9-5-2024

Regulating Algorithmic Harms

Sylvia Lu
University of Michigan Law School, sylvialu@umich.edu

Working Paper Citation
Lu, Sylvia, "Regulating Algorithmic Harms" (2024). *Law & Economics Working Papers*. 277.
https://repository.law.umich.edu/law_econ_current/277

Regulating Algorithmic Harms

Sylvia Lu[*]

FLORIDA LAW REVIEW (Forthcoming 2025)

ABSTRACT

In recent years, the rapid expansion of artificial intelligence (AI) innovations has led to a rise in algorithmic harms—harms emerging from AI operations that pose significant threats to civil rights and democratic values in today's technological landscape. A facial recognition system for improving criminal detection wrongly collected sensitive personal data and flagged racial minorities as shoplifters. A risk-prediction algorithm adopted to identify patients denied medical treatment to Black individuals with poor health conditions. A social media algorithm intended to boost social engagement exacerbated addictive behavior and mental illness in teenagers. These harms are becoming increasingly ubiquitous yet often manifest in small and invisible forms, enabling them to aggregate while eluding regulatory oversight. Secretly and cumulatively, they affect millions to billions of individuals.

This Article constructs a legal typology to categorize these harms. It argues that there are four primary types of algorithmic harms: eroding privacy, undermining autonomy, diminishing equality, and impairing safety. Additionally, it identifies two aggravating factors—accountability paucity and algorithmic opacity—that cause these seemingly minor harms to escalate into significant problems by obstructing harm detection and correction.

[*] Faculty Fellow, Visiting Assistant Professor, University of Michigan Law School. I am grateful to Evan Caminker, Daniel Deacon, Rebecca Eisenberg, Andrew Guthrie Ferguson, Alex Klass, Aaron Perzanowski, J.J. Prescott, Nicholson Price, Steven Schaus, Paul Schwartz, Lawrence Solum, and Rebecca Wexler for their helpful feedback on this project. Special thanks to Reuven Avi-Yonah, Tabatha Abu El-Haj, Eric Baudry, Sherman Clark, Ofer Eldar, Daniel Fryer, Molly Van Houweling, Nicholas Howson, Mark Jia, Vikramaditya Khanna, Chris Lewis, Margaret Lewis, Katerina Linos, Noah Kazis, Emily Prifogle, Roseanna Sommers, Chris Walker, Mark Wu, Ekow Yankah, and Jeff Zhang for generous conversation and suggestions. I also benefited from comments from Jocelyn Aqua, Maureen Carroll, Michael Carroll, Chao-Tien Chang, Keith Clayton II, Kristina Daugirdas, Samuel Erman, Anna Fisher, Sharon Franklin, Daniel Halberstam, James Hines, Dennis Hirsch, Howell Jackson, Christopher McCrudden, Peter Railton, Rachel Rothschild, Tifani Sadek, Simon Sun, Charlotte Tschider, Salome Viljoen, Li-dar Wang, Peter Yu, and Taisu Zhang. Thank you to my workshop participants at the Inaugural Michigan Law Junior Scholars Forum, the Inaugural Yale Law School Asian Law Junior Scholars Workshop, the Michigan Law Legal Theory Workshop, the 2024 Intellectual Property Scholars Conference, the 2024 Privacy Law Scholars Conference, the 2024 Michigan Junior Scholars Conference, the University of Michigan Ross Business School Symposium on *Generative AI Law, Regulation, and Innovation in the Field*, the Chengchi University *Digital Governance, Law, and Policy* Forum, and the Hong Kong University *Enhancing Personal Data Protection in the Age of Artificial Intelligence* Conference. Sincere thanks to students in Professor Paul Schwartz's *Topics in Privacy and Security Law Seminar* at Berkeley Law, Professor Lawrence Solum's Course on *Law and Artificial Intelligence* at Virginia Law, Professor Anna Fisher's *Digital Culture* Course at Michigan Digital Institute, Professor Nicholson Price's *Artificial Intelligence and Law* Seminar, and my *Artificial Intelligence Regulation* Seminar for their terrific responses.

This Article then conducts case studies of relevant legal frameworks in the United States, the European Union, and Japan to assess the effectiveness of existing responses to algorithmic harms. The case studies reveal that these regulatory examples are insufficient; they either overlook certain types of harms or fail to consider their cumulative effects, thereby allowing problematic AI practices to circumvent legal obligations.

Drawing on these findings, this Article proposes three legal interventions to address algorithmic harms, each aims to mitigate primary harms by targeting aggravating factors. Refined harm-centric algorithmic impact assessments, which impose an obligation on AI developers to address the compounded harms, serve as a starting point for enhancing algorithmic accountability. While these assessments often have a collective focus and overlook individual differences, individual rights in terms of algorithmic systems provide enhanced control over AI applications that could lead to aggregated primary harms. The success of these tools relies on a set of disclosure duties designed to reduce algorithmic opacity in favor of increased harm awareness, especially in situations where AI use is associated with intangible yet far-reaching harms. Taken altogether, this harm-centric procedural approach advances the conversation about the legal definition of algorithmic harms, the boundaries of AI law, and viable approaches to effective algorithmic governance.

TABLE OF CONTENTS

INTRODUCTION

In recent years, the rapid expansion of artificial intelligence (AI)[1] innovations has led to a rise in *algorithmic harms*—harms produced by AI innovations that impair civil rights and democratic values in today's technological landscape.[2] As data extraction becomes prevalent, computing power increases, and algorithmic capacity expands, AI has penetrated more and more aspects of civil society.[3] Through the increased impact of AI, algorithmic harms infiltrate the social fabric. While facial recognition systems were ideally designed to bolster precision in criminal detection, they overly surveilled and misidentified racial minorities as shoplifters,[4] diminishing their privacy and equality. Social media algorithms were developed to boost social engagement, but they also exacerbated addictive behaviors and mental illness in teenagers,[5] undermining their autonomy and safety. When generative chatbots were claimed to make our lives easier, they often achieve the opposite—providing misleading medical advice, generating discriminatory responses, and unlawfully collecting sensitive data.[6]

Such harms are proliferating, signaling a disturbing trend in our increasingly automated society. What has been revealed, however, is just the tip of the iceberg. Despite their growing ubiquity, algorithmic harms typically manifest in small, invisible forms without drawing our attention. Unlike recognized physical injuries, these harms often involve an

[1] This Article uses the term "AI" to refer to algorithmic systems that exhibit human-like intelligence, including but not limited to reasoning, predicting, and decision-making skills. This definition encompasses areas such as algorithmic operations driven by big data, machine learning, large language models and other AI tools, aligning with the expansive definitions of AI set forth by the Organisation for Economic Co-operation and Development (OECD) and European Union (EU) AI Act. This Article focuses on the types of harms consistently posed by these interconnected areas of AI. Some aspects discussed here are specific to a particular AI technology, while others are not. For brevity, this Article does not extensively detail the specific AI techniques involved unless it is conducive to the discussion. The terms "AI applications," "automated systems," "algorithmic systems," and "AI innovations" are used interchangeably here to denote the development, deployment, and business practices associated with AI. For the same purpose of brevity, the terms "privacy" and "data privacy" are used to refer to the field of "privacy and data protection." This Article acknowledges that these deliberately decided definitional and stylistic choices may contain drawbacks. For conceptual definitions of AI, see Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 369-62 (2016); Michael Veale, Kira Matus & Robert Gorwa, *AI and Global Governance: Modalities, Rationales, Tensions*, ANN REV. L. & SO SCI 255, 256 (2023); Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1307 (2019); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 660–61 (2017).

[2] In this Article, harms refer to impairment of a given set of interests, including setbacks and damages to diminishment of wellness. This Article does not address the highly speculative harms or existential risks related to artificial general intelligence. It acknowledges that some algorithmic harms might overlap with long-standing harms posed by computing technologies that AI has multiplied and made more harmful than ever.

[3] Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1958, 1960 (2021).

[4] Adam Satariano & Kashmir Hill, *Barred from Grocery Stores by Facial Recognition*, N.Y. TIMES (Jun. 28, 2023), https://www.nytimes.com/2023/06/28/technology/facial-recognition-shoplifters-britain.html.

[5] Karen Hao, *The Facebook Whistleblower Says Its Algorithms Are Dangerous. Here's Why*, MIT TECH REV. (Oct. 5, 2021), https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/ (last visited Oct 19, 2023).

[6] Cecilia Kang & Cade Metz, *F.T.C. Opens Investigation into ChatGPT Maker over Technology's Potential Harms*, N.Y. TIMES (July 13, 2023), https://www.nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html (last visited Jul 27, 2023).

intangible erosion of civil rights—such as privacy, autonomy, equality, and safety—that are not posing overt inconvenience or immediate suffering.[7] For instance, Meta's social media algorithms are known to unlawfully harvest user behavior data, transforming these into sensitive insights for repeatedly manipulative or biased operations.[8] Yet, until whistleblowers reported corporate scandals to the public, these harms remained mostly imperceptible to millions and even billions of victims.[9]

Given their intangible nature, algorithmic harms are often downplayed as smaller secondary problems. A human who steals our personal identification for identity theft is considered a flagrant violation of privacy, yet an algorithm doing the same thing escapes our attention. Given their elusive nature, algorithmic harms affect us in various imperceptible ways. They compromise our privacy through unauthorized data extraction that reveals our vulnerabilities and intimate details. They distort our autonomy based on highly personalized manipulation and deceptively authentic appearance. They exacerbate inequality through AI-generated content that disproportionately discriminates against minorities. Their addictive and unpredictable operations undermine our health, safety, and security.

In countless cases, these neglected harms *accumulate* quietly, undermining civil rights without evident signs. A single unauthorized collection of our personal data may not reveal many of our personal lives. But with the aid of AI, a hundred data collections may.[10] A one-time manipulation may be innocuous, yet a hundred manipulations conducted by algorithms can shape our decisions.[11] As AI applications become increasingly common, a diffused set of actors—AI designers, developers, and often companies—repeatedly generate such harms. Over time, these harms add up. Millions and even billions of individuals interacting with AI innovations are exposed to them; socially marginalized groups are said to suffer even more.

Yet, because of their intangible form, these cumulative harms are difficult to track down and redress at early stages. Victims lack knowledge of their existence; regulators lack foresight to investigate them; legislators lack awareness of their actual gravity; and companies lack incentives to fix them. Secretly and cumulatively, algorithmic harms have contributed to widespread mental health issues in teenagers, social events like the United States Capitol attack on January 6, 2021,[12] and pervasive bias generated by large language models.[13]

[7] Danielle Keats Citron & Daniel Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 818 (2022) (discussing the features of privacy harms).

[8] Carole Cadwalladr, *Fresh Cambridge Analytica Leak 'Shows Global Manipulation Is out of Control,'* GUARDIAN, Jan. 4, 2020, https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation (last visited Mar 27, 2023).

[9] Karen Hao, *The Facebook Whistleblower Says Its Algorithms Are Dangerous. Here's Why*, MIT TECH. REV. (Oct. 5, 2021), https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/ (last visited Oct 19, 2023).

[10] DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 44-47 (2004) [hereinafter SOLOVE, THE DIGITAL PERSON].

[11] Deborah Yao, *Meta Sued in 8 States for 'Addictive' Platforms that Harm Young Users*, AI BUS. (Jun. 10, 2022), https://aibusiness.com/verticals/meta-sued-in-8-states-for-addictive-platforms-that-harm-young-users.

[12] Roger McNamee, *Platforms Must Pay for Their Role in the Insurrection*, WIRED, Jan. 2021, https://www.wired.com/story/opinion-platforms-must-pay-for-their-role-in-the-insurrection/.

[13] Nitasha Tiku, Kevin Schaul & Szu Yu Chen, *These Fake Images Reveal How AI Amplifies Our Worst Stereotypes*, WASH. POST (Nov. 11, 2023), https://www.washingtonpost.com/technology/interactive/2023/ai-

Consequently, these unaddressed harms have become an urgent social issue that demands meaningful action.

The trouble with algorithmic harms raises two fundamental questions regarding the development of AI law: what kinds of harms deserve legal treatment, and how the law can meaningfully address those harms? To date, legal studies have primarily focused on specific harms arising from the most controversial AI applications. Facial recognition systems, risk scoring, deepfakes, medical AI, and generative AI are some areas that have attracted significant academic interest.[14] Legal scholars have argued that such AI tools increase risks such as mass surveillance, bias and structural inequality, and manipulation concerns.[15]

These findings are valuable and largely correct. However, they have not yet resolved the basic problems underlying these phenomena: the intangible harm stemming from AI operations, its cumulative effect, and the connection between harm and legal interests. Accordingly, the current literature lacks a thorough examination of which legal interests should be safeguarded and how to do so. Additionally, it does not adequately address how to balance these protections without unduly impeding innovation, a primary competing policy interest. These overlooked aspects obscure a realistic view of the impact of algorithmic harms and hinder the law's capacity to recognize and respond meaningfully.

This Article fills these underestimated and underrecognized gaps. Building on prior AI-regulation scholarship, this Article provides a novel and in-depth account of algorithmic harm. It investigates the kinds of algorithmic harms deserving legal treatment, and how the law should hold the perpetrators of such harms more accountable.[16] To delineate the legal

generated-images-bias-racism-sexism-stereotypes/ (last visited May 19, 2024).

[14] *See, e.g.*, Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN L. REV. 1105, 1109-24 (2021) (arguing that facial recognition enables expansive police surveillance and proposing constitutional and legislative solutions to address accountability gaps left by the Fourth Amendment doctrine); Daniel Solove, *Artificial Intelligence and Privacy*, 77 FLA L. REV 52 (2025 forthcoming) (pointing out how facial recognition tools result in surveillance and threat to personal anonymity) [hereinafter Solove, *AI and Privacy*]; Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447 (2019) (discussing a widely used healthcare risk software, designed to discern high risk patients, denied medical treatment to Black individuals with inferior health conditions.); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008); Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, CAL. L. REV. 1753 (2019) (deepfakes); Mark Lemley, *How Generative AI Turns Copyright Upside Down*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702 (generative AI's copyright implications); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1347 (2018) (software programs and machine learning applications in criminal systems); W. Nicholson Price II, Rachel Sachs & Rebecca S. Eisenberg, *New Innovation Models in Medical AI*, 99 WASH. U. L. REV. 1121 (2022); Paul Ohm, *Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI*, 25 COLUM. SCI. & TECH. L. REV. 214 (2024) (examining the four distinct stages of shaping generative AI model behavior and arguing that legal interventions may be most effective at the fine-tuning stage).

[15] *See, e.g.,* Solove, *AI and Privacy*, *supra* note 14 (AI and mass surveillance); Obermeyer et al. *supra* note 14 (algorithmic bias and inequality); W. Nicholson Price II, *Contextual Bias and Medical AI*, 33 HARV. J.L. &. TECH. 66, 68 (2019) [hereinafter Price, *Contextual Bias*] (contextual bias in medical AI); Karen Yeung, *'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 INFO. COMMUN & SOC. 118, 123 (2017) (algorithmic systems' liberal manipulation); Chesney and Citron, *supra* note 14, at 1778 (deepfakes and election manipulation).

[16] This piece is specifically concerned with the kinds of harms arising from algorithmic and AI operations

boundaries of algorithmic harms, the Article develops a basic legal typology with four categories of civil rights harms: eroding privacy, undermining autonomy, diminishing equality, and impairing safety.[17] Additionally, the typology includes two aggravating factors that can cause these intangible harms to escalate into cumulative harms: First, a deficiency in accountability mechanisms, termed *accountability paucity*, that hinders harm correction; and second, insufficient transparency due to *algorithmic opacity*, an inherent feature of AI systems that further obstructs harm detection.[18]

This conceptual typology has significant implications for AI regulation. By specifying a basic legal scope of algorithmic harms, the typology offers normative guidance to policymakers, lawmakers, and regulators for taking effective regulatory actions. Moreover, by incorporating the typology into suggested legal interventions, policymakers can carefully balance competing regulatory priorities. Over the years, a lack of legal understanding of algorithmic harms has been accompanied by insufficient harm considerations and ineffective regulatory efforts in the realm of AI.[19] Currently, countries worldwide are adopting a range

that negatively impact privacy, autonomy, equality, and safety due to their intangible, ubiquitous, and cumulative nature.

[17] The typology does not, however, offer an exhaustive list of algorithmic harms. The harms that are not discussed further in this piece, while also need legal treatment through other venues, include environmental harms generated by AI, copyrights issues caused by generative AI, and job displacement resulting from automation or generative AI models. This paper's typology stands apart from other taxonomies of algorithmic harms found in computer science literature, which do not focus on legal harms and norms in their categorization of harms. *See, e.g.*, Renee Shelby et al., *Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction*, *in* PROCEEDINGS OF THE 2023 AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY 723 (2023), https://dl.acm.org/doi/10.1145/3600211.3604673 (last visited Jun 3, 2024); Colin Watson et al., *Hostile Systems: A Taxonomy of Harms Articulated by Citizens Living with Socio-Economic Deprivation*, *in* PROCEEDINGS OF THE CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1 (2024), https://dl.acm.org/doi/10.1145/3613904.3642562 (last visited Jun 3, 2024). A recent noteworthy legal taxonomy developed in this field, particularly that of the FTC Commissioner Rebecca Slaughter et al., explicitly frames the problems arising from machine learning and other applications as algorithmic harms. The Slaughter et al. work has offered a basic legal taxonomy of algorithmic harms by categorizing these harms into two groups: those resulting from flawed designs to cause discriminatory results, as well as those stemming from sophisticated algorithms to produce systemic problems like discrimination, surveillance, and unfair competition. Unlike the influential work by Slaughter et al., which taxonomizes algorithmic harms based on algorithmic design flaws, the typology constructed by this Article situates algorithmic harms within the four distinct and interconnected abovementioned categories. This legal distinction advances the current technically based taxonomies by ensuring that an array of applications, flaws, and causes old and new are covered under a *normative* analytic framework. Rebecca Slaughter, Janice Kopec & Mohamad Batal, *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*, 23 YALE J.L. & TECH (2021).

[18] Existing literature has highlighted the opacity and accountability concerns inherent in algorithmic systems, but a thorough examination of their relation to crucial legal harms is still missing. Sonia K Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV 1183, 1225-36 (2019) [hereinafter Katyal, *Source Code Secrecy*] (opacity concerns inherence in algorithmic systems); Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOCIETY 1, 3-4 (2016); Jeanne C. Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 NYU L. REV. 706, 716 (2019); Sylvia Lu, *Data Privacy, Human Rights, and Algorithmic Opacity*, 110 CALIF L. REV. 2087, 2099-2100 (2022) (explaining how AI and algorithmic opacity contribute to large-scale privacy harms and governance problems) [hereinafter Lu, *Data Privacy*].

[19] Craig S. Smith, *Act One: Opposition Takes Center Stage Against EU AI Legislation*, FORBES (Sept. 5, 2023), https://www.forbes.com/sites/craigsmith/2023/09/05/act-one-opposition-takes-center-stage-against-eu-ai-

of legal measures to address algorithmic harms.[20] Many of these measures, however, are either ill-equipped to recognize certain types of harms or are unable to mitigate their cumulative effects, causing harmful algorithmic practices to evade regulatory oversight.

The pitfalls and strengths of existing regulatory frameworks offer rich opportunities for policymakers to further develop AI law. By examining these aspects, policymakers can gain valuable insights toward a harm-centric approach—addressing the aggravating factors of accountability paucity and algorithmic opacity to regulate the primary harms. Drawing upon these insights, this Article suggests three procedural legal interventions. First, future legislation should cover the duty to perform impact assessment that considers the cumulative impact of intangible harms to meaningfully enhance accountability. This requires AI developers to identify, evaluate, and address harms throughout the AI system lifecycle, allowing for a careful assessment of compounded harms as categorized in the typology.

Given that impact assessments often have a collective focus and overlook individual differences, policymakers should consider individual rights that give potential victims greater control and power over AI applications that could aggregate primary harms. In recognition of individuals' limited capacity to measure the impact of algorithmic harms, a scheme of combined opt-out and opt-in AI applications is recommended.[21] Algorithmic harms that inevitably violate civil rights, such as those generated by emotion recognition software, should be placed in an opt-in regime to provide citizens with substantive protection. Low-stake harms, such as commercial recommendation AI, could be considered in the opt-out regime.

The success of such interventions depends on reducing algorithmic opacity, which impedes the detection and resolution of algorithmic harm. A set of disclosure obligations should be established to enhance harm awareness. Where AI use leads to intangible yet significant harms, the law should impose transparency obligations on AI adopters, such as indicating the use of AI, providing a warning label of anticipated harms, and disclosing harm mitigation measures, thereby facilitating law enforcement and stakeholder oversight through different layers of mandated disclosures.

Based on these findings, this Article makes three central contributions. First, it advances our understanding of the characteristics and scope of algorithmic harm. The Article illuminates how key features of algorithmic harms—ubiquity, intangibility, and aggregation—turn seemingly small harms into significant legal harms while depriving victims of legal remedies. Then it further taxonomizes the types of algorithmic harm that have victimized individuals, groups, and society at large, elucidating why these harms deserve legal recognition. Such understanding is crucial for policymakers to develop regulatory strategies that take primary algorithmic harms into account.

legislation/ (last visited Oct 19, 2023).

[20] Courtney Rozen & Jillian Deutsch, *Regulate AI? How US, EU and China Are Going About It*, BLOOMBERG, Mar. 13, 2024, https://www.bloomberg.com/news/articles/2024-03-13/regulate-ai-how-us-eu-and-china-are-going-about-it (last visited May 19, 2024).

[21] Natasha Lomas, *California's Privacy Watchdog Eyes AI Rules with Opt-out and Access Rights*, TECHCRUNCH (Nov. 27, 2023), https://techcrunch.com/2023/11/27/cppa-admt-draft-rules/ (last visited Jan 1, 2024).

The second contribution is proposing a harm-based approach informed by influential regulatory frameworks in the US, EU, and Japan, where policymakers do not fully understand the nature and scope of these harms. To guide AI legislation and enforcement actions, this Article underscores the value of procedural interventions in mitigating algorithmic harms under the harm-based approach. It argues that by targeting aggravating factors with harm-centric procedural rules, policymakers can address algorithmic harm without overly restricting the development of AI innovations.

Third, building upon the first and second contributions, this Article's harm-based approach advances the conversation about viable approaches to algorithmic governance. Its legal conceptualization of the nature, scope, and solutions for algorithmic harm furthers both policy debate and legal scholarship by providing normative grounds for a regulatory shift: from treating AI law as *risk regulation*—a prevalent approach taken by policymakers—to *harm regulation*, an advanced development of AI law. A harm-based approach, starting with suggested procedural interventions, opens up the possibility of a broader array of restrictive substantive rules for effective algorithmic governance.

To that end, this Article proceeds in four parts. It starts with two descriptive sections that explore the nature and substance of algorithmic harms, followed by two normative sections that analyze legal frameworks and suggest interventions. Section I investigates the characteristics of algorithmic harms, arguing that their ubiquitous, intangible, and aggregate nature has adversely affected numerous individuals and groups, along with their significant social implications. Section II introduces a typology of algorithmic harms to delineate their legal boundaries, demonstrating how these harms substantially undermine privacy, autonomy, equality, and safety due to insufficient transparency and accountability. Building on this typology, Section III evaluates the effectiveness of three regulatory frameworks for algorithmic harms: the American consumer-centric regime, the European Union risk-based approach, and Japanese proactive ethical norms. These case studies reveal the advantages and limitations of each regime, offering a pathway to more effective regulation of algorithmic harms. Section IV synthesizes the findings of the typology and case studies, providing normative guidance for AI law to meaningfully redress algorithmic harms. It proposes procedural harm regulation that can establish a new generation of harm mitigation strategies in the process. Together, they carry the potential to advance policymakers' harm calculation, rebalancing the dynamics between innovation and civil interests with respect to robust AI governance.

