February 15, 2019

Federal Trade Commission
Ellen Connelly
Office of Policy Planning
600 Pennsylvania Avenue NW
Washington, DC 20580

RE: Competition and Consumer Protection in the 21st Century Hearings, Project Number P181201

Dear Ms. Connelly,

The Center for Data Innovation is pleased to submit these comments in response to the request for comment from the Federal Trade Commission (FTC) about its hearing on Competition and Consumer Protection in the 21st Century focusing on algorithms, artificial intelligence (AI), and predictive analytics.[1] These comments explain the importance of AI in the digital economy and offer ideas for how regulators can provide effective oversight of automated systems so as to both protect consumers and allow innovation to flourish.

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, D.C., and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

[1] "FTC Announces Agenda for the Seventh Session of its Hearings on Competition and Consumer Protection in the 21st Century; Session at Howard University to Focus on Algorithms, Artificial Intelligence, and Predictive Analytics," Federal Trade Commission, October 31, 2018, https://www.ftc.gov/news-events/press-releases/2018/10/ftc-announces-agenda-seventh-session-its-hearings-competition.

BACKGROUND ON ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS, AND APPLICATIONS OF THE TECHNOLOGIES

From a technical perspective, is it sometimes impossible to ascertain the basis for a result produced by these technologies? If so, what concerns does this raise?

It is sometimes possible to determine how algorithmic systems make decisions, either by examining static code or by building the systems so that they indicate which factors weighed heavily in their decision-making processes. However, for some advanced AI applications, these methods would be impossible, or when they are, there may be inescapable tradeoffs between interpretability of a system and its accuracy. As data scientists Max Kuhn and Kjell Johnson say in their book *Applied Predictive Modeling*, "Unfortunately, the predictive models that are most powerful are usually the least interpretable."[2] An algorithm's accuracy typically increases with its complexity, but the more complex an algorithm is, the more difficult it is to explain.[3] There are few situations where it would be appropriate to sacrifice accuracy for an increase in interpretability. For example, it would not be desirable to prioritize explainability over accuracy in autonomous vehicles, as even slight reductions in navigation accuracy or to a vehicle's ability to differentiate between a pedestrian on the road and a picture of a person on a billboard could be enormously dangerous.[4] While many researchers are exploring how to build algorithmic systems that can explain results, it is unclear with the accuracy-explainability tradeoff will be entirely eliminated. Thus, it would be imprudent to limit the use of automated systems that can not provide explainability. Moreover, consumers do not always prioritize explanations. In a representative survey of U.S. Internet users, only 25 percent of respondents agree with the statement: "Agree or disagree? If doctors use an app to diagnose patients, it is more important that the app be able to explain its diagnosis than for the diagnosis to be correct."[5]

What are the advantages and disadvantages of developing technologies for which the basis for the results can or cannot be determined? What criteria should determine when a "black box" system is acceptable, or when a result should be explainable?

---

[2] Max Kuhn and Kjell Johnson, Applied Predictive Modeling (New York: Springer-Verlag New York, 2013) 50.
[3] Jason Brownlee, "Model Prediction Accuracy Versus Interpretation in Machine Learning," Machine Learning Mastery, August 1, 2014, https://machinelearningmastery.com/model-prediction-versusinterpretation-in-machine-learning/.
[4] Joshua New and Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability," Center for Data Innovation, May 21, 2018, http://www2.datainnovation.org/2018-algorithmic-accountability.pdf.
[5] Daniel Castro, Unpublished submission to PrivacyCon 2019, Center for Data Innovation, *forthcoming*.

While it is true that the decision-making processes of some algorithmic systems can be so opaque as to be considered "black boxes," this is not a disadvantage unless the context of a particular decision requires explainability or transparency. In fact, algorithmic systems are desirable for the same properties that can make them black boxes, in that their complexity is what enables them to process large amounts of data and solve enormously complex problems better than humans.

However, this does not mean a black box algorithm could, or should, be exempt from oversight. All algorithms, "black box" or otherwise, can be acceptable so long as their operators (i.e., the party responsible for deploying the algorithm) abide by algorithm accountability—the principle that an algorithmic system should employ a variety of controls to ensure the operator can verify it acts in accordance with its intentions, as well as identify and rectify harmful outcomes.[6] Algorithmic accountability promotes desirable outcomes, protects against harmful ones, and ensures algorithmic decisions are subject to the same requirements as human decisions. This approach is technology neutral, granting operators flexibility to employ a variety of different technical and procedural mechanisms to achieve algorithmic accountability. Importantly, algorithmic accountability is relevant only when an application of algorithmic decision-making poses potential harms significant enough to warrant regulatory scrutiny, and not, for example, applications that only pose the risk of minor inconveniences should the algorithms involved be flawed.

The first step in achieving algorithmic accountability is determining whether algorithms are working the way their operators intended. If the answer is yes, and it is causing harm, then it is important to recognize that laws and regulations already exist for many commercial functions to prohibit racial discrimination, require due process, and so on. When an operator intends to cause harm, whether they use an algorithm to do so should be irrelevant. There are a variety of different technical and procedural mechanisms that can be employed, when contextually relevant, to make the determination of whether a harm is intentional. These include: transparency, explainability, confidence measures, and procedural regularity. In most cases, operators would likely have to employ a combination of several of these mechanisms in order to be confident an algorithm is acting as they intended. This is not meant to be a comprehensive list of all the ways an operator can verify an algorithm is acting as intended, as there may be methods that are only useful in niche circumstances or that have yet to be developed.

[6] Ibid.

Simply taking steps to verify an algorithm is acting as intended is not enough to ensure it is not also producing harmful outcomes. Thus, an accountable algorithmic system must also allow operators to identify and minimize harmful outcomes. This is an important capability because it allows for organizations to responsibly deploy algorithms despite not being able to predict or control for every possible harmful outcome that could arise from an algorithm's decisions—which would likely be impossible and could severely limit the utility of algorithms. There are a variety of methods to accomplish this that allow operators to take meaningful steps to minimize harms. These include, but are not limited to, impact assessment, error analysis, and bias testing. Importantly, these are not simply just post hoc controls—operators can and should apply these steps throughout the entire process of developing and deploying an algorithm, and continuously employ them throughout the time an algorithm is in use. Operators can also prioritize their efforts to address uses of algorithms where the consequences of errors pose the greatest risk to consumers.

Overall however, from a regulatory perspective, an algorithmic system's decisions should only be explainable in specific contexts where the law already requires explainability. For example, the Equal Credit Opportunity Act requires a creditor to provide consumers with an adequate explanation of why their credit application was declined, while the Fair Credit Reporting Act requires a creditor to both provide consumers with their credit report and investigate disputes concerning incorrect information and make corrections as needed.[7] These laws apply regardless of whether a creditor uses an algorithm in these processes. It would be a mistake to apply additional explainability requirements for algorithmic decisions when those requirements do not also apply to human decisions in that same context because if that decision has the potential to cause harm to the extent that explainability is desirable, it should not matter whether a human or algorithm makes the decision. Imposing general regulations on the use of algorithms would be particularly problematic because it would limit innovation even in areas where concerns about errors are significantly less consequently.

COMMON PRINCIPLES AND ETHICS IN THE DEVELOPMENT AND USE OF ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS

What are the main ethical issues (e.g., susceptibility to bias) associated with these technologies? How are the relevant affected parties (e.g., technologists, the business community, government,

[7] "Rights if Denied Credit," LegalMatch, Accessed May 9, 2018, https://www.legalmatch.com/law-library/article/rights-if-deniedcredit.html?intakeredesigned=1; "Disputing Errors on Credit Reports," U.S. Federal Trade Commission, Accessed May 9, 2018, https://www.consumer.ftc.gov/articles/0151-disputing-errors-creditreports.

consumer groups, etc.) proposing to address these ethical issues? What challenges might arise in addressing them?

For many years, policymakers have sought to address the "digital divide"—the social and economic disadvantages that may result from a lack of access to technology. Now policymakers should begin a concerted effort to address the "data divide"—the social and economic inequalities that may result from a lack of collection or use of data about an individual or community. Already gaps are appearing where certain groups of individuals do not have data collected about them or their communities because of where they live.[8] Policies that limit data collection may exacerbate the data divide.

