displayed on the LED screen at a single intersection.[141] Such capabilities could soon come to the US. In April 2018, the Wall Street Journal reported that "[s]everal technology companies are working with police departments across the U.S. to develop the capability to add artificial intelligence to video surveillance and body cameras that could identify faces in real time."[142]

The privacy laws in the US have barely begun to deal with this future. They set few limits on data collection by the private sector or by the government though its own sensor networks.[143] However, a major new phase in US privacy law was opened on June 22, 2018, when the US Supreme Court ruled in that the government must obtain a warrant issued by a judge in order to compel a cell phone service provider to turn over historical cell site records indicating the location of a mobile phone user. Previously, such data had been available to the government with a mere subpoena, issued by executive branch investigators without approval of a judge. It will take years to develop the implications of the decision for the huge quantities of other kinds of data that individuals voluntarily disclosed to corporations in the course of using modern communications and information services.[144] Meanwhile, there remain relatively few limits on use of data once collected

---

[141] Li Tao, *Jaywalkers Under Surveillance in Shenzhen Soon To Be Punished Via Text Messages*, SOUTH CHINA MORNING POST (Mar. 27, 2018), http://www.scmp.com/tech/china-tech/article/2138960/jaywalkers-under-surveillance-shenzhen-soon-be-punished-text.

[142] Shibani Mahtani & Zusha Elinson, *Artificial Intelligence Could Soon Enhance Real-Time Police Surveillance*, WALL ST. J. (Apr. 3, 2018), https://www.wsj.com/articles/artificial-intelligence-could-soon-enhance-real-time-police-surveillance-1522761813.

[143] Clare Garvie & Alvaro Bedoya, *The Perpetual Lineup: Unregulated Police Face Recognition in America* (Oct. 18, 2016), https://www.perpetuallineup.org/; Branko Marcetic, *As License-Plate Tracking Increases, Privacy Advocates Press for More Regulation*, THE AM. PROSPECT (Oct. 11, 2016), http://prospect.org/article/license-plate-tracking-increases-privacy-advocates-press-more-regulation.

[144] *Carpenter v. United States*, 585 U.S. ____ (June 22, 2018) https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf.

(until one gets charged with a crime, when the due process protections of the Bill of Rights kick in). As to corporate uses, credit scores and credit reporting agencies are regulated, but many other uses of data are not specifically regulated.

As the US grapples with these issues, policy will also likely evolve in Europe, where data protection laws emphasize the principles of fairness, purpose limitation, data minimization, and transparency and where automated processing of data has long been a concern.[145] Issues related to fairness may be addressed in Europe in the context of the General Data Protection Regulation, mentioned above, which requires that a data controller engaged in "automated decision-making" must provide the data subject "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject," GDPR, Art. 13, and must implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision, Art. 22.[146]

1. Facial Recognition

Concerns around AI-based facial recognition have grown rapidly. Among other issues, there has been concern about the overall reliability of such systems and especially about error rates when matching faces of Blacks or other persons of color. [147] In January 2020 these concerns moved from the laboratory to the real world when a Black man was falsely arrested based on a facial recognition match.[148]

---

[145] *See* DATATILSYNET, THE NORWEGIAN DATA PROTECTION AUTHORITY, ARTIFICIAL INTELLIGENCE AND PRIVACY, (Jan. 2018) https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf.

[146] See Article 29 Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (wp251rev.01) (Feb. 6, 2018). See also GDPR Arts. 14, 15.

[147] See, for example, Joy Buolamwini and Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, PROC. OF MACHINE LEARNING RESEARCH 81:1–15 (2018) http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Patrick Grother, Mei Ngan and Kayee Honaoka, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, NIST (Dec. 2019) https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf (study of 189 facial recognition algorithms from 99 developers found that most had higher error rates with Black, Native American, and Asian faces); Jacob Snow, Amazon's Facial Recognition Falsely Matched 28 Members of Congress with Mugshots, ACLU Blog (July 26, 2018) https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.

