Information Society Project
Yale Law School

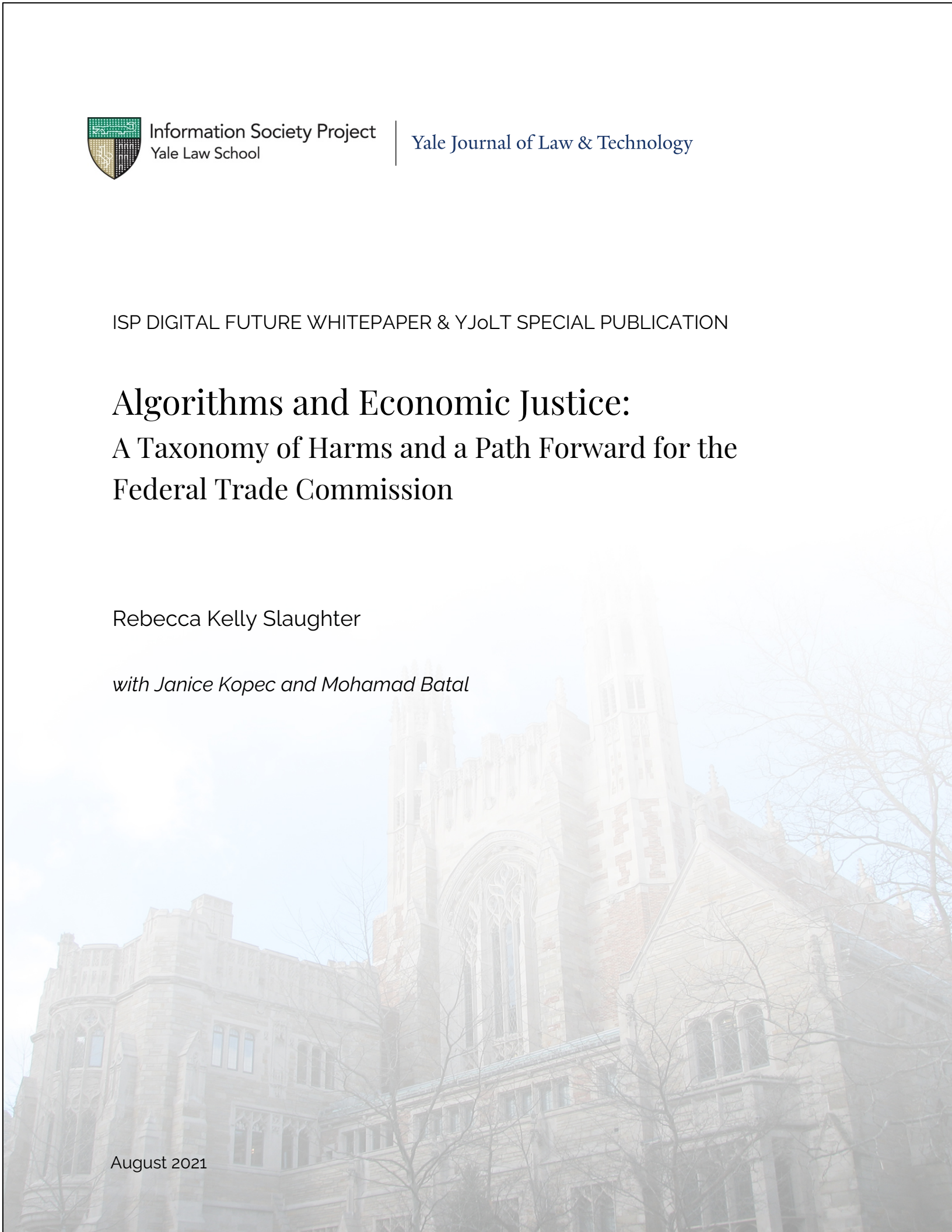
Yale Journal of Law & Technology

ISP DIGITAL FUTURE WHITEPAPER & YJoLT SPECIAL PUBLICATION

Algorithms and Economic Justice:
A Taxonomy of Harms and a Path Forward for the Federal Trade Commission

Rebecca Kelly Slaughter

with Janice Kopec and Mohamad Batal



August 2021

Contents

Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission

The proliferation of artificial intelligence and algorithmic decision-making has helped shape myriad aspects of our society: from facial recognition to deepfake technology to criminal justice and health care, their applications are seemingly endless. Across these contexts, the story of applied algorithmic decision-making is one of both promise and peril. Given the novelty, scale, and opacity involved in many applications of these technologies, the stakes are often incredibly high.

As an FTC Commissioner, I aim to promote economic and social justice through consumer protection and competition law and policy. In recent years, algorithmic decision-making has produced biased, discriminatory, and otherwise problematic outcomes in some of the most important areas of the American economy. This article describes harms caused by algorithmic decision-making in the high-stakes spheres of employment, credit, health care, and housing, which profoundly shape the lives of individuals. These harms are often felt most acutely by historically disadvantaged populations, especially Black Americans and other communities of color. And while many of the harms I describe are not entirely novel, AI and algorithms are especially dangerous because they can simultaneously obscure problems and amplify them—all while giving the false impression that these problems do not or could not possibly exist.

This article offers three primary contributions to the existing literature. First, it provides a baseline taxonomy of algorithmic harms that portend injustice, describing both the harms themselves and the technical mechanisms that drive those harms. Second, it describes my view of how the FTC's existing tools—including section 5 of the FTC Act, the Equal Credit Opportunity Act, the Fair Credit Reporting Act, the Children's Online Privacy Protection Act, and market studies under section 6(b) of the FTC Act—can and should be aggressively applied to thwart injustice. And finally, it explores how new legislation or an FTC rulemaking under section 18 of the FTC Act could help structurally address the harms generated by algorithmic decision-making.

I. Introduction

The proliferation of artificial intelligence and algorithmic decision-making[1] in recent years has shaped myriad aspects of our society. The applications of these technologies are innumerable, from facial recognition to deepfake technology, criminal justice, and health care. Across these contexts, the story of algorithmic decision-making is one of both promise and peril. Given the novelty, scale, and opacity involved, the stakes are high for consumers, innovators, and regulators.

Algorithmic decision-making, and the AI that fuels it, could realize its promise of promoting economic justice by distributing opportunities more broadly, resources more efficiently, and benefits more effectively. Pairing dramatically deeper pools of data with rapidly advancing machine-learning technology might yield substantial benefits for consumers, including by potentially mitigating the pervasive biases that infect human decision-making.[2] When used appropriately and judiciously, algorithms have also

[1] Throughout this article, I use several related, but distinct terms, including algorithms, artificial intelligence (AI), machine learning, deep learning, and neural networks. Each of these terms is a component of the term that comes before it. Commentators have often used the image of Russian nesting (Matryoshka) dolls to illustrate these relationships: An algorithm is defined as a finite series of well-defined, computer-implementable instructions—algorithms are the outermost doll, because while all AI uses algorithms, not all algorithms use AI. Next is AI, which includes machine learning, and machine learning, in turn, includes deep learning; finally, neural networks make up the backbone of deep learning. *See, e.g.*, *The Definitive Glossary of Higher Mathematical Jargon*, MATH VAULT (last accessed Mar. 4, 2021), https://mathvault.ca/math-glossary/#algo; Eda Kavlakoglu, *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?*, IBM BLOG (May 27, 2020), https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

[2] *See, e.g.*, Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Algorithms as Discrimination Detectors*, 117 PROC. OF THE NAT'L ACAD. SCI. (Dec. 1, 2020), https://www.pnas.org/content/pnas/117/48/30096.full.pdf.

transformed access to educational opportunities[3] and improved health outcomes through improved diagnostic rates and care adjustments.[4]

But the potentially transformative power of algorithmic decision-making also risks serious harm if misused. In the criminal justice system, for example, commentators note that algorithms and AI contribute to over-surveillance,[5] wrongful detainment and arrest,[6] and biased risk assessments used to determine pre-trial status and even sentencing.[7] Mounting evidence reveals that algorithmic decisions can produce biased, discriminatory, and unfair outcomes in a variety of high-stakes economic spheres including employment, credit, health care, and housing.[8]

---

[3] *See, e.g.*, Matt Kasman & Jon Valant, *The Opportunities and Risks of K-12 Student Placement Algorithms*, BROOKINGS INST. (Feb. 28, 2019), https://www.brookings.edu/research/the-opportunities-and-risks-of-k-12-student-placement-algorithms/.

[4] *See, e.g.*, Cade Metz, *London A.I. Lab Claims Breakthrough That Could Accelerate Drug Discovery*, N.Y. TIMES (Nov. 30, 2020), https://www.nytimes.com/2020/11/30/technology/deepmind-ai-protein-folding.html; Irene Dankwa-Mullan, et al., *Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here*, 22 POPULAR HEALTH MGMT. 229, 240 (2019).

[5] *See, e.g.*, Alvaro Bedoya, *The Color of Surveillance*, SLATE (Jan. 18, 2016), https://slate.com/technology/2016/01/what-the-fbis-surveillance-of-martin-luther-king-says-about-modern-spying.html; Amy Cyphert, *Tinker-ing with Machine Learning: The Legality and Consequences of Online Surveillance of Students*, 20 NEV. L. J. 457 (May 2020); Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, GEO. L. CTR. PRIVACY & TECH. (Oct. 18, 2016), https://www.perpetuallineup.org.

[6] *See, e.g.*, Kashmir Hill, *Another Arrest and Jail Time, Due to a Bad Facial Recognition Match*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.

[7] *See, e.g.*, *Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools*, ELEC. PRIVACY INFO. CTR., https://epic.org/algorithmic-transparency/crim-justice (last visited Jan. 17, 2020); Jason Tashea, *Courts Are Using AI to Sentence Criminals. That Must Stop Now*, WIRED (Apr. 17, 2017), https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/.

[8] *See infra* Section II.

The COVID-19 pandemic and its attendant social and economic fallout underscore the incredible stakes of the decisions we now delegate to technology. Even as unemployment has soared, firms increasingly use algorithms to help make employment decisions,[9] notwithstanding the questions that swirl about their compliance with nondiscrimination law.[10] Likewise, opaque algorithms used to select who receives COVID-19 vaccinations have resulted in wide distributional disparities[11] and perverse outcomes.[12]

As a Commissioner at the Federal Trade Commission—an agency whose mission is to protect consumers from unfair or deceptive practices and to promote competition in the marketplace—I have a front-row seat to the use and abuse of AI and algorithms. In this role, I see firsthand that the problems posed by algorithms are both nuanced and context-specific. Because many of the flaws of algorithmic decision-making have long-standing analogs, related to both human decision-making and other technical processes, the FTC has a body of enforcement experience from which we can and should draw.

This article utilizes this institutional expertise to outline the harms of applied algorithms and AI as well as the tools the FTC has at its disposal to address them, offering three primary contributions to the existing literature. First, it provides a baseline taxonomy

---

[9] Adam S. Forman, Nathaniel M. Glasser & Christopher Lech, *INSIGHT: Covid-19 May Push More Companies to Use AI as Hiring Tool*, BLOOMBERG L. (May 1, 2020, 4:00 AM), https://news.bloomberglaw.com/daily-labor-report/insight-covid-19-may-push-more-companies-to-use-ai-as-hiring-tool.

[10] Miriam Vogel, *COVID-19 Could Bring Bias in AI to Pandemic Level Crisis*, THRIVE GLOBAL (June 14, 2020), https://thriveglobal.com/stories/covid-19-could-bring-bias-in-ai-to-pandemic-level-crisis/.

[11] Natasha Singer, *Where Do Vaccine Doses Go, and Who Gets Them? The Algorithms Decide*, N.Y. TIMES (Feb. 7, 2021), https://www.nytimes.com/2021/02/07/technology/vaccine-algorithms.html.

[12] Eileen Guo & Karen Hao, *This is the Stanford Vaccine Algorithm that Left Out Frontline Doctors*, MIT TECH. REV. (Dec. 21, 2020), https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/.

of some of the algorithmic harms that threaten to undermine economic and civil justice.[13] I identify three ways in which flaws in algorithm design can produce harmful results: faulty inputs, faulty conclusions, and failure to adequately test. But not all harmful consequences of algorithms stem from design flaws. Accordingly, I also identify three ways in which sophisticated algorithms can generate systemic harm: by facilitating proxy discrimination, by enabling surveillance capitalism,[14] and by inhibiting competition in markets. In doing so, I show that at several stages during the design, development, and implementation of algorithms, failure to closely scrutinize their impacts can drive discriminatory outcomes or other harms to consumers.

Second, this article describes my view of how the FTC's existing toolkit—including section 5 of the FTC Act, the Equal Credit Opportunity Act (ECOA), and the Fair Credit Reporting Act (FCRA)—can and should be aggressively applied to defend against these threats. For example, I argue that we should encourage non-mortgage creditors to collect demographic data in compliance with ECOA's self-testing safe harbor to assess existing algorithms for indicia of bias. I also discuss algorithmic disgorgement, an innovative and

---

[13] Of course, the consumer protection and competition challenges posed by algorithmic decision-making go well beyond those listed here; the purpose of this taxonomy is not to be comprehensive but to provide a working framework of some of the more common and obvious concerns to facilitate a mapping of enforcement tools onto the problems.

[14] The term "surveillance capitalism" was coined by Shoshanna Zuboff in her recent book; acknowledging that there is a rich body of work on these topics, I will borrow Zuboff's shorthand for the purposes of describing this algorithmic flaw. *See generally* Shoshana Zuboff, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019); *see also* Julie E. Cohen, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L. J. 1460 (Mar. 2020). Surveillance capitalism refers to the way in which, throughout today's digital economy, a pervasive web of machine-learning algorithms collects and processes immense pools of consumer data, often in real time. Through constant, data-driven adjustments, these algorithms evolve and "improve" in a relentless effort to capture and monetize as much attention from as many people as possible. Practically speaking, these companies accomplish their goals through microtargeting and other forms of subtle behavioral manipulation. This system of "surveillance capitalism" systematically erodes consumer privacy, promotes misinformation and disinformation, drives radicalization, undermines consumers' mental health, and reduces or eliminates consumers' choices. *See infra* Section II, Part B, Surveillance Capitalism.

promising remedy the FTC secured in recent enforcement actions. Finally, in this section I identify some of the limitations on the reach of our existing enforcement tools.

Those limitations tie directly to this article's third contribution: I explore how FTC rulemaking under section 18 of the FTC Act or new legislation could help more effectively address the harms generated by AI and algorithmic decision-making. I hope to draw the attention and ingenuity of the interested public to the challenges posed by algorithms so that we can work together on creating an enforcement regime that advances economic justice and equity.

Ultimately, I argue that new technology is neither a panacea for the world's ills nor the plague that causes them. In the words of MIT-affiliated technologist R. David Edelman, "AI is not magic; it is math and code."[15] As we consider the threats that algorithms pose to justice, we must remember that just as the technology is not magic, neither is any cure to its shortcomings. It will take focused collaboration between policymakers, regulators, technologists, and attorneys to proactively address this technology's harms while harnessing its promise.

This article proceeds in three sections. Section II outlines the taxonomy of harms caused by algorithmic decision-making. Section III outlines the FTC's existing toolkit for addressing those harms, the ways we can act more comprehensively to improve the efficacy of those tools, and the limitations on our authority. Finally, Section IV discusses new legislation and regulation aimed at addressing algorithmic decision-making more holistically.

II. Algorithmic Harms

A taxonomy of algorithmic harms, describing both the harms themselves and the technical mechanisms that drive them, is a useful starting point. This section is divided into

[15] R. David Edelman (@R_D), TWITTER (Jan. 14, 2020, 10:45AM), https://twitter.com/R_D/status/1217155806409433089.

two subparts. The first addresses three flaws in algorithm design that frequently contribute to discriminatory or otherwise problematic outcomes in algorithmic decision-making: faulty inputs, faulty conclusions, and failure to adequately test. The second subpart describes three ways in which even sophisticated algorithms still systemically undermine civil and economic justice. First, algorithms can facilitate discrimination by enabling the use of facially neutral proxies to target people based on protected characteristics. Second, the widespread application of algorithms both fuels and is fueled by surveillance capitalism. Third, sophisticated and opaque use of algorithms can inhibit competition and harm consumers by facilitating anticompetitive conduct and enhancing market power.

These six different types of algorithmic harms often work in concert—with the first set often directly enabling the second—but before considering their interplay, it is helpful to describe them individually. Of course, the harms enumerated herein are not, and are not intended to be, an exhaustive list of the challenges posed by algorithmic decision-making. This taxonomy, however, does help identify some of the most common and pervasive problems that invite enforcement and regulatory intervention, and therefore is a helpful framework for consideration of potential enforcement approaches.