In addition to the risk of underrepresentation in data, there is the risk of data capturing human biases and thus perpetuated these biases in algorithms. While a valid concern and important to address, most solutions proposed thus far are typically ineffective, counterproductive, or harmful to innovation. These proposals fall into three main categories: calls for algorithmic transparency, explainability, or both; calls for the creation of regulatory bodies to oversee all algorithmic decision-making; and generalized regulatory proposals, or proposals that rely so heavily on poorly articulated or vague concepts that they are simply not viable. Most of these proposals endorse the precautionary principle and are based on the belief that algorithms, particularly AI, should be proactively regulated and proven safe before being deployed—and once deployed, should be heavily regulated. But there are also a number of people who believe government should not regulate emerging technologies and should leave industry solely responsible for addressing the potential harms of algorithmic decision-making. While many types of algorithmic decision-making do not require additional regulatory oversight, some do. It is important to note that certain aspects of these proposals do have merit, and some of these concepts are valid and useful components of algorithmic accountability. However, while they have their place in particular contexts, it would be inappropriate to apply these policies across all sectors of the economy.

Algorithmic Transparency and Explainability

The most common proposal for regulating algorithms focuses on the principle of algorithmic transparency, which requires organizations to expose their algorithms and information about their data to some degree of public scrutiny. Supporters define this principle in different ways, but the common theme is algorithmic transparency is based on the notion that the complexity and

[8] Daniel Castro, "The Rise of Data Poverty in America," Center for Data Innovation, September 10, 2014, http://www2.datainnovation.org/2014-data-poverty.pdf.

proprietary nature of algorithms can obscure how they make decisions and thus mask harmful behavior. Algorithmic transparency advocates believe that exposing the code and underlying data of these black boxes would allow the public and regulators to identify whether and how an algorithm is producing harmful outcomes.

Support for algorithmic transparency is widespread, both in the United States and abroad.[9] A Pew survey found that many technologists believe algorithmic transparency would be a good way to mitigate the risks of algorithms, while the U.S. Federal Trade Commission (FTC) has expressed support for algorithmic transparency—though it is unclear exactly how the FTC defines it.[10] Cathy O'Neil, author of the book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* and founder of O'Neil Risk Consulting & Algorithmic Auditing (ORCAA), a consultancy that helps companies manage algorithmic risks, writes, "Models that have a significant impact on our lives, including credit scores and e-scores, should be open and available to the public," and that certain potentially harmful algorithms "must also deliver transparency, disclosing the input data they're using as well as the results of their targeting."[11] Additionally, the Electronic Privacy Information Center (EPIC) states, "Algorithmic transparency should be established as a fundamental requirement for all AI-based decision-making."[12] EPIC would have regulators go even further, asserting, "The algorithms employed in big data should be made available to the public."[13] In effect, many in this camp believe any computer system that uses automated decisions should make its source code available for some degree of public scrutiny.

---

[9] "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" (Washington, D.C.: Executive Office of the President, May 2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf; Catelijne Muller, *Artificial Intelligence* (Netherlands: European Economic and Social Committee, 2017), https://www.eesc.europa.eu/our-work/opinions-information-reports/opinions/artificial-intelligence.

[10] Lee Rainie and Janna Anderson, "Code-Dependent: Pros and Cons of the Algorithm Age," Pew Research Center, February 8, 2017, http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/; Christopher Zara, "FTC Chief Technologist Ashkan Soltani on Algorithmic Transparency and the Fight Against Biased Bots," *International Business Times*, April 9, 2015, http://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-biased-1876177; "Office of Technology Research and Investigation," Federal Trade Commission, Accessed March 8, 2018, https://www.ftc.gov/about-ftc/bureaus-offices/bureau-consumer-protection/office-technology-research-investigation.

[11] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016), ORCAA, Accessed May 8, 2018, http://www.oneilrisk.com/#services-section.

[12] Ibid.

[13] "Comments of the Electronic Privacy Information Center to the Office of Science and Technology Policy," EPIC, April 4, 2014, https://epic.org/privacy/big-data/EPIC-OSTP-Big-Data.pdf.

At the same time, there are others who call for more general transparency. Market research executive Barry Chudakov thinks companies should include the equivalent of a nutrition label for their algorithms, indicating how an algorithm might make certain decisions, and the implications of those decisions.[14] Ben Wagner, director of the Centre for Internet and Human Rights, argues that companies should disclose whether decisions they make on their platforms are made by algorithms or humans.[15]

Others, such as Judith Donath, a fellow at Harvard University's Berkman Klein Center for Internet & Society, lament the opacity of complex algorithmic systems, arguing:

> The danger in increased reliance on algorithms is that the decision-making process becomes oracular: opaque yet unarguable. The solution is design. The process should not be a black box into which we feed data and out comes an answer, but a transparent process designed not just to produce a result, but to explain how it came up with that result. The systems should be able to produce clear, legible text and graphics that help the users—readers, editors, doctors, patients, loan applicants, voters, etc.—understand how the decision was made.[16]

While various players in this camp state they want "transparency," they typically mean "explainability," which are two commonly conflated terms in discussions about governing algorithms.[17] Transparency refers to disclosing an algorithm's code or data (or both), while explainability refers to the concept of making algorithms interpretable to end users, such as by having operators describe how algorithms work or by using algorithms capable of articulating the rationales for their decisions. For example, the European Union has made explainability a primary check on the potential harms of algorithmic decision-making, guaranteeing in its GDPR the right for a person to obtain "meaningful information" about certain decisions made by an

[14] Lee Rainie and Janna Anderson, "Code-Dependent: Pros and Cons of the Algorithm Age," Pew Research Center, February 8, 2017, http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.
[15] Matt Burgess, "Holding AI to Account: Will Algorithms Ever Be Free from Bias If They're Created by Humans?" *Wired*, January 11, 2016, http://www.wired.co.uk/article/creating-transparent-ai-algorithms-machine-learning.
[16] Lee Rainie and Janna Anderson, "Code-Dependent: Pros and Cons of the Algorithm Age," Pew Research Center, February 8, 2017, http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.
[17] Ibid.

algorithm.[18] Similarly, France's Secretary of State for Digital Affairs, Mounir Mahjoubi, has stated that the government should not use an algorithm if it cannot explain its decisions.[19]

While transparency and explainability are fundamentally different concepts, they share many of the same flaws as a solution for regulating algorithms. First, they hold algorithmic decisions to a standard that simply does not exist for human decisions. As EPIC describes, "Without knowledge of the factors that provide the basis for decisions, it is impossible to know whether government and companies engage in practices that are deceptive, discriminatory, or unethical. Therefore, algorithmic transparency is crucial to defending human rights and democracy online."[20] This argument fails to recognize that algorithms are simply a recipe for decision-making. If proponents of algorithmic transparency and explainability are concerned that these decisions are harmful, then it is counterproductive to only call for algorithmic decisions to be transparent or explainable, rather than for all aspects of all decision-making to be made public or explained. If blanket mandates for transparency and explainability are appropriate for algorithmic decision-making, but not human decision-making (which itself is often supported by computers), logic would dictate that human decisions are already transparent, fair, and free from unconscious and overt biases. In reality, bias permeates every aspect of human decision-making, so to hold algorithms to a higher standard than for humans is simply unreasonable. For example, research shows taxicabs frequently do not pick up passengers based on their race, and employers may eliminate job applicants with African-American sounding names despite their sufficient qualifications.[21] Yet, understandably, taxi drivers are not required to publicly report their reasons for not picking up every passenger they pass by, and employers do not have to publish a review of every resume they receive, with detailed notes explaining why they choose not to offer a particular candidate a job, because laws and regulations for these sectors focus on outcomes, not unconscious bias. If EPIC and other proponents of algorithmic transparency and explainability worry that such broad categories of decisions have the potential to be harmful due to the

[18] Nick Wallace and Daniel Castro, "The Impact of the EU's New General Data Protection Regulation on AI" (Center for Data Innovation, March 2018), http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf.
[19] "Humans May Not Always Grasp Why AIs Act. Don't Panic." *The Economist*, February 15, 2018, https://www.economist.com/news/leaders/21737033-humans-are-inscrutable-too-existing-rules-and-regulations-can-apply-artificial?frsc=dg%7Ce.
[20] "Algorithmic Transparency: End Secret Profiling," EPIC, Accessed May 8, 2018, https://www.epic.org/algorithmic-transparency/.
[21] Cornell Belcher and Dee Brown, "Hailing While Black—Navigating the Discriminatory Landscape of Transportation" (key findings from the hailing while black survey of Chicago voters, Brilliant Corners, February 12, 2015), http://www.brilliant-corners.com/post/hailing-while-black; David R. Francis, "Employers' Replies to Racial Names" (The National Bureau of Economic Research, accessed May 16, 2016), http://www.nber.org/digest/sep03/w9873.html.

influence of bias, then they should advocate for transparency and explainability in all significant decision-making, as an algorithm's involvement in those decisions is irrelevant.