[148] Kashmir Hill, *Wrongfully Accused by an Algorithm*, NEW YORK TIMES (June 24, 2020) https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html. See also Jason Koebler, *Detroit Police Chief: Facial Recognition Software Misidentifies 96% of the Time*, MOTHERBOARD (June 29, 2020) ("Detroit Police Chief James Craig admitted that the technology, developed by a company called DataWorks Plus, almost never brings back a direct match and almost always misidentifies people") https://www.vice.com/en_us/article/dyzykz/detroit-police-chief-facial-recognition-software-misidentifies-96-of-the-time.

A growing number of legislative bodies in the US have banned use of facial recognition systems because of fears about both inaccuracy and pervasive surveillance. In 2019, California adopted AB 1215, which placed a 3 year moratorium on law enforcement's use of any biometric surveillance system in connection with an officer camera or data collected by an officer camera. Penal Code, Section 832.19. "Biometric surveillance system" was defined to mean any computer software or application that performs facial recognition or other biometric surveillance, but not in-field fingerprint collection. The cities of Berkeley, CA and Somerville, MA adopted bans on all government use of facial recognition technology, while San Francisco adopted a ban and established an approval process for future adoption.[149] In March 2020, Washington state adopted what are perhaps the most detailed limits on facial recognition to date, requiring a detailed "accountability report" before any state or local government agency develops, procures, or uses a facial recognition service.[150] It also requires any vendor providing facial recognition to a state or local agency to provide an API or other technical capability to enable "legitimate, independent, and reasonable tests of those facial recognition services for accuracy and unfair performance differences across distinct subpopulations." In June 2020, Boston's City Council voted unanimously to ban the city government from using facial-recognition software.[151] In July, the New York state legislature passed a two-year moratorium on the use of facial recognition in schools.

Professional associations and tech companies have also acted. On June 30, 2020, the U.S. Technology Policy Committee of the Association of Computing Machinery issued a statement urging immediate suspension of private and governmental use of facial recognition technology in circumstances that might impact a person's human and legal rights.[152] In 2020, Amazon, Microsoft and IBM announced they would stop or pause their facial recognition offerings for law enforcement.[153]

---

[149] Tim Cushing, *California Governor Signs Bill Banning Facial Recognition Tech Use By State's Law Enforcement Agencies,* TechDirt (Oct. 18, 2019) https://www.techdirt.com/articles/20191011/18013143178/california-governor-signs-bill-banning-facial-recognition-tech-use-states-law-enforcement-agencies.shtml

[150] SB 6280, adding a new chapter to Title 43 RCW, lawfilesext.leg.wa.gov/biennium/2019-20/Pdf/Bills/Session%20Laws/Senate/6280-S.SL.pdf.

[151] See http://meetingrecords.cityofboston.gov/sirepub/cache/2/eqinlu0jap1ujw3tjrdbfztz/3067620706202005515 0790.PDF at p. 17 and http://meetingrecords.cityofboston.gov/sirepub/mtgviewer.aspx?meetid=511&doctype=MINUTES.

[152] https://www.acm.org/binaries/content/assets/public-policy/ustpc-facial-recognition-tech-statement.pdf.

[153] Kashmir Hill, *Wrongfully Accused by an Algorithm*, NEW YORK TIMES (June 24, 2020) https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (noting, however, that "The gestures were largely symbolic, given that the companies are not big players in the industry. The

L. Welfare Law and Other Civil Governmental Implementations

As governments adopt AI-based or algorithmic systems to make decisions, especially about welfare benefits, these systems are coming under challenge. Often these cases focus on the procedures by which the system was adopted:

- *K.W. ex rel. D.W. v. Armstrong*, 298 F.R.D. 479 (D. Idaho 2014) – automated decision-making system for Medicaid payments – court ordered the state to disclose the formula. As part of settlement, state agreed to develop new formula with greater transparency.
- *Ark. Dep't of Human Servs. v. Ledgerwood*, 530 S.W.3d 336 (Ark. 2017) – homecare for individuals with profound physical disabilities – complex computer algorithms - injunctive relief for plaintiffs - failure to adopt the rule according to notice and comment procedures.
- *Houston Federation of Teachers v. Houston Independent School District*, 51 F. Supp. 3d 1168 (S.D. Tex. 2017) – court ruled in favor of teachers on the ground that secrecy about the algorithm prohibited teachers from accessing, understanding, or acting on their own evaluations.
- *Barry v. Lyon*, 834 F.3d 706 (6th Cir. 2016) – state adopted matching algorithm for food assistance eligibility. More than 19,000 people were improperly matched, automatically disqualified, and given only vague notice. Court ruled the notice denied due process.