A. Algorithmic Design Flaws and Resulting Harms

The first three categories of algorithmic harms generally stem from common flaws in the design and application of specific algorithms.

Faulty Inputs

The value of a machine-learning algorithm is inherently related to the quality of the data used to develop it, and faulty inputs can produce thoroughly problematic outcomes. This broad concept is captured in the familiar phrase "garbage in, garbage out."

The data used to develop a machine-learning algorithm might be skewed because individual data points reflect problematic human biases or because the overall dataset is not adequately representative. Often skewed training data reflect historical and enduring

patterns of prejudice or inequality, and when they do, these faulty inputs can create biased algorithms that exacerbate injustice.[16]

One recent example is Amazon's failed attempt to develop a hiring algorithm driven by machine learning, an effort ultimately abandoned before deployment because the algorithm systematically discriminated against women. This discrimination stemmed from the fact that the resumes used to train Amazon's algorithm reflected the male-heavy skew in the company's applicant pool, and despite the engineers' best efforts, the algorithm kept identifying this pattern and attempting to reproduce it.[17]

Faulty inputs also appear to have been at the heart of problems with standardized testing during the COVID-19 pandemic.[18] The International Baccalaureate (IB), a prestigious global degree program for high school students, cancelled its in-person exams and instead relied on an algorithm to "predict" student test scores based on inputs such as teacher-estimated grades and past performance by students at a given school. The result? Baffling test scores with life-altering consequences. For example, relying on schools' past average test scores likely disadvantaged high-achieving students from low-income communities—many of whom had taken these courses to receive college credit and save

---

[16] *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677–87 (2016); Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS INST. (May 22, 2019), https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms; David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 676–77 (2017).

[17] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[18] Roby Chatterji, *How to Center Equity in Advanced Coursework Testing During COVID-19*, CTR. FOR AM. PROGRESS (Sept. 10, 2020), https://www.americanprogress.org/issues/education-k-12/news/2020/09/10/490198/center-equity-advanced-coursework-testing-covid-19/.

thousands of dollars in tuition.[19] According to the IB, 60 percent of US public schools that offer IB classes are Title I schools,[20] and numerous IB students reportedly saw their college scholarships or admissions offers rescinded because the algorithm assigned them unexpectedly low test scores.[21] In a similar case, the United Kingdom used an algorithm to replace its A-Level exams—which play a pivotal role in university admissions there—before ultimately retracting the scores in response to widespread protests. Critics pointed out that the inputs, which were similar to those used in the IB algorithm, unfairly stacked the deck against students at lower-performing schools.[22] Education should help enable upward social mobility, but the inputs in these instances reflected structural disadvantages

---

[19] *See e.g.*, Meredith Broussard, *When Algorithms Give Real Students Imaginary Grades*, N.Y. TIMES (Sept. 8, 2020), https://www.nytimes.com/2020/09/08/opinion/international-baccalaureate-algorithm-grades.html; Avi Asher-Schapiro, *Global Exam Grading Algorithm Under Fire for Suspected Bias*, REUTERS (July 21, 2020), https://www.reuters.com/article/us-global-tech-education-trfn/global-exam-grading-algorithm-under-fire-for-suspected-bias-idUSKCN24M29L (As one IB student describes, "I come from a low-income family—and my entire last two years [of high school] were driven by the goal of getting as many college credits as I could to save money on school . . . when I saw those scores, my heart sank.").

[20] Melissa Gordon, Emily VanderKamp & Olivia Halic, *International Baccalaureate Programmes in Title I Schools in the United States: Accessibility, Participation, and University Enrollment 2*, IBO (2015), https://ibo.org/research/outcomes-research/diploma-studies/international-baccalaureate-programmes-in-title-i-schools-in-the-united-states-accessibility-participation-and-university-enrollment-2015 ("Low-income and underrepresented minority students have less access to social and economic capital, which can hinder educational attainment and exacerbate the cycle of poverty. US schools with a high proportion of low-income students are eligible to become Title I schools, which allows for the allotment of federal resources to attempt to close this achievement gap (US Department of Education, 2014).").

[21] *See, e.g.*, Hye Jung Han, Opinion, *An Algorithm Shouldn't Decide a Student's Future*, POLITICO (Aug. 13, 2020), https://www.politico.eu/article/an-algorithm-shouldnt-decide-students-future-coronavirus-international-baccalaureate; Tom Simonite, *Meet the Secret Algorithm That's Keeping Students Out of College*, WIRED (July 10, 2020), https://www.wired.com/story/algorithm-set-students-grades-altered-futures.

[22] *See* Daan Kolkman, *'F**k the Algorithm'?: What the World Can Learn from the UK's A-level Grading Fiasco*, LONDON SCH. OF ECON. IMPACT BLOG (Aug. 26, 2020), https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/.

and socioeconomic differences.[23] As the BBC noted in its coverage, "it locks in all the advantages and disadvantages—and means that the talented outlier, such as the bright child in the low-achieving school, or the school that is rapidly improving, could be delivered an injustice."[24]

In short, when developers use faulty data to train an algorithm, the results may replicate or even exacerbate existing inequalities and injustices.

Faulty Conclusions

A different type of problem involves the feeding of data into algorithms that generate conclusions that are inaccurate or misleading—perhaps better phrased as "data in, garbage out."[25] This type of flaw, faulty conclusions, undergirds fears about the rapidly proliferating field of AI-driven "affect recognition" technology and is often fueled by failures in experimental design. Many companies claim that their affect recognition

---

[23] *See id.*; Richard Adams & Niamh McIntyre, *England A-level Downgrades Hit Pupils from Disadvantaged Areas Hardest*, GUARDIAN (Aug. 13, 2020), https://www.theguardian.com/education/2020/aug/13/england-a-level-downgrades-hit-pupils-from-disadvantaged-areas-hardest.

[24] Sean Coughlan, *Why Did the A-level Algorithm Say No?*, BBC NEWS (Aug. 14, 2020), https://www.bbc.com/news/education-53787203.

[25] There are additional implications around the concept of "faulty conclusions" to consider going forward; in particular, we need to think carefully about how AI is deployed or implemented, as well as who is impacted by those choices. One important example of this is the development and deployment of facial recognition technology, which can clearly exacerbate existing racial disparities. There is clear and disturbing evidence that these technologies are not as accurate in identifying non-white individuals, and on at least three separate occasions, Black men have been wrongfully arrested based on faulty facial recognition matches. *See* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. IN MACHINE LEARNING RES. 1 (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Brian Fung, *Facial recognition systems show rampant racial bias, government study finds*, CNN BUS. (Dec. 19, 2019), https://www.cnn.com/2019/12/19/tech/facial-recognition-study-racial-bias/index.html; Kashmir Hill, *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.

products can accurately detect an individual's emotional state by analyzing her facial expressions, eye movements, tone of voice, or even gait.[26]

The underlying algorithms attempt to find patterns in, and reach conclusions based on, certain types of physical presentations and mannerisms. But, as one might expect, human character cannot be reduced to a set of objective, observable factors. For example, consider the algorithmic analysis of facial expressions—one popular flavor of affect recognition technology. A review that analyzed more than a thousand studies on emotional expression concluded that "[e]fforts to simply 'read out' people's internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms."[27] Nevertheless, large companies[28]—plus a host of well-funded

---

[26] *See* Kate Crawford et al., *AI Now 2019 Report*, N.Y.U. AI NOW INST. 1, 50–52 (2019), https://ainowinstitute.org/AI_Now_2019_Report.pdf; Manish Raghavan et al., *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*, 2020 PROC. THE 2020 CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 469, 480 (Jan. 2020), https://arxiv.org/abs/1906.09208.

[27] Lisa Feldman Barrett et al., *Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements*, 20 PSYCHOL. SCI. PUB. INT. 1, 48 (2019); *see also id.*, Abstract, at 1, 46–51 (explaining that "how people communicate anger, disgust, fear, happiness, sadness, and surprise varies substantially across cultures, situations, and even people within a single situation. Furthermore . . . a given configuration of facial movements, such as a scowl, often communicates something other than an emotional state."); Zhimin Chen & David Whitney, *Tracking the Affective State of Unseen Persons*, 116 PROC. OF THE NAT'L ACAD. OF SCI. 1, 5 (2019) (finding that detecting emotions with accuracy requires more information than is available just on the face and body); Angela Chen & Karen Hao, *Emotion AI Researchers Say Overblown Claims Give Their Work a Bad Name*, MIT TECH. REV. (Feb. 14, 2020), https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics.

[28] In late 2018, one researcher ran Microsoft's Face API (Application Programming Interface) on a public dataset of NBA player pictures and found that it interpreted Black players as having more negative emotions than white players. *See* Lauren Rhue, *Racial Influence on Automated Perceptions of Emotions*, 2018 CJRN: RACE & ETHNICITY, 6 (Dec. 17, 2018); *see also* Isobel Asher Hamilton, *AI Experts Doubt Amazon's New Halo Wearable Can Accurately Judge the Emotion in Your Voice, and Worry About the Privacy Risks*, BUS. INSIDER (Aug. 29, 2020), https://www.businessinsider.com/experts-skeptical-amazon-halo-judges-emotional-state-from-voice-2020-8; Saheli Roy Choudhury, *Amazon Says Its*

start-ups—continue to sell questionable affect recognition technology, and it is sometimes deployed to grant or deny formative life opportunities.

A striking example of the use of affect recognition technology is in hiring. Despite growing concerns, including from policymakers,[29] a number of companies claim their products are capable of reliably extrapolating personality traits and predicting social outcomes such as job performance.[30] Their methods of "analysis" range from questionable assessments of observable physical factors, such as those described above, to what one researcher has characterized as "AI snake oil."[31]

For example, one recent study of algorithmic employment screening products highlighted a company that purports to profile more than sixty personality traits relevant to job performance—from "resourceful" to "adventurous" to "cultured"—all based on an algorithm's analysis of an applicant's 30-second recorded video cover letter.[32] Of course, not all algorithmic hiring tools are this potentially problematic, but growing evidence

---

Facial Recognition Can Now Identify Fear, CNBC (Aug. 14, 2019), https://www.cnbc.com/2019/08/14/amazon-says-its-facial-recognition-can-now-identify-fear.html.

[29] *See, e.g.*, Rebecca Heilweil, *Illinois Says You Should Know if AI Is Grading Your Online Job Interviews*, VOX RECODE (Jan. 1, 2020), https://www.vox.com/recode/2020/1/1/21043000/artificial-intelligence-job-applications-illinios-video-interivew-act (discussing a recently passed Illinois law requiring companies to notify job seekers that AI will be used to evaluate an applicant's fitness as part of a video interview); Bradford Newman, *Using AI to Make Hiring Decisions? Prepare for EEOC Scrutiny*, BLOOMBERG LAW (Jan. 15, 2021), https://news.bloomberglaw.com/us-law-week/using-ai-to-make-hiring-decisions-prepare-for-eeoc-scrutiny (describing a Dec. 8 letter from ten Senators to the EEOC).

[30] *See* Rebecca Heilweil, *Artificial Intelligence Will Help Determine If You Get Your Next Job*, VOX RECODE (Dec. 12, 2019), https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen (providing examples of companies that use AI in recruiting (Arya and Leoforce), initial contact with a potential recruit or reconnecting a prior candidate (Mya), personality assessments (Pymetrics), and video interviews (HireVue)).

[31] *See* Arvind Narayanan, Assoc. Professor Comput. Sci., Princeton University, Presentation: How to Recognize AI Snake Oil, https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf.

[32] *See* Raghavan et al., *supra* note 26, at 11.

suggests that other products in this space can suffer from major structural deficiencies.[33] Indeed, Princeton's Arvind Narayanan, a computer scientist, has criticized AI tools that claim to predict job performance based on body language and speech patterns as "fundamentally dubious."[34] Such algorithmic hiring products merit skepticism in any application, and recent studies suggest they might systematically disadvantage applicants with disabilities because they present differently than the majority of a company's applicants or employees.[35] These reports should trouble any employer using an algorithmic hiring product to screen applicants.

Pseudoscience claims of power to make objective assessments of human character are not new—consider handwriting analysis that purports to reveal one's personality or even the lie-detector polygraph testing that has long been inadmissible in court. But "AI-powered" claims can be more pernicious than their analog counterparts because they might encounter less skepticism even though opacity in algorithms can prevent objective analysis of their inputs and conclusions.[36] Despite the veneer of objectivity that comes from

---

[33] *Id.* at 5 – 13.

[34] *See* Narayanan, *supra* note 31.

[35] *See* Jim Fruchterman & Joan Mellea, *Expanding Employment Success for People with Disabilities*, BENETECH 1, 3 (Nov. 2018), https://benetech.org/wp-content/uploads/2018/11/Tech-and-Disability-Employment-Report-November-2018.pdf; Anhong Guo et al., *Toward Fairness in AI For People With Disabilities: A Research Roadmap* 125, ACM SIGACCESS ACCESSIBILITY & COMPUTING 1, 4 (Oct. 2019), https://arxiv.org/abs/1907.02227; *see also* Alex Engler, *For Some Employment Algorithms, Disability Discrimination by Default*, BROOKINGS INST. (Oct. 31, 2019), https://www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default.

[36] Additionally, many companies capitalize on the positive associations with AI, despite the fact that they do not even *use* AI in any material way for their business. One recent report found that a full 40 percent of European startups that were classified as AI companies do not accurately fit that description and that startups with the AI label attract 15 percent to 50 percent more in their funding rounds than other technology startups. *See* Parmy Olson, *Nearly Half of All 'AI Startups' Are Cashing in on Hype*, FORBES (Mar. 4, 2019), https://www.forbes.com/sites/parmyolson/2019/03/04/nearly-half-of-all-ai-startups-are-cashing-in-on-hype/#151cd215d022.

throwing around terms such as "AI" and "machine learning," in many contexts the technology is still deeply imperfect.[37]

Indeed, an employment-screening algorithm's assessment of a candidate can sometimes be less accurate or useful than the subjective (though still imperfect) impression an employer gets from conducting an interview.[38] These risks can be compounded when certain products are emphatically marketed as producing reliable or objective predictions about potential hires when their conclusions are in fact flawed and misleading.[39] This is a

---

[37] Algorithmic hiring is problematic for a number of other reasons. For example, we are already seeing the development of a market for strategies and products that are designed to "beat" different kinds of hiring algorithms. Some people will be unable to afford these services, and they will be judged against those who can, creating yet another barrier to employment that perpetuates historical wealth inequality and hinders social mobility. *See* Sangmi Cha, *'Smile with Your Eyes': How to Beat South Korea's AI Hiring Bots and Land a Job*, REUTERS (Jan. 12, 2020), https://www.reuters.com/article/us-southkorea-artificial-intelligence-jo/smile-with-your-eyes-how-to-beat-south-koreas-ai-hiring-bots-and-land-a-job-idUSKBN1ZC022; Hilke Schellmann, *How Job Interviews Will Transform in the Next Decade*, WALL ST. J. (Jan. 7, 2020), https://www.wsj.com/articles/how-job-interviews-will-transform-in-the-next-decade-11578409136; *see also* Miranda Bogen & Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, UPTURN (Dec. 2018), https://www.upturn.org/reports/2018/hiring-algorithms/.

[38] *See generally*, Alex Engler, *Auditing Employment Algorithms for Discrimination*, BROOKINGS INST. (Mar. 12, 2021), https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/.