Second, calls for the right to meaningful information about certain algorithmic decisions, as the European Union's GDPR mandates, disregard the many laws that already exist guaranteeing a right to an explanation for certain high-impact decisions, such as the reasons behind a bank refusing to grant an applicant a loan, or why a company fired an employee.[22] Existing laws would still apply to these situations regardless of whether companies use an algorithm to make the decision. In application areas where laws already exist, new requirements specifically targeting algorithms would be redundant—although the GDPR extends this requirement to all algorithmic decisions with legal or significant consequences.[23] If there are certain decisions that warrant an explanation or meaningful information, then surely it should not matter whether an algorithm was involved. But if these decisions do indeed carry potential risks, the construction of this requirement allows for companies to use humans instead of algorithms to skirt the law.[24] If the GDPR's supporters believe such decisions warrant an explanation, then it is ineffective for the GDPR to only target decisions made by algorithms.

Proponents for algorithmic transparency often justify their stance by pointing to the potential for biased and flawed algorithms in the criminal justice system to cause substantial harm to individuals. As this paper discusses, transparency, as well as other components of algorithmic accountability such as error analysis and procedural regularity, will likely be key factors to ensure the beneficial use of algorithms wherever market forces are muted, such as with the criminal justice system. However, the value of transparency in the criminal justice context does not support the conclusion that algorithmic transparency would be necessary or beneficial in most contexts. As noted above, for most applications, operators have strong incentives to minimize flaws and potential harms. But for applications that lack these incentives, whether an operator uses algorithms is irrelevant. It is also important to bear in mind that even in the criminal justice system, algorithmic transparency would not address the root causes of many of the harms that such decisions can cause. For example, algorithmic transparency alone would not solve inherent

[22] Nick Wallace, "EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence," *TechZone360*, January 25, 2017, http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#.
[23] Ibid.
[24] Ibid.

bias problems, such as the large disparity in arrest rates for blacks and whites for marijuana possession, despite marijuana use being roughly equal among blacks and whites.[25]

Third, another major flaw with more extreme demands for transparency, such as EPIC's call for all source code to be made fully public, is that while "pulling back the curtain" to allow regulators and the public to scrutinize how an algorithm might be flawed may sound reasonable, it is unrealistic to expect that even the most technologically savvy, resource-flush regulators, advocacy groups, or concerned citizens would be capable of reliably gleaning meaningful information from scrutinizing advanced AI systems and their underlying data, particularly at scale. For example, after Reddit disclosed a portion of its ranking algorithm, a group of computer scientists led by Christian Sandvig at the University of Michigan noted that "even with complete transparency about a particular part of [this] algorithm, expert programmers have been sharply and publicly divided about what exactly that part of the algorithm does. This clearly implies that knowing the algorithm itself may not get us very far in detecting algorithmic misbehavior."[26] While examining code can provide meaningful information about how some algorithmic systems make decisions, for many advanced AI systems that rely on thousands of layers of simulated neurons to interpret data, even their developers cannot explain their decision-making. For example, researchers at Mount Sinai Hospital in New York developed an AI system called Deep Patient that can predict whether a patient is contracting any of a wide variety of diseases.[27] The researchers trained Deep Patient on the health data from 700,000 patients, including hundreds of variables, which allow it to predict disease without explicitly having to be taught how.[28] The system is substantially better than other disease-prediction methods, yet its own developers do not know how its decision-making process works.[29] Thus, there is little to reason to believe a third party would be able to understand it. As Curt Levey of the Committee for Justice and Ryan Hagemann of the Niskanen Center describe, "The machine's 'thought process' is not explicitly described in the weights, computer code, or anywhere else. Instead, it is subtly encoded in the interplay between the weights and the neural network's architecture. Transparency sounds nice,

[25] "Report: the War on Marijuana in Black and White," ACLU, Accessed May 8, 2018, https://www.aclu.org/report/report-war-marijuana-black-and-white?redirect=criminal-law-reform/war-marijuana-black-and-white.
[26] Christian Sandvig et al., "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" (paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association, Seattle, Washington, May 22, 2014).
[27] Will Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[28] Ibid.
[29] Ibid.

but it's not necessarily helpful, and may be harmful."[30] The United Kingdom's Government Office for Science cautions, "Most fundamentally, transparency may not provide the proof sought: Simply sharing static code provides no assurance it was actually used in a particular decision, or that it behaves in the wild in the way its programmers expect on a given dataset."[31]

Fourth, calls for algorithmic transparency and, sometimes, for algorithmic explainability discount the value of proprietary software. Requirements to publicly disclose source code or information about the inner workings of software would reduce incentives for a company to invest in developing algorithms, as competitors could simply copy them. While copyright laws could reduce this risk in countries with strong intellectual property protections like the United States, this would make it significantly easier for bad actors in countries that routinely flout intellectual property protections, such as China, to steal source code.[32] Ardent supporters of algorithmic transparency, such as Frank Pasquale, author of *The Black Box Society: The Secret Algorithms That Control Money and Information*, dismiss this concern out of hand, claiming the argument is just a nefarious smoke screen to cover for deliberate exploitation or abuse: "They [corporations] say they keep techniques strictly secret to preserve valuable intellectual property—but their darker motives are also obvious."[33] It is not clear what these darker motives are, other than to maximize profits. But again, in almost all cases where companies stand to lose from an algorithmic system making biased decisions, the company is highly motivated to make accurate decisions—unless Pasquale is arguing that accurate decisions, like denying a loan to someone who presents a bad credit risk is a reflection of a "darker" motive, it is hard to know what the problem is.

Fifth, requiring algorithmic transparency can also create opportunities for bad actors to "game the system" and take advantage of algorithm-driven platforms. For example, for years, Google relied on an algorithm called PageRank to determine the order of search results to display based on factors such as a website's meta tags and keywords.[34] However, because these factors were

---

[30] Curt Levey and Ryan Hagemann, "Algorithms With Minds of Their Own," *The Wall Street Journal*, November 12, 2017, https://www.wsj.com/articles/algorithms-with-minds-of-their-own-1510521093.
[31] Ibid.
[32] Patrick Gillespie, "China Broke Hacking Pact Before New Tariff Fight," *Axios*, April 10, 2018, https://www.axios.com/china-broke-hacking-pact-before-new-tariff-tiff-d19f5604-f9ce-458a-a50a-2f906c8f12ab.html.
[33] Ibid; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 10.
[34] John Faber, "How to Future-Proof Your Search Ranking," Chapter Three, April 2, 2018, https://www.chapterthree.com/blog/how-to-future-proof-your-search-ranking.

widely known, any site owner could manipulate the algorithm by populating a page with hidden content that PageRank interpreted as desirable in an effort to push their website higher in the search rankings and increase views, despite it being irrelevant to a user's query.[35] Now, Google uses a combination of multiple, complex algorithms, including machine learning systems, that weighs hundreds of factors to order search results based on content quality and relevance.[36] If Google or other search engines were required to disclose how their search algorithms work, it would once again allow websites to exploit these systems—and consumers would suffer for it.

Sixth, transparency would not solve some of the key challenges in the information economy. Some, such as German Chancellor Angela Merkel, argue, "Algorithms, when they are not transparent, can lead to a distortion of our perception," contributing to the formation of filter bubbles in online platforms (i.e., situations in which users are only exposed to content that conforms to their world view) and damages public discourse.[37] Similarly, German Federal Minister of Justice Katarina Barley believes such practices contribute to the proliferation of online disinformation campaigns.[38] Both Merkel and Barley claim that algorithmic transparency would alleviate these problems by enabling users to understand how their perspectives are being influenced. However, it would likely have the opposite effect, making it easier for bad actors to game these algorithms and flood platforms with low-quality or deliberately misleading content.

Seventh, mandating algorithmic explainability could severely limit the potential benefits of algorithms, as there can be inescapable trade-offs between the interpretability or explainability of an AI system and its accuracy. As data scientists Max Kuhn and Kjell Johnson put it in their book *Applied Predictive Modeling*, "Unfortunately, the predictive models that are most powerful are usually the least interpretable."[39] An algorithm's accuracy typically increases with its

[35] Ibid.
[36] Danny Sullivan, "Google Uses RankBrain for Every Search, Impacts Rankings of 'Lots' of Them," *Search Engine Land*, June 23, 2016, https://searchengineland.com/google-loves-rankbrain-uses-for-every-search-252526.
[37] Kate Connolly, "Angela Merkel: Internet Search Engines Are 'Distorting Perception,'" *The Guardian*, October 27, 2016, https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception.
[38] Adam Segal, "Germany Wants Greater Algorithmic Transparency to Fight Disinformation, But Its Approach is Half-Baked," Council on Foreign Relations, April 22, 2018, https://www.cfr.org/blog/germany-wants-greater-algorithmic-transparency-fight-disinformation-its-approach-half-baked.
[39] Max Kuhn and Kjell Johnson, *Applied Predictive Modeling* (New York: Springer-Verlag New York, 2013) 50.

complexity, but the more complex an algorithm is, the more difficult it is to explain.[40] While this could change in the future as research into explainable AI matures, at least in the short term, requirements for explainability would only be desirable in situations where it is appropriate to sacrifice accuracy—and these cases are rare. For example, it would not be desirable to prioritize explainability over accuracy in autonomous vehicles, as even slight reductions in navigation accuracy or to a vehicle's ability to differentiate between a pedestrian on the road and a picture of a person on a billboard could be enormously dangerous. Thus, a mandate for algorithmic explainability is essentially a mandate to use less-effective AI, or, in cases where sacrifices in accuracy are prohibitive, such as with self-driving cars, a ban on the use of effective but uninterpretable algorithms.

Finally, the most fundamental flaw in these proposals is that algorithmic transparency and explainability are a means for achieving the goal of preventing algorithms from causing harm, not an end themselves. Transparency and explainability can indeed be useful mechanisms for achieving this goal, but only in select contexts. It would be unwise for regulators to treat achieving algorithmic transparency or explainability as either a panacea or an end-goal. In most contexts, mandating transparency and explainability would limit innovation and fail to prevent potential harm.

Master Regulatory Bodies to Oversee All Algorithmic Decision-making

As concerns about the potential risks of algorithms proliferate, some have advocated for governments to create new regulatory bodies specifically devoted to overseeing algorithms. For example, University of Maryland computer science professor Ben Shneiderman, in a 2017 speech at the Alan Turing Institute, proposed the creation of a "National Algorithm Safety Board" to independently oversee the use of algorithms, such as by auditing, monitoring, and licensing algorithms when a company wants to deploy one.[41] Similarly, the Oxford Internet Institute calls for the creation of an "algorithmic oversight institution" with the powers to audit algorithms and determine whether they serve the public interest.[42] Attorney Andrew Tutt proposes the creation of the equivalent of the U.S. Food and Drug Administration for algorithms,

---

[40] Jason Brownlee, "Model Prediction Accuracy Versus Interpretation in Machine Learning," Machine Learning Mastery, August 1, 2014, https://machinelearningmastery.com/model-prediction-versus-interpretation-in-machine-learning/.
[41] "Turing Lecture: Algorithmic Accountability: Professor Ben Shneiderman, University of Maryland," The Alan Turing Institute, May 31, 2017, https://www.youtube.com/watch?v=UWuDgY8aHmU.
[42] "Written Evidence Submitted By the Oxford Internet Institute," Oxford Internet Institute, April 2017, http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/69003.pdf.

which would have the power to "prevent the introduction of algorithms into the market until their safety and efficacy has been proven through evidence-based premarket trials."[43] Entrepreneur Elon Musk, speaking at a 2017 meeting of the National Governors Association, urged policymakers to take a precautionary principled approach, arguing that "the right order of business would be to set up a regulatory agency [with the] initial goal: gain insight into the status of AI activity, make sure the situation is understood, and once it is, put regulations in place to ensure public safety."[44]

These and related proposals suffer from a number of serious challenges. First, they all fail to recognize that to adequately assess an algorithmic decision, one would need to have context-specific knowledge about the type of decisions an algorithm is dealing with. What constitutes harm in consumer finance involves dramatically different criteria than what constitutes harm in health care, which is why governments have different sector-specific regulatory bodies. If it would be ill-advised to have one government agency regulate all human decision-making, then it would be equally ill-advised to have one agency regulate all algorithmic decision-making. This is why during the growth of the Internet in the 1990s, the United States did not establish a federal Internet agency to regulate all online activity, as some proposed. Instead, the Federal Communications Commission regulated the telecommunications aspects of the technology, the FTC regulated online commerce, the National Telecommunications and Information Administration regulated spectrum, and so on.

It is unclear why advocates for these proposals believe existing regulatory bodies are incapable of scrutinizing algorithms effectively. It is important for regulators to understand the technology related to issues under their purview and, given the newness of this technology, it is likely that many agencies lack the technical expertise to understand how algorithmic decision-making works. For example, after Congress proposed the U.S. National Highway Transportation Safety Administration (NHTSA) be responsible for certifying the safety of autonomous vehicles, Mike Ramsay, an automotive technology analyst at Gartner Research, lamented that "there's no way

[43] Andrew Tutt, "An FDA for Algorithms," Administrative Law Review 69, no. 83 (March 15, 2016), http://dx.doi.org/10.2139/ssrn.27479944.
[44] "Elon Musk, National Governors Association, July 15, 2017," https://www.youtube.com/watch?time_continue=245&v=b3IzEQANdHk; Camila Domonoske, "Elon Musk Warns Governors: Artificial Intelligence Poses 'Existential Risk,'" *NPR*, July 17, 2017, https://www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk.

NHTSA has the technical capability to do this right now."[45] However, agencies often lag behind the private sector in their ability to understand new technologies, and have always had to deal with the issue of staying informed about new innovations. Moreover, some agencies, such as the FTC, actively cultivate and seek out technical expertise to allow them to effectively oversee complicated technology issues in a wide array of industries.[46] Regardless, if the concern is government agencies not having sufficient technical expertise, simply establishing a new regulatory agency devoted to algorithms would not fix this, as any difficulties governments face in attracting and retaining human capital would still apply.[47]

This does not mean Congress and other legislative bodies should not support agencies in developing the needed technical expertise to manage AI-related concerns. Stanford University's One Hundred Year Study on Artificial Intelligence, or AI100, led by a group of academics and AI experts, recommends that policymakers:

> Define a path toward accruing technical expertise in AI at all levels of government. Effective governance requires more experts who understand and can analyze the interactions between AI technologies, programmatic objectives, and overall societal values. ... Absent sufficient technical expertise to assess safety or other metrics, national or local officials may refuse to permit a potentially promising application. Or insufficiently trained officials may simply take the word of industry technologists and green light a sensitive application that has not been adequately vetted. Without an understanding of how AI systems interact with human behavior and societal values, officials will be poorly positioned to evaluate the impact of AI on programmatic objectives. ... Faced with the profound changes that AI technologies can produce, pressure for "more" and "tougher" regulation is inevitable. Misunderstanding about what AI is and is not, especially against a background of scare-mongering, could fuel opposition to technologies that could benefit everyone. This would be a tragic mistake.[48]

This is not to say regulatory regimes should never change as algorithmic decision-making proliferates and AI matures—although every regulator modernizes in tandem with its sector. The

---

[45] Alan Ohnsman, "Push for Self-Driving Car Rules Overlooks Lack of Federal Expertise in AI Tech," *Forbes*, July 18, 2017, https://www.forbes.com/sites/alanohnsman/2017/07/19/push-for-self-driving-car-rules-overlooks-lack-of-federal-expertise-in-ai-tech/#2fd44c7dcbf3.
[46] Neil Chilson, "How the FTC Keeps up on Technology," U.S. Federal Trade Commission, January 4, 2018, https://www.ftc.gov/news-events/blogs/techftc/2018/01/how-ftc-keeps-technology.
[47] "Strategic Human Capital Management," U.S. Government Accountability Office, 2017, https://www.gao.gov/highrisk/strategic_human_management/why_did_study.
[48] "Artificial Intelligence and Life in 2030" (Stanford University One Hundred Year Study on Artificial Intelligence, September 2016), https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf.