AI Now is tracking litigation challenging AI, issuing periodic updates on cases.[154]

III. The Broader Policy Framework

A. National AI Development Plans

The Eurasia Group, a consulting firm, has found that there are four pre-requisites to the development of AI:

- Data – "By far the most important element is the availability of large, labelled data sets" that can be used to train algorithms to optimize.
- Computational power – "Huge data sets require significant computing power."

---

technology police departments use is supplied by companies that aren't household names, such as Vigilant Solutions, Cognitec, NEC, Rank One Computing and Clearview AI.").

[154] Richardson, Schultz and Southerland, *Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems* (AI Now Institute, Sept. 2019) https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf

- Domain-specific focus – For now, successful AI is focused on a single, clearly defined domain.
- Special human expertise – Experts are still needed to tune AI to work for a specific domain and data set and there is competition for such talent.[155]

In plans of greater or lesser degrees of specificity, governments in developed countries around the world have addressed how to take advantage of the AI revolution and mitigate its negative impacts.

1. China

The Chinese government has issued a series of policy statements, plans, and strategies intended to boost AI research and development.[156] Most notably, in July 2017, the State Council declared that China intended to become the world leader in AI by 2030. In August 2017, the National Natural Science Foundation of China released guidelines identifying a series of research priorities to receive new funding. In October 2017, the National Development and Reform Commission (NDRC) announced its AI Innovation and Development Megaproject, with priorities that included deep learning and intelligent unmanned systems and service robots and funding for a series of new AI projects, including AI chips, cloud services, and open-source platforms.

In November 2017, the Ministry of Science and Technology created a New Generation AI Development Plan Promotion Office to coordinate 15 different entities, including the NDRC, the Ministry of Industry and Information Technology (MIIT) and key defense bodies. At the same time, the New Generation AI Strategic Advisory Commission was created, convening experts from academia and key private sector companies, including Baidu, Alibaba, Tencent, iFlytek, and Horizon Robotics.

In December 2017, the MIIT released a Three-Year Action Plan to Promote the Development of New-Generation Artificial Intelligence Industry (2018-2020). The plan recognizes the importance of an AI industry "support system" to include a data resource base with standard test data sets, cloud-based training frameworks, and initial test and evaluation systems. (In this context, the availability of massive amounts of data, a natural

---

[155] Kai-Fu Lee & Paul Triolo, *China's Artificial Intelligence Revolution: Understanding Beijing's Structural Advantages* (Dec. 2017), EURASIA GROUP, https://www.eurasiagroup.net/files/upload/China_Embraces_AI.pdf. For more on the talent question, see Jeremy Kahn, *Just How Shallow is the Artificial Intelligence Talent Pool?,* BLOOMBERG (Feb. 7, 2018), https://www.bloomberg.com/news/articles/2018-05-24/andy-rubin-s-phone-maker-essential-is-said-to-consider-sale; Cade Metz, *Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent*, N.Y. TIMES (Oct. 22, 2017) https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html.

[156] The description of China's AI strategy in the next two paragraphs relies heavily on Elsa Kania, *China's AI Agenda Advances*, THE DIPLOMAT (Feb. 14, 2018), https://thediplomat.com/2018/02/chinas-ai-agenda-advances/.

feature of China's information ecosystem, could be an advantage bolstered through policy.) The plan also reaffirmed China's commitment to accelerating the development of 5G networks, seen as part of the basic foundations of an ecosystem that could create a favorable environment for AI development.[157] The 3 Year Plan proposed concentrating on seven specific technologies: connected vehicles, service robots, unmanned aerial vehicles, medical imaging diagnosis systems, video image recognition, audio intelligence, and computer translation.[158]

Some analysts have concluded that the size of available datasets is the most important source of China's competitive advantage in AI.[159] Others have noted that the AI sector in China is dominated by private companies (Baidu, Tencent, others) that are not owned nor controlled by the Chinese government, possibly freeing these companies to take an innovative approach toward their development of AI.[160]

2. European Union

In April 2018, the European Commission issued a Communication[161] setting out a European initiative on AI, aiming to:

- Boost the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors9. This includes investments in research and innovation and better access to data.
- Prepare for socio-economic changes brought about by AI by encouraging the modernization of education and training systems, nurturing talent, anticipating changes in the labor market, supporting labor market transitions and adaptation of social protection systems.
- Ensure an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU. This includes forthcoming guidance on existing product liability rules, a detailed analysis of

---

[157] Paul Triolo, Elsa Kania, & Graham Webster, *Chinese Government Outlines AI Ambitions Through 2020*, NEW AM. (Jan. 26, 2018), https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/.