[39] Algorithms are also used to deceive and manipulate consumers in other ways—for example, through dark patterns and deepfake technology. Dark patterns are digital user interfaces that are designed to deceive and manipulate consumers into taking unintended actions that may not be in their interests—often through microtargeting and personalization. *See, e.g.*, Lauren E. Willis, *Deception by Design*, HARV. J. L. & TECH. (Fall 2020), https://jolt.law.harvard.edu/assets/articlePDFs/v34/3.-Willis-Images-In-Color.pdf; Rebecca Kelly Slaughter, Acting Chair, Fed. Trade Comm'n, Opening Remarks at Bringing Dark Patterns to Light: An FTC Workshop (Apr. 29, 2021), https://www.ftc.gov/system/files/documents/public_statements/1589931/opening_remarks_of_acting_chairwoman_slaughter_at_ftc_dark_patterns_workshop.pdf; Arunesh Mathur et al., *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROC. ACM ON HUMAN-COMPUTER INTERACTION 1 (Nov. 2019), https://arxiv.org/pdf/1907.07032.pdf?mod=article_inline. Deepfakes use deep learning and neural networks to identify and reconstruct patterns in audio, video, or image data—generating new content that looks, feels, and sounds real. For example, companies can use a five-second clip of a person's

phenomenon with which we are quite familiar at the FTC: new technology, same old lack of substantiation for claims.[40]

Failure to Test

Even if an algorithm is designed with care and good intentions, it can still produce biased or harmful outcomes that are unanticipated. Too often, algorithms are deployed without adequate testing that could uncover these unwelcome outcomes before they harm

---

actual voice to generate a deepfake audio clip of the voice saying anything. This type of content can be used to manipulate and deceive consumers, including through imposter scams, fraud, and disinformation. Deepfake technology is widely accessible, can avoid detection, and is constantly improving. *See, e.g.*, Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L.R. 1753, 1769–70 (2019), https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1640&context=faculty_scholarship; Dan Boneh et al., *Preparing for the Age of Deepfakes and Disinformation*, STAN. U. (Nov. 2020), https://hai.stanford.edu/sites/default/files/2020-11/HAI_Deepfakes_PolicyBrief_Nov20.pdf. The FTC has recently held workshops on both of these important topics, bringing together enforcers, academics, and advocates to help inform the agency's law enforcement and regulatory approach. *See* Fed. Trade Comm'n, *Bringing Dark Patterns to Light: An FTC Workshop*, FTC (Apr. 29, 2021), https://www.ftc.gov/news-events/events-calendar/bringing-dark-patterns-light-ftc-workshop; Fed. Trade Comm'n, *You Don't Say: An FTC Workshop on Voice Cloning Technologies*, FTC (Jan. 28, 2020), https://www.ftc.gov/news-events/events-calendar/you-dont-say-ftc-workshop-voice-cloning-technologies.

[40] Advertisers must have a reasonable basis for their advertising claims. FTC Policy Statement Regarding Advertising Substantiation, 104 F.T.C. 839 (1984) (appended to Thompson Med. Co., 104 F.T.C. 648 (1984)). When an advertiser claims that its product is proven to work—i.e., that its efficacy has been "established"—a reasonable basis for that claim "must consist of the precise type and amount of proof that would satisfy the relevant scientific community." Removatron Int'l Corp., 111 F.T.C. 206, 306 (1988) (citations omitted), aff'd, 884 F.2d 1489 (1st Cir. 1989). "If the advertisements contain express representations regarding a particular level of support that the advertiser has for the product claim . . . the Commission expects the firm to have that level of substantiation." Thompson Med. Co., 104 F.T.C. 648, 813 (1984). To determine what constitutes a "reasonable basis" for an efficacy claim, the Commission applies the Pfizer factors: (1) the type of claim; (2) the type of product; (3) the benefits of a truthful claim; (4) the ease of developing substantiation for the claim; (5) the consequences of a false claim; and (6) the amount of substantiation experts in the field would agree is reasonable. Pfizer Inc., 81 F.T.C. 23, 30 (1972); *see also* Thompson Med. Co., 104 F.T.C. at 813; Daniel Chapter One, No. 9329, 2009 FTC LEXIS 157, at *226–27 (Aug. 5, 2009); FTC v. Direct Marketing Concepts, 569 F. Supp. 2d 285, 299 (D. Mass 2008).

people in the real world. And, as the FTC frequently cautions in the area of data security, pre-deployment testing is an important step but insufficient to prevent all problems.[41] Constant monitoring, evaluating, and retraining are essential practices to identify and correct embedded bias and disparate outcomes.[42]

The health care field provides good examples of bias that can result from failure to adequately assess the variables used in an algorithm pre-deployment *and* failure to monitor outcomes and test for bias post-deployment. A recent study found racial bias in a widely used machine-learning algorithm intended to improve access to care for high-risk patients with chronic health problems.[43] The algorithm used health care costs as a proxy for health needs, but for a variety of reasons unrelated to health needs, white patients spend more on health care than their equally sick Black counterparts do. Using health care costs to predict health needs therefore caused the algorithm to disproportionately flag white patients for additional care.[44] Researchers estimated that as a result of this embedded bias, the number

---

[41] An additional caution on the subject of data security: The vast quantities of information involved in these algorithms—particularly those developed through machine learning—can lead to serious data security concerns, such as access control implementations, proper storage, and proper disposal of data. Companies that get the security side of the equation wrong may be violating data security rules as well as causing the more AI-specific harms discussed in this article.

[42] *See generally* Turner Lee et al., *supra* note 16.

[43] Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447, 477 (2019). The medical community is increasingly scrutinizing the role of race-based algorithms in perpetuating and potentially exacerbating racial disparities in health care treatment and outcomes. *See, e.g.*, Nwamaka D. Eneanya et al., *Reconsidering the Consequences of Using Race to Estimate Kidney Function*, 322 JAMA 113 (2019), https://jamanetwork.com/journals/jama/article-abstract/2735726 (arguing that race-based equation for kidney function leads to under-provision of care to Black patients without substantial increase in diagnostic precision); Darshali A. Vyas et al., *Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms,* NEW ENG. J. MED. (Aug. 27, 2020), https://www.nejm.org/doi/full/10.1056/NEJMms2004740 (critically cataloguing racially-based algorithms from across health care fields).

[44] Sujata Gupta, *Bias in a Common Health Care Algorithm Disproportionately Hurts Black Patients*, SCI. NEWS (Oct. 24, 2019), https://www.sciencenews.org/article/bias-common-health-care-algorithm-

of Black patients identified for extra care was reduced by more than half. The potential scale of this harm is staggering: the researchers called this particular healthcare algorithm "one of the largest and most typical examples of a class of commercial risk-prediction tools that . . . are applied to roughly 200 million people in the United States each year."[45]

The researchers who uncovered the flaw in the algorithm were able to do so because they looked beyond the algorithm itself to the outcomes it produced and because they had access to enough data to conduct a meaningful inquiry.[46] Notably, when the researchers identified the flaw, the algorithm's manufacturer worked with them to mitigate its impact, ultimately reducing bias by 84 percent—exactly the type of bias reduction and harm mitigation that testing and modification seeks to achieve.[47] Still, an inquiry into the risks of using health care spending as a proxy for health care needs, including relevant social context, should have raised concerns *pre*-deployment. And, although there is no simple

---

hurts-black-patients ("because of the bias . . . healthier white patients get to cut in line ahead of black patients, even though those black patients go on to be sicker.").

[45] The researchers continue: "It should be emphasized that this algorithm is not unique. Rather, it is emblematic of a generalized approach to risk prediction in the health care sector . . . [an industry] in which algorithms are already used at scale today, unbeknownst to many." Obermeyer et al., *supra* note 43, at 447.

[46] The researchers analyzed data on patients at one hospital that used the high-risk care algorithm and focused on 40,000 patients who self-identified as white and 6,000 who identified as Black during a two-year period. The algorithm had given all patients a risk score based on past health care costs. In theory, patients with the same risk scores should be similarly sick. Instead, on average Black patients with the same risk scores as white patients had more chronic diseases. Gupta, *supra* note 44.

[47] Obermeyer et al., *supra* note 43, at 453. During a recent FTC PrivacyCon, Professor Obermeyer explained that while correcting the issue with this algorithm required more effort, "the message from [their] work is that that extra effort can be hugely valuable, because it can make the difference between a biased algorithm and one that actually works against the structural biases in our society." *See* Fed. Trade Comm'n, Transcript of PrivacyCon 2020, 106:20 – 108:7 (2020), https://www.ftc.gov/system/files/documents/public_events/1548288/transcript_privacycon_2020_virtual_event_07212020.pdf.

test to reliably detect and prevent bias, early and ongoing testing of the outcomes in this instance may have caught this flaw years earlier.[48]

Another example on this front: A few years ago, a reporter found that typing in a number of common female names on LinkedIn would result in a prompt for a similarly spelled man's name instead—for example, "Stephan Williams" when searching for "Stephanie Williams."[49] But, according to the reporter, when any of the 100 most common male names was entered, LinkedIn never prompted with a female alternative. This example of a potentially biased outcome, uncovered through user testing, might have been prevented altogether if the platform engaged in regular outcome testing. The company initially denied there was any algorithmic bias,[50] but almost a month later, LinkedIn's VP of engineering conceded that the AI-powered search algorithm did, in fact, produce biased outcomes.[51] Indeed, the executive highlighted a systematic shortcoming in LinkedIn's approach to testing that may have prevented the company from effectively detecting bias.[52]

---

[48] *See generally* Nicole Wetsman, *There's No Quick Fix to Find Racial Bias in Health Care Algorithms*, VERGE (Dec. 4, 2019) https://www.theverge.com/2019/12/4/20995178/racial-bias-health-care-algorithms-cory-booker-senator-wyden ("Algorithms that use proxy measures, for example—like health costs as a measure of sickness—need to be examined more carefully, he says, and any bias in the proxy would have to be evaluated separately.").

[49] *See* Matt Day, *How LinkedIn's Search Engine May Reflect a Gender Bias*, SEATTLE TIMES, (Aug. 31, 2016), https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias. LinkedIn discontinued this practice after these reports. *See* Matt Day, *LinkedIn Changes Search Algorithm to Remove Female-to-Male Name Prompts*, SEATTLE TIMES (Sept. 8, 2016), https://www.seattletimes.com/business/microsoft/linkedin-changes-search-algorithm-to-remove-female-to-male-name-prompts/.

[50] *See* Chris Baraniuk, *LinkedIn Denies Gender Bias Claim Over Site Search*, BBC NEWS (Sept. 8, 2016), https://www.bbc.com/news/technology-37306828.

[51] *See* Igor Perisic, *Making Hard Choices: The Quest for Ethics in Machine Learning*, LINKEDIN ENGINEERING BLOG (Nov. 23, 2016), https://engineering.linkedin.com/blog/2016/11/making-hard-choices--the-quest-for-ethics-in-machine-learning.

[52] *See id.* ("However, having this algorithm serve suggestions based solely on word search frequency, without being aware of gender, actually resulted in biased results. In retrospect it is obvious that, by removing gender from our consideration, our algorithms were actually blind to it. Since we weren't

This kind of bias can have meaningful real-world consequences: in this case, that profiles with female-identified names turned up less frequently, potentially resulting in fewer employment opportunities for women.

Perhaps one of the most troubling examples of a chronic failure to test is reflected in Dr. Safiya Noble's chronology of repeated instances of search bias and stereotyping on Google. Beginning in 2009, searches for the phrase "black girls" on Google would return multiple pornographic and sexual content results, including as the top result. This type of damaging proliferation of racist stereotypes through massive algorithms went on for years and extended to many other groups. As recently as 2016, a Google search for "three black teenagers" generated results perpetuating stereotypes of violence.[53] In each instance, when public attention identified the flaw, it was fixed, but troubling results persisted for years. Failure to test at this scale risks elevating and embedding some of our society's most persistent and pernicious stereotypes.

It is entirely possible that each of these examples of algorithmic bias was the product of multiple algorithmic flaws, as is often the case. But what stands out about all of these examples is that additional testing about the algorithm's impact across two of the most obvious protected classes, race and gender, might have detected the disparate effect much earlier and facilitated a correction. And, in some examples, the deployer of the algorithm

---

tracking that information in the first place, we couldn't use it to verify that the output of the algorithms were, in fact, unbiased."). In recent years, LinkedIn appears to have taken a more deliberate approach to testing their machine-learning algorithms and the underlying data—part of a stated effort to make the company's machine-learning algorithms more fair and equitable. The company has also created open-source tools designed to address biased or inequitable results that are uncovered through testing. These initiatives are encouraging and the goals are laudable, but success will require constant vigilance. *See, e.g.*, Ryan Roslansky, *Helping Every Company Build More Inclusive Products*, LINKEDIN BLOG (May 26, 2020), https://blog.linkedin.com/2020/may/26/helping-every-company-build-more-inclusive-products; Sriram Vasudevan et al., *Addressing Bias in Large-scale AI Applications: The LinkedIn Fairness Toolkit*, LINKEDIN ENGINEERING BLOG (Aug. 25, 2020), https://engineering.linkedin.com/blog/2020/lift-addressing-bias-in-large-scale-ai-applications.

[53] *See* Safiya Noble, *Google Has a Striking History of Bias Against Black Girls*, TIME (Mar. 26, 2018), https://time.com/5209144/google-search-engine-algorithm-bias-racism/; *see also* Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU PRESS (Feb. 2018).

was reluctant to acknowledge (or flat-out denied) the possibility of bias.[54] This problem is not limited to algorithms, but, as in other instances of unintended bias, admitting that it might occur, despite best intentions, is imperative.

B. How Sophisticated Algorithms Exacerbate Systemic Harms

The previous subsection explored problems in the specific design and application of individual algorithms. By contrast, the harms enumerated in this subsection describe more broadly the societal consequences of flawed algorithmic decision-making. Of course, these categories of harms are neither exhaustive nor mutually exclusive; the issues described in this subsection tie closely to the flaws enumerated above.

Proxy Discrimination

In addition to the flaws in algorithmic design and implementation enumerated above, the promise of algorithmic decision-making is also tempered by its systemic contributions to broader social harms. One such pernicious harm at work in recent examples of algorithmic bias is a problem scholars have termed "proxy discrimination."[55] When algorithmic systems engage in proxy discrimination, they use one or more facially neutral variables to stand in for a legally protected trait, often resulting in disparate treatment of or disparate impact on protected classes for certain economic, social, and civic opportunities.[56] In other words, these algorithms identify seemingly neutral characteristics to create groups that closely mirror a protected class, and these "proxies" are used for inclusion or exclusion.

---

[54] *See* Baraniuk, *supra* note 50.

[55] *See generally* Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020).

[56] *See id.* at 1260–61, 1269–1270, 1273 ("[P]roxy discrimination requires that the usefulness to the discriminator derives, at least in part, from the very fact that it produces a disparate impact. . . . humans can unwittingly proxy discriminate when the law prohibits 'rational discrimination' . . . a person or firm may find that discrimination based on a facially-neutral characteristic is predictive of its legitimate objectives, even though the characteristic's predictive power derives from its correlation with a legally-prohibited characteristic."); *infra* Part III, Section B, note 123.

Facebook's use of Lookalike Audiences that facilitated housing discrimination presents one of the clearest illustrations of proxy discrimination. According to allegations by the Department of Housing and Urban Development (HUD), Facebook offered customers that were advertising housing and housing-related services a tool called "Lookalike Audiences."[57] An advertiser using this tool would pick a "Custom Audience" that represented her "best existing customers," then Facebook identified users who shared "common qualities" with those customers, who then became the ad's audience.

To generate a Lookalike Audience, Facebook considered proxies that included a user's "likes," geolocation data, online and offline purchase history, app usage, and page views.[58] Based on these factors, Facebook's algorithm created groupings that aligned with users' protected classes. Facebook then identified groups that were more or less likely to engage with housing ads and included or excluded them for ad targeting accordingly. According to HUD, "by grouping users who 'like' similar pages (unrelated to housing) and presuming a shared interest or disinterest in housing-related advertisements, [Facebook]'s mechanisms function just like an advertiser who intentionally targets or excludes users based on their protected class."[59]

---

[57] Charge of Discrimination at 4, Facebook, Inc., FHEO No. 01-18-0323-8 (Mar. 28, 2019); *see also* Tracy Jan & Elizabeth Dwoskin, *HUD Is Reviewing Twitter's and Google's Ad Practices as Part of Housing Discrimination Probe*, WASH. POST (Mar. 28, 2019), https://www.washingtonpost.com/business/2019/03/28/hud-charges-facebook-with-housing-discrimination/.

[58] Charge of Discrimination at 5, Facebook, Inc., FHEO No. 01-18-0323-8 (Mar. 28, 2019).

[59] *Id.* at 5-6; *see* also Cmt. of Comm'r Rohit Chopra in the Matter of Proposed Rule to Amend HUD's Interpretation of the Fair Housing Act's Discriminatory Effects Standard (Oct. 16, 2019), https://www.ftc.gov/system/files/documents/public_statements/1549212/chopra_-_letter_to_hud_on_disparate_impact_proposed_rulemaking_10-16-2019.pdf ("These inputs do not have to be intuitive stand-ins to result in discrimination. Seemingly 'neutral' inputs, especially when analyzed in combination with other data points, can also be a substitute. Members of a protected class will likely have a wide range of other characteristics in common that can be detected with the increased collection of more and different types of information. . .With more data points and more volume, any input or combination of inputs can turn into a substitute or proxy for a protected class.").