U.S. Federal Aviation Administration will likely operate substantially differently in 50 years if flying cars become commonplace, just as the European Medicines Agency will have to adapt should cancer treatments that rely on nanorobotics become the norm. But reworking a government's entire regulatory system in response to just a single technology would be a dramatic and likely ineffective measure.

Establishing a regulator to oversee the use of algorithms also implies that all algorithms pose the same level of risk and need for regulatory oversight. However, algorithms pose a wide variety of risk depending on their application. Low-risk decisions should not be subject to regulatory oversight simply because they use an algorithm.

Finally, some of these proposals focus on serving the "public interest," even though reasonable people can differ on what this should include. Use of personal vehicles, for example, could be considered in the public interest because they provide mobility and access to economic opportunity—although they also pollute the environment, create sprawl, and kill upwards of 30,000 people a year in the United States. Rather than have a regulator decide whether use of vehicles is in the public interest, government instead regulates specific activities based on their objective benefits and harms, such as the fuel economy and safety of vehicles and land use in cities. Similarly, policymakers should not decide whether the use of algorithms are in the public interest, but instead regulate specific uses of them.

Generalized Regulatory Proposals

The third category of proposals is a disparate group of likely well-intentioned recommendations that are vague and memorable, but ultimately just meaningless slogans and buzzwords that are unworkable from a regulatory perspective. To be sure, some of these proposals could have value in certain areas of algorithmic decision-making, such as in AI development guidelines or corporate social responsibility standards, but as guides for regulation they are impractical. There are countless examples of calls to action for regulating algorithms that stress the need to rethink current approaches but fail to articulate an effective path forward. Emblematic is a January 2018 speech from British Prime Minister Theresa May, who stated that while the potential of AI is fundamental to the advancement of humanity, "This technological progress also raises new and profound challenges which we need to address. … So today I am going to make the case for how we can best harness the huge potential of technology. But also how we address

these profound concerns."[49] However, the rest of the speech failed to actually propose any meaningful solutions to potential challenges posed by AI. As one critic writes, May's argument can be boiled down to "AI can do great things, but we must be sure it's safe and ethical," and that this is "vapid, a truism."[50]

In some cases, these narratives are presented as guiding principles to help policymaking rather than as bona fide policy proposals themselves. While they are designed to serve as a reference for future efforts, they often include recommendations that are either too vague to be useful or that would restrict beneficial uses of algorithms. For example, a March 2018 report from the European Group on Ethics in Science and New Technologies, which advises the European Commission, that called for the creation of a shared ethical and regulatory framework for AI concluded, "The principle of responsibility must be fundamental to AI research and application. 'Autonomous' systems should only be developed and used in ways that serve the global social and environmental good, as determined by the outcomes of deliberative democratic processes."[51] The report's attempt to explain this is even more vague, stating, "[Autonomous systems] should be designed so that their effects align with a plurality of fundamental human values and rights."[52] While this may sound innocuous, it is incredibly problematic. Would it be acceptable to deploy algorithms to deliberately facilitate discrimination in societies where the plurality of human values is to limit rights for women or religious minorities? This lack of specificity gives policymakers little to work with when it comes to crafting regulation. The problem with such proposals is that they do not specify who decides what values to endorse or how to reconcile trade-offs, such as between job loss and economic growth. Additionally, this approach could potentially prohibit the use of algorithms for the purpose of increasing productivity, with no impact on human rights.

More importantly, such assertions ignore that democratic societies have processes by which they adjudicate these conflicting interests and values: legislatures and courts of law. Yet some in the AI community seem to think they are self-appointed guardians of ethics, as they define it. For

---

[49] "PM's Speech at Davos 2018: 25 January," U.K. Prime Minister's Office, January 25, 2018, https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january.
[50] Rowland Manthorpe, "May's Davos Speech Exposed Emptiness in the UK's AI Strategy," *Wired*, January 28, 2018, http://www.wired.co.uk/article/theresa-may-davos-artificial-intelligence-centre-for-data-ethics-and-innovation.
[51] European Group on Ethics in Science and New Technologies, "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems" (Luxembourg: European Commission Directorate-General for Research and Innovation, 2018), http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.
[52] Ibid.

example, some have argued against autonomous weapons systems, arguing developers should only create algorithms for robots that augment—but do not replace—human workers. While the intent of these proposals is to save lives and jobs, decisions like these should ideally be made through democratic processes, not by a select group of individuals who may not reflect the broad diversity of society. Nation states, for instance, are best suited to determine what defense systems they need to protect themselves from adversaries. Furthermore, social and political preferences are normally not applied to technologies, but rather to specific sectors and industries. Because of these constraints, and the need to satisfy consumers, firms are better suited to determine how to maximize innovation.

Some of these proposals attempt to take a more productive approach but are still ultimately unworkable. For example, in May 2016, the White House published a report detailing the opportunities and challenges of big data and civil rights. But rather than focus on demonizing the complex and necessarily proprietary nature of algorithmic systems, it presented the concept of "equal opportunity by design," which it defined as the principle of ensuring fairness and safeguarding against discrimination throughout a data-driven system's entire lifespan.[53] This approach, described more generally by then Federal Trade Commissioner Terrell McSweeny as "responsibility by design," recognizes that algorithmic systems can produce unintended outcomes, and encourages developers to address the root problems that could cause harms in algorithmic systems, such as failing to account for historical bias.[54] Encouraging developers to be responsible in the creation and application of algorithms is a worthwhile goal, however merely stating developers should consider "responsibility by design" is not a clear solution to the challenges algorithms pose.[55] Furthermore, these approaches focus on the developer, not the operator. While developers could wholeheartedly embrace "responsibility by design," it would have little impact if their algorithms were not viable products. Rather, if operators were exposed to regulatory incentives to deploy algorithms responsibly, the market would respond to this demand much more efficiently.

---

[53] "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" (Washington, D.C.: Executive Office of the President, May 2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

[54] "Keynote Remarks of Commissioner Terrell McSweeny," U.S. Federal Trade Commission, September 10, 2015, https://www.ftc.gov/system/files/documents/public_statements/800981/150909googletechroundtable.pdf.

[55] Daniel Castro, "How Congress Can Fix 'Internet of Things' Security," *The Hill*, October 18, 2016, http://thehill.com/blogs/pundits-blog/technology/303302-how-congress-can-fix-internet-of-things-security.

Doing Nothing Is Not the Answer

Overzealous regulation of technology, inspired by fears of worst-case scenarios or beliefs that a select group of AI developers are the only ones that can know and protect consumer interests will clearly harm innovation. Speculative fears about the potential risks of new technology are a powerful driver of advocacy efforts to restrict how a technology can be used before it matures to the point where society can fully realize its benefits and understand its impacts. Fears that preempt the proliferation of disruptive technologies have spurred regulatory proposals that seem ridiculous in retrospect after the technology becomes commonplace.[56] For example, as transistors proliferated in the 1950s and 1960s, some U.S. policymakers were so concerned about their potential to be used for surveillance that one senator proposed a law that would have required all bugging equipment to be licensed by the government.[57] Had Congress succumbed to such hysterical concerns and passed that bill, many innocuous technologies that are widely enjoyed today, such as smartphones and baby monitors, would have been greatly impeded. Thus, it is wise to be skeptical of advocates rushing to regulate new technologies due to concerns about their hypothetical harms before it is clear how market forces, technological advancement, and existing regulations would shape their use as they mature. However, with algorithmic decision-making, dismissing any and all efforts to improve governance would be problematic.