[158] *See id.*

[159] Kai-Fu Lee and Paul Triolo, *China's Artificial Intelligence Revolution: Understanding Beijing's Structural Advantages* (Dec. 2017) https://www.eurasiagroup.net/files/upload/China_Embraces_AI.pdf.

[160] Steve Dickinson, *China's Artificial Intelligence Plan – Stage 1*, CHINA LAW BLOG (March 19, 2018) https://www.chinalawblog.com/2018/03/chinas-artificial-intelligence-plan-stage-1.html.

[161] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS ON ARTIFICIAL INTELLIGENCE FOR EUROPE (Apr. 25, 2018), https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.

emerging challenges, and cooperation with stakeholders, through a European AI Alliance, for the development of AI ethics guidelines.

In the Communication, the Commission announced that it was increasing its annual investments in AI by 70% to EUR 1.5 billion for the period 2018-2020. It will:

- support research and innovation in AI technologies, in both basic and industrial research;
- strengthen AI research centers across Europe and encourage and facilitate their collaboration and networking;
- support the development of an "AI-on-demand platform" that will facilitate access of all potential users, especially small and medium-sized enterprises, companies from non-tech sectors and public administrations, to the latest technologies and encourage them to test AI;
- support the development of AI applications in key sectors.

The EU intends the infusion of public funds to stimulate private sector efforts, Under the existing public-private partnerships (for example in robotics and big data), the Commission predicted that its investment will trigger an additional EUR 2.5 billion over the same period.

The EU also recognized the importance of data. To that end, the Commission put forward a set of initiatives to grow the European data space. These are:

- an updated Directive on public sector information, e.g. traffic, meteorological, economic and financial data or business registers;
- guidance on sharing private sector data in the economy (including industrial data);
- an updated Recommendation on access to and preservation of scientific information;
- Communication on the digital transformation of health and care, including sharing of genomic and other health data sets.

Further on the question of data, the European Commission issued a data strategy in February 2020.[162]

In June 2019, the EU's High-Level Expert Group on AI issued policy and investment recommendations for "Trustworthy AI."[163] The recommendations focus on four main

[162] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *A European strategy for data*, COM(2020) 66 final (Feb. 19, 2020) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.

[163] https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

areas where Trustworthy AI can help achieving a beneficial impact, starting with humans and society at large, and continuing then to focus on the private sector, the public sector, and Europe's research and academia institutions. They address the main enablers needed to facilitate those impacts, focusing on availability of data and infrastructure, skills and education, appropriate governance and regulation, and funding and investment.

In February 2020, the European Parliament adopted a resolution on AI.[164]

Also in February 2020, the European Commission issued a white paper on AI in which it outlined principles for a new regulatory framework for AI, with the twin objectives of promoting the uptake of AI and of addressing the risks associated with certain uses of AI.[165] The paper noted that existing provisions of EU law, both with respect to product safety and liability and as regards the protection of fundamental rights and consumer rights, will continue to apply in relation to AI, but that certain updates to that framework may be necessary to reflect the digital transformation and the use of AI. The white paper suggested that new rules might be needed in six areas:

- training data (such as requirements aimed at providing reasonable assurances that the subsequent use of the products or services that the AI system enables is safe);
- data and record-keeping (for example, requiring accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected, and, in certain justified cases, the data sets themselves;
- information to be provided (ensuring clear information to be provided as to the AI system's capabilities and limitations);
- robustness and accuracy (such as requirements ensuring that the AI systems are robust and accurate, or at least correctly reflect their level of accuracy, during all life cycle phases, or requirements ensuring that outcomes are reproducible;
- human oversight (either before or after a decision is made); and
- specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification.