This problem may persist across advertising algorithms, which are designed to maximize clicks and conversions. Even when the advertiser requests a broad audience and more inclusivity, an algorithm may skew ads to demographic segments that are expected (based on historical performance) to generate more clicks. In one recent study, researchers specified an identical audience for three different job postings: a lumber industry position, a supermarket cashier position, and a taxi position.[60] Despite the request for the same audience, the lumber job went to an audience that was 72 percent white and 90 percent male, the supermarket cashier went to an 85 percent female audience, and the taxi position went to a 75 percent Black audience.[61]

The dangers of proxy discrimination, amplified by machine learning and optimization, likely affect the credit sphere as well.[62] The combination of an expanding and innovative FinTech market paired with alternative credit scoring has the potential to extend credit to more people who need it. But FinTech innovations can also enable the continuation of historical bias to deny access to the credit system or to efficiently target high-interest products to those who can least afford them.[63] Indeed, these biases can be exacerbated through the use of algorithms, because the algorithms automate decision-making—giving the appearance of impartiality—while simultaneously obscuring visibility

---

[60] Muhammad Ali et al., *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, 3 PROC. ACM ON HUMAN-COMPUTER INTERACTION 1, 12 (Nov. 2019), https://arxiv.org/pdf/1904.02095.pdf.

[61] *Id.* at 2. A prior study on Google's delivery of job ads demonstrated similar problematic results: in an identical sample that was randomly assigned a male or female identity, Google showed an ad for "$200k+ executive position" to the male group 1,852 times and just 318 times to the female group. Amit Datta, Michael Carl Tschantz & Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, 2015 PROC. ON PRIVACY ENHANCING TECH. 92 (Feb. 2015), https://arxiv.org/abs/1408.6491.

[62] *See, e.g.*, John Detrixhe & Jeremy B. Merril, *The Fight Against Financial Advertisers Using Facebook for Digital Redlining*, QUARTZ (Nov. 1, 2019), https://qz.com/1733345/the-fight-against-discriminatory-financial-ads-on-facebook.

[63] These concerns have also caught the attention of Congress. *See, e.g.*, *Examining the Use of Alternative Data in Underwriting and Credit Scoring to Expand Access to Credit: Hearing Before the Task Force on Fin. Tech. of the H. Comm. on Financial Serv.*, 116th Cong. (2019).

into both the inputs and the formulae used to make those decisions. That opacity can make the bias even harder to identify.

A recent study illustrates both the promise and residual peril of algorithmic lending decisions for credit discrimination.[64] The study found that in loans made by face-to-face lenders, Latinx and Black borrowers pay considerably more in interest for home-purchase and refinance mortgages.[65] The study also found that FinTech algorithms discriminate 40 percent less than lenders—but that significant discrimination harming the Latinx and Black borrowers still occurs.[66] The scholars could not conclude definitively what caused the discriminatory outcomes from the FinTech platforms, but they surmised it was likely due to some type of optimization based on a neutral characteristic that aligned with minority status, just as we saw in the examples above.[67]

Proxy discrimination is not a new problem—the use of facially neutral factors that generate discriminatory results is something that society and civil rights laws have been grappling with for decades.[68] In the context of algorithms, sometimes this flaw might be accidental.[69] For example, proxy discrimination was one of the reasons that the health care algorithm discussed earlier ultimately produced biased outcomes, but we have no reason

---

[64] Robert Bartlett et al., *Consumer-Lending Discrimination in the FinTech Era* (Nat'l Bureau of Econ. Research, Working Paper No. 25943, 2019).

[65] By 7.9 and 3.6 basis points, respectively. *Id.* at 5.

[66] To the tune of 5.3 basis points more in interest for purchase mortgages and 2.0 basis points for refinance mortgages. *Id.* at 6.

[67] In this case, learning that "higher prices could be quoted to profiles of borrowers or geographies associated with low-shopping tendencies." *Id.* at 20.

[68] *See* Prince & Schwarcz, *supra* note 55, at 1268-1270.

[69] *See id.* at 1270-1276 ("Big data and AI are game changers when it comes to the risk of unintentional proxy discrimination. In particular, proxy discrimination by AIs is virtually inevitable whenever the law seeks to prohibit use of characteristics whose predictive power cannot be measured more directly by facially neutral data.").

to believe that the hospital or manufacturer of the algorithm in question was trying to disadvantage Black patients. It is important to note, however, that proxy discrimination can also be intentional, and the obscurity provided by black-box decision-making can allow bad-faith actors to effectively launder bias and discrimination through their algorithms in pursuit of illegitimate profits or to maintain oppressive hierarchies.[70] Proxy discrimination that results in disparate impact is always pernicious, whether or not we can identify underlying intent, and it can and should give rise to legal liability even if it is not intentional.[71]

Surveillance Capitalism

An additional way algorithmic decision-making can fuel broader social challenges is the role it plays in the system of surveillance capitalism[72]—a business model that systematically erodes privacy, promotes misinformation and disinformation,[73] drives

---

[70] *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677–87 (2016); Alan Rubel, Clinton Castro & Adam Pham, *Agency Laundering and Information Technologies*, ETHICAL THEORY & MORAL PRAC. 22, 1017–1041 (Oct. 24, 2019), https://link.springer.com/article/10.1007/s10677-019-10030-w; Fed. Trade Comm'n, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*, FTC (Jan. 2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[71] *See* Elisa Jillson, *Aiming for Truth, Fairness, and Equity in Your Company's Use of AI*, FTC BUS. BLOG (Apr. 19, 2021), https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.; *infra* Part III, Section B.

[72] *See* Zuboff, *supra* note 14.

[73] *See, e.g.*, Dipayan Ghosh & Nick Couldry, *Digital Realignment: Rebalancing Platform Economies from Corporation to Consumer* (M-RCBG Associate Working Paper Series, Working Paper No. 155, Oct. 2020), https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/files/AWP_155_final2.pdf; Filippo Menczer & Thomas Hills, *Information Overload Helps Fake News Spread, and Social Media Knows It*, SCI. AM. (Dec. 1, 2020), https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it; Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, SCIENCE (Mar. 9, 2018), https://science.sciencemag.org/content/359/6380/1146. The proliferation of other AI-driven technologies, such as deepfakes and rapidly improving algorithmic text generation, can further exacerbate the

radicalization,[74] undermines consumers' mental health,[75] and reduces or eliminates consumers' choices.[76]

---

disinformation problem. *See supra* note 39; Ben Buchanan et al., *Truth, Lies, and Automation: How Language Models Could Change Disinformation*, CTR. FOR SECURITY & EMERGING TECH. (May 2021), https://cset.georgetown.edu/publication/truth-lies-and-automation/?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosam&stream=top.

[74] *See, e.g.*, Marc Faddoul, Guillaume Chaslot, & Hany Farid, *A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos*, 2020 ArXiv 1 (Mar. 2020), https://farid.berkeley.edu/downloads/publications/arxiv20.pdf; Manoel Horta Ribeiro et al., *Auditing Radicalization Pathways on YouTube*, 2020 PROC. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 131 (Jan. 2020), https://arxiv.org/pdf/1908.08313.pdf; Jeff Horwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, WALL ST. J. (May 26, 2020), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 ("'Our algorithms exploit the human brain's attraction to divisiveness,' read a slide from a 2018 [internal Facebook] presentation. 'If left unchecked,' it warned, Facebook would feed users 'more and more divisive content in an effort to gain user attention & increase time on the platform . . . [A] 2016 presentation stated that '64% of all extremist group joins are due to our recommendation tools' and that most of the activity came from the platform's 'Groups You Should Join' and 'Discover' algorithms: 'Our recommendation systems grow the problem.'"); Jeff Horwitz, *Facebook Knew Calls for Violence Plagued 'Groups,' Now Plans Overhaul*, WALL ST. J. (Jan. 31, 2021), https://www.wsj.com/articles/facebook-knew-calls-for-violence-plagued-groups-now-plans-overhaul-11612131374.

[75] *See, e.g.*, Vikram R. Bhargava & Manuel Velasquez, *Ethics of the Attention Economy: The Problem of Social Media Addiction*, BUS. ETHICS Q. 1, 2 (2020), https://www.cambridge.org/core/journals/business-ethics-quarterly/article/ethics-of-the-attention-economy-the-problem-of-social-media-addiction/1CC67609A12E9A912BB8A291FDFFE799 ("[A]ddicting users to social media . . . unjustifiably harms them and does so in a way that is both demeaning and objectionably exploitative . . . [T]he business model of social media companies generates a strong incentive to perpetrate this very wrongdoing"); Yu-Shian Cheng et al., *Internet Addiction and Its Relationship With Suicidal Behaviors: A Meta-Analysis of Multinational Observational Studies*, 79 J. CLINICAL PSYCHIATRY (2018); Melissa G. Hunt et al., *No More FOMO: Limiting Social Media Decreases Loneliness and Depression*, 37 J. SOC. AND CLINICAL PSYCH. 751 (2018); Christian Montag, Bernd Lachmann, Marc Herrlich & Katharina Zweig, *Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories*, 16 INT'L J. OF ENV. RES. & PUB. HEALTH. 2612 (2019), https://www.mdpi.com/1660-4601/16/14/2612/htm.

[76] *See infra* Section II, Part B, Threats to Competition.

Much of today's digital economy is fundamentally geared toward maximizing consumer attention and then monetizing it—the more eyeballs and time spent, the better. In some industries, such as broadcasting, this focus is long-standing. But, in a remarkably short time, the proliferation of machine-learning algorithms and behavioral advertising has created a staggering and fundamentally different system of "surveillance capitalism."

Unlike the attention ecosystems of old, these opaque, data-hungry algorithms are ubiquitous in our lives; it is functionally impossible to escape their reach. What is more, machine learning enables this pervasive web of algorithms to process immense pools of consumer data, often in real time.[77] Through constant, data-driven adjustments, these algorithms evolve and "improve" in a relentless effort to capture and monetize as much attention from as many people as possible. Many surveillance-capitalism enterprises are remarkably successful at using algorithms to "optimize" for consumers' attention with little regard for downstream consequences.

One of the most troubling aspects of surveillance capitalism is how it affects children. As in other contexts, many companies use machine-learning algorithms to attract, maintain, and monetize children's attention—often employing algorithmic recommendation systems and auto-play functions. Especially when children are involved, these engines of the digital economy can be dangerous and their consequences difficult to avoid. For example, a recent study of toddler-oriented content on YouTube concluded that due to the platform's recommendation system, "young children are not only able, but *likely* to encounter disturbing videos when they randomly browse the

[77] Of course, the amount of information encountered by each consumer is enormous, and her attention is a limited resource, so participants in the so-called attention economy go to great lengths to capture it. The resulting "information overload" often plays a key role in the loss of high-quality information. *See, e.g.*, Filippo Menczer & Thomas Hills, *Information Overload Helps Fake News Spread, and Social Media Knows It*, SCI. AM. (Dec. 1, 2020), https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/.

platform starting from benign videos."[78] While most of the toddler-oriented content on YouTube is innocuous, the authors highlight an influx of disturbing or inappropriate content that targets young children, as in the infamous "Elsagate" controversy. In that episode, nefarious users uploaded disturbing videos featuring well-known children's characters—such as Spiderman, Elsa, and Mickey Mouse—involved in violent or lewd conduct. This pernicious content was often masked by innocent thumbnails and video titles, making it more difficult for children and parents to avoid.[79]

Exposure to this type of content can traumatize young children and carries serious risks for early childhood development.[80] But even more troubling is that YouTube's algorithms promote these disturbing and inappropriate videos to young children and that each additional view brings the platform and the content creator additional revenue. Of course, YouTube asserts that it has taken steps to address this issue, such as by removing disturbing content, but the company has not been transparent about the effectiveness of

---

[78] Kostantinos Papadamou et al., *Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children*, 2020 PROC. FOURTEENTH INT'L AAAI CONF. ON WEB AND SOCIAL MEDIA 522 (2020), https://ojs.aaai.org/index.php/ICWSM/article/view/7320. This important study received funding from the European Union and the National Science Foundation.

[79] *See id.*

[80] Dr. Michael Rich, the Director and Founder of Harvard Medical School's Center on Media and Child Health, explains that these videos are made "more upsetting by the fact that these characters [who children] thought they knew and trusted are behaving in these ways." *See* Sapna Maheshwari, *On YouTube Kids, Startling Videos Slip Past Filters*, N.Y. TIMES (Nov. 4, 2017), https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html; *see also* Elyse Samuels & William Neff, *Sex, Drugs and Peppa Pig? How Big a Problem Is Disturbing Kids' Content on YouTube?*, WASH. POST (Mar. 14, 2019), https://www.washingtonpost.com/video/business/technology/sex-drugs-and-peppa-pig-how-big-a-problem-is-disturbing-kids-content-on-youtube/2019/03/14/d29fd339-9f62-4675-9e4d-b1f0570680eb_video.html; Craig Timberg, *YouTube Says It Bans Preteens from Its Site. But It's Still Delivering Troubling Content to Young Children.*, WASH. POST (Mar. 14, 2019), https://www.washingtonpost.com/technology/2019/03/14/youtube-says-it-bans-preteens-its-site-its-still-delivering-troubling-content-young-children/.

these efforts.[81] As the researchers suggest, these are "the dangers of crowd-sourced, uncurated content combined with engagement oriented, gameable recommendation systems. Considering the advent of algorithmic content creation (e.g., 'deepfakes') and the monetization opportunities on sites like YouTube, there is no reason to believe there will be an organic end to this problem."[82]

Beyond recommendation algorithms, companies may seek to monetize children's attention by illegally harvesting their data and using it to serve them behavioral advertisements. This conduct violates the Children's Online Privacy Protection Act (COPPA)[83] and was at the core of the Commission's recent complaints against HyperBeard[84] and Google/YouTube.[85] In the latter case, the Commission's settlement required the defendants to materially remake the YouTube platform in ways that would reduce the amount of illegal behavioral advertising on child-directed content. Specifically, whenever a new video is uploaded to YouTube, content creators now have to designate that content as child-directed or not. For videos designated as child-directed, YouTube would not serve behavioral advertisements, track persistent identifiers, or allow comments.

---

[81] Papadamou et al., *supra* note 78, at 532 ("[O]ur assessment on YouTube's current mitigations shows that the platform struggles to keep up with the problem: only 20.5% and 2.5% of our manually reviewed disturbing and restricted videos, respectively, have been removed by YouTube.").

[82] Papadamou et al., *supra* note 78, at 532.

[83] The Children's Online Privacy Protection Rule, which implements COPPA, requires that child-directed websites, apps, and other online services provide notice of their information practices and obtain verifiable parental consent before collecting personal information from children under thirteen, including the use of persistent identifiers to track a user's internet browsing habits for targeted advertising. In addition, third parties, such as advertising networks, are also subject to COPPA where they have actual knowledge they are collecting personal information directly from users of child-directed websites and online services. *See* 16 C.F.R. § § 312.1 – 312.13.

[84] *See* Complaint, United States v. HyperBeard, Inc., No. 3:20-cv-3683, 2020 WL 5535925 (N.D. Cal. June 3, 2020), https://www.ftc.gov/system/files/documents/cases/192_3109_hyperbeard_-_complaint.pdf.

[85] The Commission brought this case with the New York Attorney General. *See* Complaint, Fed. Trade Comm'n v. Google LLC and YouTube, LLC, No.: 1:19-cv-2642 (D.D.C. Sept. 4, 2019).

Despite this requirement, I voted against the settlement. My primary objection was that we should have required an enforceable commitment that YouTube would police the accuracy of channels' designations to identify undesignated child-directed content and turn off behavioral advertising on those videos.[86] As suggested in my dissent, one way to do this would be through a technological backstop. Securing such a commitment in our settlement was important to me because there are strong financial incentives to mis-designate child-directed content: Behavioral advertising is more lucrative than contextual advertising—both for YouTube and for the myriad content creators who might bet that they could escape COPPA enforcement. Since the settlement was finalized, YouTube has announced that they will use machine learning to actively search for mis-designated content and automatically apply age-restrictions.[87] This sounds like the technological backstop I had in mind, but with two major differences: first, it is entirely voluntary, and second, both its application and effectiveness are opaque. YouTube can dial back this

---

[86] *See* Dissenting Statement of Commissioner Rebecca Kelly Slaughter, In the Matter of Google LLC & YouTube, LLC, Fed. Trade Comm'n File No. 1723083 (Sept. 4, 2019), https://www.ftc.gov/system/files/documents/public_statements/1542971/slaughter_google_youtube_statement.pdf ("When it comes to fencing-in relief, the current order looks like a fence but one with only three sides. The missing fourth side is a mechanism to ensure that content creators are telling the truth when they designate their content as not child-directed. . . . A cynical observer might wonder whether in the wake of this order YouTube will be even more inclined to turn a blind eye to inaccurate designations of child-directed content in order to maximize its profit. . . . In that light, the fence looks more like a moat, giving YouTube a handy argument that it should face no COPPA liability for content mis-designated as not child-directed."); *see also* Dissenting Statement of Commissioner Rohit Chopra in the Matter of Google LLC and YouTube, LLC, Fed. Trade Comm'n File No. 1723083 (Sept. 4, 2019), https://www.ftc.gov/system/files/documents/public_statements/1542957/chopra_google_youtube_dissent.pdf.