While explicit calls for the government to not regulate any algorithms and leave it entirely to industry to self-regulate are few and far between, some do advocate for it. For example, technology reporter Tristan Greene, writing for *The Next Web*, concluded that due to the speculative nature of many of the fears about AI, the "government is clueless about AI and shouldn't be allowed to regulate it."[58] Mouloud Dey, director of innovation and business solutions at SAS France, argues that governments should not step in to regulate algorithms because of the burden regulations could have on innovation—and that industry self-regulation would be adequate to address any potential harms.[59] In many cases however, more general anti-regulation attitudes could still lend credence to the notion that the government should not regulate

---

[56] Daniel Castro and Alan McQuinn, "The Privacy Panic Cycle: A Guide to Public Fears About New Technologies" (Information Technology and Innovation Foundation, September 2015), http://www2.itif.org/2015-privacy-panic.pdf.
[57] Ibid; John Neary, "The Big Snoop: Electronic Snooping—Insidious Invasions of Privacy," *Life Magazine*, May 20, 1966, http://www.bugsweeps.com/info/life_article.html.
[58] Tristan Greene, "U.S. Government Is Clueless About AI and Shouldn't Be Allowed to Regulate It," *The Next Web*, October 24, 2017, https://thenextweb.com/artificial-intelligence/2017/10/24/us-government-is-clueless-about-ai-and-shouldnt-be-allowed-to-regulate-it/.
[59] Daniel Saraga, "Opinion: Should Algorithms Be Regulated?," Phys.org, January 3, 2017, https://phys.org/news/2017-01-opinion-algorithms.html.

algorithms at all, by overshadowing legitimate efforts to regulate the technology in an evenhanded, beneficial way. For example, Simon Constable, a fellow at the Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise, writing in *Forbes*, erroneously concluded that due to the U.S. government's failure to prevent or mitigate the 2008 financial crisis, "It's time to just say no to calls for more government regulation of the tech industry."[60]

Given the steps some governments, such as the EU's GDPR, have already taken that will clearly limit innovation, it is easy to be sympathetic to such positions. But while industry self-regulation, market forces, and tort law will likely play a large role in positively shaping the use of algorithms, there are reasons why these alone would be insufficient to protect against all potential harms of algorithmic decision-making, which likely fall into one of three categories. First, there are some potential applications of algorithms where traditional market forces that could mitigate the harms of algorithms, such as the threat of reputational damage if a company's algorithm causes harm, are diminished, making the cost of this flawed decision-making one-sided. This is particularly true with government uses of AI wherein the costs of bad decisions are indeed problematic, but not borne directly by the government agency using the algorithm. In other words, even though a discriminatory algorithm is an inferior product, there are some situations where this would not deter an operator from deploying it. Second, there are applications of algorithmic decision-making where even though incentives to minimize harms exist, the potential harms could be significant enough to warrant regulation, such as is with autonomous vehicles. And third, certain applications of algorithms could cause harms, such as exacerbating inequality, but without an operator expressly or obviously breaking the law. For example, an online jobs board could utilize a targeted advertising algorithm that does not consider race but nonetheless uses variables that inadvertently serve as proxies for race, such as zip code, thereby favoring members of a certain race for job opportunities. This harm may not be immediately obvious to the public, regulators, or even the operator. In such cases, absent public outrage, businesses have reduced incentive to scrutinize their algorithms thoroughly to prevent this harm, as there is not a strong profit motive to do so.

Are there ethical concerns raised by these technologies that are not also raised by traditional computer programming techniques or by human decision-making? Are the concerns raised by

[60] Simon Constable, "Why We Should Not Regulate the Tech Industry," *Forbes*, March 26, 2018, https://www.forbes.com/sites/simonconstable/2018/03/26/no-we-really-dont-need-government-regulation-of-the-tech-industry/#2e7ad53deb8d.

these technologies greater or less than those of traditional computer programming or human decision-making? Why or why not?

Algorithmic decision-making poses the same kinds risks as human decision-making, however these risks are modified somewhat due to the complexity and scale of algorithmic decision-making. Algorithms can perform increasingly complex tasks to help solve newer and bigger challenges in the public and private sectors far more efficiently—and sometimes more effectively—than humans. However, this unprecedented complexity and scalability has also led to fears of algorithms potentially creating substantial risks that existing laws may not be able to effectively address.

The most common criticism of algorithmic decision-making is that it is a "black box" of extraordinarily complex underlying decision models involving millions of data points and thousands of lines of code. Moreover, the model can change over time, particularly when using machine learning algorithms that adjust the model as the algorithm encounters new data. Further complicating things, in many cases, developers lack the ability to precisely explain how their algorithms make decisions, and instead can only express the degree of confidence they have in the accuracy of the algorithms' decisions.[61] The difficulty arises from the fact that while developers or operators can control what data goes into their systems, and instruct algorithms how to weigh different variables, it can be challenging, if not impossible, to program their systems to explain or justify their decisions.[62] As a result, many have labeled these algorithms as impenetrable black boxes that defy scrutiny.[63]

Complexity can be problematic for several reasons. First and foremost, it creates opportunities for bias to inadvertently influence algorithms in a number of different ways. The data algorithms train on can be flawed, such as reflecting historical biases or being incomplete, which developers or operators could fail to account for.[64] For example, if a university inadvertently denies admissions to a particular demographic at an unfair rate relative to other demographics, and then trains an algorithm to make admissions decisions based on historical admissions data, the

---

[61] Will Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[62] Cliff Kuang, "Can A.I. Be Taught to Explain Itself?" *The New York Times*, November 21, 2017, https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html.
[63] "Humans May Not Always Grasp Why AIs Act. Don't Panic." *The Economist*, February 15, 2018, https://www.economist.com/news/leaders/21737033-humans-are-inscrutable-too-existing-rules-and-regulations-can-apply-artificial?frsc=dg%7Ce.
[64] "Algorithmic Accountability" (World Wide Web Foundation, July 2017), http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf.

algorithm could interpret this bias as a relevant decision-making parameter. Similarly, if developers were to train a facial-recognition algorithm on a dataset that consists primarily of images of white men's faces, it may not be able to accurately recognize faces of black women.[65] Additionally, algorithmic systems could be subject to feedback loops that perpetuate and amplify biases over time.[66] For example, consider a court system that routinely sentences blacks more harshly than whites for the same crime.[67] If that court were to implement a decision support system for sentencing that used machine learning and historical sentencing data to inform judges' decisions, that system could recommend harsher sentences for blacks based on the examples it learned from. Over time, this could serve as confirmation for a judge's unconscious bias and thus exaggerate sentencing disparities along racial lines—which can lead to increased recidivism rates and subject more blacks to harsher sentences.[68] Compounding all of this, the lack of diversity in the developer community creates the risk of homogenous developer teams failing to consider how their own unconscious biases may influence their work, such as not recognizing their training data as not being representative.[69] It should be clear, however, that in almost all of these cases the outcomes are avoidable, as developers can account for these risks and control for bias in their algorithms.

In other cases, the complexity of algorithms causes some to fear that corporations or governments could hide behind their algorithm and use algorithmic decision-making as a cover to deliberately exploit, discriminate, or otherwise act unethically.[70] For example, in her book *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*, author Virginia Eubanks describes how policymakers in Indiana decided to implement an automated system for determining welfare eligibility.[71] While the stated goal of the switch was to increase efficiency

---

[65] Clare Garvie and Jonathan Frankle, "Facial-Recognition Software Might Have a Racial Bias Problem," *The Atlantic*, April 7, 2016, https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/.
[66] Ibid.
[67] Editorial Board, "Unequal Sentences for Blacks and Whites," *The New York Times*, December 17, 2016, https://www.nytimes.com/2016/12/17/opinion/sunday/unequal-sentences-for-blacks-and-whites.html.
[68] Dylan Matthews, "Making Prison Worse Doesn't Reduce Crime. It Increases It." *The Washington Post*, August 14, 2012, https://www.washingtonpost.com/news/wonk/wp/2012/08/24/making-prison-worse-doesnt-reduce-crime-it-increases-it/?utm_term=.b5e14537f123.
[69] Kate Crawford, "Artificial Intelligence's White Guy Problem," *The New York Times*, June 15, 2016, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.
[70] Robert D. Atkinson, "'It's Going to Kill Us!' and Other Myths About the Future of Artificial Intelligence" (Information Technology and Innovation Foundation, June 2016), http://www2.itif.org/2016-myths-machine-learning.pdf.
[71] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).

and combat fraud, the lack of evidence regarding substantial amounts of fraud in the original system, combined with the dramatic increase in erroneous benefits denials after transitioning to the automated system, led Eubanks to conclude the system had been deliberately designed to covertly cut welfare spending without the need to change policy.[72] Eubanks fears that the public sector could exploit the use of algorithms to "avoid some of the most pressing moral and political challenges of our time—specifically poverty and racism."[73] Some also worry that algorithms could be used as covers for negligence. For example, a 2017 ProPublica investigation revealed that Facebook's advertising algorithm could allow advertisers to target anti-Semitic users by automatically generating categories of users to target for ads based on topics the users liked, which included "Jew hater" and "History of 'why jews ruin the world.'"[74] Dave Lee of the *BBC* contends Facebook tried to deflect responsibility for this by faulting their algorithm, rather than owning up to a lack of oversight—although Lee offers no evidence of Facebook intentionally trying to court anti-Semitic advertisers.[75] What is more likely, however, is Facebook's system automatically pulled data about users' likes, which in some select cases included bigoted views. Requiring the rollout of every new technology to go perfectly would doom most of it to the scrap heap of history. All new technologies improve over time; as society interacts with them and identifies problems, developers improve the technology. Granted, this does not necessarily prevent organizations from denying responsibility for any misuse of their algorithm. But requiring an error rate of zero would considerably stifle innovation.

Another aspect of algorithmic decision-making that poses a challenge is its capacity to make a large number of decisions significantly faster than humans. As the public and private sectors increasingly rely on algorithms in high-impact sectors such as consumer finance and criminal justice, a flawed algorithm could potentially cause harm at higher rates. As existing legal oversight may not be sufficient to respond quickly or effectively enough to mitigate this risk, it is clear why increased risk warrants greater regulatory scrutiny.

---

[72] Ibid.
[73] Alyssa Edes and Emma Bowman, "'Automating Inequality': Algorithms in Public Services Often Fail the Most Vulnerable," *NPR*, February 19, 2018, https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab.
[74] Julia Angwin, Madeleine Varner, and Ariana Tobin, "Facebook Enabled Advertisers to Reach 'Jew Haters'," *ProPublica*, https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters.
[75] Dave Lee, "Facebook Can't Hide Behind Algorithms," *BBC*, September 22, 2017, http://www.bbc.com/news/technology-41358078.

By automating human-led processes, such as determining loan eligibility, banks could use algorithms to dramatically shorten the time it takes to evaluate applicants while reducing operating costs, and then pass those savings on to borrowers in the form of lower interest rates. However, if these algorithms are flawed, the sheer volume of their decisions could end up significantly amplifying the potential negative impact of these flaws. Compared with a single human, whose output is only a handful of loan applications per week, routinely making errors while evaluating loan applications, a flawed algorithm misevaluating hundreds of loan applications per week across an entire bank branch would clearly cause harm at a much larger scale.

In most cases, flawed algorithms hurt the organization using them. Therefore, organizations have strong incentives to not use biased or otherwise flawed algorithmic decision-making and regulators are unlikely to need to intervene. For example, banks making loans would be motivated to ensure their algorithms are not biased because, by definition, errors such as granting a loan to someone who should not receive one, or not granting a loan to someone who is qualified, costs banks money. But in other cases, where the cost of the error falls largely on the subject of the algorithmic decision, these incentives may not exist. Biased algorithms in parole decision systems, for instance, hurt individuals who are unfairly denied parole, but impose little cost on the court system. In such cases, existing legal frameworks may not be sufficiently equipped to respond quickly or effectively to mitigate this risk.

Of course, if an organization has a flawed process for human decisions, the impact could also be significant—such as when banks changed their lending practices to extend credit to borrowers who had little or no documentation of income contributing to the 2008 financial crisis.[76]

Is industry self-regulation and government enforcement of existing laws sufficient to address concerns, or are new laws or regulations necessary?

Self-regulation, provided operators adhere to the principle of algorithmic accountability, and the enforcement of existing laws will be sufficient to address most of the concerns posed by algorithms. However, in select cases where market forces, which encourage operators to ensure they adhere to algorithmic accountability, are muted and significant harm is possible, it may be appropriate for policymakers to dictate specific requirements for algorithmic accountability. This

[76] Martin Baily, Robert Litan, and Matthew Johnson, "The Origins of the Financial Crisis" (Brookings Institution, November 2008), https://www.brookings.edu/wp-content/uploads/2016/06/11_origins_crisis_baily_litan.pdf.

is particularly relevant in the criminal justice system. Caleb Watney, a technology policy fellow at the R Street Institute, argues that because the concept of transparency is central to the goals of the justice system, as indicated by countless court precedents and statutory obligations, such as the Freedom of Information Act and other "sunshine" laws, it would be appropriate to mandate all algorithms that influence judicial decision-making be open-source.[77] Though this transparency may not shed much light on how more-advanced machine learning systems work, there is likely a compelling public interest in ensuring these algorithms are nonetheless exposed to the highest degree of scrutiny possible. Similarly, it would likely be appropriate for policymakers to mandate that public agencies conduct thorough impact assessments for algorithms they intend to use in decisions with high social or economic consequences, such as the administration of entitlement programs.[78] However, any such rules should be narrow and targeted to identifiable harms that algorithmic decision-making could cause in a specific context.

CONSUMER PROTECTION ISSUES RELATED TO ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS

What choices and notice should consumers have regarding the use of these technologies?

Creating separate notice for the use of algorithms or AI sets a higher bar for the use of this technology that would discourage its use. These problems can already be seen in the GDPR. Article 22 of the GDPR confers a right for individuals "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her, or similarly significantly affects him or her."[79] In other words, if the decision is necessary to complete a contract with the customer, or if the customer has given consent to the controller to make such a decision, then the GDPR requires data controllers to give the customer "at least the right to obtain human intervention on the part of the controller."[80] That means wherever use of AI has legal or similarly significant effects—such as in deciding whether to offer a loan—the data subject has the right to have a human review that decision.

---

[77] Caleb Watney, "When it Comes to Criminal Justice AI, We Need Transparency and Accountability," R Street Institute, December 1, 2017, http://www.rstreet.org/2017/12/01/when-it-comes-to-criminal-justice-ai-we-need-transparency-and-accountability/.
[78] Ibid.
[79] Regulation 2016/679 (General Data Protection Regulation), Article 22, (see page L 119/46), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.
[80] Regulation 2016/679 (General Data Protection Regulation), Article 22(3), (see page L 119/46), accessed December19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

Having humans review an algorithmic decision is costly—and only more so as the complexity of the algorithm increases. This is because there is a trade-off between the representational capacity of a model and the ease with which a human can review the calculations it makes.[81] In other words, the more sophisticated the algorithmic model, the more time and expertise (and, therefore, resources) is needed for a human to make sense of the model's decisions. Indeed, the main point of artificial intelligence is to processes large quantities of data much more efficiently and accurately than humans—if it were no costlier for humans to repeat these calculations, there would be little economic incentive to use AI. The right to human review is essentially a tax on AI systems that are capable of making calculations that would be impractical for humans.

Faced with the cost of paying a qualified human to review an algorithmic decision and all the data it involves, companies will respond in one of two ways, depending on how critical AI is to their business. They will either limit the quantity and complexity of the data and the sophistication of the algorithm in order to minimize the cost of compliance, or simply forgo the use of AI altogether. Either outcome is ultimately bad for European competitiveness, as it ends up limiting the valuable contribution AI makes to European industry in improved efficiency.

Relatedly, Articles 13–15 of the GDPR confer a right to receive "meaningful information about the logic involved" in an algorithmic decision covered by Article 22—that is, one with legal or similarly significant effects. The phrase appears once in each of the three articles: Article 13 concerns personal data obtained from the subject, Article 14 addresses data obtained by other means, and Article 15 deals with data subjects' right to know whether somebody is processing their information, and if so, how.[82] Some scholars, such as Bryce Goodman and Seth Flaxman of Oxford University, argue that the right to "meaningful information" amounts to a "right to explanation" of algorithmic decisions, citing Recital 71's (a recital is a non-binding paragraph intended to help judges interpret the law) assertion that a data subject should have the right to "obtain an explanation of the decision" and challenge it after a human has reviewed it.[83]

---

[81] Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813.
[82] Regulation 2016/679 (General Data Protection Regulation), Article 13-14, (see page L 119/40-42), accessed December19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.
[83] Andrew D. Selbst and Julia Powles, "Meaningful Information and the Right to Explanation," *International Privacy Law,* Volume 7, Issue 4, November 1, 2017, pages 233-242, accessed January 30, 2018, https://doi.org/10.1093/idpl/ipx022; Ibid.

The GDPR provides relatively little clarification as to how it defines "meaningful information," which could lead to legal battles over the extent of the right. The articles themselves do not specify whether "meaningful information about the logic involved" refers to an explanation of how a particular algorithm generally reached decisions, or to a precise explanation of exactly how the algorithm arrived at a particular conclusion. Recital 71 seems to imply the latter when it says the data subject should be able to "obtain an explanation of the decision," but it does not specify what information would constitute an explanation or whether information about the "logic involved" should pertain to the algorithm or to the decision. Were a regulator or court to interpret the law to mean data controllers must be able to explain precisely how each individual decision was reached, it would severely inhibit AI because there is a trade-off between an algorithm's sophistication and its explicability—arising from the fact that AI systems are designed to carry out data processing tasks that would be more difficult or time-consuming for a human.[84]

OTHER POLICY QUESTIONS

What responsibility does a company utilizing these technologies bear for consumer injury arising from its use of these technologies? Can current laws and regulations address such injuries? Why or why not?