Under the Commission's vision, mandatory requirements contained in any new regulatory framework on AI would in principle apply only to those applications identified as high-risk under two key criteria discussed in the report. The Commission stressed that the White Paper and an accompanying report on the safety and liability framework marked the opening of a broad consultation of Member States, civil society, industry and academics. The consultation is open for comments until 19 May 2020.

---

[164] https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.html.

[165] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

3. France

In March 2018, Cedric Villani, an MP and renown mathematician, released his report on AI born out of a 6-month mission requested by the Prime Minister.[166] The plan recommended –

- An aggressive strategy to overcome France's lag in accumulating data in a form useful for AI research. This should involve encouraging economic players to share and pool their data, with the State acting as a trusted third party;
- Targeting four sectors: healthcare, environment, transport-mobility and defense-security;
- Promoting agile and enabling research;
- Assessing and planning for the effects of Ai on the future of work and the labor market: Setting up a public lab for labor transformations and trying out new professional training funding methods;
- Leveraging AI for a more ecological economy;
- Addressing the ethical considerations of AI, by opening up the "black boxes" of AI, implementing ethics by design, setting up an AI ethics committee; and
- Ensuring that AI supports inclusivity and diversity.

President Macron endorsed the report and announced that the French government will spend €1.5 billion ($1.85 billion) over five years to support research in the field, encourage startups, and collect data that can be used, and shared, by engineers.

4. United States

Compared to other countries, the US was slow in development of a national AI strategy.

In February 2019, President Trump issued an Executive Order on AI, declaring, "It is the policy of the United States Government to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI R&D and deployment through a coordinated Federal Government strategy … ."[167] On some key

---

[166] Cedric Villani, *For a Meaningful Artificial Intelligence: Towards a French and European Strategy* (Mar. 2018), https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf. France released its AI strategy. What does it mean for Europe? https://medium.com/@IF_Estonie/france-released-its-ai-strategy-what-does-it-mean-for-europe-c4ff40ab2ce4

[167] *Executive Order 13859, on Maintaining American Leadership in Artificial Intelligence* (Feb. 11, 2019) https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/. The Trump EO does not mention but probably supersedes the National Artificial Intelligence Research and Development Strategic Plan released in October 2016 by the Obama Administration. NAT'L SCI. AND TECH. COUNCIL, THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN (Oct. 2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf. That plan also did not define specific research agendas for individual Federal

issues, the order lacked specifics. For example, on funding, it merely directed agencies to "consider" AI as a priority:

> Heads of implementing agencies that also perform or fund R&D (AI R&D agencies), shall consider AI as an agency R&D priority, as appropriate to their respective agencies' missions, consistent with applicable law and in accordance with the Office of Management and Budget (OMB) and the Office of Science and Technology Policy (OSTP) R&D priorities memoranda.

The EO recognized the importance of data to AI development, directing heads of all agencies to review their Federal data and models "to identify opportunities to increase access and use by the greater non-Federal AI research community in a manner that benefits that community, while protecting safety, security, privacy, and confidentiality." On the regulation of AI applications, the order directed the OMB director to issue a memorandum to the heads of all agencies that shall inform the development of regulatory and nonregulatory approaches by such agencies regarding technologies and industrial sectors that are either empowered or enabled by AI and consider ways to reduce barriers to the use of AI technologies. The EO mentions a National Security Presidential Memorandum of February 11, 2019 titled Protecting the United States Advantage in Artificial Intelligence and Related Critical Technologies.

In August 2019, pursuant to a directive in the President's EO, the U.S. Department of Commerce's National Institute of Standards and Technology (NIST) released a plan for prioritizing federal agency engagement in the development of standards for artificial intelligence (AI). The plan recommends that the federal government "commit to deeper, consistent, long-term engagement" in activities to help the United States speed the pace of reliable, robust and trustworthy AI technology development.[168]

In January 2020, the White House's Office of Science and Technology Policy released for public comment a draft of 10 principles that agencies would have to meet when drafting regulations for the private sector's use of AI.[169]

---

agencies. As is also true under the Trump Administration order, agencies were left to continue pursuing priorities consistent with their missions, capabilities, authorities, and budgets. See also Press Release, White House, Artificial Intelligence and the American People (May 10, 2018), https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/.