[87] *See, e.g.*, Todd Spangler, *YouTube New 'Supervised' Mode Will Let Parents Restrict Older Kids' Video Viewing*, Variety (Feb. 24, 2021), https://variety.com/2021/digital/news/youtube-supervised-accounts-kid-controls-1234913968/; *Using Technology to More Consistently Apply Age Restrictions*, YOUTUBE OFFICIAL BLOG (Sept. 22, 2020), https://blog.youtube/news-and-events/using-technology-more-consistently-apply-age-restrictions/; *Better Protecting Kids' Privacy on YouTube*, YOUTUBE OFFICIAL BLOG (Jan. 6, 2020), https://blog.youtube/news-and-events/better-protecting-kids-privacy-on-youtube/.

mechanism—or do away with it altogether—at its discretion, and the public would be none the wiser.

This brings up a broader set of concerns about surveillance capitalism—one that extends beyond COPPA or any single platform. Certain technology companies have almost unlimited discretion over how their algorithms present information to consumers. These pervasive algorithms process an unfathomable amount of data about each of us, which makes them remarkably effective at exploiting and exacerbating our cognitive vulnerabilities.[88] The companies that deploy them often maximize and monetize user engagement through microtargeting and other forms of subtle behavioral manipulation, in order to fuel not only advertising revenue but also further data collection—the more engaged a user is, the more data she generates. It is worth emphasizing, however, that these companies can use this information to do more than merely capture our attention. By simply tweaking their code, these companies can powerfully shape our behavior and even our outlooks on the world—threatening to rob us not only of our privacy but also of our autonomy.[89]

---

[88] To understand how surveillance capitalism harms and manipulates consumers, it is crucial to understand the way a given set of algorithms exacerbates and exploits our social tendencies and cognitive vulnerabilities. *See, e.g.*, Ryan Calo, *Digital Market Manipulation*, GEO. WASH. L. REV. (2014), https://www.gwlr.org/wp-content/uploads/2014/10/Calo_82_41.pdf. If we are to successfully curtail these harms, interdisciplinary partnerships between technologists, behavioral scientists, and others will be key. Thankfully, creative initiatives have already begun to emerge in this space—for example, the partnership between the University of Warwick and Indiana University Bloomington's Observatory on Social Media (OSoMe, pronounced "awesome"), which recently published an informative primer on this topic. *See* Menczer & Hills, *supra* note 73.

[89] *See, e.g.*, Cohen, *supra* note 14; Zuboff, *supra* note 14; S. C. Matz, M. Kosinski, G. Nave, D. J. Stillwell, *Psychological Targeting in Digital Mass Persuasion*, 114 PROC. NAT'L ACAD. SCI. 12714–19 (Nov. 2017), https://www.pnas.org/content/114/48/12714 ("Building on recent advancements in the assessment of psychological traits from digital footprints, this paper demonstrates the effectiveness of psychological mass persuasion—that is, the adaptation of persuasive appeals to the psychological characteristics of large groups of individuals with the goal of influencing their behavior."); Petra Persson, *Attention Manipulation and Information Overload*, 2 BEHAV. PUB. POL'Y, 78 (2018), https://www.cambridge.org/core/journals/behavioural-public-policy/article/attention-manipulation-and-information-overload/3987E9B897AFC10CB7AD85D9E4868881.

The FTC recently announced a timely and important section 6(b) study[90] of nine social media and video-streaming services—an industry where the potential for this subtle, data-driven manipulation is clear and obvious. The 6(b) study is intended to help the agency better understand these companies' advertising and user-engagement practices; how they collect, use, track, or derive personal and demographic information; and how their practices affect children, teens, and other vulnerable populations. My joint statement with Commissioners Chopra and Wilson noted that it is alarming that we still know so little about companies that know so much about us.[91] The project seeks to understand how business models influence what Americans hear and see, with whom they talk, and what information they share. Among other topics, the study seeks to uncover how children and families are targeted and categorized, as well as whether consumers are being subjected to social-engineering experiments.[92]

Threats to Competition

The pitfalls associated with algorithmic decision-making sound most obviously in the laws the FTC enforces through our consumer protection mission. But the FTC is also responsible for promoting competition, and the threats posed by algorithms profoundly affect that mission as well; moreover, these two missions are not actually distinct, and problems—including those related to algorithms and economic justice—need to be considered with both competition and consumer protection lenses. A full discussion of the

---

[90] *See infra* Section III.

[91] *See* Joint Statement of FTC Comm'rs Chopra, Slaughter & Wilson, Social Media and Video Streaming Service Providers' Privacy Practices, Fed. Trade Comm'n File No. P205402 (Dec. 14, 2020) https://www.ftc.gov/system/files/documents/public_statements/1584150/joint_statement_of_ftc_commissioners_chopra_slaughter_and_wilson_regarding_social_media_and_video.pdf.

[92] *See* Fed. Trade Comm'n, Resolution Directing Use of Compulsory Process to Collect Information Regarding Social Media & Video Streaming Service Providers' Privacy Practices, Fed. Trade Comm'n Matter No. P205402 (Apr. 12, 2012), https://www.ftc.gov/reports/6b-orders-file-special-reports-social-media-video-streaming-service-providers; *see also* Vindu Goel, *Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry*, N.Y. TIMES (June 29, 2014), https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html.

implications of algorithms in antitrust law is well beyond the scope of this article, but I will briefly highlight a few of the ways competition can be imperiled by use or misuse of algorithms.[93] These topics include traditional antitrust fare such as pricing and collusion, as well as more novel questions such as the implications of the use of algorithms by dominant digital firms to entrench market power and to engage in exclusionary practices.

Algorithmic pricing, the practice of setting prices dynamically and automatically with algorithms, sometimes enhanced by artificial intelligence and machine learning, has become ubiquitous.[94] The body of literature about how algorithmic pricing can affect competition has grown over the past several years and includes concerns about the use of algorithms to facilitate collusion[95] and anticompetitive personalized pricing.[96] The use of

---

[93] Exec. Order No. 14036, 86 Fed. Reg. 36,987 (July 9, 2021), https://www.federalregister.gov/documents/2021/07/14/2021-15069/promoting-competition-in-the-american-economy (recognizing that unfair data collection and surveillance practices may damage competition).

[94] *See, e.g.*, Salil K. Mehra, *Antitrust and the Robo-Seller: Competition in the Time of Algorithms*, 100 MINN. L. REV. 1323, 1352 (2016); Andreas Mundt, *Algorithms and Competition in a Digitalized World*, 2020 COMPETITION POL'Y INT'L (July 13, 2020).

[95] *See, e.g.*, Emilio Calvano et al., *Protecting Consumers from Collusive Prices Due to AI*, 370 SCI. 1040, 1040 (2020), https://science.sciencemag.org/content/370/6520/1040; John Asker, Chaim Fershtman & Ariel Pakes, *Artificial Intelligence and Pricing: The Impact of Algorithm Design* (NBER, Working Paper No. 28535, March 1, 2021), https://ssrn.com/abstract=3805295; Zach Brown & Alexander MacKay, *Competition in Pricing Algorithms* (Harv. Bus. Sch., Working Paper No. 20-067, April 29, 2021), https://ssrn.com/abstract=3485024; A. Ezrachi & M. E. Stucke, *Algorithmic Collusion: Problems and Counter-Measures*, DAF/COMP/WD(2017)25 (2017), https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/WD%282017%2925&docLanguage=En.

[96] *See, e.g.*, Jean-Pierre Dube & Sanjog Misra, *Personalized Pricing and Customer Welfare* (Feb. 21, 2020), https://ssrn.com/abstract=2992257; Patrick J. Kehoe, Bradley J. Larsen & Elena Pastorino, *Dynamic Competition in the Era of Big Data* (Stanford Univ. & Fed. Reserve Bank of Minneapolis, Working Paper, Nov. 3, 2020), https://web.stanford.edu/~bjlarsen/dynamic_comp_big_data.pdf; *Pricing Algorithms, Economic working paper on the use of algorithms to facilitate collusion and personalised pricing*, 5.25-.28 (94 U.K. Competition and Mkts. Auth., Working Paper, October 8, 2018), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf.

algorithms to execute a price-fixing agreement has even given rise to criminal antitrust charges.[97]

Algorithms may enhance the ability of firms to collude, either tacitly or explicitly.[98] While there have been limited cases of enforcement against collusion facilitated by algorithms, it is unclear whether the conduct is in fact not occurring or whether it is simply very difficult for enforcers to detect. Moving forward, competition enforcers may deploy

---

[97] In 2015, the US Department of Justice brought criminal charges against two e-commerce companies in *United States v. Topkins* for executing a price-fixing agreement using algorithms. In that matter, the executives of an e-commerce seller of posters and art agreed to fix the prices of their products sold through Amazon Marketplace and then adopted specific pricing algorithms to execute their price-fixing conspiracy. In its press release announcing the matter, the DOJ noted its commitment to pursue illegal price-fixing agreements whether they occur in "a smoke-filled room or over the Internet using complex pricing algorithms." Press Release, U.S. Dep't of Justice, Former E-Commerce Executive Charged with Price Fixing in the Antitrust Division's First Online Marketplace Prosecution (Apr. 6, 2015), https://www.justice.gov/opa/pr/former-e-commerce-executive-charged-price-fixing-antitrust-divisions-first-online-marketplace.

[98] Tacit collusion, in contrast to explicit collusion, is when firms are able to coordinate their behavior and achieve anticompetitive outcomes through independent conduct and without an agreement. This conduct can enable joint profit maximization while reducing competition and harming consumers. *See* Calvano et al., *supra* note 95. In one recent study of duopoly German gasoline station markets, researchers found empirical evidence that prices increased when both stations adopted algorithmic pricing practices. Modeling has shown that high-frequency pricing algorithms can temper competition and increase profits, with the largest benefits going to the most dominant and technologically savvy firms. *See* Stephanie Assad et al., *Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market* (CESifo, Working Paper No. 8521, 2020), https://ssrn.com/abstract=3682021; Brown & MacKay, *supra* note 95, However, the fact that algorithms may make it easier for firms to predict and respond to changes in demand may also increase each firm's temptation to deviate to a lower price in times of high predicted demand. *See*, *e.g.*, Jeanine Miklós-Thal & Catherine E. Tucker, *Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?*, 65 MGMT. SCI. 1552 (Apr. 2019), https://ssrn.com/abstract=3261273; Jason O'Connor & Nathan E. Wilson, *Reduced Demand Uncertainty and the Sustainability of Collusion: How AI Could Affect Competition*, INFO. ECON. AND POL'Y (Sept. 5, 2020), https://www.sciencedirect.com/science/article/abs/pii/S0167624520301268?via%3Dihub.

their own machine-learning technology in an effort to detect collusion.[99] Indeed, the United Kingdom's Competition and Markets Authority is already deploying online price monitoring in an effort to detect illegal resale price maintenance.[100]

Even absent collusion, algorithms can fuel personalized pricing practices that may alter the competitive dynamics of a market in ways that harm consumers, for example through supra-competitive prices.[101] As more data is collected about consumers, pricing algorithms may be able to help sellers better gauge a consumer's maximum willingness to pay.[102] For example, in 2015, the job-matching firm ZipRecruiter, changed its flat $99 subscription fee to range of fees decided customer by customer based on data provided in a survey by the potential customer. By basing the price on indicia about each customer's willingness to pay, the company's profits increased 84 percent compared to the price of

---

[99] *See, e.g.*, Giovanna Massarotto & Ashwin Ittoo, *Gleaning Insight from Antitrust Cases Using Machine Learning*, 1 STAN. COMPUTATIONAL ANTITRUST 17 (2021), https://law.stanford.edu/wp-content/uploads/2021/03/Computational-Antitrust-Article-2-Gleaning-Insight-1.pdf; Thibault Schrepel, *Computational Antitrust: An Introduction and Research Agenda* (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3766960 (laying out a broad vision for the use of machine learning to enhance antitrust investigation and enforcement).

[100] Simon Nichols, *Restricting Resale Prices: How We're Using Data to Protect Customers*, U.K. COMPETITION AND MKTS. AUTH. (June 29, 2020), https://competitionandmarkets.blog.gov.uk/2020/06/29/restricting-resale-prices-how-were-using-data-to-protect-customers/.

[101] Personalized pricing is a form of what antitrust doctrine refers to as price discrimination: charging different consumers different prices. First-degree price discrimination involves presenting a personalized price exactly equal to the consumer's willingness to pay. This type of pricing is rare and extremely difficult to achieve. Second-degree price discrimination refers to pricing based on the volume of goods or services purchased; and third-degree price discrimination is pricing based on characteristics of consumers or groups of consumers that provide indicia of customers' willingness to pay. The goal underpinning all forms of price discrimination is to charge consumers the maximum price that the consumer is willing to pay for a product. *See* HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY: THE LAW OF COMPETITION AND ITS PRACTICE 769 – 72 (5th ed. 2016) (comparing the profitability of various forms of price discrimination).

[102] Rafi Mohammed, *How Retailers Use Personalized Prices to Test What You're Willing to Pay*, Harvard Business Review (Oct. 20, 2017), https://hbr.org/2017/10/how-retailers-use-personalized-prices-to-test-what-youre-willing-to-pay.

$99.[103] This practice does not always result in price increase, but it can. While in this case a majority of customers enjoyed a price reduction, future algorithmic advances could allow firms to precisely target willingness to pay and pocket all consumer surplus as profit. Furthermore, increased prices are not the only potential harm from price discrimination.

In addition to changing the competitive landscape, personalized pricing can also implicate broader discrimination concerns surrounding the use of algorithms.[104] Algorithm-enabled personalized pricing may also lead to a "fracturing" of relevant product markets for purposes of merger analysis and increase the possibility of harm to particular groups of consumers. Antitrust enforcers may need to examine numerous markets in order to fully capture the potential competitive harm to specific groups of consumers, especially those targeted to consumers who may be uniquely vulnerable to harm. Such targeted groups of consumers may also disproportionately fall into protected classes.[105]

---

[103] Dube & Misra, *supra* note 96.

[104] If pricing is personalized to the degree of discriminating against consumers based on race, religion, gender, or national origin, this could violate antidiscrimination laws. Pricing decisions based on consumer data may also have disparate impacts on protected classes. *See infra* Part III, Section B; *see also Personalized Pricing in the Digital Era – Note by the United States*, DAF/COMP/WD(2018)140 (Nov. 2018), https://www.ftc.gov/system/files/attachments/us-submissions-oecd-2010-present-other-international-competition-fora/personalized_pricing_note_by_the_united_states.pdf; Claire Kelloway, *Personalization or Price Discrimination?*, OPEN MKTS. INST. (Jan. 30, 2020) https://www.openmarketsinstitute.org/publications/personalization-price-discrimination. As a possible example, ProPublica found that when Princeton Review priced its SAT prep services based on zip code, Asians were twice as likely to get a higher price than non-Asians. *See* Julia Angwin et al., *When Algorithms Decide What You Pay*, PROPUBLICA (Oct. 5, 2016), https://www.propublica.org/article/breaking-the-black-box-when-algorithms-decide-what-you-pay ("Propublica's analysis found that Asians were nearly twice as likely to get that higher price from The Princeton Review than non-Asians. Asians make up 4.9 percent of the U.S. population overall, but they accounted for more than 8 percent of the population in areas where The Princeton Review was charging higher prices for its SAT prep packages.").

[105] McSweeny & O'Dea, *The Implications of Algorithmic Pricing for Coordinated Effects Analysis and Price Discrimination Markets in Antitrust Enforcement*, 32 ANTITRUST 75 (Fall 2017) ("A merger that might previously have required an analysis of competitive effects in one relevant product market may

Of course, the concerns of antitrust law extend well beyond pricing, especially in the kinds of data-dependent digital markets that broadly deploy algorithms. The accumulation and concentration of vast amounts of data can entrench incumbents and create barriers to entry. As these firms accumulate data, they can use it to better train algorithms that may give them an enduring advantage—to the point where new entrants may never be able to reach the scale that is needed to meaningfully compete.[106] In short, lack of access to data may become a lasting barrier to entry in algorithm-driven product markets.[107]

Algorithms also play a significant role in the types of antitrust complaints that have been raised against dominant digital platforms over exclusionary conduct, such as self-preferencing or manipulation of search results.[108] These complaints reflect the fact that digital platforms are able to deploy algorithms such that opaque business decisions that appear neutral may in fact benefit the platform at the expense of their (especially vertical)

---

instead require antitrust enforcers to examine dozens, if not hundreds, of potential relevant product markets. . . . Even if the majority of consumers would not be negatively affected by the proposed transaction, however, it may nonetheless be appropriate to define a price discrimination market for 'product consumers who live in households without a vehicle.'").