Some people are concerned that algorithmic decision-making will result in racial bias, such as financial institutions denying loans on the basis of race. However, in many cases, because flawed algorithms hurt the company using them, businesses have strong incentives to not use biased algorithms and regulators are unlikely to need to intervene. For example, banks making loans would be motivated to ensure their algorithms are not biased because, by definition, errors such as granting a loan to someone who should not receive one, or not granting a loan to someone who is qualified, costs banks money. In addition, even if some companies do not have a financial incentive to avoid biased algorithms, existing laws that prohibit such discrimination, such as the Fair Credit Reporting Act and the Equal Credit Opportunity Act, still apply.

Another argument regarding the inadequacy of privacy laws to protect consumer welfare is that the collection of large amounts of data allows companies to discriminate against consumers, including

[84] Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813; Innocent Kamwa, S. R. Samantary, Geza Jobs, "On the Accuracy Versus Transparency Trade-off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transactions on Smart Grid,* Volume 3, Issue 1, March 2012, accessed February 1, 2018, http://ieeexplore.ieee.org/abstract/document/6096427/; U Johannson, U Norinder, H Boström, "Trade-Off Between Accuracy and Interoperability for Predictive In-Silico Modelling," *Future Med Chem,* April 2011, 3(6):647-663, accessed February 1, 2018, https://www.ncbi.nlm.nih.gov/pubmed/21554073.

engaging in price discrimination, charging different consumers different prices depending upon the likelihood that they will buy a product.[85] Indeed, there is some evidence that companies are getting quite good at doing this.[86] This is often combined with the worry that disadvantaged groups will end up paying higher prices. But there are two reasons why price discrimination might not be a bad thing. First, to the extent that a platform has market power and can only set one price, its incentive is to raise prices on everyone and decrease supply. This allows the company to capture more value from the product and lowers the total benefit to society. If the company can charge different prices to different users, this social loss is reduced. Some consumers might still pay higher prices, but buyers will not purchase a product unless it makes them better off. Second, the ability to charge different prices is not limited to raising prices. Companies also have an incentive to lower prices for consumers who are reluctant to purchase the good.[87] This effect might actually be progressive. The company will charge a higher price to those users whose demand is inelastic. To the extent that lower-income consumers are more price responsive, they will benefit from price discrimination.[88]

More broadly, the FTC should recognize that consumers as a whole are going to benefit from greater use of algorithms, particularly artificial intelligence (AI). Though there are concerns about the potential harms that could arise from the use of AI, such as AI exacerbating unconscious human bias, the proposals that have gained popularity among consumer advocates to address these harms would be at best largely ineffective and at worst cause more harm than good. The two most popular ideas—requiring companies to disclose the source code to their algorithms and explain how they make decisions—would cause more harm than good by regulating the business models and the inner workings of the algorithms of companies using AI, rather than holding these companies accountable for outcomes.

The first idea— "algorithmic transparency"—would require companies to disclose the source code and data used in their AI systems. Beyond its simplicity, this idea lacks any real merits as a wide-scale solution. Many AI systems are too complex to fully understand by looking at source code alone. Some AI systems rely on millions of data points and thousands of lines of code, and decision models

---

[85] Nathan Newman, "The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google," William Mitchell Law Review 40, no. 2 (2014), http://open.mitchellhamline.edu/cgi/viewcontent.cgi?article=1568&context=wmlr.
[86] Burton G. Malkiel, "The Invisible Digital Hand," The Wall Street Journal, updated November 28, 2016, http://www.wsj.com/articles/the-invisible-digital-hand-1479168252
[87] Manne and Sperry, "The Problems and Perils of Bootstrapping Privacy and Data Into an Antitrust Framework," 7. "It is inconsistent with basic economic logic to suggest that a business relying on metrics would want to serve only those who can pay more by charging them a lower price, while charging those who cannot afford it a larger one."
[88] The White House, Big Data and Differential Pricing (Washington, DC: The White House, February 2015), 17, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf.

can change over time as they encounter new data. It is unrealistic to expect even the most motivated, resource-flush regulators or concerned citizens to be able to spot all potential malfeasance when that system's developers may be unable to do so either.[89]

Additionally, not all companies have an open-source business model. Requiring them to disclose their source code reduces their incentive to invest in developing new algorithms, because it invites competitors to copy them. Bad actors in China, which is fiercely competing with the United States for AI dominance but routinely flouts intellectual property rights, would likely use transparency requirements to steal source code.[90]

The other idea—"algorithmic explainability"—would require companies to explain to consumers how their algorithms make decisions. The problem with this proposal is that there is often an inescapable trade-off between explainability and accuracy in AI systems. An algorithm's accuracy typically scales with its complexity, so the more complex an algorithm is, the more difficult it is to explain. While this could change in the future as research into explainable AI matures—DARPA devoted $75 million in 2017 to this problem—for now, requirements for explainability would come at the cost of accuracy.[91] This is enormously dangerous. With autonomous vehicles, for example, is it more important to be able to explain an accident or avoid one? The cases where explanations are more important than accuracy are rare.

Fortunately, regulators have an alternative to these flawed approaches. Instead of pursuing heavy-handed regulations or ignoring these risks, they should adopt the tried-and-true approach of emphasizing light-touch regulation, with tailored rules for certain regulated sectors that fosters the growth of the algorithmic economy while minimizing potential harms. The challenge for regulators stems from the fact that innovation, by its very nature, involves risks and mistakes—the very things regulators inherently want to avoid. Yet, from a societal perspective, there is a significant difference between mistakes that harm consumers due to maleficence, negligence, willful neglect, or ineptitude on the part of the company, and those that harm consumers as a result of a company striving to innovate and benefit society. Likewise, there should be a distinction between a company's actions that violate regulations and cause significant harm to consumers or competitors, and those that

[89] Will Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[90] Joe Uchill, "China Broke Hacking Pact Before New Tariff Fight," *Axios*, April 10, 2018, https://www.axios.com/china-broke-hacking-pact-before-new-tariff-tiff-d19f5604-f9ce-458a-a50a-2f906c8f12ab.html.
[91] Cliff Kuang, "Can A.I. Be Taught to Explain Itself?," *New York Times Magazine*, November 21, 2018, https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html.

cause little or no harm. If regulators apply the same kind of blanket penalties regardless of intent or harm, the result will be less innovation.[92]

To achieve a balance, regulators should take a harms-based approach to protecting individuals, using a sliding scale of enforcement actions against companies that cause harm through their use of algorithms, with unintentional and harmless actions eliciting little or no penalty while intentional and harmful actions are punished more severely. Regulators should focus their oversight on operators, the parties responsible for deploying algorithms, rather than developers, because operators make the most important decisions about how their algorithms impact society.

This oversight should be built around algorithmic accountability—the principle that an algorithmic system should employ a variety of controls to ensure the operator can verify algorithms work in accordance with its intentions and identify and rectify harmful outcomes. When an algorithm causes harm, regulators should use the principle of algorithmic accountability to evaluate whether the operator can demonstrate that, in deploying the algorithm, the operator was not acting with intent to harm or with negligence, and to determine if an operator acted responsibly in its efforts to minimize harms from the use of its algorithm. This assessment should guide their determination of whether, and to what degree, the algorithm's operator should be sanctioned. Defining algorithmic accountability in this way also gives operators an incentive to protect consumers from harm and the flexibility to manage their regulatory risk exposure without hampering their ability to innovate.

This approach would effectively guard against algorithms producing harmful outcomes, without subjecting the public- and private-sector organizations that use the algorithms to overly burdensome regulations that limit the benefits algorithms can offer.

Sincerely,

Daniel Castro
Director, Center for Data Innovation
dcastro@datainnovation.org

Joshua New
Senior Policy Analyst, Center for Data Innovation
jnew@datainnovation.org

[92] Daniel Castro and Alan McQuinn, "How and When Regulators Should Intervene," (Information Technology and Innovation Foundation, February 2016), http://www2.itif.org/2015-how-when-regulators-intervene.pdf.