[168] https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[169] https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

Federal funding for AI research was initially slow in coming. For FY 2018, President Trump proposed a 7% cut in overall funding for research in the areas that included AI,[170] but Congress increased funding for research and development.[171] (However, it was hard to tell how much money was going to AI research.) In an updated FY 2020 budget request released in September 2019, the White House proposed doubling agencies' spending on AI compared to 2016.[172] According to a White House fact sheet issued in January 2020, the first ever agency-by-agency report of nondefense AI R&D spending identified $1 billion in nondefense R&D for AI in FY 2020.[173] The FY 2021 budget called for further increases: more than $830 million for AI R&D and interdisciplinary research institutes at the National Science Foundation to, which represents a more than 70 percent increase over the FY 2020 budget, and, in regards to Defense AI R&D, $459 million for DARPA, an increase of $50 million from FY 2020, and for the Department of Defense's Joint AI Center an increase from $242 million in FY 2020 to $290 million in FY 2021.[174]

In May 2018, the Trump Administration announced the creation of a Select Committee on Artificial Intelligence, under the National Science and Technology Council (NSTC).[175] The February 2019 EO stated that the AI Initiative would be coordinated through the Select Committee. There is an apparently separate National Security Commission on Artificial Intelligence, which issued an interim report in November 2019.[176]

B. The Impact of AI on Work and Employment

---

[170] NETWORKING AND INFO. TECH. RESEARCH AND DEV. PROGRAM, SUPPLEMENT TO THE PRESIDENT'S BUDGET (Oct. 2017), https://www.nitrd.gov/pubs/2018supplement/FY2018NITRDSupplement.pdf.

[171] Marina Koren, *Congress Ignores Trump's Priorities for Science Funding*, THE ATLANTIC (Mar. 23, 2018), https://www.theatlantic.com/science/archive/2018/03/trump-science-budget/556229/.

[172] https://www.whitehouse.gov/wp-content/uploads/2019/09/FY2020-NITRD-AI-RD-Budget-September-2019.pdf. See https://federalnewsnetwork.com/artificial-intelligence/2019/09/federal-cto-spending-on-ai-research-nearly-doubled-since-2016/.

[173] https://www.whitehouse.gov/briefings-statements/president-trumps-fy-2021-budget-commits-double-investments-key-industries-future/

.

[174] Id.

[175] For materials related to the committee, see https://epic.org/privacy/ai/wh-committee/. The White House AI website is https://www.whitehouse.gov/ai/.

[176] National Security Commission on Artificial Intelligence, Interim Report (Nov. 2019) https://www.epic.org/foia/epic-v-ai-commission/AI-Commission-Interim-Report-Nov-2019.pdf.

AI will have a large impact on work, employment, and employability. How much and in what direction the impact of AI will be felt is hard to predict. Some studies have warned that a large percentage of jobs will be lost to automation in the coming decades.[177] The reality is probably more complex. In a December 2017 report, the McKinsey Global Institute concluded that very few occupations—less than 5 percent—consist entirely of activities that can be fully automated.[178] However, McKinsey found that in about 60% of occupations at least one-third of activities can be automated. Taking into account various factors that will affect adoption, McKinsey estimated that as much as 30% of hours worked globally could be automated by 2030, with a midpoint of 15%. Even a 15% loss of hours worked, if translated into lower demand for workers, could impose significant hardship. Yet McKinsey also stressed that increased investment and productivity growth from automation could spur enough growth to ensure full employment.

The key message of the McKinsey study is that automation will have a huge impact, across all economies and almost all occupations: "[Al]l workers will need to adapt, as almost all occupations will evolve alongside increasingly capable machines." Some significant percentage of workers (3 to 14 percent of the global workforce) will need to change occupations. Many more will need different skills.

"To achieve good outcomes," McKinsey warned, "policy makers and business leaders will need to embrace automation's benefits and, at the same time, address the worker transitions brought about by these technologies." "Ensuring positive employment outcomes will require a laser focus on retooling the workforce, stepping up support for workers in transition, and improving how local and national labor markets function. Societies can choose to transform the coming labor market disruptions into an opportunity rather than a pitfall." McKinsey pointed to Germany as an example of how revamping labor market agencies and support for workers in times of transition can dramatically reduce unemployment.