I. THE TROUBLE WITH ALGORITHMIC HARMS

Before exploring the nature of algorithmic harms, it is crucial to define the meaning of *algorithmic harms* for conceptual clarity. In this Article, harms refer to impairment of a given set of interests, ranging from hurdles, setbacks, and damages to diminishment of wellness.[22] They make the condition of a particular interest worse than it would have been if the harms had not taken place. In the space of AI, this Article defines algorithmic harms as a host of

[22] Citron & Solove, *supra* note 7, at 799 ("Harms involve injuries, setbacks, losses, or impairments to well-being.").

issues emerging from AI applications that diminish civil rights and democratic values in today's technological landscape.

Algorithmic harms have become a pressing problem for contemporary democracy. As machine learning and other AI technologies increasingly permeate our societies, they introduce both conceptual and regulatory complexities, which primarily arise from the scale and nature of algorithmic harms. Some issues are unique to AI, while others reflect concerns that are widespread across various technologies. Because of these challenges, people may find it difficult to address the harmful effects of these AI applications.

This Section explores the problematic attributes of algorithmic harms, elucidating why they are more challenging to address than traditional harms. It then discusses *who is harmed*, elaborating on how these unresolved harms injure a wide array of victims, ranging from individuals to groups and society at large. As this Section explains, aggregation of harms leads to both individual and collective problems, undermining civil rights and broader social functions.[23]

A. The Characteristics of Algorithmic Harms

Regulating algorithmic harms presents distinct challenges owing to three distinct attributes: ubiquity, intangibility, and aggregation. The intangible feature renders these harms imperceptible, making them easily overlooked and hard to measure, with their causality often difficult to establish. When ubiquitous harms remain unaddressed, they tend to accumulate, ultimately resulting in considerable adverse impacts on victims over time.[24]

1. Ubiquity

The rise of algorithmic harm stems from our society's increasing dependence on AI. As of this writing, more than seventy percent of Americans engage with AI in their daily lives, either by choosing technologies like virtual assistants or by being subject to AI systems selected by other entities, such as those used for algorithmic assessments of credit risk.[25] The global user base of AI tools has been expanding annually. In 2024, the number of individuals utilizing AI tools exceeded 310 million, more than doubling the figure reported in 2020. This upward trend is anticipated to persist, with forecasts suggesting that the number of AI tool users will surpass 700 million by the end of the decade.[26]

In the private sector, more than half of entities have deployed AI to innovate their products.[27] According to a 2023 estimate, eighty percent of enterprises will have adopted

[23] Nathalie A. Smuha, *Beyond the Individual: Governing AI's Societal Harm*, 10 INT POL'Y REV. 1, 6 (2021).

[24] Citron & Solove, *supra* note 7, at 816 ("When these harms happen to an individual repeatedly by different actors, they become significantly more harmful.")

[25] Katherine Haan & Rob Watts, *24 Top AI Statistics & Trends In 2024*, FORBES (Apr. 25, 2023), https://connect.comptia.org/blog/artificial-intelligence-statistics-facts (last visited Mar 17, 2023).

[26] *Id.*

[27] Chirag Chauhan, *60 Artificial Intelligence Statistics You Need to Know in 2024*, RADIXWEB (Dec. 3, 2023), https://radixweb.com/blog/artificial-intelligence-statistics (last visited Jan 14, 2024).

generative AI models by 2026.[28] Presently, over ninety percent of leading companies are investing in AI.[29] With a global market valued at around 1.502 billion dollars, which is estimated to grow twentyfold in coming years,[30] AI has become a driving force for innovations that affect the daily lives of citizens.[31] Most recently, the popularity of generative AI services has increased the number of AI users. Generative AI applications span many industries, such as healthcare, entertainment, marketing, and software development, to name a few. Many AI innovations promise to optimize how we live our lives, influencing us through smartphones, websites, social media platforms, and the increasing number of sensors scattered throughout our cities.[32]

This expansive growth of AI has occurred without much AI regulation, leading to the spread of harms to countless users and the broader public. Under-regulated and poorly monitored AI systems have been employed by both public and private sectors across an indefinite spectrum of uses, including crime detection, disease diagnosis, personalized advertising, and employment decisions. As Part II will illustrate, each of these applications are fraught with potential harms such as privacy violations, bias, inaccuracy, opacity, or manipulation.[33] For instance, generative AI tools have been accused of large-scale unauthorized collection of personal data, making up false information (hallucinations), creating realistic but fake videos of public figures, AI-generated scam emails causing monetary losses, and stereotypical outputs leading to reinforced discrimination.[34] Because instances like these often occur without regulatory measures that make their harm evident to users, the incidents reported thus far represent merely a small fraction of algorithmic harms.

Although many in the media and academia have voiced concerns regarding AI systems, citizens may find it increasingly difficult to avoid exposure to their applications, even if they are known to generate harms.[35] Today, facial recognition technology is being deployed in

[28] *Id.*

[29] Marko Dimitrievski, *Artificial Intelligence Statistics 2023*, TRUELIST (2023), https://truelist.co/blog/artificial-intelligence-statistics/ (last visited Mar 17, 2023).

[30] *Id.*

[31] *See* AMBA KAK & SARAH MYERS WEST, THE AI NOW REPORT: 2023 LANDSCAPE: CONFRONTING TECH POWER 15 (2023), https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf.

[32] Brad Glosserman, *Artificial Intelligence Gets Scarier and Scarier*, JAPAN TIMES (Mar. 22, 2022), https://www.japantimes.co.jp/opinion/2022/03/22/commentary/dangerous-ai/ (last visited Mar 21, 2023).

[33] *See, e.g.*, W. Nicholson Price II, Rachel Sachs & Rebecca S. Eisenberg, *New Innovation Models in Medical AI*, 99 WASH. U. L. REV. 1121, 1123 (2022); Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1359-60 (2022) [hereinafter Kaminski, *Regulating the Risks of AI*] ("Many large companies already use AI tools in recruitment, including McDonald's, JP Morgan, Kraft Heinz . . . . [N]inety percent of Fortune 500 companies use automation . . . to screen or rank job candidates.").

[34] Kang & Metz, *supra* note 6.

[35] More than half of Americans believe that AI is more detrimental than beneficial in maintaining individual privacy. The FTC has also reported increasing consumer concerns regarding the harms associated with AI applications. A recent survey indicates a growing public apprehension about the rapid expansion of AI technologies, with individuals expressing worries about the misuse of their data for AI training purposes and AI-driven crimes, such as scams and fraud. Consumers Are Voicing Concerns About AI, FEDERAL TRADE COMMISSION (2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing-concerns-about-ai (last visited Jul 23, 2024); Alec Tyson and Emma Kikuchi, *Growing Public Concern*

public spaces like streets, airports, schools, and retail stores; social media platforms have become crucial for many to gather information and stay connected with distant friends; search engine and generative AI tools have gradually become indispensable in our professional and daily lives. As AI continues to automate tasks in both the private and public domains without sufficient legal oversight, citizens in these scenarios have very few if any options to avoid exposure to their potential harms.

2. Intangibility

Despite the ubiquity of algorithmic harms, they often appear intangible and thus seemingly negligible. In many situations, civil rights harms concerning privacy, autonomy, and equality do not lead to immediate economic loss or physical injury, but rather harm to dignity and liberties as an intangible interest.[36] Examined separately, intangible harms may not seem problematic. Identifying these problems from an individual perspective can thus be challenging. For example, whoever is subject to facial recognition AI does not perceive inconvenience or danger from the collection of their biometric data until identity theft occurs.[37] A deepfake video of a political leader making controversial statements could be spread to incite public outrage or mistrust, yet this manipulation is often hard to discern before experts or the political leader expose the deception.

With such intangibility, algorithmic harms can lead to hard-to-detect injury in many interconnected scenarios, including biometric identification systems that not only enable identity theft, but also bring about psychological harms like fear and anxiety.[38] Similarly, emotional recognition tends to give wrong assessments of one's inner self, contributing to reputational harm and inaccurate personal insights, leading in turn to biased automated decisions.[39]

However, even if algorithmic harms are attributable to various types of injury, the law struggles to recognize intangible impairment of interests or future injury. Courts, for instance, have been reluctant to consistently recognize intangible future privacy harms.[40]

This intangibility allows entities to ignore algorithmic harms. Recent corporate scandals show that firms often design AI systems to maximize business success, even if their

about the Role of Artificial Intelligence in Daily Life, PEW RESEARCH CENTER (Aug. 28, 2023), https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ (last visited Jul 23, 2024).

[36] *Id.*

[37] Kaminski, *Regulating the Risks of AI*, *supra* note 33, at 50 ("Some privacy harms, such as identity theft, can be readily observed and measured. Others, such as harms to dignity or autonomy, or for the more concretely minded, exposure to future risks of unauthorized disclosure or identity theft, cannot.").

[38] Danielle Citron, *The Privacy Policymaking of State Attorneys General*, 92 NOTRE DAME L. REV. 747, 798-99 (2016) ("For most courts, privacy and data security harms are too speculative and hypothetical, too based on subjective fears and anxieties, and not concrete and significant enough to warrant recognition.").

[39] Kate Crawford, *Artificial Intelligence Is Misreading Human Emotion*, ATLANTIC (Apr. 2021), https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/ (last visited Mar 21, 2023).

[40] Citron & Solove, *supra* note 7, at 817, 834, 843.

business model sacrifices individual privacy, health, and other interests.[41] Furthermore, because of proclaimed trade secrecy involved in commercial use of AI, the harmful aspects of corporate practices can operate opaquely and escape external review.[42] This in turn prevents users and regulators from discerning the potential harms of seemingly well-intentioned AI services, thereby externalizing the cost of harms.[43]

3. Aggregation

An essential feature of algorithmic harms is that they may seem negligible in isolation but are cumulatively significant.[44] Many harmful algorithmic practices are repetitive in nature, enabling smaller harms to add up to a substantial injury. To take one example, some social media algorithms continuously push engaging content tailored to user interests, manipulating behavior to increase user attachment to their platform.[45] Users are not immediately troubled by their impaired autonomy and prolonged usage in a given day. Yet over weeks or months, excessive usage driven by algorithm-pushed content consumption changes user behavior, leading to more serious addiction and other mental health problems.[46] Similarly, personal data collected by a single tracking app may appear harmless when examined in isolation. However, as numerous websites and apps extract and collate individuals' activities, data brokers and other entities can form a fine-grained picture of their personal lives.[47] When numerous actors engage in a host of harmful practices and repeatedly disseminate these harms, individuals suffer plenty of harms that undermine their interests. Gradually, the sum of these detriments becomes significant.

The imperceptibility of harm along with the opacity of algorithmic practices makes it difficult for individuals to trace their causes. In recent years, the growing number of

[41] *See, e.g.*, Carole Cadwalladr, *Fresh Cambridge Analytica Leak 'Shows Global Manipulation Is out of Control,'* GUARDIAN (Jan. 4, 2020), https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation; Billy Perrigo, *Cambridge Analytica Whistleblower Christopher Wylie Tells All*, TIME (Oct. 8, 2019), https://time.com/5695252/christopher-wylie-cambridge-analytica-book/; Justin McCurry, *South Korean AI Chatbot Pulled from Facebook after Hate Speech towards Minorities*, GUARDIAN (Jan. 14, 2021), https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook.

[42] Fromer, *supra* note 18, at 720-24.

[43] Sylvia Lu, *Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence*, 23 VAND. J. ENT. & TECH. L 99, 117-27 (2020) (noting that algorithmic opacity hinders the detection of algorithmic harms to privacy, safety, and equality [hereinafter Lu, *Corporate Social Disclosures*].

[44] Citron & Solove, *supra* note 7. While both privacy harms and algorithmic harms manifest in features like aggregation and intangibility—and there are overlapping areas of these two harms (i.e. facial recognition that generates both privacy and algorithmic harms)—the scope of algorithmic harms is broader than that of privacy harms. Given the unique nature of AI, algorithmic harms often cause a series of damage to more than one interest, which includes privacy but exceeds the realm of privacy to others like equality, safety, and more.

[45] *See, e.g.*, Andrzej Cudo et al., *Dysfunction of Self-Control in Facebook Addiction: Impulsivity Is the Key*, 91 PSYCHIATR Q 91 (2020).

[46] Jonathan Stempel, Diane Bartz & Nate Raymond, *Meta's Instagram Linked to Depression, Anxiety, Insomnia in Kids - US States' Lawsuit*, REUTERS (Oct. 25, 2023, 8:33 AM EDT), https://www.reuters.com/legal/dozens-us-states-sue-meta-platforms-harming-mental-health-young-people-2023-10-24/.

[47] *Id.*

teenagers with psychiatric disorders is said to be correlated to the influence of social media algorithms designed to promote addiction to these platforms.[48] However, because of the intangibility of algorithmic harms and the diffuseness of the actors involved, identifying causal linkages between psychiatric disorders and intangible harms derived from algorithmic applications has been challenging. This leads to difficulty in determining causation, whereby individuals are likely to suffer unknowingly from various forms of algorithmic harm.[49]

Owing to their indefinite causality, many individual harms remain overlooked, which can lead to graver consequences. Consider the following scenario. Posting personal thoughts on social media might provide clues for AI to infer one's emotional stability and mental health,[50] yet very few social media users are aware of this and can seldom sense the associated harm. These insights might be used to predict one's health conditions and feed decision-making algorithms (algorithms that make decisions to displace human decision-makers) in various contexts.[51] For instance, in a professional setting,[52] a job applicant may be denied a job interview based on health predictions, which would be difficult to trace.[53]

Currently, these unresolved problems continue to have prolonged effects on individuals. Multiple entities collect data about a consumer's daily activities, generate insights about them, manipulate their feelings and choices, and make decisions against that person's best interests. The consumer is bombarded with hundreds or thousands of algorithmic harms. Eventually, this combined effect of these harms severely undermines individual civil rights and interests. As underregulated AI applications reach more users, a large portion of citizens become victims of algorithmic harms.

B. Who Is Harmed?

Building upon an understanding of the nature of algorithmic harms, this subsection seeks to identify their victims. Consumers are often perceived as one of the primary sufferers of harmful AI innovations. After all, AI-driven products and services are pervasive, making consumers more likely to encounter and be affected by them. However, algorithmic harms

[48] Jonathan Haidt, *The Dangerous Experiment on Teen Girls*, ATLANTIC (Nov. 2021), https://www.theatlantic.com/ideas/archive/2021/11/facebooks-dangerous-experiment-teen-girls/620767/ (last visited Mar 22, 2023); Elena Bozzola et al., *The Use of Social Media in Children and Adolescents: Scoping Review on the Potential Risks*, 19 INT. J. ENVIRON. RES. PUBLIC. HEALTH 9960 (2022).

[49] Kaminski, *Regulating the Risks of AI*, *supra* note 33, at 1365 ("When an AI system causes harm, it can be particularly hard to determine causality ex post, because such systems are often technically and legally opaque.").

[50] Alexa Hagerty & Alexandra Albert, *AI Is Increasingly Being Used to Identify Emotions–Here's What's at Stake*, CONVERSATION (2021), http://theconversation.com/ai-is-increasingly-being-used-to-identify-emotions-heres-whats-at-stake-158809 (last visited Mar 22, 2023) (illustrating how AI may be used to identify an interviewer's emotion in a biased manner.

[51] Amanda Parsons & Salomé Viljoen, *Valuing Social Data*, 124 COLUM. L. REV. 993, 997 (2024) (describing this as the prediction value of social data).

[52] Citron & Solove, *supra* note 7, at 818 ("Sharing an innocuous piece of data with another company might provide a key link to other data or allow for certain inferences to be made").

[53] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 99-101 (2014) (discussing predictive privacy harms arising from discriminatory applications); Solove, *AI and Privacy*, *supra* note 14, at 40.

can extend their impact beyond consumers, affecting even non-consumers, various groups, and society at large.

1. Individuals

Consumers often experience direct harm when using services and products powered by AI systems that carry algorithmic harms. At first glance, these AI applications may appear reliable, even if their underlying datasets and models may contain errors, biases, and other design flaws.[54] The problematic nature of these applications is often masked by the apparent benefits they promise, making them seemingly innocuous. This disguise is particularly effective because the benefits are tangible and immediate,[55] whereas the harms are often intangible or not readily apparent. As a result, consumers may be exposed to algorithmic harms without adequate awareness or understanding of the harms involved.

Even non-consumers—those who choose to minimize their exposure to AI-driven services or products—are not immune to algorithmic harms.[56] Both the public and private sectors have increasingly deployed algorithmic decision-making and other AI applications. Colleges and universities use algorithmic systems to decide which applicants should be admitted.[57] Predictive policing algorithms target individuals based on biased crime data.[58] Facial recognition systems installed in public spaces can harvest data from anyone within their range.[59] Data scraping technologies can process images that are publicly available online, capturing individuals who do not intend to interact with these technologies.[60] These applications impact individuals who are not in a consumer setting, so even those not actively using these services can become victims of algorithmic harms. Again, the intangible nature of these harms often hinders these non-consumer victims' harm awareness, leading to greater collective damage to civil rights and democratic values.[61]

[54] Slaughter, Kopec, & Batal, *supra* note 17.

[55] *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008); Chesney & Citron, *supra* note 14, at 1769-71 (introducing beneficial uses of automated systems and deepfakes).

[56] As Part II will indicate, these non-consumers include the general public, such as residents and pedestrians affected by AI systems and algorithmic operations, as well as role-specific individuals impacted by these technologies, such as prospective job seekers, patients, students, and more.

[57] Denisa Gándara et al., *Inside the Black Box: Detecting and Mitigating Algorithmic Bias Across Racialized Groups in College Student-Success Prediction*, 10 AERA OPEN (Jan 2024).

[58] *See* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA L. REV. 109 (2017).

[59] Paul Bischoff, *Facial Recognition Technology (FRT): Which Countries Use It?*, COMPARITECH (Jun. 8, 2021), https://www.comparitech.com/blog/vpn-privacy/facial-recognition-statistics/ (last visited May 19, 2024); Nadia Kanwal, *Facial Recognition Technology Could Soon Be Everywhere – Here's How to Make It Safer*, THE CONVERSATION (2023), http://theconversation.com/facial-recognition-technology-could-soon-be-everywhere-heres-how-to-make-it-safer-205040 (last visited May 19, 2024); Kim Hart, *Facial Recognition Surges in Retail Stores*, AXIOS (2021), https://www.axios.com/2021/07/19/facial-recognition-retail-surge.

[60] Johana Bhuiyan, *Clearview AI Uses Your Online Photos to Instantly ID You. That's a Problem, Lawsuit Says*, LOS ANGELES TIMES, Mar. 9, 2021, https://www.latimes.com/business/technology/story/2021-03-09/clearview-ai-lawsuit-privacy-violations (last visited Oct 4, 2023); Billy Perrigo, *Why Regulators Can't Stop Clearview AI*, TIME, May 2022, https://time.com/6182177/clearview-ai-regulators-uk/.

[61] Smuha, *supra* note 23, at 5.

2. Groups

In addition to their reach to individuals, algorithmic harms can befall a group whose members are related to each other or belong to the same social network. AI systems that collect considerable data on a given user are empowered to gather a correspondingly large amount of data about the user's social network. They can obtain information about their family members, partners, friends, colleagues, and acquaintances through what the user has shared on social media platforms.[62] Any information the user exposes to the public, including data about any other people mentioned by or interacting with the user, may be used to collect information and gain insights about those people as well.[63]

In many other cases, group victims are not related to each other, but are rather individuals who display similar behavioral patterns identified by AI systems. Clustering algorithms can profile and categorize those who have shown similar inclinations, estimating which behavior patterns best signal an individual's personal traits and preferences.[64] Based on the detected behaviors, the algorithms can correlate detected traits to particular mental, physical, or economic conditions.[65] This phenomenon raises concerns over the unauthorized disclosure of a range of sensitive data, including tendencies toward suicidality, [66] Alzheimer's, [67] pregnancy, [68] or shopaholism. [69] To take one example, data broker MedBase200 was found to be monitoring online activities of Facebook users, categorizing them into groups of "sexual violence victims," "survivors of intimate partner violence," or "AIDS patients."[70] Although these inferences involve sensitive personal insights into these potentially vulnerable individuals, entities further use these insights to target them without authorization.[71] Because a person's data can be used to target a group of people without

[62] Daniel J. Solove, *The Limitations of Privacy Rights*, 98 NOTRE DAME L. REV. 975, 978 (2023), ("Individuals make privacy choices that have effects not just for themselves but for many others. For example, sharing one's genetic data also shares the genetic data of one's family members.")

[63] *Id.*

[64] Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L. J. 573, 607 (2021).

[65] Anya E. R. Prince, *Location as Health*, 21 HOUS. J. HEALTH L. & POL'Y 43 (2021) (noting that location data can unveil a massive amount of sensitive information about an individual).

[66] Daniel D'Hotman & Erwin Loh, *AI Enabled Suicide Prediction Tools: A Qualitative Narrative Review*, 27 BMJ HEALTH CARE INFORM. 1, 1 (2020).

[67] Charles R. Marshall & Ijeoma Uchegbu, *Artificial Intelligence for Detection of Alzheimer's Disease: Demonstration of Real-World Value Is Required to Bridge the Translational Gap*, 4 LANCET DIGIT. HEALTH e768 (2022).

[68] Diego Jemio, Alexa Hagerty, & Florencia Aranda, *The Case of the Creepy Algorithm That 'Predicted' Teen Pregnancy*, WIRED (Feb. 2022), https://www.wired.com/story/argentina-algorithms-pregnancy-prediction/.

[69] Amanda L. Giordano, *Online Shopping and "Compulsive Buying-Shopping Disorder,"* PSYCHOLOGY TODAY (Aug. 16, 2022), https://www.psychologytoday.com/us/blog/understanding-addiction/202208/online-shopping-and-compulsive-buying-shopping-disorder (last visited Mar 22, 2023).

[70] Kashmir Hill, *Data Broker Was Selling Lists Of Rape Victims, Alcoholics, and "Erectile Dysfunction Sufferers,"* FORBES, Dec. 19, 2013, https://www.forbes.com/sites/kashmirhill/2013/12/19/data-broker-was-selling-lists-of-rape-alcoholism-and-erectile-dysfunction-sufferers/ (last visited May 19, 2024).

[71] Gary Mortimer & Michael Milford, *When AI Meets Your Shopping Experience It Knows What You Buy – And What You Ought to Buy*, CONVERSATION (Aug. 2018), http://theconversation.com/when-ai-meets-your-shopping-experience-it-knows-what-you-buy-and-what-you-ought-to-buy-101737 (last visited Mar 22, 2023).

directly gathering information from them,[72] such algorithmic practices can trigger harms to multiple interests of these group members, ranging from data privacy to equality and more.

With the rise of automated decision-making, group harms also befall groups of people with shared backgrounds in terms of income, race, gender, nationality, and more.[73] These background-based victims, particularly social minorities, are at a higher risk of experiencing unfair applications resulting from algorithmic bias. Many scholars have emphasized the problem with AI systems reflecting historical discrimination against marginalized groups.[74] One of the most cited examples concerns a widely used health scoring AI that consistently categorizes African Americans into a lower health risk group, regardless of their actual (often poorer) health conditions.[75] This inaccurate and unfair scoring is partly caused by under-representative datasets that generate a negative sampling bias.[76] As a result, AI reproduces the bias problem, leading to a collective harm of discrimination against members of the disadvantaged community. Many other AI applications have been found to disproportionately create group harms to less wealthy groups, raising the risks of disparate impacts and other forms of mistreatment. For marginalized groups, these harms may be life-changing, such as denial of housing, insurance, healthcare, and other vital services.

The existence of group victims multiplies the number of those exposed to algorithmic harms, but most of them have no idea that their interests have been eroded or compromised. Like individual victims, group victims have difficulty being identified as plaintiffs when they are not aware of or lack access to the opaque algorithmic activities that generate intangible harms. Even if group victims are aware of these harms, prevention can be challenging as people have little control over how AI generates insights about them.[77]

3. Society

One of the central issues surrounding algorithmic harms is their social dimension, whereby the broader society suffers from the sum of both individual and group harms. With these overlooked problems, unaddressed harms can befall numerous individuals and groups

[72] Viljoen, *supra* note 64.

[73] Smuha, *supra* note 23, at 5.

[74] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

[75] Obermeyer et al., *supra* note 6.

[76] *Id.*

[77] In spite of reduced exposure to notorious algorithmic practices, many actors can still use AI to extract valuable personal insights from other sources, smart devices, websites, and sensors across cities. Years ago, a notable Target case revealed that a teenager who purchased household products enabled a supermarket to identify the likelihood of her pregnancy through her data. Today, people cannot control how such AI-generated insights may be fed into decision-making algorithms that make prejudicial decisions about our lives. Keith Wagstaff, *How Target Knew a High School Girl Was Pregnant Before Her Parents Did*, TIME (Feb. 2012), https://techland.time.com/2012/02/17/how-target-knew-a-high-school-girl-was-pregnant-before-her-parents/ (last visited Mar 23, 2023); In theory, victims with similar backgrounds may detect the correlation and causation between an AI application and its discriminatory effect. The bonds among group members, such as people of the same color, also consolidate the power of victims to fight against biased algorithmic practices through class actions. Nevertheless, in practice, they are often less wealthy groups who have fewer resources to combat harmful practices.

repeatedly, compromising their interests on a larger scale. At the same time, these harms tend to be intangible and indiscernible, making their aggregate effect difficult to measure. As our democratic society becomes progressively automated, significant numbers of people may find themselves both individual and group victims. The cumulative effect, as elaborated below, can be socially consequential.[78]

A tiny harm such as an individual being subjected to a single occurrence of targeted ads, algorithmic tracking, or a discriminatory chatbot remark is far from a significant issue.[79] But algorithmic harms can have a social impact since they cause a significant number of citizens to suffer from their aggregate effect.[80] The individual algorithms of big tech companies can impact billions of people over time.[81] Additionally, the expanding use of generative AI and other AI tools by various entities also holds the potential to affect millions or even billions, repeatedly inflicting numerous harms on their users. If present trends in AI adoption continue, citizens will be subject to a significant number of unaddressed harms. Group harms further expand the reach of such problems, leading to group victims—particularly marginalized groups—suffering extra compounded injuries. This in turn can ultimately decrease both individual and social expectations of democratic values, leading to substantial impairments to civil liberties like privacy, autonomy, equality, and safety.

Take data privacy, a crucial social value in democratic society, as one illustration. Privacy scholars have recognized the central importance of privacy in preserving the standards of civility that constitute both individuals and the collective community.[82] Privacy is considered a constitutive element of civil society that sets adequate boundaries to protect personal information from various forms of external scrutiny for essential democratic dialogue and self-determination.[83] Today, the rise of invasive AI techniques has challenged permitted levels of external observation of information access. Ubiquitous algorithmic processing results in a society with immense quantities of personal data being collected, processed, and transferred by private entities without knowing authorization, compromising the privacy interests of a vast number of individuals. This leads to a reality in which citizens may be expected to share more personal data given reduced expectations of privacy.[84]

[78] Smuha, *supra* note 23, at 7.

[79] Jennifer Horton, *Companies Are Tracking Your Personal Data Without Your Consent, What You Need to Know*, WBRC (Oct. 5, 2021), https://www.wbrc.com/2021/10/06/companies-are-tracking-your-personal-data-without-your-consent-what-you-need-know/ (last visited Mar 22, 2023) (documenting the phenomenon of omnipresent algorithmic tracking).

[80] For instance, the Cambridge Analytica scandal involved the misuse of the personal data of 87 million Facebook users to create detailed psychographic profiles. Cambridge Analytica used AI algorithms to deliver highly personalized and targeted political advertisements to individuals based on their psychographic profiles, leading to socially consequential manipulation of voter behavior in the 2016 U.S. presidential election.

[81] Big tech companies like Meta have 3.6 billion users. Yao, *supra* note 11.

[82] *See* Joel Reidenberg, *Privacy Wrongs in Search of Remedies*, 54 HASTINGS J. 877, 882-83 (2003) ("Society as a whole has an important stake in the contours of the protection of personal information."); Paul Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1607, 1613 (1999) [hereinafter Schwartz, *Privacy and Democracy*]; Citron & Solove, *supra* note 7, at 818-19; Robert C. Post, *The Social Foundations of Privacy: Community and Self in the Common Law Tort*, 77 CALIF. L. REV. 957, 959 (1989); Citron & Solove, *supra* note 7, at 818-1

[83] *See* Schwartz, *Privacy and Democracy in Cyberspace*, at 1664-65, 1667.

[84] Maurice E. Stucke & Ariel Ezrachi, *How Digital Assistants Can Harm Our Economy, Privacy, and Democracy*, 32 BERKELEY TECH. L.J. 1239, 1279-87 (2017).

Historically, expectation of privacy has been an important metric in deciding how much privacy protection should be granted to citizens.[85] In the United States, as long as there is a reasonable expectation of privacy, the Fourth Amendment mandates that governments obtain search warrants for enforcement purposes.[86] Courts also interpret that citizens shall not have an expectation of privacy for data shared with a third party, such as AI-powered innovations.[87] While doctrinal concerns about diminished expectations of privacy apply primarily to individuals aware of these issues, emergent corporate data scandals, often involving the use of AI, have also brought greater awareness of privacy losses to the public. Additionally, the FTC uses consumers' expectation of privacy as a standard for its enforcement action.[88] As algorithms constantly automate tracking, citizens may assume that their behavior might be monitored to generate more insights about them. This may result in a scenario where citizens suffer from a decreased expectation of privacy and thus a lower level of privacy protection. In this vein, AI not only reduces the scope of privacy protection;[89] it can be both individually and socially harmful due to its harm to privacy as a fundamental element of democratic society.

Other civil rights interests may suffer from reduced protection due to algorithmic harms. As machines become a vital part of modern societies, citizens may find it challenging to eliminate their negative impacts. Because of the pervasiveness of AI developed without regulatory restraints, social members unknowingly suffer from algorithmic harms. These victims barely understand whether and how their private activities are being monitored and utilized by AI in a safe, legal, and responsible way. Although their interests are part of essential democratic values, a range of AI innovations are increasingly empowered to compromise these interests. The resulting harms can be easily neglected by innovators developing AI to pursue tangible benefits like financial profits. Without adequate interventions, citizens are likely to live in a society with a lowered expectation of civil rights and democratic interests.

II. A Typology of Algorithmic Harms

Although algorithmic harms affect numerous victims and societies at large, too little is known about the legal boundaries of these harms. To bridge this gap, this Section outlines the core areas of algorithmic harms that warrant legal recognition. Expanding on Section I's discussions of the features and victims of algorithmic harms, this Section offers a basic typology of the legal problems at stake, illuminating their subsets and causes from technical, commercial, or legal standpoints. Drawing from insights extracted from cases, policy papers,

[85] *See* Matthew B. Kugler & Lior Jacob Strahilevitz, *Actual Expectations of Privacy, Fourth Amendment Doctrine, and the Mosaic Theory*, 2015 S. Ct. Rev. 205 (2016); Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 Mich. L. Rev. 311 (2012).

[86] *Id.*

[87] *See, e.g.*, *United States v. Miller*, 425 U.S. 435, 443 (1976); *Smith v. Maryland*, 442 U.S. 735, 743–44 (1979).

[88] Chris Jay Hoofnagle, Federal Trade Commission Privacy Law and Policy 145, 146 (2016).

[89] Shlomit Yanisky-Ravid & Sean K. Hallisey, *Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbor Regimes*, 46 Fordham Urb. L.J. 428, 470 (2019). ("industry can nonetheless violate expectations of privacy as they exist normatively, and can also erode those expectations over time as technology evolves.").

and legal instruments across the US, EU, and Asia, this Article contends that algorithmic harms involve four primary types and two aggravating factors: (1) eroding privacy, (2) undermining autonomy, (3) diminishing equality, (4) impairing safety, (5) accountability paucity, and (6) algorithmic opacity. Categories one to four encompass *primary harms* to civil rights and democratic values. Categories five and six are *aggravating factors* that compound and prolong those harms, including the absence of transparency or accountability regimes, which hinders the detection, documentation, and eradication of primary harms.

This typology encompasses a broad spectrum of AI practices, though it does not exhaustively cover all dimensions of algorithmic harm. Primary harms mainly cover those that have uniform effects on individual civil rights interests. Non-universal harms such as copyright infringement, unfair competition, and job displacement are not covered in this typology because they are more role-specific and less concerned with universally applicable civil rights interests. This typology also does not consider highly speculative harms such as those associated with artificial general intelligence, a theoretical type of AI that surpasses human cognitive capabilities.[90] In recognition of its limited scope, this typology aims to advance our conceptualization of algorithmic harms, which has been challenging due to their variety and dynamics. The following theory of algorithmic harm seeks to give lawmakers, regulators, innovators, and other stakeholders a new perspective to recognize these universal harms as a starting point.

A. Primary Harms

AI applications commonly give rise to four substantive algorithmic harms: eroding privacy, undermining autonomy, diminishing equality, and impairing safety, which respectively put privacy, autonomy, equality, and safety at risk. Depending on the legal regime, one specific interest may overlap with another. However, each interest can be evaluated and addressed separately based on its distinct meanings. The concept of privacy under the American legal regime, for instance, represents the notion of autonomy in contexts like decisional privacy.[91] Yet autonomy can also be examined separately because it pertains to an individual's freedom of thoughts and right to make personal choices without unwanted external pressure, interference, or manipulation.[92]

These harms are varied, but they are also commonly interlinked. A single AI application can generate multiple algorithmic harms and thus jeopardize multiple interests. For example, AI-powered cancer prediction can detect cancerous cells and estimate survival outcomes at an early stage, yet it may also potentially lead to a series of harms undetectable to those subject to these AI systems. It might begin with eroding privacy caused by the

[90] *Id.*

[91] *See, e.g.*, Griswold v. Connecticut, 381 U.S. 479 (1965) (noting "a right to privacy in the 'penumbras' and 'emanations' of other constitutional protections."). Beyond the U.S. context, the European Convention on Human Rights (ECHR) also recognizes that privacy represents a right to personal autonomy. Guide on Article 8 of the European Convention on Human Rights, European Court of Human Rights (Aug. 31, 2021), https://www.echr.coe.int/documents/guide_art_8_eng.pdf (last visited Mar 28, 2023).

[92] Nicole Legate & Richard M. Ryan, *Individual Autonomy*, *in* ENCYCLOPEDIA OF QUALITY OF LIFE AND WELL-BEING RESEARCH 3233 (Alex C. Michalos ed., 2014), https://doi.org/10.1007/978-94-007-0753-5_140.

transfer of such health prediction data to third parties. While the predictions may not always be accurate, these third parties might utilize such health assessment data for algorithmic decision-making, resulting in discriminatory treatments and diminished equality. Entities may further employ AI-assisted nudges that undermine autonomy by pushing people to purchase excessive health supplements or unproven preventive treatments. This can lead not only to unnecessary financial burdens but also to misleading health decisions that compromise safety.[93] The subsequent part details the main specifics and subcategories of these harms.

1. Eroding Privacy

Eroding *privacy*—an individual right concerning the collection, processing, and disclosure of personal data—involves algorithmic practices that impair our privacy or data protection interests, including large-scale data collection, processing, and generation.

The origin of eroding privacy is closely tied to AI's fundamental reliance on personal data.[94] From a technical perspective, AI systems are built upon large datasets that feed and train machine learning algorithms and other software. This critically drives entities to gather vast amounts of personal data from various sources, as well as frequently use algorithmic methods to collect data from websites, a practice known as scraping. The availability of large amounts of data allows AI systems to generate increasingly precise insights about us.[95] As noted by a number of experts, big data analytics and machine learning algorithms have significantly expanded the capacity to generate a wider scope of personal data.[96] Specifically, this technology not only gathers vast amounts of data, but also draws inferences about highly sensitive, intimate, and private aspects of individuals' lives. This reliance on data, coupled with AI's ability to generate personal insights, motivates entities to harvest and process personal data on a large scale.[97] Driven by data's immense economic value, these practices have encroached upon people's private spheres, extracting and processing huge amounts of personal data without meaningful consent. The following discussion shows how eroding privacy impinges on our privacy interests in ubiquitous, intangible, and cumulative ways.[98]

[93] Nudges, as a concept from behavioral economics studied by Thaler and Sustein, refer to "any aspect of choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives." RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE, PENGUIN BOOKS (2008).

[94] *See, e.g.*, Griswold v. Connecticut, 381 U.S. 479 (1965) (noting "a right to privacy in the 'penumbras' and 'emanations' of other constitutional protections."). Beyond the U.S. context, the European Convention on Human Rights (ECHR) also recognizes that privacy represents a right to personal autonomy. Guide on Article 8 of the European Convention on Human Rights, European Court of Human Rights (Aug. 31, 2021), https://www.echr.coe.int/documents/guide_art_8_eng.pdf (last visited Mar 28, 2023).

[95] *See* ARTIFICIAL INTELLIGENCE: WHAT IT IS AND WHY IT MATTERS, SAS INST., https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html.

[96] Crawford & Schultz, *supra* note 53, at 94.

[97] Justin Sherman, *How Shady Companies Guess Your Religion, Sexual Orientation, and Mental Health*, SLATE (Apr. 2023), https://slate.com/technology/2023/04/data-broker-inference-privacy-legislation.html.

[98] Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J. L. TECH. 106 (2019) (explaining the risks IoT technologies pose to data privacy protection).

The first harm concerns the ubiquity of algorithmic extraction of personal data. As AI systems are increasingly operating our smartphones, the internet, chatbots, driving assistants, unmanned vehicles, and more, their ubiquity enables them to harvest and process significant amounts of personal data. In the United States, digital lenders commonly collect consumer data through social media platforms, websites, and various apps.[99] Similar practices are also adopted by landlords that use AI to identify qualified tenants and retailers that extract data through mobile phone activities to analyze customer shopping habits.[100] Many firms have been applying tracking algorithms to surveil the behaviors and activities of consumers.[101] Scraping has also become increasingly common, resulting in the gathering of massive amounts of publicly available data.[102] From the private corners of our homes to the public records, websites, and areas of a city, AI captures a wide range of information. Its extraction scope includes, but is not limited to, many personal aspects of an individual such as locations, emotions, income, relationships, and health conditions, to name a few. As many cases have indicated, entities use this technology to track individual activities without authorization.[103] While these data processing activities have become more and more ubiquitous, they come in intangible forms. This enables many entities to engage with such data collection without obtaining meaningful consent, as individuals often have no idea that their data has been accumulated, by whom, or for what purposes.

Algorithmic extraction of personal data results in lasting harm when it involves biometric information because biometric identification systems process personal data that is highly sensitive in nature. Given the distinctness of such information, when biometrics are leaked, the damage may be irreparable. There is no way to replace biometrics like fingerprints to protect individuals from future identity theft and other security risks.[104] However, as shown in real-life cases, entities can easily gather extensive sensitive data with the aid of this technology.[105] Clearview AI, for example, obtained facial geometry information of individuals to identify people by scraping billions of photos from the internet.[106] British

[99] *See* Jennifer Valentino-Devries, Natasha Singer, Michael H. Keller & Aaron Krolik, *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

[100] James Green, *3 Ways Customer Data Allows for Pinpoint Marketing*, ENTREPRENEUR (July 24, 2015), https://www.entrepreneur.com/article/247372.

[101] Janakiram MSV, *Why AIoT Is Emerging as the Future of Industry 4.0*, FORBES (Aug. 12, 2019), https://www.forbes.com/sites/janakirammsv/2019/08/12/why-aiot-is-emerging-as-the-future-of-industry-40/?sh=31a56ce5619b.

[102] *See* Daniel J. Solove & Woodrow Hartzog, *The Great Scrape: The Clash Between Scraping and Privay*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4884485.

[103] *See* Jennifer M. Urban, Chris Jay Hoofnagle, & Su Li, *Mobile Phones and Privacy*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103405; Michael Ehret & Jochen Wirtz, *Unlocking Value from Machines: Business Models and the Industrial Internet of Things*, 33(1) J. MKTG. MGMT. 111 (2016).

[104] Matthew Fennell, *Korea's Biometric Data Dilemma*, ASIA SOCIETY (Dec. 2021), https://asiasociety.org/korea/koreas-biometric-data-dilemma (last visited Mar 21, 2023).

[105] Satariano & Hill, *supra* note 4. ("Those subject to facial recognition have little way of knowing they are on the watchlist or how to appeal. In a legal complaint last year, Big Brother Watch, a civil society group, called it "Orwellian in the extreme.")

[106] United States: United States District Court Northern District of Illinois et al., *In Re: Clearview AI, Inc., Consumer Privacy Litigation. 1:21-cv-00135* (2022), https://www.govinfo.gov/app/details/USCOURTS-ilnd-1_21-cv-00135 (last visited Mar 21, 2023).

retailers have also widely adopted Facewatch, a facial recognition system, to conduct biometric scans and identify potential shoplifters among thousands of supermarket shoppers.[107] Unauthorized extraction of sensitive information in these examples not only compromises privacy interests but also paves the way for security incidents like identity theft.[108]

Following algorithmic data extraction, a second form of eroding privacy occurs: algorithmic generation of personal data. As previously stated, AI systems can potentially acquire comprehensive information and broaden the scope of data that can identify a person.[109] Many firms have applied predictive analytics to make predictions and assessments about consumers. Combining multiple sources of data collected from websites, public records, and smart devices, AI can identify underlying patterns in data to generate various insights about specific individuals.[110] These insights increasingly involve sensitive personal data. For instance, with linguistics prediction models that assess language use on social media posts, entities can estimate the psychological conditions of users,[111] or they can predict one's likelihood of developing skin cancer through photo analysis performed by machine learning algorithms;[112] generative AI also features useful visual analysis techniques to assess people's facial expressions, genders, and emotions.[113] The consequence of algorithmic inferences is revealing substantially more insights about individuals as countless pieces of data aggregate.[114] This data generation, however, typically bypasses privacy laws, preventing individuals from controlling which significant insights about them are known by entities.[115]

When algorithmic extraction and data generation converge, they lead to ubiquitous surveillance that undermines our privacy. The prevalence of data harvesting brings increasing data about our personal lives and social activities into the algorithmic space, leading to prolonged monitoring of our actions and thoughts. However, unlike traditional tangible surveillance tools such as cameras, AI-driven surveillance operates invisibly. Before an individual is aware of it, their data is stored for unknown future use by unknown actors. Over

[107] Satariano & Hill, *supra* note 4.

[108] Manheim & Kaplan, *supra* note 98, at 123.

[109] Nicholson Price, *Problematic Interactions Between AI and Health Privacy*, 2021 UTAH L. REV. 925, 928-29 (2021) (examining how AI, with its powerful data inference capabilities, can potentially compromise a patient's privacy by analyzing their health data).

[110] Kashmir Hill, *How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012), https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-beforeher-father-did/#546582fa6668.

[111] Starpre Vartan, *Racial Bias Found in a Major Health Care Risk Algorithm*, SCIENTIFIC AMERICAN (Oct. 24, 2019), https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/; Chris Poulin et al., *Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes*, 9 PLOS ONE (2014).

[112] Amanda Capritto, *4 Ways to Check for Skin Cancer with Your Smartphone*, CNET (Jan. 1, 2020), https://www.cnet.com/health/personal-care/how-to-use-your-smartphone-to-detect-skin-cancer/; Ryen W. White, Murali Doraiswamy & Eric Horvitz, *Detecting Neurodegenerative Disorders from Web Search Signals*, NPJ DIGITAL MEDICINE 1 (2018); Bo Zhang, Huiping Shi & Hongtao Wang, *Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach*, 16 J MULTIDISCIP HEALTHC 1779 (2023).

[113] Kashmir Hill, *OpenAI Worries About What Its Chatbot Will Say About People's Faces*, N.Y. TIMES (July 18, 2023), https://www.nytimes.com/2023/07/18/technology/openai-chatgpt-facial-recognition.html.

[114] SOLOVE, THE DIGITAL PERSON, *supra* note 10, at 44-47.

[115] *See* Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. L. REV. 357 (2022).

time, AI can transform minor monitoring into mass surveillance, empowering private entities and governments to identify, target, and control individuals on a larger scale.[116]

Yet one cannot anticipate the ultimate privacy harm involved in an innocuous data collection practice such as liking a Facebook page or participating in a cancer research survey,[117] which results in inferences of sensitive information ranging from health conditions to political views. Data brokers and entities exploit this by gathering a wide range of data from various sources to create detailed individual profiles that contain sensitive information, intruding upon the secluded sphere that data privacy law was intended to safeguard.[118] Such collected and generated data can be used for extensive surveillance and can undermine autonomy, as discussed in the following section. Given the intangibility of these harms, individuals cannot reasonably perceive or anticipate the cumulative effect of each particular data processing activity,[119] making algorithmic harms to privacy hard to address.[120]

2. Undermining Autonomy

Autonomy refers to an individual's liberty to secure sovereignty over their free will, such as making personal choices without undue intervention.[121] In algorithmic settings, undermining autonomy involves activities that leverage AI to intervene one's freedom of choice, including manipulating or impairing a person's cognition and decision-making processes. Two essential sets of algorithmic techniques have been linked to manipulative practices that undermine individual autonomy: (1) trickery, deception, or pressure that reduces one's decision-making capacity; and (2) personalized content that targets one's personal trait to influence one's decision-making process.[122]

A major cause of undermining autonomy stems from AI's ability to significantly enhance precise personalization. Among firms that prioritize value creation, personalization has long been a widely adopted practice, tailoring services to match people's individualized preferences and enhancing customer satisfaction and attachment to the service.[123] Based on insights gathered from personal data, AI enables a higher degree of precision and scalability of intelligent personalization.[124] However, in various scenarios, personalization transforms

[116] Solove, *AI and Privacy*, *supra* note 14, at 50.

[117] Crawford & Schultz, *supra* note 53, at 106.

[118] Sherman, *supra* note 97.

[119] Crawford & Schultz, *supra* note 53, at 106.

[120] *See generally*, Jeffrey M. Skopek, *Untangling Privacy: Losses Versus Violations*, 105 IOWA L. REV. 2169, 2229-30 (2020).

[121] The legal construct of autonomy is often grounded in the legal concept of liberty. MORTIMER SELLERS, AUTONOMY IN THE LAW 2 (2007) ("If law seeks justice (as it should), then law will protect liberty, and autonomy will always be a central element in law."); James E. Fleming, *Securing Deliberative Autonomy*, 48 STAN. L. REV. 1 (1995) (arguing autonomy manifests as an unenumerated fundamental right as a type of substantiative liberties).

[122] Yeung, *supra* note 15.

[123] David C. Edelman & Mark Abraham, *Customer Experience in the Age of AI*, HARVARD BUSINESS REVIEW, Mar. 2022, https://hbr.org/2022/03/customer-experience-in-the-age-of-ai (last visited May 19, 2024).

[124] How AI Can Scale Personalization and Creativity in Marketing, HARVARD BUSINESS REVIEW, Aug. 2023, https://hbr.org/sponsored/2023/08/how-ai-can-scale-personalization-and-creativity-in-marketing.

into customized algorithmic manipulation, diminishing an individual's ability to think and act rationally.[125] AI utilizes insights extracted from an individual's activities to understand their personal needs and vulnerabilities. As discussed in the section on eroding privacy, massive amounts of data have been used to construct a detailed profile of a person. These insights serve as useful ingredients for machine learning algorithms to capture individual preferences, predict their behaviors, and target their weaknesses for manipulative purposes.[126]

In addition to personalization enhanced by personal insights, undermining autonomy also involves architectural techniques that deliver tailored content in ways that are either repetitive, frequent, or subtly manipulative. Through delicate control of the frequency, content, and format presented to users, AI systems can selectively filter out information that contradicts the interests of the entities they serve.[127] As the following part will explain, people's cognition, preferences, and decisions can be subtly molded by algorithmic practice.[128] This intervention, being pervasive, intangible, and often repetitive, cumulatively undermines one's freedom of thought. Over time, subtle manipulation causes substantial harms to personal autonomy for the ultimate benefit of the manipulators.[129]

The first common form of autonomy harm concerns algorithmic synthetic content that undermines our ability to recognize what is real. Problems emerge when generative AI can create convincing content that appears authentic and legitimate. Fake videos and images produced by deepfakes have manipulated people into misjudging the trustworthiness of content providers.[130] The images of child sexual abuse created by generative AI are so realistic that people find it challenging to distinguish fabricated content from real criminal reporting.[131]

Manipulative AI also undermines freedom of personal choice in the marketplace, as exemplified by pervasive behavioral AI applications.[132] Machine learning techniques gather real-time information about potential consumers to estimate how to motivate them to take

[125] Yeung, *supra* note 15.

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

[127] Dirk Helbing et al., *Will Democracy Survive Big Data and Artificial Intelligence?*, SCI. AM. (Feb. 25, 2017), https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/.

[128] Silvio Palumbo & David Edelman, *What Smart Companies Know About Integrating AI*, HARV. BUS. REV. (2023), https://hbr.org/2023/07/what-smart-companies-know-about-integrating-ai.

[129] ROSTAM J. NEUWIRTH, THE EU ARTIFICIAL INTELLIGENCE ACT: REGULATING SUBLIMINAL AI SYSTEMS 86 (2022), https://www.routledge.com/The-EU-Artificial-Intelligence-Act-Regulating-Subliminal-AI-Systems/Neuwirth/p/book/9781032333755 (last visited Mar 23, 2023) ("the explicit reference to subliminal perception must be regarded as an important step in the protection of the human right to freedom of thought and personal mental privacy, as a way to maintain an environment free from manipulation and harmful interference with a person's control over her own mind").

[130] Chesney & Citron, *supra* note 14, at 1776.

[131] Issie Lapowsky, *The Race to Prevent 'the Worst Case Scenario for Machine Learning,'* N.Y. TIMES (Jun. 24, 2023), https://www.nytimes.com/2023/06/24/business/ai-generated-explicit-images.html (last visited Jul 27, 2023); Leonardo Nicoletti & Dina Bass, *Humans Are Biased. Generative AI Is Even Worse*, BLOOMBERG (July 27, 2023), https://www.bloomberg.com/graphics/2023-generative-ai-bias/.

[132] *See generally* Chris Jay Hoofnagle et al., *Behavioral Advertising: The Offer You Cannot Refuse*, 6 HARV L. & POL'Y REV 273 (2012).

a particular action for the benefits of certain manipulators.[133] This phenomenon commonly involves a range of profiling activities operated by AI systems,[134] starting with algorithmic analysis of individual behavior patterns for prediction of personal desires, income levels, and more. AI systems then refine one's environment choices based on insights extracted from individual and population-wide surveillance, recommending content tailored to estimated needs.[135] Behavioral AI techniques are influential due to their remarkable ability to discern the motivating factors behind human behaviors. Through dynamic continuous interpretation of cognitive disposition, AI can offer a set of tailored content that caters to individual real-time preferences.

While such personalization seems to benefit users, it can nonetheless be harmful to individual autonomy. Algorithmic tailoring is not solely operated to help users make optimal decisions; it is deployed to nudge users toward making decisions that are optimal to businesses or other actors. Many algorithmic systems operate to shape individual preferences, compelling them to become attached to specific services or certain public figures. By altering the choices people see, algorithmic systems influence individuals to make decisions they may not otherwise make.[136] This autonomy harm, also termed "hypernudge," stems from the use of big data that subtly influences users, whereby advertisers steer consumer decisions in alignment with commercial objectives.[137] Another way autonomy is undermined involves repeated annoyances that compel users to make decisions they would not otherwise make. Streaming service providers like YouTube use algorithmic systems to set up advertisements played with maddening frequency until users agree to subscribe. Although algorithmic manipulation initially manifests as soft power, its impact is augmented by the evolving, dynamic, and ubiquitous characteristics of machine learning algorithms that make correlations unobservable to human cognition.[138] The algorithms modify the structure of choices received by a user, predictably changing one's behavior without preventing them from making a decision freely or apparently altering one's economic incentives.[139] Their strong power in turn systematically influences the behaviors of individuals to serve the interests of private entities without adequate legal restraints, leading to illegitimate ends such as deceptive exploitation of cognitive vulnerabilities and/or distortions in decision-making.[140] Altogether, AI harms the ability to make informed, rational, and meaningful decisions. When there is no other platform offering the same service without obscure manipulative practices, users must endure such manipulation in many digital contexts.

Finally, and perhaps most importantly, algorithmic manipulation is socially harmful due to its detrimental effects on democracy. In the realm of privacy, firms design their

[133] Azati Team, *How Artificial Intelligence (AI) Is Used in Targeted Marketing*, AZATI: UNITING EXPERTS TO FULFIL IMPORTANT PROJECTS (2020), https://azati.ai/ai-targeted-marketing/ (last visited Mar 21, 2023).

[134] Calo, *supra* note 126, at 1017.

[135] Yeung, *supra* note 15, at 122.

[136] *See* Sophie C. Boerman et al., *Online Behavioural Advertising: A Literature Review and Research Agenda*, 46(3) J. AD 363 (2017).

[137] Yeung, *supra* note 15.

[138] *Id.* at 122.

[139] *Id.* at 120.

[140] *Id.* at 124.

websites or platforms to actively encourage visitors to share more personal information. A common practice involves the use of algorithmic systems to extract data and make sharing a default setting. To take one example, Facebook was found to push users to add details to their personal profiles. The firm stated that they would not sell their information but then gave over a hundred firms access to that information.[141] Another example is seen in the context of political elections. In the United States and elsewhere, politicians have deployed AI systems to conduct covert election manipulation.[142] As machine learning models are embedded in search engine and social media services, they have been used to determine the range of content their users receive.[143] These AI systems filter the information one reads, constructing and constricting one's understanding of what is truly happening in the world without giving users the chance to challenge their authenticity or objectivity. In South Korea, search engine algorithmic filtering has been found to favor certain political parties or candidates.[144] This results in unequal exposure, with some candidates receiving more favorable coverage and others less, thereby distorting public perception and affecting voting behavior.[145] Similarly, during the 2016 US presidential election period, machine learning systems were employed to identify swing voters, sending them a set of tailored search results, news feeds, and images.[146] When implemented on a large scale, content customized by algorithms significantly influences swing voters' decisions and consequently the outcomes of elections. These diminishments of autonomy are often invisible and subtle, although it has led to consequential manipulation at both personal and collective levels.[147]

3. Diminishing Equality

Equality often denotes treating people without favoritism, bias, or discrimination, ensuring that individuals receive equal treatment and are considered equals.[148] In AI settings, this piece focuses on three dimensions of diminishing equality, where AI use has been widely found to unequally allocate benefits and risks among users and other stakeholders.[149] These

[141] Gabriel J. X. Dance, Michael LaForgia & Nicholas Confessore, *As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants*, N.Y. TIMES (Dec. 19, 2018), https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html (last visited Jan 28, 2024).

[142] *See* Sunny Yoon, *Techno Populism and Algorithmic Manipulation of News in South Korea*, 18 J. CONT. E. ASIA 33 (2019).

[143] Sonia K Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 91 (2019) [hereinafter Katyal, *Private Algorithmic Accountability*].

[144] Hyun-woo Nam, *Naver Fined W26.7 Bil. for Manipulating Search Algorithm*, KOREA TIMES, Oct. 6, 2020, https://www.koreatimes.co.kr/www/tech/2023/02/133_297112.html (last visited Feb 25, 2023).

[145] Sunny Yoon, *Techno Populism and Algorithmic Manipulation of News in South Korea*, 18 JOURNAL OF CONTEMPORARY EASTERN ASIA 33 (2019).

[146] *Id.*

[147] Chesney & Citron, *supra* note 55, at 1772.

[148] Stefan Gosepath, *Equality*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., Summer 2021 ed. 2021), https://plato.stanford.edu/archives/sum2021/entries/equality/.

[149] A notion often related to equality is equity, which involves providing individuals with the resources they need. The concepts of equality and equity are often used interchangeably, despite being distinct and contested. While some AI applications aim to promote equity, this piece is less concerned with the promotion of equity. Rather, it focuses on the harm narrative of those that result in unequal conditions or outcomes, often linked to the legal concept of unfairness.

dimensions include AI applications that benefit people differently, produce biased content, and render discriminatory decisions against certain groups.

Diminishing equality typically results from the technical traits of AI systems, which are reflections of the data on which they are fed, trained, and tested. The data, along with AI models and their training processes, tend to replicate historical stereotypes and inequalities with deep social roots.[150] Scholars in computer science, law, psychology, and other areas have identified a host of bias arising from AI applications. Unrepresentative training data in supervised learning often leads to the underrepresentation and exclusion of certain groups. Unconscious biases embedded by the designs of these AI systems tend to detrimentally impact certain communities. A range of other biases—including selection, categorization, cognitive, position, and web-based biases, along with the exclusion of outliers—further risks unfair outcomes and can escalate into structural discrimination. The biased operating results of AI systems tend to unequally distribute gains and perils, subtly diminishing equality.[151]

The first type of diminishing equality concerns unequal design applications—that is, AI designs that are applied unequally to different users. Many models of AI systems are said to contain design flaws such as faulty inputs or a failure to test.[152] For instance, numerous AI systems are trained on data that underrepresents certain groups of people, leading to unequal applications of these systems or biased outcomes. This type of harm has emerged in many AI-based services or products for quite a long time, where they tend to characterize people in a discriminatory manner or disfavor certain groups. Facial recognition systems recognize the opened eyes of Asian Americans as blinking;[153] image categorization labels African Americans as gorillas;[154] translation AI associates female engineers with being male; generative AI depicts African cultural images as ruined buildings.[155] Some instances reveal that AI applications perform better for certain groups but are worse for others because of datasets that exclude marginalized groups, including individuals from lower socioeconomic backgrounds, people of color, Native Americans, and immigrants, among others. This exclusion diminishes equality in non-obvious ways. For example, a facial recognition system that can accurately identify 85% of users may fail to identify the whole Black population, yet the misidentified individuals often lack data to demonstrate a disparate impact or treatment. Moreover, unequal AI applications often result from a contextual disconnect between where AI is trained and where it is used. For instance, AI trained in high-resourced settings may perform poorly in low-resourced contexts. This discrepancy prevents individuals in less

[150] *See* Anupam Chander, *The Racist Algorithms?*, 115 MICH. L. REV. 1023, 1023-45 (2017).

[151] *Id.*

[152] Slaughter, Kopec, and Batal, *supra* note 17.

[153] *See* Selina Cheng, *An Algorithm Rejected an Asian Man's Passport Photo for Having "Closed Eyes,"* QUARTZ (Dec. 7, 2016), https://qz.com/857122/an-algorithm-rejected-an-asian-mans-passport-photo-for-having-closed-eyes (last visited Mar 21, 2023).

[154] *See* Jessica Guynn, *Google Photos Labeled Black People "Gorillas,"* USA TODAY (July 1, 2015), https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/ (last visited Mar 21, 2023).

[155] Zachary Small, *Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History*, N.Y. TIMES (July 4, 2023), https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html. Nicoletti & Bass, *supra* note 131.

developed areas from receiving high-quality AI services, such as healthcare recommendations.[156] When AI designs are applied unequally to different users, they prevent less resourced and marginalized groups from enjoying the same benefits of AI applications. This can also lead to errors and exclusion of services, imposing additional life hurdles on these victims. Many of these problems, however, are hard to discern and seem minor when evaluated in isolation, concealing their tendency to disadvantage certain groups.[157]

The second category of diminishing equality pertains to the extensive circulation of bias whereby AI contributes to inequality in the real world. Since machine learning bots began to be widely adopted, chatbots have served as a superspreader of human bias.[158] In 2016, Microsoft's Tay, a machine-learning chatbot,[159] became a sexist, racist, and genocidal Nazi within a day of interaction with users.[160] Another prominent chatbot, ChatGPT, still faces this issue six years later. After chatting with around 100 million users, it turned into a sexist and racist, encouraging terrorist-style racism. The chatbot suggested torturing Iranians and wrote biased reviews that discriminated against women and African Americans.[161] Elsewhere, a South Korean chatbot also learned to call lesbian, Black, and disabled people "disgusting" after chatting with users.[162] Even if the discriminatory remark is partly due to user input, it can lead to amplified racist, sexist, and discriminatory inputs affecting countless other AI systems and users.[163] In today's generative AI era, chatbots and other generative tools have become one of the most commonly adopted AI innovations with which individuals interact. In various contexts, from texts to images, generative tools tend to interpret minority groups as less socially desirable, subjecting them to inferior or distorted social images that are far from reality.[164] While these harms may seem subtle and attributable to users, AI's widespread circulation of bias can become socially harmful over time as it repetitively reinforces stereotype, microaggressions, or dehumanization across algorithmic systems and among numerous users who increase their reliance on these services. These users receive countless biased and misleading content without taking these intangible harms seriously. With their increasing user base, they play a crucial role in distributing discriminatory ideology on a significant scale, which may ultimately impede our ongoing efforts to eradicate prejudice.

[156] Price, *Contextual Bias*, *supra* note 15, at 68 (identifying that contextual bias emerges when algorithmic systems are shifted from one context to another).

[157] Small, *supra* note 155.

[158] *Id.*

[159] Daniel Victor, *Microsoft Created a Twitter Bot to Learn from Users. It Quickly Became a Racist Jerk.*, N.Y. TIMES (Mar. 24, 2016), https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html (last visited Mar 15, 2023).

[160] *Id.*

[161] Kieran Snyder, *We Asked ChatGPT to Write Performance Reviews and They Are Wildly Sexist (and Racist)*, FAST COMPANY (2023), https://www.fastcompany.com/90844066/chatgpt-write-performance-reviews-sexist-and-racist (last visited Mar 15, 2023).

[162] *Korea's Controversial AI Chatbot Luda to Be Shut Down Temporarily*, PULSE (Jan. 12, 2021), https://pulsenews.co.kr/view.php?year=2021&no=34618.

[163] Ananya, *AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed?*, 627 NATURE 722 (2024).

[164] Tiku, Schaul, and Chen, *supra* note 13.

Last but not least, as human decision-making is increasingly delegated to AI systems, they undermine equality through unfair algorithmic choices that generate unequal consequences across various social contexts. Algorithmic decision-making has been deployed in many important social settings such as employment, health, housing, education, and more.[165] The seemingly objective decisions made by machines have been found to exhibit bias against groups. In healthcare scenarios, under-represented or overrepresented patient cohorts have led to discriminatory AI decisions that limit racial minorities' access to medical treatment.[166] In professional settings, many firms have adopted hiring algorithms that tend to filter out female candidates, stemming from the training data's unfavorable perception of female applicants.[167] Automated decision-making also disadvantages certain groups based on algorithmic correlations that rely on unrelated factors. Mortgage lenders have used predictive loan assessment services that discriminate against low-income and minority applicants.[168] Advertising algorithms assign consumers rankings based on their gender and income, restricting the range of options that lower-income users see in real life,[169] such as postings for higher-paying jobs. Those adversely impacted by these algorithmic decision-making operations often lack the resources to detect intangible bias or unfairness in decision-making processes, and are thus not adequately identified as victims.

Through biases in design systems, generative models, and decision-making, these seemingly beneficial AI applications perpetuate stereotypes and may continue to generate equality harms associated not only with race, gender, and income, but also age, class, nationality, sexual orientation, disability, and many other categories. The reinforcement of stereotypes seems negligible in many cases, yet its cumulative impact continuously shapes how citizens see themselves and others through a biased lens. This has blocked under-represented groups from certain social services and opportunities, subjecting them to increased harms from undue denial of services, exclusion from social opportunities, and undesirable social images. In a reality in which AI reproduces wrongful misrepresentation, embedding these biases into a growing array of AI-generated decisions and content, these distortions can solidify into lasting, compounded inequality.

4. Impairing Safety

Impairing safety refers to AI operations that threaten people's safety and security, encompassing both mental and physical health. While physical injuries, psychological harms, and security breaches represent different dimensions of safety interests, their commonality

[165] *See* Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMMUN. ACM 44–54 (2013).

[166] Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447 (2019); Kara Manke, *Widely Used Health Care Prediction Algorithm Biased Against Black People*, BERKELEY NEWS (Oct. 24, 2019), https://news.berkeley.edu/2019/10/24/widely-used-health-care-prediction-algorithm-biased-against-black-people/?fbclid=IwAR21ND23XtA6GZXKLBe15SajborPwJaGi2gksEek7o5Ju1Kea9JM1lf3IiE.

[167] Council of Europe, DISCRIMINATION, ARTIFICIAL INTELLIGENCE, AND ALGORITHMIC DECISION-MAKING 10 (2018), https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

[168] Edmund L. Andrews, *How Flawed Data Aggravates Inequality in Credit*, STANFORD HAI (Aug. 6, 2021), https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit (last visited Mar 29, 2023).

[169] Katyal, *Private Algorithmic Accountability*, *supra* note 143, at 91.

lies in the often invisible threats they pose to public safety, caused by the increasing reliance on and delegation to AI systems as they grow in intelligence.

As machines grow in intelligence and interactive capability, their influences have grown significantly. While these applications ought to be safe, several technical and business factors have generated more safety and security problems for the general public. From a technical standpoint, design flaws like ambiguity, bias, and error often lead to flawed results.[170] Deficiencies and misalignments of interests in algorithmic programming cause AI systems to struggle when reacting to unanticipated scenarios and, in some cases, leading them to pursue harmful alternatives rather than desirable solutions to achieve AI's preset goals.[171] Even with high-quality algorithmic programming, the inner workings of machine learning algorithms change based on their post-design experiences, introducing additional complexity,[172] unpredictability,[173] and uncontrollability[174] in managing subsequent risks.[175]

For AI systems that control hardware, incidents can cause immediate harm to physical safety. While fully AI-powered lethal systems are still years away, advancements in AI are making these systems increasingly powerful and potentially dangerous if they fall into the wrong hands. The ongoing conflict in Ukraine has highlighted the use of automated weaponry such as loitering munitions—drones capable of hovering in the sky to attack specific targets. [176] Simultaneously, Pentagon is developing drone swarms, potentially numbering in the thousands, for both surveillance and combat purposes. [177] The potential for individuals with malicious intent to exploit autonomous systems, which can operate with minimal human oversight, raises public safety concerns.[178]

Aside from powerful autonomous weapons, a range of other AI-enabled autonomous devices, including driverless cars, buses, and assistive robots, are emerging with safety issues. Privately designed autonomous vehicles have caused a number of severe injuries and even fatalities.[179] A female pedestrian was killed by Uber's driverless car that

[170] *See, e.g.*, Barocas & Selbst, *supra* note 74, at 680, 688; Willem Sundblad, *Data Is the Foundation for Artificial Intelligence and Machine Learning*, FORBES (Oct. 18, 2018, 10:30 AM), https://www.forbes.com/sites/willemsundbladeurope/2018/10/18/data-is-the-foundation-for-artificial-intelligence-and-machine-learning/#65bca29451b4.

[171] *See* Scherer, *supra* note 1, at 366-69.

[172] Anjanette H. Raymond et al., *Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias*, 15 NW. J. TECH. & INTELL. PROP. 215, 220–21 (2018).

[173] Scherer, *supra* note 1, at 365–66.

[174] *See id.* at 366.

[175] *See* Burrell, *supra* note 18, at 5; *id.* at 365.

[176] Eric Lipton, *From Land Mines to Drones, Tech Has Driven Fears About Autonomous Arms*, THE NEW YORK TIMES, Nov. 21, 2023, https://www.nytimes.com/2023/11/21/us/politics/drones-ai-weapons-war.html.

[177] *Id.*

[178] *See* Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CAR. L. REV.; Rebecca Crootof, *Autonomous Weapon Systems and the Limits of Analogy*, 9 HARV. N'TL SEC. J. 51 (2018).

[179] Faiz Siddiqui, *Silicon Valley Pioneered Self-Driving Cars. But Some of Its Tech-Savvy Residents Don't Want Them Tested in Their Neighborhoods*, WASH. POST (Oct. 3, 2019, 10:16 AM), https://www.washingtonpost.com/technology/2019/10/03/silicon-valley-pioneered-self-driving-cars-some-its-tech-savvy-residents-dont-want-them-tested-their-neighborhoods/.

neglected to actually *see* her.[180] Similarly, Tesla's autopilot vehicles have already caused some fatal crashes and other serious accidents.[181] During the 2020 Paralympic Games in Tokyo, a visually impaired athlete was hit by an autonomous bus that lost control.[182] Beyond transportation, AI innovations intended to enhance safety have also proven problematic. In one case, an advanced chatbot designed to assist visually impaired people gave wrong instructions on the position of buttons on a remote control. If visually impaired users rely on incorrect instructions in high-stakes settings, such as operating an oven or stove, they may face hazardous situations. This could lead to accidental activation of the wrong burner or setting the temperature too high, resulting in a fire or severe burn.[183]

In addition to issues with AI systems that control hardware, software-related problems also contribute to impairing safety.[184] As a 2023 experiment with AI chatbots reveals, uncensored chatbots have guided users to hurt other individuals, use drugs, and commit suicide.[185] Generative chatbots that give suggestions in a persuasive manner have also contributed to various risks to users' health and well-being.[186] These chatbots have provided personal life advice without proper supervision or license, raising concerns about the wellness of individuals who heavily depend on answers offered by AI, especially disadvantaged populations who cannot afford to hire lawyers or visit doctors, to make choices about their personal lives.[187]

The third source of safety concerns is the design of AI business models that impair users' health. In the United States, social media algorithms are engineered to enhance user engagement with few legal restrictions. These algorithms harvest data on individual behavior, send personalized content based on these insights, and enhance user reliance on their service through this personalization.[188] Initially, these algorithmic practices were deemed beneficial

[180] Sam Levin & Julia Carrie Wong, *Self-Driving Uber Kills Arizona Woman in First Fatal Crash Involving Pedestrian*, GUARDIAN (Mar. 19, 2018, 6:48 PM), https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe.

[181] Sean O'Kane, *Tesla Hit with Another Lawsuit over a Fatal Autopilot Crash*, VERGE (Aug. 1, 2019, 5:59 PM), https://www.theverge.com/2019/8/1/20750715/tesla-autopilot-crash-lawsuit-wrongful-death.

[182] Paul MacInnes, *Toyota Pauses Paralympics Self-Driving Buses After One Hits Visually Impaired Athlete*, GUARDIAN (Aug. 28, 2021), https://www.theguardian.com/technology/2021/aug/28/toyota-pauses-paralympics-self-driving-buses-after-one-hits-visually-impaired-athlete (last visited Mar 21, 2023).

[183] Hill, *supra* note 113. Furthermore, the healthcare industry that increasingly depends on AI for faster diagnoses and cheaper medical treatment is facing the risk of widespread misdiagnoses and other problems due to inherent flaws or biases in AI-enabled devices. Price, *Contextual Bias*, *supra* note 156, at 90–99 (2019); Robert David Hart, *When Artificial Intelligence Botches Your Medical Diagnosis, Who's to Blame?*, QUARTZ (May 23, 2017), https://qz.com/989137/when-a-robot-ai-doctor-misdiagnoses-you-whos-to-blame/.

[184] Stuart A. Thompson, *Uncensored Chatbots Provoke a Fracas Over Free Speech*, N.Y. TIMES (July 2, 2023), https://www.nytimes.com/2023/07/02/technology/ai-chatbots-misinformation-free-speech.html.

[185] *Id.*

[186] Charlotte Tschider, *Humans Outside the Loop*, YALE J. L. & TECH forthcoming, https://papers.ssrn.com/abstract=4580744 ("Although GAI might seem harmlessly expressive, they are being positioned to power chat and other communication-based tools that involve interacting with humans and directing human behavior").

[187] Nico Grant, *Google Tests an A.I. Assistant That Offers Life Advice*, N.Y. TIMES (Aug. 16, 2023), https://www.nytimes.com/2023/08/16/technology/google-ai-life-advice.html (last visited Aug 19, 2023).

[188] Chris Weller, *A Group of Former Facebook and Apple Employees Are Teaming Up to Warn Kids About Tech Addiction*, BUSINESS INSIDER (Feb. 2018), https://www.businessinsider.com/ex-facebook-and-google-

to users, enabling them to enjoy the benefits of a world of information and connections useful to their professional and personal life. Yet for users to enjoy such services, they are subjected to algorithmic operations that constantly send them addictive content, urge them to take impulsive actions, and increase their unhealthy dependence on platforms.[189] Facebook and Instagram, for instance, are accused of designing social media algorithms in a way that has harmfully "physiologically entrapped" young users,[190] causing higher incidences of suicide, self-harm, and mental illness.[191] The U.S. Surgeon General Vivek Murthy describes the mental health crisis in teenagers as "an emergency," treating social media algorithms as a significant contributing factor.[192] These problematic algorithmic practices affect not only younger users but also adults within and beyond the United States. According to a 2018 British study, the use of social media leads to disrupted sleep, depression, and symptoms of physical sickness like headaches.[193] AI-driven feeds, along with interfaces like autoplay and infinite scroll, are designed for excessive use by users, but their harmful effects on brains and wellness are often underestimated. Due to the minor harm of each single newsfeed feature that prolongs usage time, their cumulative impact on health was neglected until the emergence of a widespread mental health crisis.

Last but not least, major safety challenges have been posed to the existing security landscape.[194] With machine-learning techniques, cybercriminals can adopt AI that learns from experience and becomes smarter, making cybercrime harder to detect and control at an early stage.[195] For now, machine learning algorithms can help attackers target the vulnerable parts of a network or AI system, such as by detecting a network without any firewall to steal data from databases.[196] Based on the intelligence of adversarial AI, hackers can also attack online networks and manipulate AI systems.[197] Biometric systems are likely

employees-launch-anti-tech-addiction-campaign-2018-2 (last visited Mar 21, 2023).

[189] Megan Mccluskey, *How Addictive Social Media Algorithms Could Finally Face a Reckoning in 2022*, TIME (Jan. 2022), https://time.com/6127981/addictive-algorithms-2022-facebook-instagram/ (last visited Mar 21, 2023).

[190] Yao, *supra* note 11.

[191] *Id.*

[192] Vivek Murthy, Why I'm calling for a warning label on social media platforms, new York time opinion.

[193] *Here's How Social Media Affects Your Mental Health*, MCLEAN HOSPITAL, https://www.mcleanhospital.org/essential/it-or-not-social-medias-affecting-your-mental-health.

[194] *See generally* GREG ALLEN & TANIEL CHAN, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY (2017); AI-powered cyberattack like malware and adversarial machine learning… Ondrej Kubovič, Peter Košinár & Juraj Jánošík, *ESET Whitepaper: Can Artificial Intelligence Power Future Malware?*, https://www.eset.com/me/whitepapers/can-artificial-intelligence-power-future-malware/; *See Cyber Grand Challenge*, DEF CON 24, https://www.defcon.org/html/defcon-24/dc-24-cgc.html (last visited Sept. 18, 2017); *see also "Mayhem" Declared Preliminary Winner of Historic Cyber Grand Challenge*, DEF. ADVANCED RES. PROJECTS AGENCY (Aug. 4, 2016), https://www.darpa.mil/news-events/2016-08-04; A deep-learning based malware, DeepLocker, can hide itself to hit a target and usurp the system. Ha Hwang & Min-Hye Park, *The Threat of AI and Our Response: The AI Charter of Ethics in South Korea*, 9 ASIAN J. INNOV. & POLICY 56 (2020).

[195] For a general discussion of how hackers can target the vulnerabilities of AI systems, see Sarah Kessler & Tiffany Hsu, *When Hackers Descended to Test A.I., They Found Flaws Aplenty*, N.Y. TIMES (Aug. 16, 2023), https://www.nytimes.com/2023/08/16/technology/ai-defcon-hackers.html (last visited Aug 19, 2023).

[196] Jennifer Gregory, *AI Security Threats: The Real Risk Behind Science Fiction Scenarios*, SEC. INTEL. (May 15, 2021), https://securityintelligence.com/articles/ai-security-threats-risk/.

[197] Robert Walters & Marko Novak, *Cyber Security*, *in* CYBER SECURITY, ARTIFICIAL INTELLIGENCE, DATA PROTECTION & THE LAW 21 (Robert Walters & Marko Novak eds., 2021), https://doi.org/10.1007/978-981-

to become the target of future AI-powered cyberattacks for identity theft and other types of crime.[198] By adding inaccurate information to poison training data, cybercriminals can even cause false algorithmic predictions using machine learning systems.[199]

Like other harms discussed in this typology, the impairment of safety resulting from AI-controlled hardware and software is becoming increasingly common, affecting not just individual users but the broader public as well. Because of the problematic nature of these algorithmic harms, those affected often have little to no control over their occurrence or impact. The following part will explore two factors that may exacerbate these harms, making it harder for legal systems to address them effectively.

B. Aggravating Factors

Unlike primary harms, which result in substantive injuries to legal interests, aggravating factors are procedural issues that intensify and prolong the effects of primary harms. These factors include: (1) a deficiency in both internal and external accountability mechanisms, termed *accountability paucity*; and (2) obstructed transparency due to *algorithmic opacity*, an inherent feature of AI systems. As will be explained in the following subsections, given the ubiquitous and intangible nature of algorithmic harms, these factors contribute to the increasing severity of primary harms by impeding awareness and mitigation efforts. Furthermore, they limit regulatory investigations and restrict victims' ability to hold perpetrators accountable, thereby allowing algorithmic harms to escalate both individually and systemically without warning.

1. Accountability Paucity

The first aggravating factor is absence of sufficient accountability in AI applications. The AI revolution has brought new challenges in conventional governance structures, causing pervasive paucity in accountability systems that ought to cope with novel issues arising in AI contexts. Algorithmic accountability is associated with the governance scheme that holds entities accountable for their AI design and deployment. Internally, algorithmic accountability enables entities to manage risks and minimize harms arising from their practices, ensuring that their AI deployment fulfills anticipated goals. Externally, accountability holds entities, especially businesses, responsible for the derived harms, providing victims with channels to seek and receive remedies.

The first cause of accountability paucity is insufficient internal accountability among AI developers and adopters. Although AI systems have been integrated into technological innovations, most AI adopters, particularly firms, have yet to establish robust accountability mechanisms in response to algorithmic harms. From an internal viewpoint, entities commonly adopt conventional thinking to address the trouble with AI. The accountability

16-1665-5_2, 3 (last visited Mar 23, 2023).

[198] Fennell, *supra* note 104.

[199] Noam Dror, *Top 5 Security Threats Facing Artificial Intelligence and Machine Learning*, HUB SECURITY (May 31, 2021), https://hubsecurity.io/top-5-security-threats-facing-artificial-intelligence-and-machine-learning/.

system is closely embedded in the broader scheme of governance, involving how they assess their business models, compliance with law, and risk management frameworks in light of potential harms.[200] However, most firms have not established such a governance system for their developing AI business[201]—that is, they have not been required or instructed to establish governance systems or managerial strategies to address algorithmic harms.[202] For instance, a tech firm that develops a generative chatbot is not required to monitor algorithmic harms, establish AI policies, and conduct impact assessment measures.[203]

The diffuseness of AI developers makes harm correction even more challenging under traditional accountability mechanisms. An AI-driven application is typically developed by multiple staff from various departments, including dataset providers, model trainers, and business model designers.[204] This diffuseness impedes the tracing of intangible algorithmic harms that arise from various actors located in different departments and even different jurisdictions. In business settings, an AI project often involves distinct employees at different stages of the project, lacking a manager who oversees the harms involved throughout the life cycle of AI businesses. In the absence of a clear obligation to account for algorithmic harms, multiple participants, whether involved in collecting data, training datasets, or business models, may tend to neglect harm mitigation.[205] Even if firms consider developing trustworthy products and services a critical business goal, the intangible harms associated with these offerings stemming from a broad network of actors necessitate the establishment of accountability measures tailored to the unique nature of AI. Such measures, often burdensome and costly, can contradict corporate profit maximization goals and deter firms from implementing them. This lack of sufficient accountability creates significant obstacles in tracing algorithmic harms within the organization and assigning responsibility for its occurrence, allowing such harms to accumulate until they escalate into major scandals.[206]

The second cause of accountability paucity is external, arising from inadequate legal accountability. External to the AI development and deployment processes, users subject to AI systems typically lack the legal tools necessary to identify ubiquitous, intangible algorithmic harms and hold wrongdoers accountable.[207] Many factors have contributed to this situation. As Section III will further illustrate, lawmakers have struggled to craft legislation that addresses the unique nature of algorithmic harms; legal precedent concerning AI applications is too scant to ensure accountability in the face of rapidly expanding AI

[200] Darrell Rigby, Zach First & Dunigan O'Keeffe, *How to Create a Stakeholder Strategy*, HARVARD BUSINESS REVIEW, May 2023, https://hbr.org/2023/05/how-to-create-a-stakeholder-strategy (last visited Aug 3, 2023).

[201] Roberto Tallarita, *AI Is Testing the Limits of Corporate Governance*, HARVARD BUSINESS REVIEW, Dec. 2023, https://hbr.org/2023/12/ai-is-testing-the-limits-of-corporate-governance (last visited May 19, 2024).

[202] *Id.*

[203] *See* Maryline Laurent & Claire Levallois-Barth, *4-Privacy Management and Protection of Personal Data*, *in* DIGITAL IDENTITY MANAGEMENT 137, 137–205 (Maryline Laurent & Samia Bouzefrane eds., 2015).

[204] Scherer, *supra* note 1.

[205] Tim Fountaine, Brian McCarthy & Tamim Saleh, *Building the AI-Powered Organization*, HARV. BUS. REV. (July 2019), https://hbr.org/2019/07/building-the-ai-powered-organization (last visited Sep 11, 2023).

[206] *Id.*

[207] Katyal, *Private Algorithmic Accountability*, *supra* note 143, at 99-107 (noting the inadequacy of principles regarding discrimination, privacy, and algorithmic due processes).

adoption;[208] and existing legal doctrines are ill-suited to address harms arising from algorithmic practices involving information exchanges among private parties.[209] Information privacy laws, for instance, constitute one of the primary frameworks regulating AI systems that process personal information. But since AI systems can collect a wide range of data not originally considered personal information, they tend to circumvent privacy laws that are not applicable to harms generated by algorithmic data processing. Additionally, broadly defined privacy notices give entities a legal justification to collect a range of personal data for indefinite purposes.[210] Yet many unanticipated harmful applications do not usually fall under the scope of information privacy laws that govern data collection from first or third parties. As previously discussed,[211] at the point of data collection, no one can precisely foresee whether a particular data collection may end up yielding personal information. Even if AI is used to produce inferences that reveal our personal insights, they do not involve the provision of personal data through any individual person, falling outside the law's regulatory scope. Most of these harms, as Section III will discuss more broadly, elude legal scrutiny due to their unclear legal nature.

A related concern pertains to corporate policies, which traditionally serve as a crucial vehicle to ensure a certain level of accountability in numerous business practices. These policies not only educate staff about how to implement corporate missions, but also inform consumers of how companies interact with them to legally offer services. However, such a policy is commonly missing from AI contexts, where many firms have not crafted a well-established policy designed for their AI-based business models. In practice, firms usually publish privacy policies on their official websites to explain the data practices associated with AI systems they have adopted.[212] Nevertheless, the disclosed information seldom includes how entities measure and manage a variety of harms derived from AI systems. Conventional corporate notices and disclosures required under existing regulations are also not structured to enhance awareness and mitigation of algorithmic harms. Existing policies and rules thus cannot offer adequate accountability in an AI context.

Both internally and externally, governance systems fail to keep up with AI revolutions,

[208] *See* Tiffany Hsu, *What Can You Do When A.I. Lies About You?*, N.Y. TIMES (Aug. 3, 2023), https://www.nytimes.com/2023/08/03/business/media/ai-defamation-lies-accuracy.html (last visited Aug 19, 2023).

[209] *See e.g.,* W. Nicholson Price II & I. Glenn Cohen, *Locating Liability for Medical AI*, sec. II (2023), https://papers.ssrn.com/abstract=4517740 (last visited Sep 11, 2023) (Exploring the limitations of tort law as a means to address and resolve issues associated with medical AI problems); Barbara Evans & Frank Pasquale, *Product Liability Suits for FDA-Regulated AI/ML Software*, *in* THE FUTURE OF MEDICAL DEVICE REGULATION: INNOVATION AND PROTECTION 22, 29 (I. Glenn Cohen, Timo Minssen, W. Nicholson Price II, Christopher Robertson & Carmel Shachar eds. 2022).

[210] Solove, *AI and Privacy*, *supra* note 14, at 29-31.

[211] *See* Sections III.A.2. & II.A.1.

[212] The information disclosed in such notices includes types of personal data collected, processing choices provided to users, methods for access and data control settings, procedures to bring a complaint, contact information, effective date of the notice, and scope of the notice. (US) FTC, *Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress*, Federal Trade Commission (May 2000), https://www.ftc.gov/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission.

leading to overall accountability paucity. This deficiency makes internal staff less motivated to trace and manage the intangible harms generated by AI applications. Simultaneously, it prevents regulators from investigating ubiquitous harms that should be addressed.

2. Algorithmic Opacity

In addition to accountability paucity, another aggravating factor is algorithmic opacity. As this term suggests, algorithmic opacity refers to the lack of transparency in AI systems' inner workings and operational outcomes,[213] which obstructs the identification and mitigation of primary harms. Algorithmic opacity is caused by various factors, ranging from technical complexity and trade secrecy to invisibility in governance systems. The interplay among these causes has created a solid barrier to the oversight needed to assess harms arising from AI innovations.[214] Consequently, this hampers effective inspection of the dark side of those algorithmic practices that entities such as firms are reluctant to reveal to the broader public.

The first source of algorithmic opacity is technical complexity, which involves the difficulty of comprehending the intricate inner workings of AI systems.[215] This opacity is tied to machine learning algorithms, a widely adopted subset of AI systems that continually change their internal structures, making the dynamics of their computational progress and outcomes challenging to understand.[216] Deep learning, a subcategory of machine learning algorithms, undergoes continuous transformation of multilayered neuron networks as they learn from experiences.[217] Over time, this modified logic and altered structures become ever more incomprehensible to human understanding,[218] making their operating consequences difficult to estimate.[219] The unpredictability, dynamics, and complexity inherent in machine learning systems create technical challenges for developers and deployers in correcting deficiencies in decision-making patterns and their derived operations.[220]

As with technical complexity, algorithmic opacity is attributable to legal provisions structured to protect secrecy interests in AI applications. A classic example of this concerns intellectual property laws crafted to protect confidentiality in technological innovations.[221]

[213] Tschider, *supra* note 186, part III.A. (arguing that the inscrutable and sometimes unpredictable nature of AI somewhat intensifies the issues posed by computer-based technologies.)

[214] Burrell, *supra* note 18, at 6; Tarleton Gillespie, *The Relevance of Algorithms, in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167 (Tarleton Gillespie et al. eds., 2014).

[215] IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING 2 (2016).

[216] *Id.*

[217] *See id*

[218] *See also* Davide Castelvecchi, *Can We Open the Black Box of AI?*, NATURE (Oct. 5, 2016), https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731.

[219] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. LAW & TECH 889, 897 (2018).

[220] Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 532, 539 (2015); Siddhartha Mukherjee, *A.I. Versus M.D.*, NEW YORKER (Apr. 3, 2017), https://bit.ly/39iOxG7; EUROPEAN COMMISSION, *A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines* 5 (2019), https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

[221] Fromer, *supra* note 18, at 709-11.

Within the intellectual property regime, trade secret law is said to be the current default scheme whereby innovators seek to protect the commercial interests of their AI applications.[222] Trade secrets law aims to protect a wide spectrum of information with potential business value and confidentiality maintained by the efforts of holders of trade secrets.[223] According to the definition of trade secrets, a wide array of AI innovations that have business value and are concealed by firms can be possibly claimed as trade secrets.[224] However, one cannot confirm if a given AI application meets the criteria for a trade secret without disclosing critical parts of the applications such as source code for a court's examination.[225] Such disclosure risks compromising the very confidentiality trade secret law aims to protect. This dilemma motivates many firms to claim trade secrecy, although the information involved might not meet the criteria for trade secret protection. Today, this widespread practice allows firms to shield their information-based applications, including their problematic algorithmic practices, from outside review.[226] This legal opacity can conceal many sources of ubiquitous, intangible algorithmic harms such as under-representativeness of databases, programming deficiencies, cognitive biases instilled by AI developers, unauthorized collection of sensitive data, manipulative news feeds,[227] and algorithmic filtering systems designed to cause unhealthy addiction. With such legal seclusion, transparency in the privately developed and deployed systems has been largely obscured, despite these being the major areas where algorithmic harms are generated.[228]

The invisibility of private governance is the third instance of opacity that precludes adequate harm identification and correction.[229] As AI develops into a crucial technology for many businesses, the private governance of AI innovation has grown in significance for stakeholders who wish to mitigate algorithmic harms, including individuals, shareholders, communities, and regulators. The invisibility of such governance hinders harm investigation when external stakeholders lack critical information about the governance regimes for harms arising from AI applications.[230] This governance opacity generally stems from a status quo that has not responded to the intangible nature of algorithmic harms and the commercial secrecy caused by trade secret protection. The absence of legally mandated disclosure of information about private governance of algorithmic operations leads to extensive invisibility in the adequacy of harm mitigation schemes.[231] As AI develops into an essential technology for many businesses, how private entities build governance to address harms has become

[222] *Id.*

[223] *See* 18 U.S.C. § 1839 Defend Trade Secrets Act.

[224] Fromer, *supra* note 18.

[225] Katyal, *Private Algorithmic Accountability*, *supra* note 143, at 125.

[226] *Id.*

[227] Yeung, *supra* note 15, at 124.

[228] *See* David Levine & Ted Sichelman, *Why Do Startups Use Trade Secrets?*, 94 NOTRE DAME L. REV. 751 (2019).

[229] Lu, *Data Privacy*, *supra* note 10, at 2099.

[230] *See generally*, Margot E. Kaminski, *Understanding Transparency in Algorithmic Accountability*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 121 (Woodrow Barfield ed., 2020) [hereinafter Kaminski, *Transparency*].

[231] While entities may be asked to conduct disclosures for the review of regulatory authorities, this disclosed information is often not publicly accessible. Lu, *Data Privacy*, *supra* note 10, at 2099-2100.

critical to the legitimacy of AI innovations. However, without access to information about algorithmic governance, stakeholders lack the resources required for necessary oversight.[232]

Given the intangible nature of algorithmic harms and increasing stakeholder interest in understanding these ubiquitous harms, invisibility in private governance is becoming an urgent issue. As many corporate scandals have indicated, firms—one of the primary AI users—often have the best knowledge of the detected or anticipated harms associated with the AI applications they develop or deploy, such as their AI applications' biased assessment of someone's likelihood of criminal behavior or unauthorized diagnosis of mental illnesses. Without their disclosures of algorithmic governance details, including anticipated harms and corresponding mitigation measures, users and other stakeholders cannot identify the known intangible harms at stake. From a business viewpoint, disclosure of governance information takes substantial time, cost, and labor while also sparking stakeholder inquiries that may require further changes in algorithmic governance or damage a business's reputation by revealing existing issues. As a result, firms are often disincentivized to share such information unless mandated. This lack of disclosure makes it difficult for stakeholders to discern intangible harms arising from the design of business models, assess the primary harms recognized by businesses, or measure the adequacy of resources allocated to harm prevention.[233] Governance invisibility thus becomes the norm, shielding the status of harm mitigation efforts from social oversight.

Together, as insufficient accountability grants algorithmic harms a free pass, algorithmic opacity further conceals these harms to external stakeholders, worsening their cumulative effects as a result. The rise of generative AI makes this problem even more pervasive, acute, and urgent. As the Chairperson of the FTC has indicated, there is a lack of checks on generative AI like ChatGPT developed by OpenAI.[234] As this AI company claims, despite crafted policies to mitigate flawed results from ChatGPT, their service may continue to spew misidentifications and falsehoods to users in opaque conditions. The issue of aggravating factors is not specific to this firm. Given that algorithmic opacity and accountability paucity befall many AI applications that generate algorithmic harms, victims have little awareness of the harms and lack effective legal tools to hold entities liable. With little reason to expect firms to internalize the costs of harm mitigation, the law must respond to these challenges.

III. Regulatory Frameworks for Algorithmic Harms

The rise of algorithmic harms has become a critical issue for legal systems. Since 2016, many countries have begun to formulate national policies and regulatory approaches

[232] When governance information partly concerns trade secrets, governance invisibility in AI systems may overlap with its legal opacity counterparts. Nevertheless, as stated previously, not everything firms claim as trade secrets deserve this legal protection. Although firms tend to assert that most if not all of their business practices fall under trade secrets, the legitimacy of these claims is a legal matter for the courts to decide. Additionally, legal doctrine allows a space where legitimate public interest can override protection of trade secrets. Katyal, *Source Code Secrecy*, *supra* note 18, at 1228 ("without first disclosing and examining the source code, it is impossible to know whether it even qualifies as a trade secret.")

[233] *Id.* at 2100.

[234] Kang & Metz, *supra* note 6.

to the emerging harms posed by AI applications. Among democratic states, policymakers continue to emphasize the importance of building trustworthy AI systems and protecting democratic values. The U.S. federal government currently adopts a regulatory option that primarily addresses algorithmic harms through existing statutes, as exemplified by its continued reliance on information laws that tend to focus on consumer interests. The EU represents an example that introduces a risk regulation to fix algorithmic harms, as shown by its AI Act designed to address AI risks to fundamental rights. Japan represents a regulatory option that maximizes the value of soft laws to respond to algorithmic harms, adopting new standards and guidance as fluid, inclusive principles to protect a broad group of stakeholders.

The following part discusses how these regulatory approaches address the algorithmic harms outlined in Part II. Each of these frameworks serves a practical purpose. They have either been adopted or are under consideration by policymakers in democratic states as viable regulatory options for algorithmic governance. The three prototypes exhibit distinct examples, including varying levels of adaptability, versatility, and inclusivity, along with inherent limitations in their capacity to address algorithmic harms. Given the potential influence of these examples, evaluating their feasibility provides practical insights for the development of AI regulations. As these case studies reveal, these influential frameworks are all insufficient; they either fail to recognize certain intangible harms or neglect to address their cumulative effects, enabling algorithmic harms to evade legal liability. As private industry continues to dominate AI deployment in this regulatory gap, victims are left vulnerable to countless harms that remain unaddressed.

A. American Legal Consumerism

As of today, the U.S. federal government has largely leveraged existing laws rather than new regulations to address algorithmic harms. This regulatory strategy utilizes adaptable regulations to protect victims while avoiding or delaying the introduction of new restrictive rules to encourage innovation. The U.S. government agencies maintain that their consumer-centric regulations can adapt to emerging AI problems, reflecting an AI regulatory option termed "legal consumerism."[235] The following section shows that policymakers are only partly right about the adaptability of these existing regulations. It first illustrates where this approach is adaptable, followed by where it is poorly suited to address many aspects of algorithmic harms.[236]

The American privacy and AI regulatory framework often sees consumers as the primary beneficiaries and victims of information-driven innovations. As privacy scholars observe, the United States adopts a market-driven structure that views individuals as consumers in the data privacy space.[237] Under the privacy regime, consumers maintain control over their information through a conventional notice and choice mechanism, also

[235] This piece focuses on the discussed regulations because they cover the most types of harms outlined in the typology in Section II and serve as a good example to measure their adequacy in addressing these harms.

[236] Paul M. Schwartz & Karl-Nikolaus Peifer, *Transatlantic Data Privacy*, 116 GEO. L. J. 115, 147 (2017) (viewing the US vision of privacy protection as protecting consumer privacy in the marketplace).

[237] *Id.* at 132.

known as a "self-management" approach, to ensure fairness in exchanges of data between consumers and entities.[238] Recent regulatory efforts in algorithmic settings often echo this view. For instance, the recently reintroduced Algorithmic Accountability Act of 2023 terms individual victims as consumers, focusing on the mitigation of harms caused by private entities.[239] At the state level, leading jurisdictions such as California and Colorado have a similar focus, proposing or enacting state bills that regulate business applications of AI to protect consumers from algorithmic discrimination.[240]

Since as of this writing, the passage of federal AI legislation is not anticipated in the coming years,[241] AI applications are primarily governed by applicable information laws, such as privacy and antidiscrimination laws, as well as regulatory actions from federal agencies that are largely focused on consumer protection.[242] Existing sparse and varying regulations are typically implemented on an ad hoc basis when requirements specified by a certain sectoral law are met. Under this regime, normative rules are triggered when certain consumer interests have been impacted by illegal business practices. The American framework emphasizes the role of AI in driving innovative benefits for consumers, industries, and society.[243] To avoid stifling innovation that yields tremendous market value, the law refrains from intervening in its technological deployment unless AI results in tangible harm to consumers. This legal framework creates a significant gray area in facilitating the development of AI applications.

1. Consumer-Centric Governance

The American legal consumerism features a combination of sectoral laws, agency enforcement actions, and self-regulation.[244] This regulatory framework, in contrast to the EU's tradition of government interference, favors a certain extent of self-regulation that enables firms to establish governance structures for their technological innovations.

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

[239] The Algorithmic Accountability Act of 2023 was introduced in September 2023 after the Algorithmic Accountability Act of 2022, its previous version, was rejected in January 2023. The Bill of the Algorithmic Accountability Act Sec. 2 (6).

[240] Titus Wu, *California Seeks to Be First to Regulate Business Use of AI*, BLOOMBURG L. (Apr. 19, 2023), https://news.bloomberglaw.com/in-house-counsel/california-seeks-to-be-first-to-regulate-business-use-of-ai; Colorado General Assembly, *Consumer Protections for Artificial Intelligence*: Bill SB24-205 (2024).

[241] Müge Fazlioglu, US Federal AI Governance: Laws, Policies and Strategies, IAPP (Nov. 2023), https://iapp.org/resources/article/us-federal-ai-governance/ (last visited Jul 24, 2024).

[242] State laws are another critical area of AI regulation. Colorado, for instance, has introduced a comprehensive AI Act addressing algorithmic bias, which may inspire other states to enact similar legislation. This article recognizes the significant role state laws play in shaping U.S. AI governance standards. A thorough discussion of the impact of state legislation is beyond the scope of this piece. As of this writing, state law has not established a universal standard and there remains a possibility of being preempted by future federal legislation in this area.

[243] On the issue of information capitalism, *see generally* JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019).

[244] HOOFNAGLE, *supra* note 88, at 145 ("The FTC has been a key force for the protection of online privacy because it fills the gaps left by the US "sectoral" regulatory approach.").

To date, the federal government's major legal milestones have largely centered on issuing guidelines to guide AI developers and adopters.[245] For instance, the Trump Administration's 2019 executive order[246] announced its vision to both maintain global leadership in AI[247] and protect civil liberties, privacy, and American values, yet it made an explicitly deregulatory turn to achieve these goals.[248] This led to its 2020 Guidance for Regulation of Artificial Intelligence Applications, encouraging federal agencies to remove unnecessary regulatory barriers to AI innovations.[249] The Biden Administration has gone further in terms of civil rights protections, releasing the Blueprint for an AI Bill of Rights, a detailed set of ethical norms devised to safeguard privacy and other civil rights.[250] According to the AI Bill of Rights, the designers and deployers of automated systems are advised to protect privacy, ensure safety, avoid discrimination, offer explanations, and provide human consideration.[251] Following this, the administration issued an executive order that further requires safety mandates, asking firms to conduct and report safety testing—termed "red teaming," for the most advanced AI systems.[252] Despite these progressive moves, none of these high-level policies guarantees meaningful legislative action, such as passing a comprehensive federal AI law that would define algorithmic harms under a robust governance framework.[253]

Against this backdrop, the American regulatory approach seeks to leverage the value of existing laws that policymakers believe are adaptable to address algorithmic harms.[254] This holds true in cases where the contours of algorithmic harms are readily recognizable and overlap with issues envisaged by lawmakers, such as unfair decision-making in critical

[245] U.S. EXEC. OFFICE OF THE PRESIDENT, AI BILL OF RIGHTS, https://www.whitehouse.gov/ostp/ai-bill-of-rights/; NIST, AI RISK MANAGEMENT FRAMEWORK (Airmf 1.0) (2023).

[246] U.S. EXECUTIVE OFFICE OF THE PRESIDENT, PROMOTING THE USE OF TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE FEDERAL GOVERNMENT, EXECUTIVE ORDER OF 13960 (2020), https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

[247] *Id.*

[248] *Id.*

[249] U.S. Executive Office of the President Office of Management and Budget, *Guidance for Regulation of Artificial Intelligence Applications*, (2020), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf. The Guidance illustrates the US aim of building AI applications that are trustworthy in the eyes of US communities. The Guidance also proposes ten principles that should be considered by policymakers intending to regulate AI applications in the private sector.

[250] U.S. EXEC. OFFICE OF THE PRESIDENT, AI BILL OF RIGHTS, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (last visited Feb 24, 2023).

[251] *Id.* at 5-7.

[252] U.S. EXECUTIVE OFFICE OF THE PRESIDENT, EXECUTIVE ORDER ON SAFE, SECURE, AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE, (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

[253] Joe Biden, *Republicans and Democrats, Unite Against Big Tech Abuses*, WALL ST. J. (Jan. 11, 2023), https://www.wsj.com/articles/unite-against-big-tech-abuses-social-media-privacy-competition-antitrust-children-algorithm-11673439411?mod=hp_opin_pos_3#cxrecs_s (last visited Feb 27, 2023).

[254] U.S. FTC et al., Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

regulated areas. In some instances, sectoral laws can be applied to address AI's disparate impacts or treatments.[255] For instance, the Equal Credit Opportunity Act (ECOA) bans credit discrimination on the basis of protected characteristics such as age, skin color, religious views, and more.[256] Under the ECOA, an entity should be liable for "disparate impact" if they use AI systems to give someone lower credit scores because of any of these protected characteristics. Statutes like this cover unfair automated decision-making processes in housing, credit, and other areas to address diminishing equality.[257]

For consumer harms that fall outside of the sectoral statutory scope, Section 5 of the FTC Act fills the gap. This FTC rule prohibits "unfair or deceptive acts or practices," applying to a range of algorithmic practices.[258] According to the FTC, an "unfair" act or practice is defined as one that "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition."[259] In a diminishing equality situation, this may cover unequal algorithmic design applications such as facial recognition software that disproportionately misidentifies racial minority as shoplifters. "Deceptive" acts or practices refers to a "representation, omission or practice that is likely to mislead a consumer . . . acting reasonably in the circumstances . . . to the consumer's detriment."[260] In this context, when a firm has claimed in its statement that its AI product is bias-free, then once its operations are revealed to discriminate against certain groups, this would be regarded as a deceptive practice and thus constitute a breach.[261]

2. Incompatible Harm Concept

While some existing regulations are adaptable enough to address issues ranging from discrimination to other aspects of algorithmic harms, many fall short in comprehensively tackling intangible, aggregating, and ubiquitous harms. Often, they focus on specific types of harm within specific contexts, leaving gaps in protection against broader algorithmic impacts.

For instance, information laws are ill-suited to mitigate AI-driven privacy erosion. AI systems enable the intelligent aggregation of data collected from multiple sources without user awareness. However, current information laws often do not protect against the misuse of these AI-generated insights, as they focus on the protection of personal data directly

[255] Office of Public Affairs USA Department of Justice, *Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising*, DEP. OF JUS. (Jun. 21, 2022), https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known (last visited Feb 24, 2023).

[256] U.S. Equal Credit Opportunity Act §1691(a).

[257] U.S. Genetic Information Nondiscrimination Act of 2008, 122 Stat. 881, secs. 201-02 (May 21, 2008), https://www.govinfo.gov/content/pkg/PLAW-110publ233/pdf/PLAW-110publ233.pdf https://www.eeoc.gov/statutes/genetic-information-nondiscrimination-act-2008 ; U.S. Fair Housing Act, 42 U.S.C. 3601 et seq., sec. 804 (2015), https://www.justice.gov/crt/fair-housing-act-2 (last visited Mar 24, 2023).

[258] 15 U.S.C. § 45(n) *45(a)(1).*

[259] 15 U.S.C. § 45(n).

[260] FTC, FTC Policy Statement on Deception (1983), appended to Cliffdale Assocs., Inc., 103 F.T.C. 110, 174 (1984).

[261] Slaughter, Kopec, & Batal, *supra* note 17, at 40-41.

collected from individuals rather than the cross-context aggregation of data that leads to mass surveillance, which can be more invasive.

Algorithmic harms to autonomy also frequently fall outside the scope of existing legal protections. The insidious effects of algorithms that intelligently manipulate individuals, often stemming from AI-generated insights, are underregulated by information privacy laws and other legal areas. For example, when hypernudges diminish people's rational decision-making capacity based on algorithmic correlations, federal law does not provide consumers with the right to challenge the use of correlations that algorithms draw from their data.[262] The lack of regulations targeting AI-driven manipulation allows entities to subtly distort people's autonomy over time.

Existing laws are also inadequate in addressing equality harms. Anti-discrimination laws are designed to address discrimination, yet they can be poorly equipped to deal with AI-powered equality harms. AI applications can use proxy variables that indirectly correlate with protected characteristics without explicitly referencing them, leading to discriminatory outcomes while avoiding direct regulatory scrutiny. In practice, the range of potential inputs for algorithmic discrimination often exceeds the traditional understanding of protected characteristics, rendering the notion of a protected class practically ineffective. [263] Underregulated equality harms like these can impact victims across different settings. Biased data used in one domain can reinforce inequalities in another, but the laws fail to recognize these interconnected, compounded harms.

This deficiency is also evident in addressing algorithmic harms to safety, where applicable laws generally lack provisions that specifically address the less immediate impacts of algorithmic systems. AI-based interfaces can be designed for addictive use without providing users with warnings or other protective measures. The intangible harms to mental health continue to accumulate until they become widespread issues, such as child addiction and mental illnesses. The underlying harms of algorithmically generated health data can lead to misuse in other systems, such as generating misleading medical recommendations. However, the laws often struggle to clarify whether these concerns constitute sufficient harm to justify legal remedies. Algorithmic opacity makes it unlikely for people to perceive the safety risks or prove intent, whether it be unsafe, misleading, manipulative, or deceptive. [264]

While the FTC Section 5 Act has the potential to capture algorithmic harms not explicitly defined in sectoral regulations, its enforcement effectiveness may be hindered by the characteristics of algorithmic harms. Recently, the FTC Chair has signaled the agency's intent to regulate AI more aggressively, leading to high-profile enforcement actions that require harsh sanctions like data and algorithm disgorgement. [265] Scholars argue that the FTC

[262] *Id. at* 95; Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 4-5 (2014).

[263] Katyal, *Private Algorithmic Accountability*, *supra* note 143, at 101-03.

[264] Lu, *Data Privacy, supra* note 10, at 2010-07.

[265] To date, the FTC has primarily targeted cases involving high-profile AI technologies or firms, such as Clearview AI's facial recognition data scraping or Amazon Alexa, an AI-powered virtual assistant's unauthorized data processing. Relatedly but not the primary focus of this piece, algorithmic manipulation may also bring about legal issues involving competition interests, falling into the purview of anti-market

approach could be adaptable to AI's discriminatory harms and establish a common law standard, guiding firms on how to develop and deploy AI systems.[266] The FTC's investigative and enforcement scope is potentially broad, enabling it to address a wide array of algorithmic harms, even those affecting large groups and representing collective interests.[267] However, the FTC's investigative capabilities are hampered by the intangible problems that operate in opaque conditions.[268] The inherent nature of algorithmic harms, coupled with the lack of detailed AI policies and the associated promises made by firms, makes identifying these harms increasingly challenging. While these harms are also urgent, regulators may enforce the law to address only a small number of the most salient harms in capturing public attention.

Finally, a fundamental problem with the current regulatory approach in American is its current focus on individual consumers, which occasionally hinders the regime from adequately addressing group and social harms. As outlined in Section I, the victims of algorithmic harms often extend beyond individual consumers to include non-consumers, groups, and the broader society. These collective harms are not sufficiently addressed by information laws centered on individual interests.[269] Although algorithmic harms have group and social dimensions, the invisibility of intangible harms makes them difficult for regulators and victims to perceive and address. Reliance on traditional statutory law and enforcement actions renders many harms unlikely to be mitigated. Although existing laws can be useful in certain algorithmic settings, they are largely incompatible with capturing algorithmic harms that elude their limited harm concept, leaving those injuries neglected.[270]

B. European Legal Fundamentalism

In contrast to the American consumer-centric framework, which is often focused on harms to consumers, the EU sees algorithmic harms as threats to fundamental rights, thereby enacting tailored regulations in response to their damaging impacts, reflecting a regulatory option termed "legal fundamentalism." As algorithmic applications generate harms that threaten individual interests, lawmakers actively devise new targeted rules to protect citizens. This regulatory regime governs AI systems based on their potential harms posed to affected individuals. From this perspective, the law should broadly safeguard crucial legal interests, particularly fundamental rights and safety, which deserve a high level of protection. While innovation is also considered a policy goal, this prototype regulates algorithmic harms

manipulation provisions. For an evaluation of the US anti-market manipulation provisions applicable to AI applications, see Gina-Gail Fletcher, *Deterring Algorithmic Manipulation*, 74 VAND. L. REV. 259, 281-86 (2021).

[266] *See* Andrew Selbst & Solon Barocas, *Unfair Artificial Intelligence: How FTC Intervention Can Overcome the Limitations of Discrimination Law*, 171 U. PA. L. REV 1023 (2023).

[267] For issues concerning unfair acts or practices, the agency recognizes that minor harms dispersed among numerous consumers constitute a substantial injury. FTC, FTC POLICY STATEMENT ON UNFAIRNESS (1980), appended to Int'l Harvester Co., 104 F.T.C. 949, 1073 (1984). Citron & Solove, *supra* note 14, at 814.

[268] *Id.*

[269] Scholars have argued that American privacy statutes with their individual focus struggle to address collective harms and fall short of tackling group harms that have evolved from private information exchanges. Schwartz & Peifer, *supra* note 236, at 135-36; Citron & Solove, *supra* note 7, at.

[270] Kang & Metz, *supra* note 6.

through relatively burdensome rules, even if it may potentially override interests in innovation in some instances. This creates a sharp contrast to the U.S.'s comparatively light regulation approach.

The EU is said to be a vanguard norm-setter in the field of data and AI regulation. Over the past decade, the bloc has progressively formulated AI policies and new regulations governing algorithmic applications that threaten the fundamental rights of citizens.[271] Its targeted regulations feature various legal tools designed to address risks and harms arising from information technologies, leading to a versatile framework for managing algorithmic harms. The following discussion will focus on provisions mainly devised to address harms arising from AI applications, particularly those under the AI Act.[272] The analysis will start by pointing out where this approach works well, and then explain where its harm concept is incomplete in addressing the trouble with algorithmic harms.

1. Risk-Based Governance

As a major step in mitigating the harms caused by AI applications, the EU has adopted the AI Act in March 2024.[273] The AI Act is one of the first sets of AI regulations designed by a democratic institution.[274] Derived from EU product safety regulations, it regulates AI systems and general-purpose AI, also known as foundational models, based on four levels of risk involved in harming fundamental rights and safety: (1) unacceptable risks that must be banned, (2) high risks that must be regulated through conformity assessments and other accountability duties, (3) transparency risks that ought to carry out disclosure

[271] *See, e.g.*, EU Civil Law Rules on Robotics, (2017), https://www.europarl.europa.eu/doceo/document/ta-8-2017-0051_en.html (Last Visited Mar 24, 2023) (Resolution asking the Commission to take more actions in the field of AI).

[272] As a significant regulatory milestone in this regard, the General Data Protection Regulation (GDPR) took effect in 2018, notably incorporating rights and obligations regarding automated decision-making. This is followed by proposals or adoptions of a set of regulations that contain provisions applicable to certain AI systems such as the Digital Services Act, Digital Markets Act, Data Governance Act, and the Machinery Regulation. Essential to the development of European AI governance is the passage of the AI Act, which addresses AI risks arising from AI applications. EU General Data Protection Regulation (GDPR); EU AI Act (2024), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206; Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance) PE/30/2022/REV/1; Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance) PE/17/2022/REV/1; Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM/2020/767 final; Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (Text with EEA relevance) PE/6/2023/REV/1.

[273] European Parliament, *Artificial Intelligence Act: MEPs Adopt Landmark Law* (Mar. 8, 2024), https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law (last visited Jul 22, 2024).

[274] *Id.* EUROPEAN COMMISSION, REGULATORY FRAMEWORK PROPOSAL ON ARTIFICIAL INTELLIGENCE (Oct. 14, 2021), https://digitalstrategy.ec.europa.eu/en/policies/regulatory-framework-ai.

duties, and (4) minimal risks that are not subject to regulation.[275]

The AI Act covers many AI applications generating algorithmic harms, ranging from facial recognition to social scoring AI. The prohibited AI applications are those that violate privacy, safety, health, and other fundamental rights, such as covered social scoring, subliminal manipulation, and untargeted scraping of facial images from the internet or CCTV footage.[276] High-risk AI applications are those posing significant threats to fundamental interests and safety, including those used as products or safety components of products governed by certain EU laws, as well as those listed under the eight areas laid out in the Act's Annex III. [277] A few examples of high-risk applications include remote biometric identification, which impairs privacy due to its extraction of highly sensitive biometric data, as well as AI systems that make unfair decisions that diminish equality in one's work opportunities.[278] High-risk AI practices must comply with a series of accountability measures related to mitigating algorithmic harms.[279] These involve a duty to conduct risk management, [280] perform fundamental rights impact assessments,[281] ensure appropriate data governance,[282] guarantee human oversight, [283] and automatically record the occurrence of situations.[284] For AI systems generating limited transparency risks, the AI Act requires a baseline level of disclosure. AI applications classified as "limited risk" cover AI that interacts with humans, recognizes human emotion, and manipulates digital content, also known as deepfakes. [285] The categorization of risks and establishment of governance measures enhance accountability and transparency across a variety of AI applications prone to generating harms. This risk-based governance motivates covered entities, including AI developers and adopters, to collaboratively address anticipated harms.

2. Incomplete Harm Concept

[275] AI that carries risks that are too high to be permitted; high-risk AI that should be subject to an array of legal duties and surveillance; AI applications of limited risks that necessitate transparency requirements; AI systems with minimal risks that are excluded from the application of the Regulation. Beyond transparency duties, AI with limited or minimal risks are exempted from the application of the AIA. The AIA plans to ban AI systems that generate unacceptable risk due to their harm to fundamental rights protected by EU law, such as harmful subliminal techniques and social scoring. For high-risks AI, the AI Act intends to impose a range of duties and restrictions on providers of such AI systems due to their potential harm to safety or fundamental rights. *Id.* Council of the EU, *Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights*, COUNCIL OF THE EU AND THE EUROPEAN COUNCIL (2022), https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/.

[276] EU AI Act, *supra* note 274, art. 5 (prohibited AI practices).

[277] *Id.* art. 6(2) & Annex III (biometric; critical infrastructure; educational and vocational training; Employment or self-employment; accessing essential public or private services; law enforcement; migration and asylum management; judicial or democratic processes).

[278] *Id.* Annex III 1. & 4.

[279] EU AI Act, *supra* note 274, arts. 7, 8 & 16.

[280] *Id.* art. 9.

[281] *Id.* art 27.

[282] *Id.* art. 10.

[283] *Id.* art. 14.

[284] *Id.* arts. 12.

[285] *Id.* art. 50.

Despite its progressive categorizing of risk-level of AI applications and their corresponding regulatory rules, its risk categorization shows an incomplete understanding of algorithmic harms. Considering algorithmic harm to privacy, for instance, the AI Act treats data privacy as a fundamental right that deserves a high level of protection. Nonetheless, certain forms of privacy-invasive applications are considered less harmful due to policymakers' lack of consideration of their cumulative impact. One example is the private sector's use of population-scale facial recognition. According to the European Consumer Organization, a privacy advocacy group and civil society organization, the AI Act has not properly addressed the risks posed by corporate facial recognition in public places. [286] The European Data Protection Supervisor and the European Data Protection Board have asserted that the AI Act fails to ban the general use of biometric identification systems in public. [287] While the public use of remote biometric identification for law enforcement purposes is generally banned, remote biometric identification systems likewise process citizens' sensitive features to surrender their privacy interests at scale.[288]

Protections for autonomy within the EU's framework are also significantly limited as manipulative AI practices are only considered unacceptable if they are reasonably likely to cause "significant harm."[289] This allows for the undermining of individual autonomy over time so long as a given manipulative AI practice does not generate physical, psychological, or other perceivable harm that is proved to be significant. The framework neglects that the harm arising from algorithmic manipulation tends to invisibly aggregate in force over time without an appearance of gravity in a single instance of its practice.[290] Large-scale applications of hypernudging, targeted AI, and dark patterns across platforms can lead to a significant erosion of individual autonomy over time. However, these cumulative impacts, stemming from the subtle and ubiquitous presence of manipulative AI, are not adequately captured by the Act's risk categorization or its accountability measures.[291]

Overlooking the nature of algorithmic harms is also evident in the Act's treatment of equality. AI systems that spread discriminatory remarks or make biased decisions in contexts like online platforms or customer interactions often fall into lower risk categories. These systems, despite their potential to perpetuate discrimination and reinforce bias in many other areas, are not subject to substantial and rigorous regulatory oversight under this regime. AI systems designed in ways that might inadvertently favor certain groups—such as favoring content or ads that appeal more to one demographic over another—can perpetuate social

[286] European Data Protection Board & European Data Protection Supervisor, EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) (June 18, 2021), https://edpb.europa.eu/system/files/2021-06/edpbedps_joint_opinion_ai_regulation_en.pdf.

[287] The earlier working versions of the AI Act proposed to impose a ban on the general use of biometric identification but the officially proposed AI Act only forbids the adoption of biometric identification systems operated in the public for law enforcement purposes.

[288] Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 4 COM L. REV. INT'L 97, 101 (2021).

[289] EU AI Act, *supra* note 274, art. 5.

[290] Veale & Borgesius, *supra* note 301, at 99.

[291] *Id.* (identifying the overlooked cumulative harm posed by manipulative practices).

disparities in visibility and access to opportunities. The EU AI Act does not specifically address the social harms of these forms of algorithmic favoritism, leaving a regulatory void where discriminatory impacts can arise and accumulate.[292]

Regarding safety, the EU AI Act primarily targets AI systems that pose immediate and tangible dangers, such as those in critical infrastructure or medical devices. However, AI systems that operate in everyday contexts—such as smart home devices, mobile apps, or AI-driven content recommendation systems—can lead to widespread harms like security risks, addiction, misinformation, or other harmful outcomes. These harms to individuals and society can evade the Act's accountability framework. The EU AI Act's focus on individual systems fails to recognize how harms can aggregate and evolve across different contexts, escalating into broader public safety concerns.[293]

Due to the AI Act's nature as a risk regulation, its current structure does not provide sufficient mechanisms for individuals or groups to challenge AI-driven applications that are believed to be harmful. While the AI Act aims to reduce opacity in lower-risk AI, it does not require entities to disclose evidence of harms to external users, who lack proof of flaws in the entities' AI practices.[294] When the harm is subtle or systemic rather than overt, individual and group victims may find it challenging to mitigate aggregate harms without sufficient legal remedies. Although the EU recognizes the socially harmful impact of AI applications, the AI Act's resulting risk-based focus has not thoroughly considered the nature and actual severity of algorithmic harms.

While the AI Act is not the only regulation applicable to AI, these omissions become a significant problem when countries view it as the first comprehensive AI regulation claiming to encompass thorough risk categorization and consideration. These limitations restrict the versatility of the EU's scheme in addressing primary algorithmic harms, particularly in capturing intangible and cumulative problems.[295] The AI Act is also considered practically important in addressing algorithmic harms within the broader EU regulatory framework. Due to the intangibility of algorithmic harms, along with the opacity that hinders harm detection, the Act's misaligned harm categorization may not be effectively remedied by other regimes, whose enforcement largely relies on harm awareness. Consequently, while the EU has chosen a progressive risk-based approach to regulate AI systems, its incomplete

[292] EU AI Act, *supra* note 274, art. 5.

[293] For example, a minor vulnerability in one AI system could be exploited across a network of devices, resulting in a large-scale security incident.

[294] EU AI Act, *supra* note 274, art. 50. In the absence of a precise definition, there is a substantial risk that firms may present their algorithmic explanations in the most innocuous manner possible. Another inadequacy in this context concerns the problem of trade secrecy. This problem has emerged under the GDPR scheme. Researchers have noted that courts in Germany and Austria have allowed companies to restrict their explanations of algorithmic processes to prevent the disclosure of proprietary trade secrets. But the AI Act has not fixed this regulatory gap. There is also a problem with lack of required transparency in algorithmic governance, as the European framework does not establish strengthened disclosure rules that reduce governance invisibility. Without transparency duties to make potential victims aware of algorithmic harms, many problems are likely to go unchecked. Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FOR. L. REV. 1085, 1, 37, 38 (2018).

[295] The AI Act has a pre-emptive legal effect, creating a ceiling that prevents EU countries from enacting higher levels of protection. Veale & Borgesius, *supra* note 301, at 108.

harm concept does not sufficiently empower victims to avoid exposure to countless harms, undermining its policy goal of establishing a high level of trustworthiness in the AI ecosystem.

C. Japanese Ethical Activism

In contrast to the EU, which views AI applications as imminent threats necessitating legal interventions, Japan perceives algorithmic harms as social concerns that can be largely addressed through soft norms and small-scale legal interventions.[296] The main goal of this approach (termed "ethical activism") is to dynamize a competitive and socially beneficial AI ecosystem.[297] To achieve this goal, it leans toward lenient rules and maximizes the value of soft laws, treating innovative applications as fairly unbounded in their potential.[298] Like the U.S., Japan sees regulations as impediments to realizing the benefits of AI innovations, avoiding the introduction of hard norms that risk stifling AI development. Unlike the U.S. and the EU, it operates under the presumption and unique cultural norm that innovators often voluntarily adhere to well-crafted ethical standards to develop socially beneficial AI systems. Rather than a purely reactive approach, this approach considers imposing additional rules only in exceptional situations while continuing to strengthen the depth and coverage of ethical norms established by a variety of stakeholders.[299] Through this collaborative process, they legitimize a soft social norm approach by involving a diverse range of participants and covering a broad array of topics.[300] The exchange of opinions helps create a consensus on community standards that are both inclusive and proactive. This section discusses the strengths of its inclusive coverage and subsequently examines the fundamental issues inherent in this approach, such as the lack of legal specificity and enforceability in addressing algorithmic harms.

1. Soft Law Governance

Japan exemplifies a regulatory approach that develops and implements AI soft laws to address social problems, such as low birth rates, an aging population, and labor shortages. Its National AI Policy outlines how AI should be used to increase the country's long-term prosperity and solve social problems. The 2019 Social Principles of Human-Centric AI is a foundational ethical standard that defines the future society Japan aspires to build. By

[296] Hiroki Habuka, *Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, CSIS (2023), https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency.

[297] Japan Council for Social Principles of Human-centric AI, *Social Principles of Human-Centric AI*, (2019), https://ai.bsa.org/wp-content/uploads/2019/09/humancentricai.pdf.

[298] Ryuichi Sato & Hirotaka Kuriyama, *Japan Looks to Take Lead on AI Regulation*, JAPAN NEWS (Aug. 3, 2023), https://japannews.yomiuri.co.jp/politics/politics-government/20230803-127362/.

[299] *See, e.g.*, Japan Council for Social Principles of Human-centric AI, *supra* note 311; Japan Ministry of Economy, Trade and Industry, *Governance Guidelines for Implementation of AI Principles Ver. 1.1*, (2022), https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf (last visited Feb 25, 2023); Japan Ministry of Internal Affairs, *AI Utilization Guidelines: Practical Reference for AI Utilization*, THE CONFERENCE TOWARD AI NETWORK SOCIETY (2019).

[300] Government agencies, firms, industrial organizations, labor unions, AI experts, or a mix thereof participate in establishing ethical standards for AI developers. *Id.*

specifying social principles,[301] this instrument serves as a list of main ethical norms that should be implemented across Japanese society.[302]

The government subsequently issued the 2019 AI Utilization Guidelines, an updated set of norms with a broader range of stakeholder considerations,[303] applicable to developers, users, and data providers in their respective social contexts.[304] The Guidelines specify ten principles, some of which are relevant to legal interests, including privacy, autonomy, fairness, safety, security, transparency, and accountability; the others are connected to innovation values, such as proper utilization, data quality, and collaboration.[305] Based on feedback from multiple stakeholders,[306] Japan issued the AI Governance Guidelines in 2022, incorporating input from industries, scholars, legal experts, and auditors, outlining the ideal AI governance framework for innovators to implement the 2019 Social Principles for Human-Centric AI.[307]

The example of ethical activism exhibits inclusivity in addressing algorithmic harms. It serves as a prime example of how proactive ethical standards are leveraged to widely cover primary harms, aggravating factors, potential wrongdoers, and stakeholders. The inclusivity strength is particularly notable in its approach to algorithmic opacity. For example, the 2019 Social Principle on Human-Centric AI emphasizes reducing opacity in algorithmic decision-making.[308] This foundational standard suggests that AI developers ensure "transparency in decision-making" and furnish appropriate explanations "on a case-by-case basis" depending on the specific applications of AI.[309] The principle also outlines the items that AI developers should explain about their algorithmic decisions,[310] encourages individuals to seek additional information, and asks developers to establish a mechanism to support public trust in AI.[311]

The 2019 AI Utilization Guideline delves more extensively into addressing issues of technical complexity.[312] Unlike previous social principles that applied solely to AI developers, the guideline encourages adherence not only from AI developers but also from business users and data providers.[313] Furthermore, this instrument offers more specific recommendations for mitigating technical complexity.[314] Instead of merely suggesting the

[301] Japan Council for Social Principles of Human-centric AI, *supra* note 311.

[302] *Id.* at 7.

[303] Japan AI R&D Guidelines of 2017, *supra* note 299; Japan AI Utilization Guidelines of 2019, *supra* note 299, at 2.

[304] Japan AI Utilization Guidelines of 2019, *supra* note 299, at 2.

[305] *Id.* at 11-12.

[306] Japan Ministry of Economy, Trade and Industry, *Governance Guidelines for Implementation of AI Principles Ver. 1.1*, *supra* note 299.

[307] *Id.*; Japan Social Principles of Human-Centric AI, *supra* note 299, at 3.

[308] Japan Social Principles of Human-centric AI, *supra* note 299, at 10.

[309] *Id.*

[310] *Id.* This includes "when AI is being used, how the data is obtained and used by AI, and what measures have been taken to ensure the appropriateness of results obtained from AI operations."

[311] *Id.*

[312] Japan AI R&D Guidelines of 2017, *supra* note 299; Japan AI Utilization Guidelines of 2019, *supra* note 299.

[313] *See id.*

[314] *See id.*

provision of appropriate explanations for algorithmic decisions, the AI Utilization Guidelines recommends that those innovators ensure (1) the verifiability of input and output of AI through recording and retaining logs;[315] and (2) that their algorithmic decisions are explainable according to social context, as long as such decisions have a significant impact on individual rights and interests.[316] The inclusion of social context consideration in the latter recommendation demonstrates a more progressive stance than other provisions on algorithmic decision-making.

A few years later, in 2022, the Governance Guidelines for Implementation of AI Principles took a step further in addressing governance invisibility. This guideline expands the scope of items that firms are encouraged to disclose, making Japan one of the first countries to promote non-financial disclosures of AI governance-related information.[317] The guideline specifically points out that investors have been curious about how a firm develops and implements AI ethics.[318] This instrument emphasizes that disclosure can serve to initiate constructive dialogue with stakeholders.[319] Businesses are thus encouraged to measure algorithmic applications' impact on various stakeholders and offer information that goes beyond what is required by law. In particular, entities are advised to disclose risk assessments related to algorithmic applications in their non-financial disclosures.[320] This shift toward outward disclosure marks a distinctive advancement compared to the previous emphasis on inner verification, as suggested in the 2019 AI Utilization Guideline.

Overall, the AI ethical guidelines proposed by Japan have gradually expanded their inclusivity in many aspects: targets (from AI developers to AI users, developers, and data providers), disclosure items (from explanations of decision-making to disclosures on AI governance), addressed opacity (from technical complexity to legal secrecy and governance invisibility), context consideration (from individual case context to social context), encouraged actions (from pay attention to ensure), and spanned from inner verification to outward disclosure. The proposed transparency scheme incorporates perspectives from both individuals and society, thus potentially mitigating harm at both the individual and collective levels.

2. Ambiguous Harm Concept

Japan establishes progressively inclusive social norms that provide a foundation for addressing the aggravating factors of algorithmic harms, but these norms fall short of addressing algorithmic harms with sufficient specificity and legal force.

[315] Japan AI Utilization Guidelines of 2019, *supra* note 299, at 26.

[316] *Id.* at 27.

[317] *Id.* at 43.

[318] *Id.* at 44.

[319] *Id.* ("we learned from interviewees that they had received inquiries regarding AI governance from European institutional investors. We believe the backdrop to these inquiries is the gradual rise in interest in AI governance among investors. For example, Hermes EOS (Equity Ownership Services) has stated to the board of directors of Google's parent company, Alphabet, that '[i]nvestors are looking to' them 'to display leadership in the responsible use of AI.'").

[320] *Id.* at 43-44.

Despite Japan's efforts to establish progressively inclusive ethical norms for algorithmic transparency, its social norms governance has not adequately specified whether, how, and which primary harms should be addressed and by whom. Additionally, its governance principles primarily target the immediate harms of specific AI applications without considering the cumulative harms that arise from repetitive or interconnected AI use. Many AI systems operate in networked environments where their harms accumulate across entities' multiple operations. While the guidelines are progressive in encouraging consideration of social context, they do not provide a guiding framework for mitigating the harms that accumulate in these situations. These crucial aspects may be missing in the suggested documentation, making the sources of harm difficult to trace and address.

While the guidelines suggest explanations for algorithmic decisions and encourage non-financial disclosures, these measures may not go far enough in ensuring that critical information about potential harms is documented and disclosed. For example, transparency might be limited to certain positive aspects of AI governance, such as the adoption of privacy-protective technologies, while leaving out crucial information about how algorithmic systems are designed in ways that could lead to algorithmic harms.

The guidelines encourage firms to disclose how their governance mitigates harms and engage in constructive dialogue with stakeholders, yet these soft norms serve as principles for AI developers and users rather than legal tools to hold harmful algorithmic applications accountable. Japan has emphasized that these ethical guidelines and principles are not legally binding. [321] In other words, without mandatory requirements or strong enforcement mechanisms, there is little to ensure that companies will sufficiently disclose the negative impacts of their AI systems. AI adopters are free to select the level of implementation from zero to ten, which significantly limits accountability. Although the norms encourage business users to allow individuals to seek additional information about algorithmic decisions, private entities may not be willing to internalize the cost of disclosures, especially when doing so could damage their reputation or involve legal risks. When harms have been detected but not corrected—such as algorithmic operations favoring certain political groups, causing manipulative effects on consumers, or being vulnerable to cyberattacks—entities are unlikely to disclose these issues.

The ethical standards do not guarantee that those individuals affected by algorithmic applications have the right to receive further explanations of algorithmic decisions and verified outcomes.[322] The same applies to other algorithmic harms to individual victims. At the collective level, these soft norms also lack provisions to protect group victims. The ethical standards have not clarified which group victims should be considered in stakeholder impact assessments and non-financial disclosures. For AI developers and users, it remains unclear what should be covered in their impact assessments and governance systems to disclose algorithmic harms to group members. Group victims mostly lack the legal tools to protect themselves from algorithmic harms and thus may continue to be excluded from datasets, as well as the corresponding benefits and social status they deserve.

[321] Japan Governance Guidelines for Implementation of AI Principles Ver. 1.1, *supra* note 299, at 3.
[322] Japan AI Utilization Guidelines of 2019*, supra* note 299, at 12, 26-27.

Additionally, while ethical norms encourage consideration of social contexts when developing algorithmic innovations, they have not clarified the social dimensions that should be taken into account or identified what socially harmful practices should be considered. Social members have to rely on developers' voluntary efforts, varying interpretations, and inconsistent practices. Absent robust accountability mechanisms, it remains questionable how such guidance can inclusively address a broad set of algorithmic harms and victims to a practical extent.

Without a legal baseline specified for algorithmic harms, it can be challenging for entities to consistently and responsibly address these harms by merely following ethical guidelines. Entities can be confused about what should be done, how much they should do, and the consequences of refusal to do so. Although this approach continually expands its ethical principles and uses existing regulations to consider the contextual nature of algorithmic harms, it currently lacks a binding legal strategy to address them. While this prototype grants innovators the freedom to experiment with algorithmic applications in pursuit of specific social goals, it largely neglects the consequences of algorithmic harm as a social problem.

IV. Regulating Algorithmic Harms

Following Section III's findings of strengths and flaws in regulatory frameworks, this Section explores how a deeper understanding of algorithmic harms, with greater specificity informed by their typology and features, can bring the law into closer alignment with the actual activities it seeks to regulate. As demonstrated throughout this Article, civil rights concerns related to privacy, autonomy, equality, and safety have emerged across various algorithmic contexts. Part of the solution involves meaningful actions from policymakers to regulate algorithmic harms. While there is no universal solution to the complex problems with AI, insights into both the types of harms and the legal tools adopted by the case study jurisdictions open up new regulatory options.

Drawing on insights from earlier sections, this Section develops a harm-centric approach through consideration of the nature of harms, the types of harm in question, and the affected victims. It offers three legal interventions essential for regulating algorithmic harms, aiming to mitigate primary harms by targeting their aggravating factors.

The first regulatory intervention starts from the premise that refined algorithmic impact assessments, which impose an obligation on AI developers to address compounded harms, serve as a starting point for enhancing algorithmic accountability.

While existing assessments often evaluate harms from a collective perspective, the provision of individual rights for automated decision-making represents a crucial second intervention to address this limitation. This approach acknowledges individual differences and provides victims with enhanced protection from harmful AI applications, potentially preventing the aggregation of primary harms.

The success of these tools relies on a third intervention—a set of disclosure duties designed to reduce algorithmic opacity and increase harm awareness, especially in situations where AI use is associated with intangible yet far-reaching harms.

These interventions cannot solve all the problems, but they offer legal incentives for AI adopters and designers to address harms that cannot be detected by victims or regulators. An AI regulation that encompasses these interventions is better equipped to internalize the costs of the ubiquitous, intangible, and accumulative harms arising from AI applications. This harm-centric approach advances the current conversation on AI regulation by incorporating the characteristics and types of algorithmic harms into its crucial provisions. This in turn potentially sets the scope and depth of harm regulation and seeks to enrich strategies for algorithmic governance and radical reforms by doing so.

A. Algorithmic Impact Assessments

To mitigate algorithmic harms, addressing the aggravating factors of accountability paucity and algorithmic opacity offers a logical starting point. Impact assessment is generally considered a major regulatory strategy to enhance algorithmic accountability.[323] In recent years, academics and policymakers have proposed its use as a crucial AI governance tool for addressing algorithmic harms.[324] Algorithmic impact assessments often require evaluating the adverse consequences of AI applications at an early stage, mitigating harms prior to AI system deployment. The entities involved are compelled to document their risk- or harm-mitigation considerations as well as associated testing results, producing documentation that may be subject to regulatory review.[325]

As discussed in Section III's case study, EU policymakers have integrated this tool into their AI governance regime. Under the AI Act, high-risk AI applications are supposed to complete fundamental rights impact assessments before releasing their products to the market.[326] The EU's data protection law, the General Data Protection Regulation (GDPR), also requires covered entities to undertake data protection impact assessments for data practices that involve a significant threat to individual rights and freedoms, explicitly covering the use of AI systems.[327] Legal scholars have acknowledged that this scheme has established a practical structure for impact assessment components, which essentially contributes to enhanced accountability. Covered entities are required to describe anticipated processing, assess the necessity and extent of compliance measures, evaluate risks to "individual rights and freedoms," and propose risk mitigation measures.[328] Along the way, they are responsible for detecting the origin, nature, and severity of anticipated harms throughout the lifecycle of

[323] Many impact assessments and statements draw on regulatory examples and proposals from environmental law, such as the Environmental Impact Statements required under the National Environmental Policy Act. *See generally* 42 U.S.C. § 4321 (2012); A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements*, 2015 U. ILL. L. REV. 1713, 1755; Jessica Erickson, Comment*, Racial Impact Statements: Considering the Consequences of Racial Disproportionalities in the Criminal Justice System*, 89 WASH L. REV. 1425, 1463 (2014); Andrew D. Selbst*, Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017); Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J. L. & TECH. 117, 117 (2021) [hereinafter Selbst, *Algorithmic Impact Assessments*].

[324] Selbst, *Algorithmic Impact Assessments*, *supra* note 323.

[325] *Id.* at 122.

[326] EU AI Act, *supra* note 274, art. 27.

[327] Data Protection Impact Assessment (DPIA), GDPR.EU (2018), https://gdpr.eu/data-protection-impact-assessment-template/ (last visited Dec 17, 2020).

[328] General Data Protection Regulation (GDPR), *supra* note 274, art. 35.

their automated processing operations.[329] These obligations enable entities to develop harm correction strategies based on their development contexts, thereby addressing the issue of accountability paucity.[330]

To date, the virtue of algorithmic impact assessments has attracted both academic and legislative interest. Scholars and advocates have agreed that a substantive and procedural corporate commitment to impact assessments is both viable and crucial to algorithmic governance.[331] A duty for AI adopters to conduct algorithmic impact assessments has also become an AI measure commonly considered or included by AI legislation at both state and federal levels.

What has not been sufficiently articulated by existing legislation and scholarship is the substantive content—the specific algorithmic harms—that should be assessed, and how to incorporate the nature of intangible, ubiquitous, and cumulative harms into impact assessments for effective harm mitigation. The typology developed in this Article fills this conceptual gap by specifying which harms should be assessed by entities, considering the characteristics of algorithmic harms, and identifying affected parties. Entities should thus be specifically tasked with discovering, documenting, and solving these harms at individual, group, and social levels, paying particular attention to intangible and cumulative harms. Building upon the important prior scholarship on impact assessments,[332] a harm-focused algorithmic impact assessment should cover the following items:

1. Detect potential users who may be affected by the use of AI systems, specifying the number of affected individuals.
2. Examine whether the AI application is likely to negatively affect individual interests such as privacy, autonomy, equality, and safety.
3. Assess how AI applications might affect certain groups on the basis of their background, traits, or behaviors.
4. Evaluate the aggregation of anticipated harms from the specific AI application, consider how these may aggregate with harms from other applications, and assess their cumulative risks to individuals, communities, and society.
5. Investigate offering less harmful alternatives, including revising target variables, modifying business models, and employing less intrusive data processing techniques.
6. Analyze each potential alternative option in sufficient detail for an evaluation of their respective feasibilities.

[329] GDPR.EU, Data Protection Impact Assessment, *supra* note 341.

[330] Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 SOUTH. CALIF. L. REV. 1529, 1574 (2019) (suggesting the establishment of officers who are responsible for algorithmic decision-making and the implementation of *ex ante* impact assessments).

[331] *See, e.g.,* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017); David Wright & Charles D. Raab, *Constructing A Surveillance Impact Assessment*, 28 COMPUTER L. & SECURITY REV. 613 (2012).

[332] Katyal, *Private Algorithmic Accountability*, *supra* note 143, at 115-17.

7. Articulate a preferred choice from the range of options, explaining the rationale behind the selection.
8. Explain the adopted harm mitigation tools and their effectiveness. This involves detailed documentation of training data, algorithms, and operating results, along with descriptions of harm identification and mitigation measures.

The nature of algorithmic harms also sheds light on when, how, and by whom these assessments should be implemented by entities to ensure effective harm mitigation. Given the evolving and aggregate nature of algorithmic harms, impact assessment should be a dynamic, ongoing process.[333] The commitment must be performed both *ex ante* and *ex post*, thereby establishing the expectation that entities formulate harm control strategies at an early stage. Entities should be required to implement such assessments before releasing their products or services into the market. *Ex post* assessments, however, are also necessary because the dynamics of machine learning systems mean there may be unpredictable harms that entities cannot cover in their envisioned plans. Such entities should therefore be required to continually detect harms and update their mitigative measures based on the evolved applications of AI systems in their *ex post* assessments.

Regarding the responsible parties, due to the complexity of tracing harms that often involve various actors and activities, entities should designate a staff member to oversee and implement impact assessment duties. It is recommended that an entity should assign algorithmic controllers specifically responsible for evaluating, documenting, and correcting harms. This can separate their roles and responsibilities from those of a diffuse set of engineers and other personnel handling algorithmic practices. By empowering controllers to perform oversight over various actors, these entities can commit to their harm mitigation obligations in collaboration with other actors involved.

Furthermore, to ensure effective harm mitigation, entities should be required to periodically report AI applications that involve significant harms for regulatory review.[334] Under the EU regime, impact assessment duties suffer from weaknesses due to their lack of external accountability, including the absence of public disclosure, regulatory scrutiny, and public comment processes.[335] This hinders regulatory investigations, preventing regulators from understanding the harm identified, how it accumulates, and the cost to its victims. An obligation to report can create external pressure that motivates entities to treat harm correction more seriously. Its operation also provides regulators with valuable resources to detect the existence and severity of intangible harms, measuring their cumulative injuries to victims. While the details of these assessments constitute internal documentation legally required of entities, as the subsequent subsection on algorithmic disclosures will explain, such documentation is intended to be reported to regulators or disclosed to individuals in due course. These assessments will make the occurrence of harms more traceable, providing clues for regulators and potential victims in tracing the cause of a detected problem and thus

[333] *See* General Data Protection Regulation (GDPR), *supra* note 274, art. 35.

[334] Selbst & Barocas, *Unfair AI*, *supra* note 268, at 1093.

[335] DILLON REISMAN ET AL., AI NOW INST., ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 7 (2018), https://ainowinstitute.org/aiareport2018.pdf.

enhance both internal and external accountability for effective law enforcement.

B. Algorithmic Individual Rights

Impact assessments tend to have a general and collective focus,[336] evaluating harms by considering the situation of an average person and ignoring variations among individuals.[337] Individuals who wish to avoid risks do not have a say; they must rely on entities' harm evaluation efforts that usually cannot fully consider individual differences. This lack of individual focus underscores a missing element in the development of AI regulation: the right of individuals, inherent in a consumer rights-focused regulatory approach, to decide whether they wish to be exposed to certain algorithmic harms. This gap has often led to AI systems operating without fully considering harms to specific individuals. To tackle this regulatory deficiency, the second proposed legal intervention is ensuring that a regulatory regime includes an individual right to opt in or opt out. Such a right would allow potential victims to mitigate the harms that they may suffer and take part in harm correction processes. This is especially necessary given that individuals have no way of monitoring whether entities have corrected such harms.

Individual rights regarding automated decision-making are still a relatively new area of law in the United States. While these rights have been incorporated into EU data protection law for years, legislators in the United States have also started to incorporate them, primarily at the state level.[338] In the United States, legislators in California are actively working on establishing individual opt-out rights. The California Privacy Protection Act (CPPA) has recently proposed that several AI applications should be subject to a set of opt-out rights.[339] The CPPA draft's opt-out rights cover automated decisions that raise significant civil rights concerns.[340] This includes AI applications that erode privacy (i.e., facial recognition and emotion assessment)[341] and diminish equality (i.e., profiling employees, contractors, job applicants, or students).[342] Additionally, the draft covers autonomy and safety harms arising from behavioral advertising practices, profiling children, and using

[336] Margot E Kaminski & Gianclaudio Malgieri, *Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations*, 11 INT. DATA PRIV. L. 125, 137 (2021).

[337] *See* William Boyd, *Genealogies of Risk: Searching for Safety, 1930s-1970s*, 39 ECOLOGY L.Q. 895, 927 (2012); Kaminski, *Regulating the Risks of AI*, *supra* note 33, at 1392.

[338] The EU GDPR's rights-based scheme has notably established procedural rules that enable individuals to challenge algorithmic AI-made decisions, that is, decisions that are made entirely by automated processes and have significant consequences for individuals' legal interests, such as decisions about one's eligibility for a loan. While this right marks a progressive move, this provision has faced criticism for its narrow scope and lack of substantive requirements or concrete procedural rules. The regulated AI-made decisions are limited to those that are made entirely by automated processes and have significant consequences for individuals' legal interests, such as decisions about one's eligibility for a loan. For assessments on the provision's procedural flaws and narrow focus, see Kaminski & Urban, *supra* note 3, at 1977-78, 1981-82 & sec. III.B.

[339] Wu, *supra* note 240.

[340] Lomas, *supra* note 21.

[341] State of California, *A New Landmark for Consumer Control Over Their Personal Information: CPPA Proposes Regulatory Framework for Automated Decisionmaking Technology*, (2023), https://cppa.ca.gov/announcements/2023/20231127.html (last visited Jan 1, 2024).

[342] *Id.*

consumer information to train automated decision-making. As of this writing, it is unclear how the planned regulatory framework might be finalized;[343] however, its progressive proposals allow us to think more broadly and thoroughly about how individuals play a role in mitigating algorithmic harms.

Currently, existing literature offers limited guidance on the specific factors that should be considered to create an algorithmic individual rights framework that is feasible in the U.S. The following part leverages the typology of algorithmic harms to develop a framework suitable for U.S. settings and elsewhere. It suggests adopting procedural interventions that protect substantive legal interests and establish victim-friendly procedural rules, empowering individuals to mitigate algorithmic harms through opt-in and -out rights.

Because many instances of eroding privacy and undermining autonomy involve the use of personal data, an individual subjected to an automated system should have the right to demand that their personal data be excluded from use in AI applications. Individuals should be able to opt into the utilization of data for activities that involve significant AI-driven privacy harms. This includes algorithmic extraction or processing of personal data through applications like biometric surveillance, biometric identification and categorization, and emotion recognition.[344] In cases where such use could lead to substantial harm to people's autonomy, such as the profiling of vulnerable groups like children, it should also be subject to an opt-in scheme. Other types of undermining of autonomy, including subtle targeting like behavioral marketing and the practice of hypernudging, should be regulated through an opt-out scheme. For equality harms, individuals should be provided an opt-in right to object to automated decisions in critical areas such as employment, health care, housing, judicial decisions, and education. To address algorithmic harms to safety, individuals should be able to opt into algorithmic practices that leverage user vulnerabilities to exacerbate health problems among users, including manipulative features of algorithmic designs that prioritize content related to cyberbullying or inciting materials.

In commercial settings, for instance, firms should ensure that those who decide against using certain applications are not disadvantaged or subjected to unequal treatment, such as refusing to provide goods or services, charging different prices, or offering a different quality of goods or services. This approach underscores the importance of maintaining fairness and inclusivity in technological offerings, catering to the diverse needs and choices of all users. Additionally, there should be a set of clear and functional procedural rules that make it easy for affected consumers to opt out at any time. For the protection of individual interests, this could include an easy-to-understand notice with rejecting a potentially harmful practice as the default option, requiring entities to address opt-out requests in a timely manner. An opt-out procedure without a timeline would delay prevention of harms, making such a right meaningless. If the procedure is too complicated or time-consuming, individuals may find it too challenging or tedious to opt out of such applications.

This is not to suggest that individual rights are without their shortcomings. Ideally, a

[343] *Id.*

[344] This requirement should be irrespective of the timing (real-time or retrospective) and location (public or private spaces).

legal regime banning or restricting certain AI applications for causing algorithmic harm can offer better protection for those affected by AI systems. However, before the real impact of algorithmic harms is extensively known, the pro-innovation culture of the American legal system, as demonstrated by its consumer-oriented AI regime, suggests that a restrictive legal regime may face strong resistance from industries due to its potential stifling effects on innovation. In contrast, the individual rights regime may be more in line with current trends in state legislation in this field, making it more likely to be adopted in future AI regulations. Under either scheme, individual rights would set a foundational expectation for AI developers and adopters. By giving individuals the chance to opt in or opt out, it rebalances the power dynamics between firms and users. Because consumers' opting out entails a loss in the economic value of what can be extracted from AI applications, firms would be motivated to avoid adopting AI in potentially harmful ways. Otherwise, corporate scandals may trigger a substantial number of individuals exercising such a right. On the other hand, this Article also acknowledges the weaknesses of an individual rights mechanism, such as individuals' limited capacity to detect and evaluate intangible harms to make meaningful decisions, the perceived consumer interests intertwined with harms, and firms' reluctance to disclose anticipated harm due to reputation concerns.[345] Certain aspects of these issues will be further addressed by a subsequent legal intervention: reducing algorithmic opacity that hinders harm detection and correction through obligations regarding algorithmic social disclosures.

C. Algorithmic Social Disclosures

The success of individual rights relies on individual awareness of the harms involved in AI applications. Crucial to this awareness are disclosure duties that enable impacted victims to evaluate potential harms. While impact assessments require the establishment of foundational mechanisms to address accountability paucity, algorithmic opacity is another aggravating factor that remains unaddressed. Transparency obligations are essential for individuals to perceive harm that was originally intangible and indiscernible. A vital transparency framework for individuals and other stakeholders to monitor AI practices is thus mandatory for effective harm correction.

The importance of algorithmic transparency has been widely recognized by policymakers, scholars, and civil organizations,[346] but few have thoroughly examined its connection to the intangible and cumulative impact of algorithmic harms. This subsection addresses this gap by arguing that in light of the characteristics of algorithmic harms, we need to establish disclosure schemes that fulfill at least three functions in harm correction.

To begin with, disclosures should be indicative, marking specific AI applications as generating significant yet intangible harms. For instance, an indication of impairing safety

[345] *See* Daniel J. Solove & Woodrow Hartzog, *Kafka in the Age of AI and the Futility of Privacy as Control*, 104 B. U. L. REV. 1021 (2024).

[346] *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015). Danielle Keats Citron, *Open Code Governance*, 2008 U. CHI. LEGAL F. 355; Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1250-79 (2019); Peter Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA L. REV. 331, 371-378 (2020).

such as a label warning of the health impacts of social media algorithms disclosed to kids and their parents can make individuals aware of such harms. For deepfakes and other generative AI applications that can produce manipulation harm, providers should be required to disclose the use of AI and, whenever possible, the source of the content.[347] Clear identification of such content ensures that users and consumers are adequately informed about the nature and origin of the materials with which they interact. The implementation of these rules, most of which have been adopted in EU and are being considered in the United States,[348] would avoid certain cognitive and manipulative harms, contributing to greater clarity and trust in AI applications.

The second role is instructive, enabling impacted individuals to make informed decisions about opting in or out. Because of the intangible nature of algorithmic harms, individuals cannot usually identify such harms by themselves. To enhance harm awareness, entities should be obligated to issue *ex ante* notifications to individuals regarding the anticipated use of AI applications, how their data will be used, and the potential direct and cumulative harms involved, along with the right to receive more detailed data regarding the intended application of AI.[349] For more effective harm evaluation, in response to access requests, the provided information should include decision-making processes, its logic and key parameters, anticipated harms and potential outcomes, and how to opt-out at any time or file complaints about the use of AI systems.[350]

Disclosures of such information can be instructive for both individual and collective harm mitigation. Scholars have observed that when a law requires entities to evaluate and disclose the same set of information under both the duty of impact assessments and the duty to disclose to impacted individuals, the information disclosed to individuals could include parts of the impact assessment documentation.[351] This piece suggests that future legislation should follow this direction. The preceding impact assessment section recommends that the documentation is supposed to explain how AI affects other group members. Here, it further suggests that entities should be required to inform individuals impacted by algorithmic decision-making about such information, allowing them to evaluate structural harms as group victims. Even if some individuals may not read every notice carefully, or choose to opt in despite a notice indicating that an AI practice is biased against a racial group because it does not directly affect them, a regime that includes individual rights, disclosure duties, and other oversight mechanisms will equip those who care about civil rights and collective welfare with the tools to protect themselves. This approach will incentivize entities to develop products and services that better meet the needs of these individuals, representing an improvement over the status quo.

[347] Regulators have called for explicit rules requiring labeling AI-generated content and disclosure of the involvement of copyrighted materials. Supantha Mukherjee et al., *EU Proposes New Copyright Rules for Generative AI*, REUTERS (Apr. 28, 2023), https://www.reuters.com/technology/eu-lawmakers-committee-reaches-deal-artificial-intelligence-act-2023-04-27/ (last visited Jan 1, 2024).

[348] *Id.*

[349] General Data Protection Regulation (GDPR), *supra* note 274, arts. 12-23.

[350] Similar requirements have been proposed under the CPPA scheme. State of California, *supra* note 357.

[351] Kaminski & Malgieri, *supra* note 350, at 134; CCPA, *supra* note 297.

The third role is supervisory. The victims of algorithmic harms extend beyond consumers, including non-consumers and other groups in society who lack legal remedies. The breadth of affected victims necessitates additional social oversight of private entities' harm mitigation efforts. Japan's ethical activism approach provides valuable lessons for algorithmic social disclosures with an inclusive focus.[352] Its advocacy, as evidenced by investor and stakeholder demands for more non-financial disclosures, suggests that Japan's regulatory example could offer practical guidance. These transparency obligations can take the form of non-financial disclosures to facilitate social oversight of AI applications that harm individuals, groups, and societies.[353] Policymakers should require such algorithmic disclosures, making them more accessible to stakeholders. At a minimum, firms should be required to disclose the types of algorithmic applications with harmful effects on civil rights and the impact of identified harms on specific populations. They could utilize their harm assessment documentation, which should conform to the standards outlined in Section IV. A. of the impact assessment proposal to disclose a summary of the assessment. This summary should disclose whether and how a particular AI application individually and cumulatively harms individuals, groups, and society. Disclosure of impact assessment both before and after implementation of AI applications can be a valuable source of justification and co-creation of harm mitigation strategies.

Importantly, these transparency rules cannot eliminate algorithmic opacity, especially technical complexity that involves the unpredictability of machine learning systems. What can be achieved by proposed disclosures is asking entities to report harms arising from such systems for broader social oversight. While these proposals cannot prevent private entities from claiming trade secret protections, enhanced disclosure duties would require them to disclose anticipated harms, promote an ongoing discussion of harm mitigation strategies, and be more vigilant about the harms that may arise from their AI practices. Those impacted by AI and other stakeholders can conduct a more thorough harm-benefit analysis. Under the status quo, people primarily perceive the consumer benefits brought by AI. With the proposed disclosure, they can better consider the harms arising from AI when making individual choices or enforcing monitoring. Together, algorithmic social disclosures provide a greater external understanding of how any adverse impacts on victims and society are addressed, motivating more actors to address algorithmic harms more proactively.

D. Policy Considerations

Regulating algorithmic harms involves a deeper yet fundamental tension between two competing policy goals: the promotion of innovation interests and the protection of civil rights. Since countries began to formulate AI policies and legislation, these two issues have been the consistent focus of policy discussion. While terminology varies from one nation to another, the interests surrounding innovation revolve around achieving the benefits of economic competitiveness or public interests;[354] democratic values involve preserving social

[352] Japan Ministry of Economy, Trade and Industry, *Governance Guidelines for Implementation of AI Principles Ver. 1.1*, *supra* note 313.

[353] *See, e.g.*, Lu, *Corporate Social Disclosures*, *supra* note 43, at IV.

[354] U.S. EXECUTIVE OFFICE OF THE PRESIDENT, MAINTAINING AMERICAN LEADERSHIP IN ARTIFICIAL

trust based on the protection of civil rights and liberties.[355] In the United States and many other countries, policymakers frequently emphasize the significance of civil rights protection in their policy frameworks. However, regulatory actions too often favor concrete innovation-driven interests, blunting the effect of laws dedicated to safeguarding individual civil rights.[356] This tendency stems from the clear benefits associated with innovation and the influence wielded by key industry stakeholders.[357] In contrast, AI's harmful effects on civil rights are more intangible, varied, and obscure.[358]

The underestimation of algorithmic harms creates a disparity between policy goals and regulatory outcomes, impeding the formulation of comprehensive legal strategies.[359] This piece's typology of algorithmic harms aims to realign goals with outcomes through specifying those harms arising from AI innovations, offering both theoretical and empirical justification for a harm correction legal framework and heightened policy emphasis on safeguarding these interests.

Furthermore, to rebalance the dynamics between innovation and civil rights, this Article has presented a refined harm-centric approach that intends to safeguard individual interests without unduly restricting the use of AI. The proposed solution primarily focuses on procedural rules such as harm-centric impact assessments, individual rights, and algorithmic disclosures rather than substantial interventions (e.g., an outright ban on deepfakes) for the following instrumental reasons.[360]

To begin with, procedural tools are considered valuable when AI applications involve impacts on individuals and society that are difficult to measure.[361] As previously discussed, algorithmic harms are intangible, varied, and dynamic. Establishing a one-size-fits-all

INTELLIGENCE, EXECUTIVE ORDER 13859, (2019), https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence; EUROPEAN COMMISSION, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, (2020), https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en; Habuka, *supra* note 310.

[355] Council of the European Union, *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 - General Approach*, (2022), https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf (last visited Feb 25, 2023); (US) Executive Office of the President, *supra* note 247.

[356] Betsy Vereckey, *Does Regulation Hurt Innovation? This Study Says Yes*, MIT SLOAN (2023), https://mitsloan.mit.edu/ideas-made-to-matter/does-regulation-hurt-innovation-study-says-yes.

[357] Billy Perrigo, *Big Tech Is Already Lobbying to Water Down Europe's AI Rules*, TIMES (Apr. 21, 2023 1:52 PM EDT), https://time.com/6273694/ai-regulation-europe/.

[358] Craig S. Smith, *Act One: Opposition Takes Center Stage Against EU AI Legislation*, FORBES (Sept. 5, 2023), https://www.forbes.com/sites/craigsmith/2023/09/05/act-one-opposition-takes-center-stage-against-eu-ai-legislation/ (last visited Oct 19, 2023).

[359] Cecilia Kang, *In the U.S., A.I. Regulation Is in Its 'Early Days'*, N.Y. TIMES (July 21, 2023), https://www.nytimes.com/2023/07/21/technology/ai-united-states-regulation.html.

[360] Because of the limited scope of this paper, the typology does not provide an exhaustive list of harmful activities, nor does it include important interests like environmental harms. How the proposals operate in conjunction with a broad range of risk regulation tools like auditing, licensing, certification, and alternative solutions awaits further research.

[361] Selbst, *Algorithmic Impact Assessments*, *supra* note 337, at 123 ("Impact assessments are most useful when projects have unknown and hard-to-measure impacts on society.").

substantive intervention such as a ban is challenging and even impractical for harms that are highly contextual. The proposed procedural rules can be adapted to specific application contexts, thereby accommodating the nature of algorithmic harms.

Next, to balance these two potentially competing policy goals, the procedural rules do not disregard the benefits that can be brought about by potentially harmful AI systems. Rather than arguing for a ban or overly restrictive use, such rules facilitate the development of harm mitigation strategies and necessary governance structures. These rules are devised to channel innovators' efforts into developing AI applications that both minimize harms to civil rights and ensure trustworthiness at an early stage.

Moreover, procedural rules play a constructive role in fostering the development of effective substantive rules. The proposed framework serves as a foundational standard for policymakers, supporting rather than hindering the potential adoption of substantive rules. By mandating harm mitigation schemes and reporting harmful practices, the proposed impact assessments and transparency duties enable policymakers to better understand the harms and develop appropriate substantive rules.

Each of the proposed solutions works to create a synergistic effect for harm correction. Specifically, establishing an individual right against algorithmic harms could potentially act as a deterrent, prompting entities to proactively engage in harm mitigation; implementing harm-centric algorithmic impact assessments and transparency rules would lead to a more holistic interplay among potential wrongdoers, victims, and regulators. The proposal provides a framework for understanding and correcting harms without overly impeding the development of AI, seeking to strike a careful balance between innovation and protection of civil rights. A profound understanding of these harms, shaped by these regulatory proposals, can lay the groundwork for more comprehensive reforms. Such reforms could extend beyond procedural efforts and consider more radical changes in areas such as tort law, remedies, and more. By addressing these issues with the nuanced understanding required to meet the unique challenges posed by algorithmic harms, legal frameworks can evolve to better protect civil rights in an increasingly automated society.

CONCLUSION

The proliferation of algorithmic harms constitutes a looming crisis in democratic systems. Their negligible form leads to a dangerous underestimation of their actual impact, hindering efforts toward creating a safer society. Existing statutory regimes with their major focus on tangible and individual harms cannot adequately address these challenges. Future development of AI law should be more attentive to the intangible, collective nature of algorithmic harms to civil rights. The law must offer meaningful incentives for entities to minimize the primary harms that impair the fundamental trustworthiness of AI systems at both individual and social levels.

Adopting harm-centric rules by addressing aggravating factors to mitigate primary harms is an important starting point. This approach does not unduly restrict the use of AI but instead encourages the development of harm mitigation strategies. Recognizing algorithmic harms as harms and not mere risks is another crucial step in developing future AI law. This harm-based turn opens up possibilities to ensure algorithmic harms are defined more clearly, assessed more holistically, and regulated more seriously for individuals, vulnerable groups, and our society.

To be sure, the costs required to implement a harm-based framework need further research, including a thorough examination of harm-benefit analysis. Despite the costs involved, however, the proposed solution aids in adequately calculating harm, ensuring that legal interventions are not reduced to symbolic tools but can truly achieve the desired policy goals. To that end, a normative turn to harm regulation can be socially beneficial. It yields greater long-term value in internalizing the costs of harms to establish robust algorithmic governance. It also incentivizes innovators to ensure individuals enjoy beneficial, sustainable, and safer AI innovations without harms that should not be tolerated. The proposed algorithmic impact assessments, individual rights, and social disclosures are just a few examples of harm regulation. More work has yet to be done, this work offers a logical start.