[106] Terrell McSweeny & Brian O'Dea, *Data, Innovation, and Potential Competition in Digital Markets – Looking Beyond Short-Term Price Effects in Merger Analysis*, 2018 CPI ANTITRUST CHRONICLE 3 (Feb. 2018) (noting that data and analytics capabilities can create "self-reinforcing" barriers to entry in digital markets and that competition enforcers thus "should pay particularly close attention to whether a merger would enhance data-related barriers to entry – even if short-term price effects are unlikely").

[107] KATHARINA PISTOR, THE CODE OF CAPITAL: HOW THE LAW CREATES WEALTH AND INEQUALITY 126–31 (2019).

[108] *See* H. Antitrust Report on Competition in Digital Markets, https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519; U.K. Competition & Markets Authority, *Algorithms: How They Can Reduce Competition and Harm Consumers*, GOV.UK (Jan. 19, 2021), https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers.

rivals. This may harm competition as well as consumers who might be inhibited from accessing a broader range of content.[109]

As algorithms and AI systems become more sophisticated and pervasive, competition enforcers must closely monitor how they affect the competition in a wide variety of markets, particularly in digital markets already dominated by powerful incumbents. Identifying and preventing these harms may require different investigative strategies and further collaboration among enforcers, academics, and advocates.[110] As these technological tools evolve, enforcement strategies must keep pace.[111]

III. Using the FTC's Current Authorities to Better Protect Consumers

There is no question that the critical algorithmic problems identified—faulty inputs, faulty conclusions, failure to adequately test, proxy discrimination, surveillance capitalism, and threats to competition—undermine rather than advance economic justice. But if these algorithmic problems can be addressed or mitigated by effective solutions, consumers and competition might benefit on net from algorithmic innovations. Throughout this article, I have suggested that the pitfalls of AI and algorithmic decision-making are not wholly different from other problems regulators have confronted for many years. In this section,

---

[109] *See, e.g.*, Complaint at 70, Colorado v. Google, No. 1:20-cv-03715 (D.D.C. Dec 17, 2020), https://coag.gov/app/uploads/2020/12/Colorado-et-al.-v.-Google-PUBLIC-REDACTED-Complaint.pdf ("Google has the incentive, power, and control to utilize this systematic multipronged discriminatory attack against specialized vertical providers operating in any vertical market of Google's choosing. Google's misconduct undermines competition, harms advertisers who wish to buy general search advertising, and hurts consumers who both face unjustified obstacles in reaching content that may be valuable to them and ultimately assume costs of higher advertising that are passed along to them.").

[110] *See, e.g.*, U.K. Competition & Markets Authority, *supra* note 108; Schrepel, *supra* note 99; James Niels Rosenquist, Fiona M. Scott Morton & Samuel N. Weinstein, *Addictive Technology and its Implications for Antitrust Enforcement*, Y. SCH. MGMT. (Sept. 2020), https://som.yale.edu/sites/default/files/Addictive-Technology.pdf.

[111] In addition to threats from algorithms and AI, other emerging technologies also merit close examination from competition enforcers. *See, e.g.*, Thibault Schrepel, *Collusion by Blockchain and Smart Contracts*, 33 HARV. J.L. & TECH. 117 (2019), http://jolt.law.harvard.edu/assets/articlePDFs/v33/03-Schrepel.pdf.

I consider how enforcers—and the FTC in particular—can use current authority to address these new fact patterns.

Civil rights laws are the logical starting point for addressing discriminatory consequences of algorithmic decision-making. Our state and federal civil rights laws already prohibit discrimination in each of the areas discussed—health care, employment, housing, and credit.[112] None of these laws specifically contemplates discrimination arising in the context of automated decisions relying on vast fields of proxy-rich data. Nor do they allow discrimination simply because it involved an algorithm. "Because AI" is neither an explanation nor an excuse. It is incumbent on law enforcers to think creatively about how to apply existing civil rights law to these new fact patterns.

But not all relevant law enforcement agencies have explicit civil-rights authorities. And, in many cases, existing civil-rights jurisprudence may be difficult to apply to algorithmic bias precisely because black-box opacity makes demonstrating discrimination (already a high bar) even more difficult. So, we must consider what other legal protections currently exist outside of direct civil rights statutes.

The FTC has four types of enforcement authority that provide the agency with some ability to protect consumers and promote economic justice in the face of algorithmic harms: our general authority under the FTC Act; sector-specific rules and statutes, such as FCRA, ECOA, and COPPA; the study authority of section 6(b); and the rulemaking authority of section 18.

A. Section 5 of the FTC Act

Most of the enforcement activity conducted by the FTC is brought under the general authority provided to the Commission by section 5 of the FTC Act, which prohibits unfair or deceptive acts or practices. The Act is more than a century old, and since its passage,

---

[112] *See, e.g.*, 42 U.S.C. § § 2000d – d-7 (2018) (prohibiting discrimination in health care); *id.* § § 2000e – 2000e-17 (prohibiting discrimination in employment); *id.* § § 3601 – 3691 (prohibiting discrimination in housing); 15 U.S.C. § § 1691 – 1691f (2018) (prohibiting credit discrimination).

the agency has been able to apply the statute's general language to meet new enforcement challenges. That same approach urgently needs to be applied to algorithms.

One innovative remedy that the FTC has recently deployed is algorithmic disgorgement. The premise is simple: when companies collect data illegally, they should not be able to profit from either the data or any algorithm developed using it. This novel approach was most recently deployed in the FTC's case against Everalbum in January 2021.[113] There, the Commission alleged that the company violated its promises to consumers about the circumstances under which it would deploy facial-recognition software.[114] As part of the settlement, the Commission required the company to delete not only the ill-gotten data but also any facial recognition models or algorithms developed with users' photos or videos. The authority to seek this type of remedy comes from the Commission's power to order relief reasonably tailored to the violation of the law.[115] This innovative enforcement approach should send a clear message to companies engaging in illicit data collection in order to train AI models: Not worth it.[116]

---

[113] Decision and Order at 4–5, Everalbum, Inc., 2021 WL 118892 (F.T.C. Jan. 11, 2021), https://www.ftc.gov/system/files/documents/cases/everalbum_order.pdf; *see also* Final Order at 4, Cambridge Analytica, LLC, F.T.C. File No. 1823107 (Dec. 6, 2019), https://www.ftc.gov/system/files/documents/cases/d09389_comm_final_orderpublic.pdf.

[114] Complaint at 6–7, Everalbum, Inc., 2021 WL 118892 (F.T.C. Jan. 11, 2021), https://www.ftc.gov/system/files/documents/cases/everalbum_complaint.pdf.

[115] The Commission may seek injunctions containing provisions "that are broader than the conduct that is declared unlawful." Telebrands Corp. v. FTC, 457 F.3d 354, 357 n.5 (4th Cir. 2006). FTC orders are not limited to prohibiting the "narrow lane" of a wrongdoer's past violations but may effectively "close all roads to the prohibited goal." FTC v. Ruberoid Co., 343 U.S. 470, 473 (1952).

[116] For a more thorough explanation of my views on the importance of specific and general deterrence for effective enforcement, *see, e.g.*, Dissenting Statement of Commissioner Rebecca Kelly Slaughter Regarding the Matter of *FTC vs. Facebook*, (F.T.C. July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536918/182_3109_slaughter_statement_on_facebo ok_7-24-19.pdf.

The agency can also use its deception authority in connection with algorithmic harms where the marketers of products or services represent that they can use machine-learning technology in unsubstantiated ways, such as to identify or predict which job candidates will be successful or will outperform other candidates. Deception enforcement is well-trodden ground for the FTC; when a company makes claims about the quality of its products or services, whether or not those products are related to AI, the law requires such statements to be supported by verifiable substantiation.[117]

Finally, the FTC can use its unfairness authority to target algorithmic injustice. The unfairness prong of the FTC Act prohibits conduct that causes or is likely to cause substantial injury to consumers, where that injury is not reasonably avoidable by consumers and not outweighed by countervailing benefits to consumers or to competition.[118] A number of factual predicates could give rise to an unfairness claim in connection with algorithmic harms. For example, secretly collecting audio or visual data—or any sensitive data—about an individual to feed an algorithm could give rise to an unfairness claim.[119] In addition, if an algorithm is used to exclude a consumer from a benefit or an opportunity based on her actual or perceived status in a protected class, such conduct could also give rise to an unfairness claim.[120]

The FTC can and should be aggressive in its use of unfairness to target conduct that harms consumers based on their protected status. Unfairness is an imperfect tool, introducing the hurdles of "reasonable avoidability" and "countervailing benefits" into what can already be a complicated question of the specific injury caused by disparate outcomes. That it has limitations does not mean, however, that the FTC's unfairness

---

[117] *See supra* note 25.

[118] 15 U.S.C. § 45(n) (2018).

[119] *See, e.g.*, Complaint at ¶¶ 31–35, FTC v. VIZIO, Inc., No. 2:17-cv-00758-SRC-CLW, 2017 WL 7000553 (D.N.J. Feb. 6, 2017) https://www.ftc.gov/system/files/documents/cases/170206_vizio_2017.02.06_complaint.pdf.

[120] *See, e.g.*, Jillson, *supra* note 71.

authority cannot be used to combat the fundamentally unfair phenomenon of unlawful discrimination, as well as some of the other algorithmic harms discussed above.

B. Vigorous Enforcement of ECOA & FCRA

The FTC also enforces two sector-specific laws that afford protections related to the extension of credit and their credit information, both of which are relevant to consumers navigating algorithms related to the credit sphere.[121]

First, the FTC enforces the Equal Credit Opportunity Act, which prohibits credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because an applicant receives income from public assistance or has in good faith exercised any right under the Consumer Credit Protection Act.[122] Everyone who regularly participates in a credit decision, including setting the terms of that credit, and those who arrange financing (such as real estate brokers), must comply with ECOA's antidiscrimination protections.[123] Under this framework, if a creditor uses proxies to determine which consumers to target for high-interest credit, and such proxies correlate

---

[121] *See generally* Fed. Trade Comm'n, *Big Data: A Tool for Inclusion or Exclusion?*, FTC (Jan. 2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[122] 15 U.S.C. § § 1691 – 1691f (2018).

[123] To prove a violation of ECOA's antidiscrimination protections, plaintiffs typically must show "disparate treatment" or "disparate impact." Disparate treatment occurs when a creditor intentionally treats an applicant differently based on a protected characteristic. Disparate impact, on the other hand, occurs when a company employs facially neutral policies or practices that have a disproportionate adverse effect on a protected class—regardless of the company's intent—unless those practices further a legitimate business need that cannot reasonably be achieved by means that are less disparate in their impact. Even if evidence shows the company's decisions are justified by a business necessity, if there is a less discriminatory alternative, the decisions may still violate ECOA. *See* Fed. Trade Comm'n, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? (Jan. 2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf; *supra* Section II, Part B, Proxy Discrimination.

with protected class membership, the creditor may be violating ECOA.[124] The FTC should investigate such conduct and, if appropriate, vigorously pursue enforcement.

In addition to enforcement, one useful approach would be to encourage creditors to make use of the ECOA exception that permits the collection of demographic information to test their algorithmic outcomes. Regulation B, which implements ECOA, presumptively prohibits the collection of protected-class demographic information, unlike Regulation C, which implements the Home Mortgage Disclosure Act (HMDA), and generally requires collection of all demographic data for mortgages. The result of these rules is that mortgage credit is monitored closely in a race-conscious way, but the demographic information of all other credit is supposed to go unmonitored. The benevolent idea behind ECOA, of course, was that gender- and race-blind lending would eliminate gender and race disparities. If only that were the case; experience shows that gender and race disparities substantially persist, often because of proxy discrimination.[125] I believe that as with mortgage data, all other kinds of credit should be monitored by creditors consciously for disparities on the basis of protected status.

Regulation B already contains an exception that permits collecting demographic data when it is "for the purpose of conducting a self-test,"[126] which is defined as any inquiry "designed and used specifically to determine the extent or effectiveness of a creditor's compliance with the Act or this part."[127] In short, ECOA permits, and the FTC

---

[124] If, for example, a company made credit decisions based on consumers' zip codes, resulting in a "disparate impact" on members of a protected class, the FTC could challenge that practice under ECOA. *See* Andrew Smith, *Using Artificial Intelligence and Algorithms*, FTC (Apr. 8, 2020), https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms.

[125] *See supra* Section II, Part B, Proxy Discrimination.

[126] 12 C.F.R. § 1002.5(b)(1).

[127] *Id.* at § 1002.15(b)(1)(i). Regulation B sets various requirements for the self-test. *See generally id.* at § 1002.15.

should encourage, non-mortgage creditors to collect demographic data on most borrowers and use it to test algorithmic systems to reduce disparities.[128]

Vanishingly few creditors take advantage of this exception. We do not know why this is the case but can speculate that perhaps it is because they fear that their collection of the data will validate or exacerbate claims that their decisions are biased. In other words, a creditor's visibility into demographics might be seen as crediting the existence of their bias; this perspective is in line with the idea that race-blindness is the same as race-neutrality. Creditors often find it much easier to never ask about race or gender, or to use (as civil-rights enforcers and private plaintiffs generally must for non-HMDA credit) the Bayesian Improved Surname Geocode algorithm to proxy for race, national origin, and gender in datasets of borrowers to self-test for disparities and fair-lending risk.[129] But the collection of demographic data for the purpose of self-testing is not a sign of bias, as long as it is clear that the data is actually and only being used for that purpose. Enforcers should see self-testing (and responsive changes to the results of those tests) as a strong sign of good-faith efforts at legal compliance and a lack of self-testing as indifference to alarming credit disparities. Of course, if creditors do collect this data to conduct self-testing, they must be able to show that they are not also using it for impermissible purposes such as marketing.

The FTC also enforces the Fair Credit Reporting Act, which applies to consumer reporting agencies (CRAs) that compile and sell consumer reports containing consumer information that is used or expected to be used for credit, employment, insurance, housing, or other similar decisions about consumers' eligibility for certain benefits and transactions.

---

[128] *See, e.g.*, Grace Abuhamad, *The Fallacy of Equating 'Blindness' with Fairness: Ensuring Trust in Machine Learning Applications to Consumer Credit*, MIT (May 15, 2019), https://dspace.mit.edu/bitstream/handle/1721.1/122094/1117710058-MIT.pdf?sequence=1&isAllowed=y; Miranda Bogen, Aaron Rieke, Shazeda Ahmed, *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination*, 2020 PROC. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 492, (Jan. 2020), https://dl.acm.org/doi/abs/10.1145/3351095.3372877.

[129] *See* Marc N. Elliott et al., *Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities*, 9 HEALTH SERVS. & OUTCOMES RES. METHODOLOGY 69 (2009).

Currently, FCRA provides several important protections for consumers. As with ECOA, companies that rely on consumer report information in making credit, housing, and other decisions must provide adverse action notices following a negative decision—for example, when a housing application is denied. These notices tell consumers about their right to see information reported about them and to dispute inaccurate information.[130]

FCRA also requires CRAs to apply reasonable procedures to ensure maximum possible accuracy when preparing consumer reports. Several companies that provide tenant screening services, including by using automated systems,[131] have faced serious consequences for allegedly failing to adhere to this standard in recent years.[132] Ensuring the accuracy of program inputs will be particularly important as companies adopt even more complex decision-making platforms, including those driven by algorithms.[133]

Still, more research is needed to understand the limitations of adverse action notices under ECOA and FCRA in providing sufficient information to consumers about application denials, particularly in the context of AI.[134] The complexity of algorithmic

---

[130] 15 U.S.C. § 1681m(a) (2018); *see also* Smith, *supra* note 124 (FCRA adverse action requirements would apply where a third-party vendor is a CRA).

[131] *See, e.g.*, FTC v. RealPage, Inc., No. 3:18-cv-02737-N (N.D. Tex. 2018).

[132] *See, e.g.*, United States v. AppFolio, Inc., No. 1:20-cv-03563 (D.D.C. Dec. 8, 2020); RealPage, Inc., No. 3:18-cv-02737-N.

[133] *See* Smith, *supra* note 124.

[134] The Consumer Financial Protection Bureau has also been exploring these issues and recently held a tech "sprint" aimed at improving consumer adverse action notices. *See* Albert Chang, Tim Lambert & Jennifer Lassiter, *CFPB's First Tech Sprint on October 5– 9, 2020: Help Improve Consumer Adverse Action Notices*, CONSUMER FIN. PROTECTION BUREAU BLOG (Sept. 1, 2020), https://www.consumerfinance.gov/about-us/blog/cfpb-tech-sprint-october-2020-consumer-adverse-action-notices/; Patrice Alexander Ficklin, Tom Pahl & Paul Watkins, *Innovation Spotlight: Providing Adverse Action Notices when Using AI/ML Models*, CONSUMER FIN. PROTECTION BUREAU BLOG (July 7, 2020), https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models.

decision-making poses unique challenges in this area.[135] Expanding data-reporting requirements under FCRA—for example, broader reporting on the existence and correction of errors, the rates of adverse action notices, and the volume and nature of error complaints—could also help mitigate problems that arise in algorithmic decision-making by providing visibility into the effects of those decisions. A significant limitation of the adverse action notice is that a consumer has access only to information about her own negative outcome, making it difficult to know if there are systematic denials taking place across protected classes.[136]

C. COPPA

The FTC also enforces COPPA,[137] which can be used to protect children from certain data abuses. The law empowers the FTC to write rules that mandate disclosures by websites directed at children, that prohibit website operators from coercing children to disclose excessive data, and that require website operators to use certain safeguards to protect children's data.[138] The FTC is currently reviewing the COPPA Rule. In addition to hosting a workshop on the topic,[139] the agency requested public comments to help inform this effort, producing much thoughtful and substantive input.[140] Interested readers should

---

[135] *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841 (2018); Louise Matsakis, *What Does a Fair Algorithm Actually Look Like?*, WIRED (Oct. 11, 2018), https://www.wired.com/story/what-does-a-fair-algorithm-look-like/.

[136] *See* Selbst & Barocas, *supra* note 135, at 1105.

[137] 15 U.S.C. § § 6501 – 6506 (2018).

[138] *Id.* at § 6502(b). *See also supra* note 83.

[139] *See* Federal Trade Commission, *Future of the COPPA Rule: An FTC Workshop*, FTC (Oct. 9, 2019), https://www.ftc.gov/news-events/events-calendar/future-coppa-rule-ftc-workshop.

[140] Fed. Trade Comm'n, Request for Public Comment on the Federal Trade Commission's Implementation of the Children's Online Privacy Protection Rule (July 25, 2019), https://www.regulations.gov/document/FTC-2019-0054-0001/comment.

see the Federal Register notice to get a sense of some of the questions under consideration.[141]

D. Section 6(b) of the FTC Act

Another tool at the FTC's disposal is the ability to write reports informed by studies conducted under section 6(b) of the FTC Act.[142] This provision gives the FTC the opportunity to study in depth how algorithms and related technologies are being deployed and how we can effectively adapt to combat their harms. This provision also empowers the Commission to require an entity to file "annual or special . . . reports or answers in writing to specific questions" to provide information about the entity's "organization, business, conduct, practices, management, and relation to other corporations, partnerships, and individuals."[143] In addition to collecting information

---

[141] Request for Public Comment on the Federal Trade Commission's Implementation of the Children's Online Privacy Protection Rule, 84 Fed. Reg. 35842 (proposed July 25, 2019), https://www.federalregister.gov/documents/2019/07/25/2019-15754/request-for-public-comment-on-the-federal-trade-commissions-implementation-of-the-childrens-online.

[142] *See, e.g.*, Fed. Trade Comm'n, Order to File Special Report, FTC Matter No. P144504 (Feb. 11, 2021), https://www.ftc.gov/system/files/documents/reports/order-file-special-section-6b-report-e-cigarette-products-calendar-year-2021-generic-text-version/generic_e-cigarette_order_2021-02-11.pdf (report to study advertising practices for e-cigarette products); *See* Fed. Trade Comm'n, Order to File Special Report, FTC Matter No. P104518 (Apr. 12, 2012), https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-orders-alcoholic-beverage-manufacturers-provide-data-agencys-fourth-major-study-alcohol/120412alcoholreport.pdf (report to study advertising practices for alcoholic beverages).

[143] 15 U.S.C. § 46(b) (2018). Such reporting requirements have the potential to focus corporate attention and resources on issues that were previously neglected. They also provide information that allows consumers, other firms, and investors to encourage companies—through both advocacy and consumer spending and investment decisions—to improve their practices. *See, e.g.,* California Transparency in Supply Chains Act, CAL. CIV. CODE § 1714.43 (requiring businesses to disclose their efforts to eradicate slavery and human trafficking from their direct supply chains); Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 amending Directive 2013/34/EU as regards to disclosure of non-financial and diversity information by certain large undertakings and groups, 2014 O.J. (L 330) (requiring companies meeting certain criteria to publish information on their social and environmental practices

about specific businesses, the FTC can also collect information about industry-wide phenomena.[144]

The study of social media and video-streaming services discussed above is one exciting use of this important tool.[145] The FTC should continue to use its 6(b) authority to deepen its expertise on the use and impact of algorithms in our modern economy, focusing on the potential harms to consumers and competition.[146]

Whether in enforcement or with industry studies, the agency always strives to keep pace with emerging technologies and changing markets.[147] But particularly given the scale, opacity, and rapid proliferation of algorithmic decision-making in our economy, there is room for improvement. Specifically, the agency needs more resources and a broader range of in-house expertise. Improved accountability in this space requires a sophisticated understanding of the underlying technologies and the business models that employ them. The FTC could more effectively study and hold companies accountable if it could hire additional staff, including those who would expand the agency's analytical framework, such as technologists and market specialists. This increased capacity would support more effective and systematic investigations into data abuses, including those perpetrated through algorithms.

---

including sustainability, treatment of employees, respect for human rights, anti-corruption/bribery, and diversity on company boards).

[144] *See, e.g.*, FED. TRADE COMM'N, PATENT ASSERTION ENTITY ACTIVITY: AN FTC STUDY 1 – 2 (2016), https://www.ftc.gov/system/files/documents/reports/patent-assertion-entity-activity-ftc-study/p131203_patent_assertion_entity_activity_an_ftc_study_0.pdf (broadly studying the Patent Assertion Entities and the industry around them).

[145] *See supra* Section II, Part B, end of Surveillance Capitalism subsection.

[146] States and localities are also studying the harms associated with AI and algorithms, as are some of our foreign counterparts. *See, e.g.*, U.K. Competition & Markets Authority, *supra* note 108; *State Artificial Intelligence Policy*, ELEC. PRIVACY INFO. CTR., https://epic.org/state-policy/ai/ (last visited Mar. 1, 2021) (identifying three states and one city conducting studies of AI).

[147] *See supra* Section II; *see also* Jillson, *supra* note 71.

IV. New Legislative and Regulatory Solutions

In addition to comprehensively and aggressively using all the enforcement tools currently at the FTC's disposal, it is also worth considering where there are gaps in its authority that can and should be addressed by legislative and regulatory solutions. Fortunately, many academics, advocates, and policymakers across the country—and around the globe—are grappling with these same questions. This section identifies some elements of promising legislative and regulatory solutions, which can powerfully complement the tools described above.

A. Guiding Principles

Many of the harms that seem novel have long-standing analogs, but algorithmic decision-making presents special risks because it can simultaneously obscure the problems and amplify them, all while giving the impression that they do not or could not possibly exist. For these reasons and others, I believe that any viable system for addressing algorithmic harms should require, at a minimum, three critical principles: transparency, fairness, and accountability.

In nearly every problematic example highlighted,[148] it is unclear precisely which inputs and decisions produced the biased or otherwise harmful outcome. Proprietary algorithmic models are often cloaked in secrecy and have limited human input,[149] and frustration with the opacity of the "black box" can lead consumers to feel powerless and distrustful.[150] At the same time, the patina of neutral technology making decisions leads to a sense that deployers or developers of bad algorithms should not be responsible for the results. The combination of black-box obscurity with the widespread application of

[148] *See supra* Section II.

[149] *See, e.g.*, Danielle Keates Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 5 (2014).

[150] *See, e.g.*, Jennifer Cannon, *Report Shows Consumers Don't Trust Artificial Intelligence*, FINTECH NEWS (Dec. 4, 2019), https://www.fintechnews.org/report-shows-consumers-dont-trust-artificial-intelligence/.

complicated and facially neutral technology provides a false sense of security in the objectivity of algorithmic decision-making.

Increasing transparency lifts the curtain on these opaque processes. A longtime staple of our regulatory system,[151] transparency requires the developers and deployers of algorithmic systems to make sure that automated decisions are as explainable and defensible as possible. With the benefit of sunlight, advocates, academics, and other third parties can more widely test for discriminatory and harmful outcomes.[152] Transparency about companies' data practices can also enable consumers to "vote with their feet."[153] And in some cases more transparency may empower consumers and advocates to challenge incorrect or unfair outcomes.

This type of transparency can be effectively incorporated into a regulatory framework. The European Union's General Data Protection Regulation (GDPR), for example, anchors its AI protections in increased transparency requirements. Under GDPR, the use of automated decision-making about individuals, including profiling, that produces legal or similarly significant effects triggers certain obligations for data controllers.[154] Controllers must give individuals specific information about the process, and they must take steps to prevent errors, bias, and discrimination. GDPR also gives individuals the right to challenge and request a review of the decision—sometimes referred to as "the right to an explanation."[155]

---

[151] LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 94 (1914).

[152] This work is already ongoing but is dependent on the accessibility of data.

[153] *See* Rebecca Kelly Slaughter, Acting Chair, Fed. Trade Comm'n, Protecting Consumer Privacy in a Time of Crisis, Remarks to Future of Privacy Forum 3 (Feb. 10, 2021), https://www.ftc.gov/system/files/documents/public_statements/1587283/fpf_opening_remarks_210_.pdf.

[154] Commission Regulation 2016/679, art. 22, 2016 O.J. (L 119) (EU).

[155] *Id.* The State of Illinois recently passed a law that seeks to introduce similar transparency into certain hiring decisions. Artificial Intelligence Video Interview Act, H.B. 2557, 101st Gen. Ass. (Ill. 2019) (enacted) (requiring employers to (1) notify each applicant that AI may be used to analyze the applicant's video interview and consider the applicant's fitness for the position; (2) provide each applicant with information

Effective transparency, however, must provide meaningful and intelligible information; it cannot simply overwhelm a user with information and trigger decision fatigue. The effects of extensive notice provisions in privacy laws like GDPR must be studied carefully to ensure they fall into the former and not the latter category. If the result of frequent pop-up notices is to nudge a user into simply accepting the practice about which she is being informed, with no opportunity to exercise choice, that nominal transparency may have no benefit whatsoever.

In addition to requiring transparency, we must also endeavor to limit—or, even better, prohibit—unfair and discriminatory applications of algorithms. In this context, fairness is sometimes difficult to define,[156] but action in multiple jurisdictions reflects an understanding that addressing discrimination is critical to any framework for regulating AI.[157] For example, the EU has issued guidelines that list seven key requirements that AI systems should meet to be trustworthy, including transparency, diversity, nondiscrimination and fairness, and accountability.[158] In the United States, the Biden Administration has already taken steps to encourage the federal government to prioritize

---

explaining how the AI works and what general types of characteristics it uses to evaluate applicants; and (3) obtain consent from each applicant to be evaluated by the AI program.). *See also* Online Privacy Act of 2019, H.R. 4978, 116th Cong. § 105 (2019) (establishing a right to human review of automated decisions).

[156] *See, e.g.*, Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, 54 ACM Computing Surveys 1 (July 2021), https://arxiv.org/pdf/1908.09635.pdf; Osonde A. Osoba et al., *Algorithmic Equity: A Framework for Social Applications*, RAND CORP. (2019), https://www.rand.org/pubs/research_reports/RR2708.html.

[157] For example, nondiscrimination is a core element of the European Commission's recently proposed regulations on AI and algorithms. *See Commission Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* COM (2021) 206 final (Apr. 21, 2021).

[158] *See* High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, EUR. COMMISSION 2 (2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation; *see also* High-Level Expert Group on Artificial Intelligence, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, EUR. COMMISSION (July 17, 2020), https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

the increase of transparency, equity, and accountability in its work—including through Executive Orders.[159]

Prioritizing transparency and fairness is necessary, but not sufficient; regulation of algorithmic decision-making must also involve real accountability and appropriate remedies. Increased accountability means that companies—the same ones that benefit from the advantages and efficiencies of algorithms—must bear the responsibility of (1) conducting regular audits and impact assessments and (2) facilitating appropriate redress for erroneous or unfair algorithmic decisions.

The principles of transparency, fairness, and accountability can inform much of the FTC's case-by-case enforcement work that implicates algorithmic decision-making. But given the breadth and depth of the algorithmic harms described above, these principles should also animate FTC rulemaking under section 18 or congressional action.

B. Section 18 Rulemaking Initiative

The FTC already possesses the means to address algorithmic harms on a forward-looking basis: our rulemaking authority under section 18 of the FTC Act, added by the Magnuson-Moss Warranty—Federal Trade Commission Improvement Act.[160] This tool,

---

[159] *See* Exec. Order No. 14036, 86 Fed. Reg. 36987 (July 9, 2021), https://www.federalregister.gov/documents/2021/07/14/2021-15069/promoting-competition-in-the-american-economy; Exec. Order No. 13,985, 86 Fed. Reg. 7009 (Jan. 20, 2021), https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/. These values were also espoused in relation to AI in an OMB memorandum to executive departments under the previous administration. This document set forth ten principles for agencies to weigh when considering regulatory and non-regulatory approaches to the design, development, deployment, and operation of AI applications. One of these key principles is "Fairness and Non-Discrimination," and specifically, OMB advised agencies to consider "whether the AI application at issue may reduce levels of unlawful, unfair, or otherwise unintended discrimination as compared to existing processes." *See* Memorandum from Russell T. Vought, Director, Office of Mgmt. & Budget, to the Heads of Executive Departments & Agencies (Nov. 17, 2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

[160] Pub. L. No. 93-637, 88 Stat. 2183 (1975).

in conjunction with the help and advice of subject-matter experts, empowers the FTC to address AI-driven harms prospectively.

With the 1975 passage of Magnuson-Moss, the FTC's rulemaking procedures diverged from those of sister agencies. Beginning in the 1960s, the FTC promulgated Trade Regulation Rules using procedures established under the Administrative Procedure Act (APA).[161] While these efforts were successful,[162] they were shrouded in questions of whether the FTC Act delegated authority to the Agency to promulgate binding regulations.[163] Magnuson-Moss provided definitive legislative affirmation of the FTC's rulemaking power.[164]

The procedures required to issue a rule under section 18 are more cumbersome than under the APA. The statute requires the additional steps of a pre-rulemaking advance notice-and-comment period, special notifications of Congress, and "informal hearings" to consider disputed issues of material fact, among other logistical hurdles.[165] But these challenges are not insurmountable: Initially, the Commission was successful at

---

[161] Note, *The Federal Trade Commission: Modes of Administration*, 80 HARV. L. REV. 1063, 1091 (1967).

[162] *See, e.g.*, *Unfair or Deceptive Advertising and Labeling of Cigarettes in Relation to the Health Hazards of Smoking*, 29 Fed. Reg. 8324, 8325 (July 2, 1964) (requiring cigarette labels and advertising to clearly disclose the health hazards of smoking). Note that Congress eventually intervened with a statute to supplant the FTC regulation. Federal Cigarette Labeling and Advertising Act of 1965, 15 U.S.C. § § 1331-41 (2018).

[163] *Modes of Administration*, *supra* note 161, at 1092. These legal questions were ultimately decided in the agency's favor. *See* Nat'l Petroleum Refiners Ass'n v. FTC, 482 F.2d 672, 698 (D.C. Cir. 1973).

[164] *Oversight Hearings into the Fed. Trade Comm'n—Bureau of Consumer Protection Before the Subcomm. of the Comm. On Government Operations*, 94th Cong. 60 (1976) (statement of Paul Rand Dixon, Acting Chairman, Fed. Trade Comm'n).

[165] 15 U.S.C. § 57a(b) (2018). Note that while much of the FTC's ability to make rules now falls under section 18, Congress has granted the FTC the ability to promulgate rules under standard APA notice-and-comment procedures in discrete arenas. *See, e.g.*, 15 U.S.C. § 45a (2018) (granting notice-and-comment rulemaking authority to regulate the use of "Made in the U.S.A." or "Made in America" labels); 12 U.S.C. § 5519(d) (2018) (authorizing notice-and-comment rulemaking authority to regulate "motor vehicle dealer[s]").

promulgating rules under section 18, resulting in a variety of rules that protect consumers.[166] In recent years, however, the Commission has shied away from extensive section 18 rulemaking.[167]

The new Democratic majority at the Commission has already taken action to make section 18 rulemaking more viable by bringing Commission procedures in line with statutory requirements and congressional intent. At its first open meeting in several decades, the Commission adopted changes to its rules of practice to remove self-imposed procedural hurdles to section 18 rulemaking.[168] These changes will help unlock section 18 rulemaking from its decades-long term in procedural prison and allow the Commission to

---

[166] *See Advertising of Ophthalmic Goods and Services*, 43 Fed. Reg. 23992, 24000 (July 3, 1978) (a successful section 18 rulemaking regulating the provision of ophthalmological goods and services). The Commission initiated more than a dozen new Magnuson-Moss rulemakings in five years following Magnuson-Moss's passage. MILES W. KIRKPATRICK ET AL., AM. BAR ASS'N, REPORT OF THE AMERICAN BAR ASSOCIATION SECTION OF ANTITRUST LAW SPECIAL COMMITTEE TO STUDY THE ROLE OF THE FEDERAL TRADE COMMISSION, 139 (1989), https://www.americanbar.org/content/dam/aba/administrative/antitrust_law/report_1989-ftc.pdf.

[167] *See* Kurt Walters, *Reassessing the Mythology of Magnuson-Moss: A Call to Revive Section 18 Rulemaking at the FTC*, 16 HARV. L. & POL'Y REV. (forthcoming 2022), https://ssrn.com/abstract=3875970.

[168] Revisions to Rules of Practice, 86 Fed. Reg. 38542 (July 22, 2021), https://www.federalregister.gov/d/2021-15313; Statement of the Commission Regarding the Adoption of Revised Section 18 Rulemaking Procedures, Fed. Trade Comm'n (July 9, 2021), https://www.ftc.gov/system/files/documents/public_statements/1591786/p210100commnstmtsec18rulesofpractice.pdf; Press Release, Fed. Trade Comm'n, FTC Votes to Update Rulemaking Procedures, Sets Stage for Stronger Deterrence of Corporate Misconduct (July 1, 2021), https://www.ftc.gov/news-events/press-releases/2021/07/ftc-votes-update-rulemaking-procedures-sets-stage-stronger ("'These changes show the FTC is turning the page on decades of self-imposed red-tape and returning to the participatory and dynamic process for issuing Section 18 rules that Congress envisioned. Clear rules help honest businesses comply with the law and better protect consumers and workers against bad actors. They will also lead to substantial market-wide deterrence due to significant civil penalties for rulebreakers. Streamlined procedures for Section 18 rulemaking means that the Commission will have the ability to issue timely rules on issues ranging from data abuses to dark patterns to other unfair and deceptive practices widespread in our economy.'").

fulfill its statutorily directed mission. One important area for the Commission's attention is data abuses.[169]

To be clear, rulemaking cannot target conduct that does not otherwise violate the law; in other words, the FTC cannot proscribe through rule conduct what it could not pursue through ex-post enforcement under the FTC Act. The value that rulemaking has is that it clarifies the boundaries of the law for the markets so that prohibited conduct is not exclusively identified in enforcement actions that take place after harm has already occurred.

The threats to consumers arising from data abuse,[170] including those posed by algorithmic harms, are mounting and urgent. It is imperative for the FTC to take all action within its existing authority to protect consumers. This authority includes section 18 rulemaking, which, although slow and imperfect, is available to help better protect consumers. At the very least, initiating such a rulemaking would significantly advance the public debate through targeted study, thoughtful commentary, and nuanced proposals.

---

[169] *See* Exec. Order No. 14036, 86 Fed. Reg. 36987 (July 9, 2021), https://www.federalregister.gov/documents/2021/07/14/2021-15069/promoting-competition-in-the-american-economy (recommending that the Commission exercise its statutory rulemaking authority "in areas such as unfair data collection and surveillance practices that may damage competition, consumer autonomy, and consumer privacy").

[170] *See supra* Section II; Rebecca Kelly Slaughter, Commissioner, Fed. Trade Comm'n, The Near Future of U.S. Privacy Law, Remarks to Silicon Flatirons – University of Colorado Law School (Sept. 6, 2019), https://www.ftc.gov/system/files/documents/public_statements/1543396/slaughter_silicon_flatirons_remarks_9-6-19.pdf ("Rather than simply thinking narrowly about data privacy, I want us to be thinking in terms of data abuses more broadly. Privacy generally refers to limits on the collection or sharing of data that an individual would prefer to keep private. But we cannot and should not separate problems involving collecting data about individuals from problems involving the targeting of information to individuals or other decisions made for individuals (often based on the collected data)."); Rebecca Kelly Slaughter, Commissioner, Fed. Trade Comm'n, FTC Data Privacy Enforcement: A Time for Change, Keynote Address at NYU Law School's Program on Corporate Compliance and Enforcement (Oct. 16, 2020), https://wp.nyu.edu/compliance_enforcement/2020/10/17/ftc-data-privacy-enforcement-a-time-of-change/.

In the area of algorithmic justice, a section 18 rule might affirmatively impose requirements of transparency, fairness, and accountability. As noted above, this is not an easy endeavor; it will require input and activism from the interested public.[171] I strongly urge attorneys, technologists, state attorneys general, academics, advocates, policymakers, and all other stakeholders to help the FTC craft rules to address these urgent issues.[172] A well-drafted rule could do so in a way that accounts for context and relative risk, while charting a new path to better protect consumers.[173]

C. Legislative Proposals

Finally, legislatures could craft their own solutions to the problem by implementing the necessary transparency and accountability framework to hold developers and deployers of AI and algorithmic decision-making accountable. Congress and state legislatures are presently considering bills that if passed, could make a meaningful difference in the regulatory landscape.

While several legislative proposals specifically address the types of transparency and accountability requirements I have discussed, the Algorithmic Accountability Act is one comprehensive example.[174] The proposed bill would impose a number of new requirements on companies using automated decision-making, mandating that they (1) assess their use of automated decision systems, including training data, for impacts on accuracy, fairness,

---

[171] As a predicate to a Section 18 rulemaking, the Commission must have reason to believe that the unfair or deceptive acts or practices which are the subject of the proposed rulemaking are prevalent. 15 U.S.C. § 57a(b)(3) (2018).

[172] The FTC accepts rulemaking petitions from the public. *See, e.g.*, *Petition for Rulemaking Concerning Use of Artificial Intelligence in Commerce,* ELECTRONIC PRIVACY INFO. CTR. (Feb. 3, 2020), https://epic.org/privacy/ftc/ai/EPIC-FTC-AI-Petition.pdf. I strongly encourage individuals and groups with well-considered proposals to submit them to the Office of the Secretary.

[173] Rebecca Kelly Slaughter, Acting Chair, Fed. Trade Comm'n, Keynote Remarks at the Consumer Federation of America's Virtual Consumer Assembly (May 4, 2021), https://www.ftc.gov/system/files/documents/public_statements/1589607/keynote-remarks-acting-chairwoman-rebecca-kelly-slaughte-cfa-virtual-consumer-assembly.pdf.

[174] Algorithmic Accountability Act, H.R. 2231, S. 1108, 116th Cong. (2019).

bias, discrimination, privacy, and security; (2) evaluate how their information systems protect the privacy and security of consumers' personal information; and (3) correct any issues they discover during the impact assessments. The proposed bill also authorizes the FTC to create rules requiring companies under its jurisdiction to conduct impact assessments of highly sensitive automated-decision systems.[175] The core insight of the proposed bill, through required impact assessments, is that vigilant testing and iterative improvements are the fair and necessary cost of outsourcing decisions to algorithms.

In addition, Congress is currently contemplating a federal privacy law. While privacy legislation may not seem directly applicable to the problems we are discussing today, it can in fact play an important role in addressing algorithmic justice—and it is worth noting that the algorithmic-justice requirements imposed in Europe were done as a part of its privacy law, the GDPR. I have been a vocal advocate for a federal privacy law,[176] and I believe that such a bill should incorporate specific protections, including civil rights provisions, to limit the dangers of algorithmic bias and require companies to be proactive in avoiding discriminatory outcomes.

The privacy bill proposed by Senator Cantwell and several colleagues, the Consumer Online Privacy Rights Act, includes a civil rights provision that seeks to accomplish this type of broader protection.[177] The bill prohibits the processing or transfer

---

[175] *Id.* at § 3(b).

[176] *See, e.g.*, *Hearing on Oversight of the Fed. Trade Comm'n: Before the S. Comm. on Commerce, Science, and Transportation,* 116th Cong. (Aug. 5, 2020) (statement of Commissioner Rebecca Kelly Slaughter, Fed. Trade Comm'n), https://www.ftc.gov/system/files/documents/public_statements/1578979/opening_statement_of_commissioner_rebe cca_slaughter_senate_commerce_oversight_hearing.pdf; Rebecca Kelly Slaughter, Commissioner, Fed. Trade Comm'n, The Near Future of U.S. Privacy Law, Remarks to Silicon Flatirons – University of Colorado Law School (Sept. 6, 2019), https://www.ftc.gov/system/files/documents/public_statements/1543396/slaughter_silicon_flatirons_remarks_9-6-19.pdf.

[177] *See* Consumer Online Privacy Rights Act, S. 2968, 116th Cong. (2019). The Algorithmic Justice and Online Platform Transparency Act also provides strong civil rights provisions, building upon the transparency requirements of the Algorithmic Accountability Act and adding the civil rights protections

of data on the basis of an individual's actual or perceived protected status for the purpose of marketing in a manner that unlawfully discriminates or otherwise makes the opportunity unavailable to the individual or class of individuals.[178] The proposed bill also prohibits the processing or transfer of data in a manner that unlawfully segregates, discriminates against, or otherwise makes unavailable the goods, services, or facilities of any place of public accommodations. History teaches that there is no substitute for strong civil rights laws that outlaw discrimination outright.

Finally, alongside federal efforts, the states—the great laboratories of democracy—will likely continue to propose and adopt innovative approaches.[179]

---

similar to those in the Consumer Online Privacy Rights Act. *See* Algorithmic Justice and Online Platform Transparency Act S. 1896, 117th Cong. (2021).

[178] The Consumer Online Privacy Rights Act protects a wider range of classes than some other civil rights laws. Those protected classes include actual or perceived race, color, ethnicity, religion, national origin, sex, gender, gender identity, sexual orientation, familial status, biometric information, lawful source of income, and disability. *See* Consumer Online Privacy Rights Act, S. 2968, 116th Cong. § 108(a) (2019).

[179] For example, the California Consumer Privacy Act (CCPA)—which was passed in 2018 and came into effect in 2020—establishes various data-related rights and protections for residents of the country's most populous state. CAL. CIV. CODE § § 1798.100 – 1798.199 (2018). The California Privacy Rights Act (CPRA)—approved via ballot proposition in 2020 and going into effect in 2023—further expands the CCPA and also establishes a dedicated state privacy enforcer, the California Privacy Protection Agency. Proposition 24: The California Privacy Rights Act of 2020, in Cal. Sec'y of State, Official Voter Information Guide 42–75 (Aug. 10, 2020), https://vig.cdn.sos.ca.gov/2020/general/pdf/complete-vig.pdf. Also noteworthy is Illinois's 2008 Biometric Information Privacy Act (BIPA), which imposes a range of requirements on companies' collection and retention of biometric information. 740 ILL. COMP. STAT. 14/5 (2008). Despite being more than a decade old, BIPA continues to be highly relevant; it targets a practice that is increasing in prevalence and sophistication (for example, facial recognition), and it is buttressed by a private right of action that allows private litigation to complement enforcement efforts by under-resourced enforcers.

V. Conclusion

The growth of algorithmic decision-making presents immense opportunity and risk to society. Algorithms could promote economic justice by helping distribute opportunities more broadly, resources more efficiently, and benefits more effectively. But this article documents the perilous potential for algorithms to amplify injustice while simultaneously making injustice less detectable. Developers create algorithms with faulty inputs and flawed conclusions. They fail to test their models and rely on proxies that foster and often exacerbate discrimination. They create powerful engines that monetize attention, surveil consumers, and manipulate behavior without regard for the societal consequences. Their deployment of algorithms also imperils competition. If left unaddressed, these algorithmic flaws will repeatedly and systematically harm consumers. These harms are often felt most acutely by already vulnerable or historically disadvantaged populations, especially Black Americans and other communities of color.

The FTC's tools are still capable of addressing some of the problems posed by algorithms and AI because algorithmic decision-making shares many features with problematic innovations of generations past. The Agency must deploy section 5 of the FTC Act, FCRA, ECOA, COPPA, and section 6(b) studies creatively to mitigate algorithmic harms.

But confronting the challenges of algorithmic decision-making will also require new tools and strategies. There also may be certain applications of AI and algorithms that pose such a profound risk of injustice to vital life functions or opportunities that a moratorium might be appropriate and necessary. Society's goal, as urgent as it is achievable through collaboration, study, and creativity, should be to limit the downside risks of algorithms without unduly constraining their upside rewards. We need to consider context- and consequence-specific applications and tailor our enforcement and policy responses appropriately.

Acknowledgements

I would like to thank the many colleagues and experts who generously helped this article take shape through their insights and contributions. I am especially grateful to Austin King and David Berman for their many substantive contributions. For their expert research assistance, keen questioning, and deft wordsmithing, I would like to thank Caroline Holland, Synda Mark, Gaurav Laroia, Tori Finkle, Elena Goldstein, Rita Xia, Elise Phillips, Maggie Yellen, Chris Suhler, and Josh Banker. I am particularly lucky to have had access to the rich expertise and patient support of so many colleagues throughout the Federal Trade Commission: the Office of Policy and Planning, Bureau of Consumer Protection, Bureau of Competition, Bureau of Economics, Office of General Counsel, Office of International Affairs, Office of the Chief Information Officer, and the Presidential Innovation Fellows all shared their time and wisdom. In particular, this effort greatly benefited from the generous and insightful engagement of Aaron Alva, Josephine Liu, and Richard Gold. Special thanks to Andrew Burt and the Yale Information Society Project as well as Ben Rashkovich, Spurthi Jonnalagadda, and their colleagues in the Yale Journal of Law and Technology for their careful edits and thoughtful guidance. Finally, I must acknowledge that all efforts in my office are greatly enhanced by the administrative and moral support of Kristin Greer.

Information Society Project
Yale Law School

Yale Journal of Law & Technology

A joint publication of the Information Society Project at Yale Law School and the Yale Journal of Law & Technology.

IMMUTA

The Digital Future Whitepaper Series is made possible thanks to the support of Immuta.

About the Author

Rebecca Kelly Slaughter is a Commissioner of the United States Federal Trade Commission. The views expressed in this article are her own; they do not necessarily reflect the views of the Federal Trade Commission or any other Commissioner.

Janice Kopec, an Attorney Advisor in the Office of Commissioner Slaughter, and **Mohamad Batal,** an Honors Paralegal in the Office of Commissioner Slaughter, contributed to this article.

Digital Future Whitepaper Series

The Digital Future Whitepaper Series, launched in 2020, is a venue for leading global thinkers to question the impact of digital technologies on law and society.

The Digital Future Whitepaper Series is led by ISP visiting fellow Andrew Burt and co-edited by ISP executive director Nikolas Guggenberger and visiting fellows Artur Pericles Lima Monteiro and Nabiha Syed. Spurthi Jonnalagadda (Yale Law School '22) served as the research assistant for this whitepaper.

Information Society Project

The Information Society Project (ISP) is an intellectual center at Yale Law School, founded in 1997 by Professor Jack Balkin. Over the past twenty years, the ISP has grown from a handful of people gathering to discuss internet governance into an international community working to illuminate the complex relationships between law, technology, and society.

Yale Journal of Law & Technology

The Yale Journal of Law & Technology (YJoLT) is the only law review at Yale focused on the interaction between law and technology and the only law review at Yale Law School to offer a fully interactive publication environment. YJoLT publishes articles related to law and technology semiannually. Ben Rashkovich (Yale Law School '21) served as the YJoLT editor for this special publication.

Information Society Project
Yale Law School

Yale Journal of Law & Technology

127 Wall Street
New Haven, CT 06511

203.432.4992
www.law.yale.edu