Specifically, McKinsey recommended that governments --

---

[177] A 2013 study from Oxford University suggested that 47% of total United States employment could be lost to automation in the next two decades. CARL B. FREY & MICHAEL A. OSBORNE, THE FUTURE OF EMPLOYMENT: HOW SUSCEPTIBLE ARE JOBS TO COMPUTERISATION? 38 (2013), http://www.oxfordmartin.ox.ac. uk/downloads/academic/TheFutureofEmployment.pdf. A 2014 joint study between Oxford University and Deloitte concluded that around 35% of jobs in the UK were at "high risk" of computation over the next 20 years. Deloitte LLP Agiletown: the relentless march of technology and London's response (2014). See also Deloitte, From brawn to brains: The impact of technology on jobs in the UK (2015). The same study found, however, that while technology had potentially contributed to the loss of 800,000 lower-skilled jobs, there was equally strong evidence to suggest that it had helped create nearly 3.5 million new higher-skilled and higher-paid jobs.

[178] McKinsey Global Institute, *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation* (December 2017).

- Radically expand midcareer training opportunities to make lifelong learning a reality.
- Modernize educational systems for the 21st century.
- Expand transition support measures for workers.
- Create income support measures consistent with the new wage realities.

McKinsey concluded that "[t]echnology adoption can and often does cause significant short-term labor displacement'' – with painful consequences for some workers – "but history shows that, in the longer run, it creates a multitude of new jobs and unleashes demand for existing ones, more than offsetting the number of jobs it destroys even as it raises labor productivity."

While the overall impact of automation will be less catastrophic and more slow-moving than some predict, the dislocation could be severe and abrupt for some occupations. Full adoption of driverless vehicles could put at least 2.5 million drivers in the US out of work.[179] According to the research firm CB Insights, 4.3 million cooks and servers in US fast food chains, cafeterias and restaurants face a high risk of automation.[180]

McKinsey concluded that the impact of automation would be felt differently in developed versus developing countries. In advanced economies, demand for work requiring completion of secondary school or less will likely decline, but in developing countries, significant demand will be created for workers with a secondary school education. "According to our analysis, as many as 100 million new jobs could be created for Indians with secondary education—even after accounting for the effect of automation—as rising prosperity will create a surge of new labor demand for construction, retail, and health care and education jobs, among others."

C. Principles for Ethical Development and Governance of AI

A wide range of entities have issued principles or guidelines for governance of AI. In addition to the questions associated with the ethical use of AI, these may address questions concerning the design of AI and when is it responsible or irresponsible to publish AI breakthroughs that may be misused.[181]

Several voluntary and informal groups have issued principles for AI development intended to address the ethical issues posed by AI:

---

[179] Darrell West, THE FUTURE OF WORK: ROBOTS, AI, AND AUTOMATION (Brookings, 2018).

[180] CBI, AI Will Put 10 Million Jobs At High Risk — More Than Were Eliminated By The Great Recession (Oct. 6, 2017) https://www.cbinsights.com/research/jobs-automation-artificial-intelligence-risk/.

[181] Rebecca Crootof, *Artificial Intelligence Research Needs Responsible Publication Norms*, Lawfare (Oct. 24, 2019) https://www.lawfareblog.com/artificial-intelligence-research-needs-responsible-publication-norms.

- Asilomar AI Principles.[182]
- Accountable Algorithms, by Joshua Kroll et al.[183]
- IEEE Global Initiative on Ethics of Autonomous and intelligence Systems.[184]

Likewise, a variety of governmental and intergovernmental bodies have issued principles or recommendations:

- In 2019, the G20 issued principles for responsible stewardship of trustworthy AI and recommendations.[185]
- In April, 2019, the European Commission High-Level Expert Group on Artificial Intelligence released the final version of its "Ethics Guidelines for Trustworthy Artificial Intelligence."[186]
- In October 2019, the U.S. Defense Innovation Board issued principles for the ethical use of AI by the Department of Defense,[187] and in February 2020, the DoD itself adopted ethical principles for AI.[188]
- In May 2019, the OECD, including the US, adopted a recommendation that governments promote and implement a set of principles for responsible stewardship of trustworthy AI.[189]
- In November 2019, the Council of Europe's Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence issued draft recommendations on the human rights impacts of algorithmic systems, including guidelines for both member states and private sector actors.[190]

---

[182] https://futureoflife.org/ai-principles/.

[183] https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/.

[184] https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

[185] https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf.

[186] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.

[187] Defense Innovation Board, *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense* (Oct. 2019) https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF.

[188] https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

[189] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[190] https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf.

- In 2018, the UN's Special Rapporteur on the promotion and protection of the right to freedom of expression and opinion issued a report, with recommendations, on the impact of AI on freedom of expression and opinion, considering the role of AI in search engines, social media platforms, messaging applications, and public information mechanisms.[191]

Among the OECD principles:

- AI actors (including corporations and individual researchers) should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.
- AI actors should commit to transparency and responsible disclosure, to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors and the logic that served as the basis for the prediction, recommendation or decision
- AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.
- Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.
- Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programs along the working life, support for those affected by displacement, and access to new opportunities in the labor market.

Meanwhile, a growing community of academics and industry experts has begun studying and debating issues of AI transparency and accountability.[192] Many such researchers convene in an annual conference, Fairness, Accountability and Transparency in ML.[193] Facebook, Alphabet, and Microsoft have formed ethics teams around AI[194] and several

---

[191] https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx.

[192] See, for example, Tolga Bolukbasi et al., *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*.

[193] Fairness, Accountability, and Transparency in Machine Learning, https://www.fatml.org/.

major companies have created the Partnership for AI to specifically address the ethical issues associated with AI, including transparency, fairness, and accountability. The AI Now Institute https://ainowinstitute.org/ is producing research on a range of social and ethical issues posed by AI, as is the AFOG Working Group at UC Berkeley. The Defense Advanced Research Projects Agency (DARPA) has a project on Explainable AI.[195]

How effective these principles for AI will be is an open question. With major corporations and government agencies adopting ethical principles, there is concern with "ethics washing," the instrumental use of the language of ethics as a façade that justifies deregulation, self-regulation, or market driven governance.[196] In a beginning to answering the question of effect, in May 2020, the Center for Long-Term Cybersecurity at UC Berkeley issued a report using three case studies to assess the implementation of AI principles.[197]

IV. Conclusion

AI poses challenges for law, corporate and public policy, and ethics. Across a wide range of disciplines (product liability, intellectual property, fraud, criminal law, discrimination, privacy, and many others) courts will have to apply traditional legal doctrines to complex and sometimes unexplainable systems. Policymakers must consider whether to modify existing regulatory structures to specifically address AI deployments. The resulting framework is likely to combine a variety of governance tools: case-by-case adjudication, legislative rulemaking, regulatory agency action, and deference to industry standards and voluntary best practices. Development is likely to be uneven, with false starts. As specific issues are addressed, the legal and policy system will face traditional tensions between supporting innovation and protecting public safety, between incentivizing investment and promoting equity.

---

[194] *See* Jordan Novet, *Facebook forms a special ethics team to prevent bias in its A.I. software*, CNBC (May 3, 2018), https://www.cnbc.com/2018/05/03/facebook-ethics-team-prevents-bias-in-ai-software.html.

[195] David Gunning, *Explainable Artificial Intelligence*, Defense Advanced Research Projects Agency, https://www.darpa.mil/program/explainable-artificial-intelligence.

[196] See Karen Yeung, Andrew Howes, and Ganna Pogrebna, *AI Governance by Human Rights-Centered Design, Deliberation and Oversight: An End to Ethics Washing* (June 21, 2019), in M. Dubber and F. Pasquale (eds.) THE OXFORD HANDBOOK OF AI ETHICS (2019) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3435011; Karen Hao, In 2020, let's stop AI ethics-washing and actually do something, TECHNOLOGY REVIEW (Dec. 27, 2019) https://www.technologyreview.com/s/614992/ai-ethics-washing-time-to-act/. For a definition of ethics washing, see Elletra Bietti, *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy* (December 1, 2019), proceedings of the ACM FAT* Conference (FAT* 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3513182.

[197] Jessica Cussins Newman, DECISION POINTS IN AI GOVERNANCE (May 2020) https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf