2023

The Prediction Society: Algorithms and the Problems of Forecasting the Future

Hideyuki Matsumi

Daniel J. Solove

THE PREDICTION SOCIETY: ALGORITHMS AND THE PROBLEMS OF FORECASTING THE FUTURE

by

Hideyuki Matsumi

&

Daniel J. Solove

Draft: July 30, 2023

ABSTRACT

Predictions about the future have been made since the earliest days of humankind, but today, we are living in a brave new world of prediction. Today's predictions are produced by machine learning algorithms that analyze massive quantities of personal data. Increasingly, important decisions about people are being made based on these predictions.

Algorithmic predictions are a type of inference. Many laws struggle to account for inferences, and even when they do, the laws lump all inferences together. But as we argue in this Article, predictions are different from other inferences. Predictions raise several unique problems that current law is ill-suited to address. First, algorithmic predictions create a fossilization problem because they reinforce patterns in past data and can further solidify bias and inequality from the past. Second, algorithmic predictions often raise an unfalsifiability problem. Predictions involve an assertion about future events. Until these events happen, predictions remain unverifiable, resulting in an inability for individuals to challenge them as false. Third, algorithmic predictions can involve a preemptive intervention problem, where decisions or interventions render it impossible to determine whether the predictions would have come true. Fourth, algorithmic predictions can lead to a self-fulfilling prophecy problem where they actively shape the future they aim to forecast.

More broadly, the rise of algorithmic predictions raises an overarching concern: Algorithmic predictions not only forecast the future but also have the power to create and control it. The increasing pervasiveness of decisions based on algorithmic predictions is leading to a prediction society where individuals' ability to author their own future is diminished while the organizations developing and using predictive systems are gaining greater power to shape the future.

Privacy and data protection law do not adequately address algorithmic predictions. Many laws lack a temporal dimension and do not distinguish between predictions about the future and inferences about the past or present. Predictions about the future involve considerations that are not implicated by other types of inferences. Many laws provide correction rights and duties of accuracy that are insufficient to address problems arising from predictions, which exist in the twilight between truth and falsehood. Individual rights and anti-discrimination law also are unable to address the unique problems with algorithmic predictions.

We argue that the use of algorithmic predictions is a distinct issue warranting different treatment from other types of inference. We examine the issues laws must consider when addressing the problems of algorithmic predictions.

THE PREDICTION SOCIETY: ALGORITHMS AND THE PROBLEMS OF FORECASTING THE FUTURE

by Hideyuki Matsumi[1] & Daniel J. Solove[2]

[1] Hideyuki ("Yuki") Matsumi. PhD candidate/researcher at the Research Group on Law Science, Technology and Society (LSTS) as well as at the Health and Ageing Law Lab (HALL) of the Vrije Universiteit Brussel (VUB). Member of the New York Bar. I would like to thank everyone who patiently listened and waited for me.

[2] Eugene L. and Barbara A. Bernard Professor of Intellectual Property and Technology Law, George Washington University Law School. Thanks to my Travis Yuille for excellent research. We both want to thank Dan Bouk, Dan Burk, Jessica Eaglin, Oscar Gandy, Talia Gillis, Woodrow Hartzog, Mireille Hildebrandt, Alicia Solow-Niederman, the participants at the Privacy Law Scholars Conference 2023 for very helpful comments, and everybody in the LSTS.

INTRODUCTION

"Prediction is very difficult, especially if it's about the future."

-- Neils Bohr[3]

Humans have always tried to predict the future.[4] History and myth abound with oracles, soothsayers, psychics, Tarot cards, and crystal balls. Well-being, fortune, and even life itself can hinge upon correctly guessing the future.

Since the beginning of civilization, there have been at least two distinct and contrasting methods for making predictions: looking to "individuals who have an intrinsic gift or ability to predict the future" or developing "systems that provide rules for calculating futures."[5]

The first method is *prophecy* – based on superstition, turning to oracles, shamans, or prophets to foretell the future. These individuals were viewed as possessing the special ability to see the future or interpret divine messages.

The second method is *forecasting* – based on calculation.[6] Perhaps one of the earliest instances of this approach was the development of the solar calendar. The ancient Egyptians measured the Nile's floods to predict the following year's harvest.[7]

[3] For example, Quote Investigator has an interesting post on who should be credited for this quote "it's difficult to make predictions, especially about the future" and its variants. See https://quoteinvestigator.com/2013/10/20/no-predict/.

[4] Hideyuki Matsumi, *Prediction as Defamation: Is Predictive AI Asserting a Fact or Expressing an Opinion?*, presented at We Robot 2022 [hereinafter *Prediction as Defamation*]; Hideyuki Matsumi, *The Failure of Rectification Rights*, presented at PLSC 2022 [hereinafter *Rectification Rights*]; Hideyuki Matsumi, *Predictions and Data Protection: Future Forecasting and Challenges for European Law*, master's thesis for LL.M. in International and European Law (2020) [hereinafter *Predictions and Data Protection*]; Hideyuki Matsumi, *Predictions and Privacy: Should There be Rules About Using Personal Data to Forecast the Future?*, 48 Cumb. L. Rev. 149 (2017) [hereinafter *Predictions and Privacy*]; Hideyuki Matsumi, *Do I Have Privacy Rights over Predictive Information?: Information Privacy in Ubiquitous Robotic Society*, presented at PLSC 2015; Hideyuki Matsumi, *Information Privacy Analysis of Forthcoming Personalized Products or Services: Are There Privacy Rights over Collateral or Future Information in Ubiquitous Robotic Society?*, master's thesis for LL.M. in Intellectual Property Law at George Washington University (2012) supervised by Professor Daniel J. Solove.

[5] Amanda Rees, "The History of Predicting the Future," Wired, Dec. 2021, https://www.wired.com/story/history-predicting-future/.

[6] JAMIE L. PIETRUSKA, LOOKING FORWARD: PREDICTION AND UNCERTAINTY IN MODERN AMERICA 11 (2017) ("Nineteenth-century writers used the words *prediction, prophecy,* and *forecast* interchangeably, and *prophecy* and *forecast*—and their twentieth-century connotations of religion and science—were not yet separated by a wide intellectual or ideological divide. But the late nineteenth century signaled the divergence of the meanings of *prophecy* and *forecasting*.").

[7] *See* TOBY WILKINSON, THE NILE: TRAVELLING DOWNRIVER THROUGH EGYPT'S PAST AND PRESENT 8 (2015) ("The annual measurements of the inundation were pored over by priests and bureaucrats alike, for they gave an unnervingly accurate prediction of the following year's harvest.").

For a long time, the calculation method was difficult, time-consuming, expensive, and only occasionally successful. But major advances were made in the nineteenth and twentieth centuries with the rise of statistics and probabilities.[8] The calculation method became easier, faster, and more accurate—at least sometimes.

Modern prediction is actuarial in nature, using mathematics to make calculations in a large-scale manner. Statistics and the power of modern computing have fueled a dramatic rise in the use of algorithmic predictions. These algorithms were fueled by personal data, and the digital age has provided an unprecedented glut of it.

Today, algorithmic predictions about people are increasingly being made about matters of great significance. Is a person likely to commit a crime in the future? Is a person likely to be a productive and honest employee? Is a person likely to pay back loans on time? Algorithmic predictions are prevalent in finance, education, employment, and insurance – and they continue to spread to other critical domains of people's lives.

The technology behind algorithmic predictions also fuels artificial intelligence (AI). Ajay Agrawal, Joshua Gans, and Avi Goldfarb refer to AI as "prediction machines.[9] They note that "the new wave of artificial intelligence does not actually bring us intelligence but instead a critical component of intelligence—*prediction.*"[10]

Evangelists for these algorithmic predictions tout them as superior to human prognostications. Algorithmic predictions are based on massive quantities of personal data that far outstrip the limited and anecdotal human range of experience and knowledge. Cary Coglianese and Lavi Ben Dor, proclaim that algorithms "are making highly accurate predictions that often outperform humans in executing important tasks."[11] Cass Sunstein contends that algorithms "prevent unequal treatment and reduce errors."[12]

But we must be cautious before embracing algorithmic predictions about human behavior. Algorithmic predictions are not as accurate and unbiased as the hype suggests, and they raise significant problems that the law currently is ill-suited to address.

This Article explores the problems of algorithmic predictions. We live today, in what Alicia Solow-Niederman aptly terms, the "inference economy."[13]

[8] THEODORE M. PORTER, THE RISE OF STATISTICAL THINKING 1820-1900 (1986); GERD GIGERENZER, THE EMPIRE OF CHANCE: HOW PROBABILITIES CHANGED SCIENCE AND EVERYDAY LIFE (1989).

[9] AJAY AGRAWAL, JOSHUA GANS, AND AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE (2022).

[10] *Id.* at X.

[11] Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 Brook. L. Rev. 791 (2021).

[12] Cass R. Sunstein, *Governing by Algorithm? No Noise and (Potentially) Less Bias*, 71 Duke L.J. 1175, 1177 (2022).

[13] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 Nw. L. Rev.

Algorithms abound, making countless inferences about people. Laws struggle to regulate algorithmic inferences, and a robust literature examines these challenges.[14] The focus is often generally on algorithmic inferences.

But in this Article, we contend that a special focus must be given to algorithmic *predictions about the future.* Algorithmic predictions are significantly different from other types of algorithmic inferences because they involve the element of *time*. The temporal dimension dramatically changes the implications of algorithmic predictions and creates a rather unique set of problems.

As we use the term in this Article, a "prediction" involves an inference or guess about the future.[15] Sometimes, commentators use "prediction" as a synonym for any inference or guess, such as making a "prediction" that someone currently has cancer. In contrast, we use the term "prediction" more precisely to refer to a type of inference that involves forecasting future events that can't be verified in the present, such as a prediction that someone is likely to get cancer in the future. The future orientation of predictions changes everything because matters asserted in predictions are presently unvested and contingent. For example, an inference about where a person is located at the present or in the past has vested. There are ways to verify if the inference is correct. In contrast, a prediction about a person's location tomorrow is unvested. Until tomorrow, the prediction remains unverifiable.

We contend that algorithmic predictions raise at least four major problems that many laws concerning privacy, data protection, and anti-discrimination fail to address adequately. First, algorithmic predictions lead to what we call the "fossilization problem" because the predictions are based on data about the past. Decisions involving algorithmic predictions can reinforce patterns from past data and can further entrench discrimination, inequality, and privilege.

A second difficulty with algorithmic predictions is the "unfalsifiability problem." Because future forecasting is about a probable but uncertain future (i.e., contingent and unvested), the matter asserted cannot be verified or falsified when predictions are made. Because the law allows individuals

357 (2022).

[14] See Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (2016); Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (2015); Solow-Niederman, Inference Economy, supra note X; Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 Harv. J.L. & Tech. 621 (2021); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Cal. L. Rev. 671 (2016).

[15] Originally, the taxonomy in *Predictions and Privacy* provided three types of inferences or predictive information that is: (1) "vested"; (2) "not vested but will eventually vest" ("forecasting information" or "future forecasting"); and (3) "will not vest" or "unlikely to vest" (subjective information). Prime examples of the first type are making inferences about present age, current location, or pregnancy status. The matters asserted are vested at the present moment. Examples of the second type are a bus driver's likeliness to get involved in an accident within a three-month period, the likelihood of divorcing in two years, or an individual's risk score of recidivism. They tend to assert that a particular event, occurrence, conduction, or status is likely to happen in the future or within a given time frame. In this Article, however, "prediction" and "future forecasting" are used synonymously.

mainly to challenge inaccurate data, individuals lack the ability to meaningfully contest predictions because they exist in the twilight between truth and falsity.[16]

Third, in what we call the "preemptive intervention problem," when preemptive decisions or interventions are made based on future forecasting, the feedback loop to assess whether the predictions are accurate dissipates, which can reinforce potentially inaccurate future forecasting.

Fourth, algorithmic predictions can lead to a "self-fulfilling prophecy problem." Predictions do not just forecast the future; they actively shape it. Decisions made based on algorithmic predictions can make them more likely to come true.

All of these problems involve time. The problems stem from dealing with a probable (or possible) but uncertain future that is contingent and unvested. Nevertheless, decisions are made based on future forecasting, which can change people's lives. In many cases, the accuracy of predictions can't be assessed or will be skewed by decisions made before the future unfolds.

More broadly, the rise of algorithmic predictions raises an overarching concern: *Algorithmic predictions not only forecast the future but also have the power to create and control it.* The increasing pervasiveness of decisions based on algorithmic predictions is leading to a prediction society where individuals' ability to author their own future is diminished while the organizations developing and using predictive systems are gaining greater power to shape the future.

Consider, for example, the 2002 science fiction movie *Minority Report,* based upon a story by Philip K. Dick. A special precrime unit of the police department has had tremendous success in predicting future crimes.[17] Using a group of enslaved psychics called "precogs," the precrime unit is able to foresee crimes before they occur. One day, a prediction is made that John Anderton, an officer in the precrime unit, will commit murder. The prediction ultimately turns out only to be partially right; the vision of the precogs is correct but misinterpreted.

The movie depicts a dystopian future, where people are punished before they do anything wrong. People are at the mercy of predictions about their future, with no ability to contest them. When Anderton learns about the prediction that he will commit murder, he doesn't try to argue because he knows it is futile. Instead, he runs.

We are not yet living in the harrowing world of *Minority Report,* but we are well along the path toward it – and in some cases, we are coming uncomfortably close to it. Every day, incarceration decisions are made based on algorithmic predictions of future crimes. Every day, people are denied

[16] The GDPR allows people to challenge "solely" automated decisions, but many decisions involving algorithmic predictions are hybrids involving humans. Moreover, as we discuss later on, it is quite difficult for people to challenge predictions under the GDPR. *See supra* Part III.

[17] Minority Report (2002).

jobs or loans based on predictions about things they haven't yet done. The escalating collection of vast quantities of personal data, gathered from a burgeoning number of connected devices, is being fed into more and more algorithms, generating countless predictions on a scale hitherto unimaginable.

Today, algorithmic predictions are being used with increasing frequency, in an ever-expanding range of domains, with too much confidence, and not enough accountability. Ironically, future forecasting is occurring with far too little foresight.

The law often fails to address the problems that emerge from algorithmic predictions. Privacy law is built around a true-false dichotomy and provides rights of correction and duties to maintain accurate records. But predictions are neither true nor false and don't fit into this dichotomy. Various other duties and rights, such as transparency and rights to object or opt out also fall short to address the problems with algorithmic predictions. Laws that provide special protections against automation are also not sufficiently focused on the problems of algorithmic predictions, which emerge not just from the use of algorithms but also from attempts to make decisions about people based on predicting their future behavior. Only by addressing the special problems caused by deciding based on *predictions* will the law find a productive way forward.

This Article proceeds as follows. In Part I, we discuss the different types of inferences and the role that algorithms play in making them. We explore why time is a central issue for understanding the impact of inferences and why predictions raise special considerations apart from other types of inferences. We chronicle the history of the rise of algorithmic predictions in several domains, and we discuss how their use is prevalent and increasing.

In Part II, we discuss four unique problems that algorithmic predictions create – the fossilization problem, the unfalsifiability problem, the preemptive intervention problem, and the self-fulfilling prophecy problem. We then discuss how these problems raise an essential issue affecting the fabric of society: *Who controls our future?*

In Part III, we examine why current laws fail to address the problems of algorithmic predictions. The law lacks the necessary tools, approaches, and concepts.

In Part IV, we discuss the issues the law should consider when regulating algorithmic predictions.

I. UNDERSTANDING ALGORITHMIC PREDICTIONS

Algorithms are increasingly playing a major role in our lives. They are being used to analyze enormous quantities of personal data to make inferences about us, which are then used to make decisions that have profound consequences for our lives and for society as a whole.

An oft-neglected dimension of algorithmic inferences is *time*. This temporal dimension has enormous implications. The distinction between past, present, and future matters. It makes an enormous difference because algorithms used to predict the future cause unique problems. It is thus imperative that algorithmic predictions be addressed differently from other inferences.

In this Part, we discuss key terminology, such as inferences, profiling, and predictions. Additionally, we contend that algorithmic predictions are a distinctive type of inference that should be treated differently from other types of inferences.

A. INFERENCES, PROFILING, AND PREDICTIONS

An "inference" involves using known true facts to make guesses about other facts.[18] The California Consumer Privacy Act (CCPA) defines "inference" as "the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data,"[19] and "inferences drawn" is included as one of the types of information that constitutes personal information.[20] Inferences are attempts to figure out from what is known about what is unknown.

Today, an unprecedented amount of data about individuals is collected and used, but as in the gold rush, greed for more abounds. Through inference, personal data is conjoined with other personal data to breed even more personal data. Inference thus involves the analysis of personal data as well as the creation of it.

"Profiling" involves making inferences about people.[21] Profiles are

[18] According to Merriam-Webster's dictionary, an inference is "the act of passing from one proposition, statement, or judgment considered as true to another whose truth is believed to follow from that of the former." https://www.merriam-webster.com/dictionary/inference.

[19] Civ. Code, § 1798.140(r).

[20] Civ. Code, § 1798.140(o)(1)(K): "Inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes."

[21] The definition of profiling here is different from how the GDPR defines profiling, as the GDPR focuses specifically on automated profiling, which it defines as "any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location

constructed from comparing facts known about a person to facts known about other people. People engage in profiling all the time – and often, they do not even know they are doing it.[22] People make inferences or assumptions about people based on patterns from their own experiences. These inferences are often crude generalizations and stereotypes; they are often highly inaccurate, limited to a person's individual range of experiences (which is typically small), and not subjected to considerable reflection, testing, and improvement.[23]

Beyond this informal profiling, a more professional form of profiling occurs when people develop profiles based upon more systematic study.[24] Police investigators, for example, may construct a profile of a serial killer based on commonalities with other serial killers. Profiles range in their sophistication; the best ones are typically based on more extensive study and more data.

In the information age, inferences are made based on vast quantities of data. For example, insurance companies possess a massive trove of data about people's life expectancy based on certain characteristics, health conditions, and behaviors. The proliferation of data has occurred in several dimensions, involving more people and more details about them.

Modern algorithms have increased the scale, speed, and sophistication of profiling. An "algorithm" is a procedure to solve a mathematical problem and to generate a particular output.[25] An "algorithmic" prediction uses probabilities and statistics in making a forecast. We thus speak of "algorithms" quite broadly – they need not be limited to sophisticated computer algorithms. Even if calculated by a human, when probabilities based on statistical data are used to make forecasts, these predictions are also "algorithmic."

Of course, our greatest concern involves modern machine learning algorithms, currently one of the most prevalent types of algorithms being used in AI. Machine learning algorithms operate at a staggering level of complexity, far beyond human capabilities.[26] They are able to identify patterns in vast quantities of data, far more than a human mind can process. These patterns can be unexpected, ones that humans might never have been

or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her." GDPR art. 4(4), Recital 71.

[22] Mireille Hildebrandt, *Defining Profiling: A New Type of Knowledge?, in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES 17, 25 (Mirielle Hildebrandt & Seth Gutwirth eds. 2008).

[23] FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 6, 96-97 (2003) "[M]odern research has shown that actuarial assessments turn out to be more reliable than clinical ones.").

[24] *See* Hildebrandt, *Profiling, supra* note X, at 23-24.

[25] According to Merriam-Webster's dictionary, an algorithm is "a procedure for solving a mathematical problem (as of finding the greatest common divisor) in a finite number of steps that frequently involves repetition of an operation" https://www.merriam-webster.com/dictionary/algorithm.

[26] For more detail about how machine learning algorithms work, see TOM M. MITCHELL, MACHINE LEARNING (1997); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. Davis L. Rev. 635 (2017).

able to discover on their own. As more data is fed into these algorithms, they learn more and hone their ability to find patterns and make inferences.

Traditional profiling involved humans figuring out the patterns in data and choosing the relevant data points to construct a profile from which to make inferences. But with machine learning algorithms, the profiling is done by the machines. Even when humans are in the loop, the machines are playing a larger role. Algorithms are becoming more complex and evolving on their own.

Nevertheless, humans are almost always involved in the process.[27] Humans often determine risk classification tiers and the significance of various scores that predictive models generate.[28] Humans use algorithmic predictions to make decisions about people. Algorithmic predictions are thus a tool used by humans (often in large organizations) to achieve their aims. There are humans behind every algorithmic prediction, much like the Wizard of Oz was a man operating a machine.

This Article is focused not on all predictions, but on algorithmic predictions about *people.* Such predictions can involve things that will happen to people as well as things that people might do. Although our concerns apply to both types of algorithmic predictions about people, our concerns are at their zenith when predictions are made about future human choices and behavior.

B. PAST OR PRESENT VS. FUTURE

Not all inferences and profiling are the same.[29] A key dimension involves time. Is the matter asserted in the inference about the past or present? Or is it about a probable but uncertain future? It is essential to appreciate the temporal dimension of algorithmic inferences, as predictions about the future are especially speculative, unverifiable, and impactful.

Consider an inference about the past, such as an inference about who was the culprit in a murder mystery. This inference can be evaluated by established methods of proof and evidence. The legal system has developed elaborate rules and procedures for adjudicating the past. Similarly, an inference about the present – such as a person's religion or political affiliation – can be verified. But algorithms that make predictions about what a person will do or what will happen to a person in the future are far more speculative. There is no comparable legal architecture for litigating the

[27] KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE 53-87 (2021) (describing the extensive human labor involved with AI); Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem,* 70 Hastings L.J. 1389, 1400 (2019); Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 Vand. L. Rev. 429, 443 (2023) ("It's humans all the way down [with AI].").

[28] For example, a classification threshold (also called a decision threshold), which sets the range of scores that will be classified as "high risk group" or "users interested in buying a bicycle," is determined by humans. E.g., Google, *Classification: Thresholding, Machine Learning*, https://developers.google.com/machine-learning/crash-course/classification/thresholding (last visited 29th April 2023).

[29] Matsumi, *Predictions and Privacy*, *supra* note _, at 191.

future.

Before discussing the problems, however, we note that making the temporal distinction and singling out predictions can be difficult and is not always clearcut.[30] It is theoretically easy, but it is tricky in practice because guesses about the present can be recast as guesses about the future, and vice versa.

Many predictions can be recast as statements about the present. A prediction that a person will suffer a heart attack in the future could be recast as a set of inferences about the person's present state of health and present heart disease. Certainly, facts about a person's current health such as high blood pressure, cholesterol numbers, and other data are verifiable, but this information by itself is not a prediction. Nor would an inference that a person currently has heart disease be a prediction, as this involves a conclusion about the present. A prediction involves a probabilistic conclusion about a future event (a heart attack) based on this data.

The practice that sparks our concerns in this Article involves making decisions based upon probabilistic future events. If one were to deny a person a job or a loan and claim that it is based on a current heart disease, we would characterize the decision likely to be based on a prediction – that the heart disease will likely lead to a future heart attack, stroke, or other adverse occurrence. The prediction may be unstated, but it is the animating factor in the decision. In contrast, suppose a person is blind and is not hired to be a pilot. The decision is not based on a prediction but on a current health condition that makes the person unable to do the job in the present.

Ultimately, the boundaries of prediction can be blurry at points, but this is endemic to many categories. In writing about property rules and liability rules, for example, Guido Calabresi and Douglas Melamed noted that "[t]he categories are not, of course, absolutely distinct."[31] Nearly all distinctions are imperfect, but this fact doesn't render them useless or unworkable.

Our primary concerns involve making decisions based on predicting future behaviors or conditions of individuals. In this Article, our focus is on algorithmic predictions because algorithms have taken prediction to a new level of systemization, data usage, and prevalence. The problems we discuss later on are rooted in the act of making decisions about humans based on predictions about their future behavior. The use of modern algorithms to enhance these predictions has exacerbated these problems to an alarming degree.

C. The Dramatic Rise of Algorithmic Predictions

Algorithmic predictions are the latest chapter in a long history of attempts at predicting the future. They are the product of the movement to forecast with

[30] Matsumi, *Predictions and Privacy*, *supra* note _, at 194; Matsumi, *Predictions and Data Protection*, *supra* note _at 25.

[31] Guido Calabresi & A Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 Harv. L. Rev. 1089 (1972).

probabilities and statistics, an approach that requires gathering quantifiable and standardized data. The growing availability of data and technologies to store and analyze it accelerated this movement, which in turn encouraged the collection of more data and the development of more powerful technologies, a loop that continues to self-propel itself to this day.

Today, algorithmic predictions about future human behavior, activity, and happenings abound. In some areas, such as life insurance and credit scoring, algorithmic predictions have long been made. In the past few decades, there has been a dramatic rise in the use of algorithmic predictions. In this section, we will briefly discuss how algorithmic predictions arose in several domains and how they are currently being used.

Credit Scoring. Credit scoring has long been determined by proprietary algorithms used by consumer reporting agencies. As Josh Lauer notes in his history of credit scoring, "the authors of early scoring systems were at pains to identify which variables even predicted creditworthiness."[32] The most common scoring system in the United States was devised in 1956 by Fair, Isaac & Co. (FICO), a score ranging from 300 to 850 to predict a person's likelihood of paying back debts.[33] As a Fair Isaac representative once declared in 1972:

> For hundreds of years, the lending of money has been an art form in the sense that judgments have had to be based on the intuitive consideration of qualitative information. Only in the last two decades have innovations in technology changed the money lending activity from an art from to a scientific process, which enables people to reach decisions based on quantitative data.[34]

Credit scoring has long been beset with problems, such as lack of transparency, inaccuracy, bias, and unfairness.[35] Abuses and inadequate consumer protection prompted the U.S. Congress to pass the Fair Credit Reporting Act (FCRA) in 1970.[36]

Notwithstanding the problems with credit scoring, its use skyrocketed as well as spread to a wide array of decisions beyond credit. Credit scoring grew in part because it was fast, cheap, and consistent.[37] Today, as Oscar Gandy notes, the "use of credit scores has expanded well beyond its initial applications, finding extensive use in housing, insurance, residential services, and employment decisions."[38]

[32] JOSH LAUER, CREDITWORTHY: A HISTORY OF CONSUMER SURVEILLANCE AND FINANCIAL IDENTITY IN AMERICA 205 (2017).

[33] Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1, 9 (2014).

[34] H.J.H. Roy, quoted in LAUER, CREDITWORTHY, *supra* note X, at 213.

[35] *Id.* at 10-16.

[36] DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 67 (2004).

[37] LAUER, CREDITWORTHY, *supra* note X, at 207.

[38] OSCAR H. GANDY, JR., COMING TO TERMS WITH CHANCE: ENGAGING RATIONAL DISCRIMINATION AND CUMULATIVE DISADVANTAGE 104 (2009).

More recently, as Talia Gillis observes, credit scoring is shifting from a reliance on a "few variables" and "human discretion" to using a broader array of personal data and machine learning.[39] Lenders are using personal data about consumer behavior, social media behavior, education, and standardized test scores.[40]

Criminal Justice. Prediction in criminal justice has early roots. Bernard Harcourt observes that actuarial methods emerged in criminal justice in the early twentieth century ironically "out of a new aspiration to individualize punishment." [41] The idea was to predict whether certain rehabilitative measures would work and to determine whether inmates should be released on parole.[42] In the United States, an actuarial prediction assessment was used in the 1930s, but such predictive tools did not begin to proliferate until the 1980s.[43]

In the 1970s in the U.S., states and the federal government instituted sentencing guidelines in an attempt to make sentencing decisions more uniform and reign in judicial discretion.[44] Sentencing guidelines resulted in more standardized data, a substantial step along the path toward more automation in sentencing.

In our times, algorithmic predictions are widely used to make decisions related to incarceration – bail, probation, and parole.[45] Originating in 1998, a widely-used system is called Correctional Offender Management Profiling for Alternative Sanctions (COMPAS).[46] The algorithm calculates risk scores for general recidivism and violent recidivism.[47] Today, "most states have adopted some measure of actuarial prediction in sentencing or parole determinations"[48] Judges "routinely rely on risk assessment instruments to predict future dangerousness before deciding on release conditions."[49]As Jessica Eaglin notes, "Predictive technologies are spreading through the criminal justice system like wildfire."[50]

[39] Talia B. Gillis, *The Input Fallacy*, 106 Minn. L. Rev. 1175, 1204 (2022).

[40] *Id.* at 1206.

[41] BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 41 (2007).

[42] *Id.* at 41-45.

[43] *Id.* at 41.

[44] Jessica M. Eaglin, *Predictive Analytics' Punishment Mismatch*, 14 I/S: A J. of L. & Pol'y, 87, 100-01 (2017).

[45] Carmen Cheung, *Making Sense of the Black Box: Algorithms and* Accountability, 64 CRIM. L.Q. 539, 540 (2017).

[46] Cheung, *supra* note X, at 543.

[47] Megan T. Stevenson and Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 Wash. U. L. Rev. 681, 688-89 (2018).

[48] Ferguson, *Predictive Policing, supra* note X, at 1120.

[49] *Id.*

[50] Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 Emory L.J. 59, 61 n.1 (2017); *see also* Sandra G. Mayson, *Bias In, Bias Out*, 128 Yale L.J. 2218, 2221 (2019) ("Over the last five years, criminal justice risk assessment has spread wildly.").

One study of COMPAS in 2017 revealed that the algorithm disfavored black defendants.[51] Another study indicated that age – in particular, youthfulness – was a heavy factor in high recidivism scores.[52] Equivant, the company that created COMPAS, only provides limited information about COMPAS; the algorithm is secret.[53]

Algorithmic predictions are increasingly being used for pre-trial detention decisions, such as granting pre-trial release and setting the amount of bail.[54] These decisions focus on the likelihood a defendant will commit crimes before trial or fail to show up in court at trial. Proponents of these risk-assessment tools hail them as effective and objective, but critics raise worries about entrenched discrimination.[55]

Algorithmic predictions are being used beyond incarceration decisions. Andrew Ferguson observes that "police are adopting predictive policing strategies that promise the holy grail of policing—stopping crime before it happens."[56] Algorithms are used to predict generally where or when crime will occur.[57] Algorithmic predictions are also being made about whether specific individuals will be committing a crime.[58] There are models for "predicting individuals most likely to be involved in gun violence" and models for "identifying law enforcement officers most likely to engage in risky behavior."[59] These predictions lead to increased police surveillance of certain areas or people.[60]

In one example, the Chicago police department used an algorithmic prediction to identify people likely to be involved in violent crimes and sent them a stern warning letter.[61] The program was eventually halted, and the City of Chicago's Inspector General found the risk scores to be "unreliable."[62] The failure of this program, however, is not stopping the steady march toward using more algorithmic predictions in the criminal justice system. As

[51] Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, at 1 (2017), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[52] Stevenson & Slobogin, *Algorithmic Risk, supra* note X, at 690.

[53] *Id.* at 690.

[54] Electronic Privacy Information Center, *Liberty at Risk: Pre-Trial Risk Assessment Tools in the U.S.* 2-8 (2020), https://archive.epic.org/LibertyAtRiskReport.pdf; Ngozi Okidegbe, *Discredited Data*, 107 Cornell L. Rev. 2007 (2022); Ngozi Okidegbe, *The Democratizing Potential Of Algorithms?*, 53 Conn. L. Rev. 739 (2022).

[55] *Id.* at 1.

[56] Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. UNIV. L.R. 1109, 1112 (2017).

[57] Albert Meijer & Martijn Wessels, *Predictive Policing: Review of Benefits and Drawbacks*, 42 Int'l J. Pub. Admin. 1031, 1031 (2019).

[58] *Id.* at 1035.

[59] SARAH BRAYNE, PREDICT AND SURVEIL: DATA, DISCRETION, AND THE FUTURE OF POLICING 23 (2021).

[60] Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 Harvard L. & Pol'y Rev. 15, 15-18 (2016).

[61] Orla Lynskey, *Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing,* 15 Int'l J. L. in Context 162, 167 (2019).

[62] Elizabeth E. Joh, *Reckless Automation in Policing*, 2022 Berkeley Tech. L. J. 116, 126-27 (2022).

Michael Rich posits, the next step is the use of "Automated Suspicion Algorithms" to determine whether there is reasonable suspicion that a person is engaged in criminal activity.[63]

Employment. In the workplace, Ifeoma Ajunwa observes, there has been a growing quantification and control of workers since the ideas of Fredrick Winslow Taylor.[64] Taylor's *Principles of Scientific Management* in 1911 helped usher in a more specialized, systematic, and quantifiable method for managing employees.[65] The idea was to improve worker efficiency through the "tactical measuring of employee activity."[66] Later on, Henry Ford took Taylorism in a new more totalitarian direction, creating a "Sociological Department" that spied on his workers' private lives. His system was not primarily about predicting worker success; it was to "create model people."[67] Workers lived "continually in fear of being discharged and blacklisted for joining unions."[68]

Today, we're in the middle of another revolution in the employment context. Algorithms are increasingly being used to screen the resumes of job candidates.[69] According to a LinkedIn survey from 2018, almost two-thirds of organizations used predictive AI in the hiring process.[70] Ajunwa notes that "nearly all Global 500 companies us algorithmic tools for recruitment and hiring."[71] To predict a job candidate's likelihood of being a successful employee, algorithms scan through application materials such as resumes and cover letters; they also use other data. Some employers are using algorithms to predict whether certain employees are likely to quit.[72]

Employers are also turning to automated video interviews. Job candidates record themselves answering questions, and an algorithm analyzes their facial expressions, movements, and vocal data (such as tone of voice and speed of talking).[73]

The turn to algorithmic predictions of employee success offers some promising benefits. As Orly Lobel notes, "employers can screen for qualities

[63] Michael L. Rich *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. Pa. L. Rev. 871, 878 (2016).

[64] IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE 9 (2023).

[65] *Id.* at 20-27.

[66] *Id.* at 182. Taylor's "scientific management" was far from scientific; he as more of an ideologue than an open-minded researcher. *See* HARRY BRAVERMAN, LABOR AND MONOPOLY CAPITAL: THE DEGRADATION OF WORK IN THE TWENTIETH CENTURY 61-63 (1998) (originally published 1974).

[67] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 66.

[68] *Id.* at 66.

[69] Pauline T. Kim, *Manipulating Opportunity*, 106 Va. L. Rev. 867, 871-72 (2020).

[70] See Rebecca Heilweil, *Artificial Intelligence Will Help Determine if You Get Your Next Job*, Vox (Dec. 12, 2019), https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen.

[71] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 76.

[72] *Id.*

[73] *Id.* at 139.

that go beyond what applicants have put on paper—beyond the dry facts of their lives—and into assessments of cognitive ability, social skills, work ethic, drive, passion, ethics, and resilience."[74] Ajunwa notes that "the algorithmic turn to hiring" is motivated in part by a desire to achieve fairness and diversity.[75] But despite these good intentions, bad results have occurred, entrenching bias rather than eliminating it.[76]

Education. In education, for decades, data has been gathered and used to assess and make decisions about student academic performance. For a long time, educational institutions have engaged in "ability grouping," which involves categorizing students by their academic ability in various subjects and educating them differently.[77] This practice long has been debated, with supporters claiming it enhances educational effectiveness and critics contending it widens achievement gaps, leads to inequality of resources, negatively affects self-esteem, has discriminatory effects on poor and minority students, and is not effective.[78]

In the U.S., the No Child Left Behind Act of 2002 greatly accelerated the turn to quantifiable data in K-12 education.[79] The Act focused rather obsessively on standardized test scores and measurable metrics of student performance, leading many schools to "teach to the test."[80]

As has been the case in many other domains, algorithmic predictions are increasingly being used throughout the educational system. For example, algorithmic predictions are being made to identify students at risk for dropping out. Widely used in the U.S., early warning systems make algorithmic predictions with the goal of improving graduation rates.[81] By 2015, more than half of public high schools were using them.[82] For example, schools in Wisconsin use Dropout EWS (DEWS) to predict how likely each sixth through ninth grade student is to graduate from high school on time.[83] DEWS uses past data, "such as students' test scores, disciplinary records, free or reduced lunch-price status, and race."[84] The system spits out a score

[74] ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHER, MORE INCLUSIVE FUTURE 71 (2022).

[75] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 77.

[76] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 77-101.

[77] Yoni H. Carmel and Tammy H. Ben-Shahar, *Reshaping Ability Grouping Through Big Data*, 20 Vand. J. Ent. & Tech. L. 87, 94-95 (2020).

[78] *Id.* at 96-103.

[79] Pub. L. 107–110 (text) (PDF), 115 Stat. 142 (Jan. 8, 2002).

[80] LaTefy Schoen & Lance D. Fusarelli, *Innovation, NCLB, and the Fear Factor: The Challenge of Leading 21st-Century Schools in an Era of Accountability*, 22 Educational Policy 181, 190 (2008).

[81] U.S. De't of Educ., *Issue Brief: Early Warning Systems* (Sept. 2016), https://www2.ed.gov/rschstat/eval/high-school/early-warning-systems-brief.pdf.

[82] *Id.* at 2.

[83] Todd Feathers, *False Alarm: How Wisconsin Uses Race and Income to Label Students "High Risk"* – The Markup, (2023), https://themarkup.org/machine-learning/2023/04/27/false-alarm-how-wisconsin-uses-race-and-income-to-label-students-high-risk.

[84] *Id.*

between 0 to 100 ("DEWS Score"). The Department of Public Instruction translates DEWS scores into risk groups. If a student gets a DEWS score below 78.5, they are "labeled high risk of not graduating on time."[85] A similar system, called Navigate, is used in universities in United States.[86] Navigate's creator claims the system predicts students' "likelihood of academic success."[87]

These early warning systems have been praised as leading to rising graduation rates.[88] But they have also been criticized for over-targeting black and Latino students, biasing how teachers think of students, and not raising the graduation rates of students in the high-risk category.[89]

In higher education, a majority of institutions are using algorithmic predictions of likelihood to enroll for making strategic decisions to award scholarships.[90] Concerns have been raised that these algorithms are "susceptible to the possibility of biased outcomes—such as against racial minorities, women, people with disabilities, or other protected groups."[91]

In one instance, the International Baccalaureate program cancelled its exam in 2020 due to the Covid pandemic. It then used an algorithm to predict how the students would have scored on the exam. Headquartered in Switzerland and used by 170,000 students around the world each year, this two-year high school diploma program affects admissions decisions and scholarships. When the program suddenly switched to predicting grades through the algorithm, the formula and inputs weren't disclosed. As one German student said after receiving an unexpectedly low score: "I basically cannot study what I want to anywhere anymore."[92]

Insurance. Insurance companies have long used actuarial methods to predict future events, such as the likelihood of accidents or when and how a person might die. In the nineteenth century, this data merely consisted of a "mortality table" with life expectancies at each age and no other variables.[93] There was a boom in life insurance in the second half of the century, with the amount of insurance growing fivefold between 1865-1870 and the number of insurance companies tripling to 129.[94] After a massive wave of insurance

[85] *Id.*

[86] *Navigate*, https://eab.com/products/navigate/ (last visited 29-April-2023).

[87] Predictive Model Reports, https://www.documentcloud.org/documents/20494040-predictive-model-reports (last visited April 29, 2023).

[88] Emma Brown, *Can 'Early Warning Systems' Keep Children from Dropping Out of School?*, Wash. Post. (June 28, 2016).

[89] Todd Feathers, *How Wisconsin Uses Race and Income to Label Students 'High Risk',* Chalkbeat (Apr. 27, 2023).

[90] Alex Engler, *Enrollment Algorithms are Contributing to the Crisis of Higher Education,* Brookings (Sept. 14, 2021).

[91] *Id.*

[92] Tom Simonite, *Meet the Secret Algorithm That's Keeping Students Out of College,* Wired (July 10, 2020).

[93] DAN BOUK, HOW OUR DAYS BECAME NUMBERED: RISK AND THE RISE OF THE STATISTICAL INDIVIDUAL 6 (2015).

[94] *Id.* at 6-8.

company failures, with more than half failing between 1871-77, the industry began to look for more individualized and accurate ways to predict lifespan.[95] The 1880s also saw an expansion into new communities that insurers had previously neglected, such as the middle class and African Americans.[96] To grow into these new markets and take on riskier people, insurers began to charge different premiums to different risk groups.

By the early twentieth century, life insurance companies were employing doctors, who examined records and recorded a few details that were standardized and recorded on index cards.[97] The process was still rather crude – "[i]nsurers sought cheap signs of a bad risk."[98] Insurers began to use more data, such as credit reporting data, and they also began to exchange data with each other.[99] As insurers sought to expand, they offered insurance to individuals they otherwise would have rejected but at higher premiums.[100] As insurers sought to individualize risk calculations, they had to gather and analyze more data and analyze it in a more systematic way.

Insurers originally sought just to predict the future to offer more individualized premiums, but they soon wanted to control it. Their reasons were laudable; some insurance officials realized that if they made interventions to encourage better health habits, people might live longer – a benefit to both the company and the individuals.[101]

As Alberto Cevolini and Elena Esposito note, "the laws of statistics showed that in the mass and over the long run, an order could be found in large numbers, and this made it possible to separate the rational and foresighted attitude of insurance from the temerity and unreasonableness of gamers."[102] Insurance, they contend, traditionally involved pooling risk across large populations.[103] Insurance "has always oscillated between two opposing needs: on the one hand, the aggregation of all cases for compensatory purposes; on the other hand, the segmentation of the pool of policyholders on the basis of certain differences (such as gender and age) which enable more homogenous risk classes to be defined."[104] Modern algorithmic predictions can produce more individualized risk profiles, and it can "radicalize the principle of segmentation."[105]

Nevertheless, as Oscar Gandy notes, "Despite the absence of solid empirical data, insurers continue to rely on group membership, and questionable

[95] *Id.* at 24.

[96] *Id.* at 32-35.

[97] *Id.* at 44-52

[98] *Id.*at 63.

[99] *Id.*at 54-80.

[100] *Id.* at 82.

[101] *Id.*at 127-28.

[102] Alberto Cevolini and Elena Esposito, *From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance,* Big Data & Society 1 , 2 (2020).

[103] *Id.* at 2-3.

[104] *Id.* at 3.

[105] *Id.* at 3-4.

assumptions about the relationship between behaviors they believe to be indicative of the moral status and character of an individual."[106]

* * *

The stories with each of these domains have similar themes. Over time, a shift in the approach to dealing with risk and uncertainty occurred, with an embrace of probabilities and statistics. Through this approach, more data had to be gathered about larger groups of people, and it had to be standardized and quantified. There wasn't much room for qualitative and idiosyncratic details.[107]

The early approaches at prediction through calculation were simple because there were limits to the available data. For life insurers, for example, one official at a large insurance company would routinely visit cemeteries to gather data about people who died because many municipalities failed to maintain reliable death records.[108]

At least in some industries, the use of quantifiable data grew as views of prediction as a "science" replaced views of prediction as just a game of chance. Probabilities and statistics challenged old ideas of death as the unforeseeable product of chance; "actuaries and statisticians discovered regularities accompanying death."[109] This new view of death "made death not only predictable and understandable, but controllable too."[110]

The study of data led to new understandings. For example, life insurers had thought underweight people were a greater health risk (low weight was a sign of tuberculosis); they discovered that overweight people had a higher risk of mortality.[111] The turn to data was illuminating; it dispelled biases and challenged old beliefs. But other biases and beliefs were not eradicated; they were part of the data itself.

We witnessed the "taming of chance," to use Ian Hacking's phrase.[112] With probabilities and statistics, life became understood as less about luck, as more controllable. The future was seen as less opaque and uncertain.

This century, with the proliferation of personal data, the use of algorithmic predictions has increased, and no end appears to be in sight. The trend appears to be quite clear – more data, more powerful algorithms, more predictions. Connected devices, such as cars, medical devices, appliances,

[106] GANDY, COMING TO TERMS WITH CHANCE, *supra* note X, at 117.

[107] Dan L. Burk, *Algorithmic Legal Metrics*, 96 Notre Dame L. Rev. 1147, 1158 (2021) (noting that compiling data for algorithms will "necessarily strip away much of the unique formatting and context of the original source" and lead to "a radical decontextualization on the data, paring away extraneous information and meanings").

[108] BOUK, HOW OUR DAYS, *supra* note X, at 116-122.

[109] *Id.* at 125.

[110] *Id.* at 127.

[111] *Id.* at 122-23.

[112] IAN HACKING, THE TAMING OF CHANCE (1990).

and others, will increasingly generate vast data streams that will be too tantalizing for prediction-makers to resist. Of course, this is just a prediction, but it is one well on its way to coming true.

D. The Limitations of Algorithmic Predictions

Current algorithmic predictions certainly have tremendous predictive power, but we should be cautious not to become swept up in the hype. Algorithmic predictions can appear to be clairvoyant, but they are not.

First, algorithmic predictions rest on certain assumptions which are not ineluctably true. These assumptions are that (1) the past repeats itself and thus the future will be similar to the past; (2) an individual will continue to say and do things similarly as in the past; and (3) groups of individuals sharing similar characteristics or traits act similarly. The forecasted future of an algorithmic prediction is not really the future that will happen; quite to the contrary, algorithmic predictions are just a projection of a possible future from the viewpoint of the past and the present.

Second, inevitably, algorithmic predictions are never 100% accurate because the future is never 100% certain. Consider an example from 1898 which still holds important lessons for today. In that year, delegates from many urban areas gathered in the New York City to discuss a solution to one of the greatest problems facing their cities: horses and their copious poop.[113] The manure issue had become particularly acute due to rapid growth of horse populations. Delegates thought their cities would be buried under manure within a few decades. Their prediction and fears seemed quite founded and likely, but the future threw a curveball. Nobody at the conference foresaw the advent of automobiles. This invention made their predictions and discussions moot.

This story carries a key lesson for present times – we must have humility when predicting the future. Anything can happen. People can change. New inventions can alter the way people think and behave. Technology can revolutionize human capabilities. Events such as the Covid pandemic and the great plagues of the past can fundamentally transform the fabric of society and dramatically reshape social norms.

We should be cautious about claims regarding the accuracy of algorithmic predictions. Research is demonstrating that many algorithmic predictions are turning out not to be quite unreliable.[114] Critics have conducted studies to conclude that algorithmic predictions fail quite spectacularly to be

[113] The First Global Urban Planning Conference Was Mostly About Manure - Atlas Obscura, https://www.atlasobscura.com/articles/the-first-global-urban-planning-conference-was-mostly-about-manure.

[114] Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 Computer Law & Security Rev. 1, 1 (2022) (concluding based on an overview of studies on inaccuracies in algorithms that the "algorithms used in practice tend to be rife with errors and biases, leading to decisions that are based on incorrect information and that exacerbate inequities.").

accurate and will likely never achieve an acceptable degree of accuracy.[115] In one study, 457 artificial intelligence researchers collaborated to build models to predict six life trajectories, such as a child's GPA and whether a family would be evicted from their home, by using dataset from the Fragile Families and Child Wellbeing Study, collected by social scientists over 15 years.[116] None of the predictions turned out to be very accurate. In another study, researchers found that COMPAS, commercial software that is widely used to predict recidivism, is "no more accurate or fair than the predictions of people with little to no criminal justice expertise who responded to an online survey."[117] In the largest-ever study of mortgage data, economists found that algorithmic predictions were less accurate for minorities than they were for the majority.[118] The researchers explained the accuracy discrepancy based on bias plus the fact that lower income families tend to have less data in their credit histories.[119]

Certainly, algorithmic predictions can be superior to human predictions in some ways (though not all ways), and they can turn out to be accurate in some circumstances. But it is often impossible to determine whether algorithmic predictions are accurate in individual cases, and accuracy is only one issue with algorithmic predictions. As we discuss in the next Part, the problems with algorithmic predictions extend far beyond accuracy.

II. THE PROBLEMS WITH ALGORITHMIC PREDICTIONS

Algorithmic predictions present a set of problems that diverge from those caused by algorithmic inferences about the past and present. These problems are unique to predictions or manifest in different ways when predictions are involved. The problems are:

- *The Fossilization Problem.* Algorithmic predictions reify certain facts from the past by casting them into the future, making the past persist and harder for people to escape from the past.

- *The Unfalsifiability Problem.* Algorithmic predictions are often unverifiable because the events they are predicting haven't yet

[115] Angelina Wang, Sayash Kapoor, Solon Barocas, and Arvind Narayanan, *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy* 3 (2022), https://ssrn.com/abstract=4238015.

[116] Matthew J. Salganik et al., Measuring the predictability of life outcomes with a scientific mass collaboration, 117 Proceedings of the Nat. Academy of Sci. 8398 (2020), https://www.pnas.org/doi/10.1073/pnas.1915006117. See Karen Hao, *AI Can't Predict How A Child's Life Will Turn Out Even With A Ton of Data*, MIT Technology Review (2020), https://www.technologyreview.com/2020/04/02/998478/ai-machine-learning-social-outcome-prediction-study/.

[117] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 Science Advances eaao5580 (2018), https://www.science.org/doi/10.1126/sciadv.aao5580; *see also* Julia Angwin et al., *Machine Bias*, ProPublica (2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[118] *See id.*

[119] *See id.*

occurred. Algorithms can't be challenged for falsity, which is the law's main vehicle for allowing individuals to contest algorithms.

- *The Preemptive Intervention Problem.* When preemptive decisions or interventions are made based on future forecasting, the feedback loop to assess whether or not the forecasting was accurate dissipates, making it difficult or impossible to evaluate the accuracy of a prediction.

- *The Self-Fulfilling Prophecy Problem.* Algorithmic predictions can turn into a self-fulfilling prophecy because decisions based on them further what they predict.

In this Part, we discuss each of these problems and why they warrant algorithmic predictions to be treated differently from other types of algorithmic inferences.

The problems we discuss in this Part are not an exhaustive list of problems with algorithmic predictions, which also share problems with other algorithmic inferences. Our focus here is on problems that are unique or uniquely acute with algorithmic predictions.

A. THE FOSSILIZATION PROBLEM

Algorithmic predictions lead to what we call the "fossilization problem" – they can reify the past and make it dictate the future. The fossilization problem is about treating individuals unequally based on their past behaviors or conditions with an assumption that their past is likely to repeat in the future.[120]

Consider a case where a teenager was facing familial, financial, or any other difficulties, and didn't perform well in her studies. If these difficulties are no longer present, then her past performance might not reflect her future academic performance. Algorithmic predictions often weigh prior data too heavily. There are certainly times when history repeats itself, but there are also many times when it doesn't.

An oft-discussed example of fossilization is Amazon's attempt from 2014 to 2019 to use an AI hiring algorithm.[121] Although the algorithm was established in part to eliminate bias, it demonstrated consistent bias against women job candidates. These results were so troubling that the algorithm was put to rest in 2019. The reason why the algorithm exhibited bias was that it was trained to look for resumes similar to existing successful employees. Because existing employees were disproportionately male, the algorithm favored males. By using past data, the algorithm fossilized the bias in the

[120] Mireille Hildebrandt terms this problem "freezing the future and scaling the past." Mireille Hildebrandt, *Code-driven Law: Freezing the Future and Scaling the Past, in* IS LAW COMPUTABLE? CRITICAL PERSPECTIVES ON LAW AND ARTIFICIAL INTELLIGENCE (Simon Deakin and Christopher Markou, eds. 2020).

[121] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 83-84.

data rather than eliminated it.

Because algorithmic predictions are backward-looking rather than forward-looking, they make decisions about the future based upon data from the past.[122] In this way, algorithmic predictions can create a world akin to Greek tragedy, a Sophoclean world where all is fated.

Algorithmic predictions often assume a static version of human nature. Although people are often creatures of habit, they also change and evolve. Philosopher John Dewey aptly stated that a person is not "something complete, perfect, finished" but is "something moving, changing, discrete, and above all initiating instead of final."[123]

There is a value in not tethering people to their past.[124] A fundamental dimension of freedom is the preservation of free will. People need space to change and grow – even if they fail to do so. A world without such space is a constrained world, where a person could become a "prisoner of [their] recorded past."[125]

Of course, there are times where it is appropriate—even desirable—to look to people's past to make decisions about the future. But the choice of when and how to do so involves ethical considerations that decisions involving algorithmic predictions fail to incorporate.

Algorithms are not adept at handling unexpected human swerves. For an algorithm, such swerves are noise to be minimized. But swerves are what make humanity different from machines.

The reified past is not just a particular individual's past. Algorithmic predictions involve facts based on the past of many people. Algorithmic predictions shackle people not just to their own past but also to the past of others. For example, the effects of discrimination and bias can dramatically affect finances, health, education, and careers of entire population groups. Data from the past can be tainted from these effects. When used to predict the future, such data further entrenches these effects and perpetuates them into the future. Discrimination and bias are so marbled throughout past data that they cannot readily be extricated.

[122] Serge Gutwirth and Paul De Hert, *Regulating Profiling in a Democratic State, in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES 271, 289 (Mirielle Hildebrandt & Seth Gutwirth eds. 2008) (noting that profiling involves "identification of patterns in the past" to generate probabilistic knowledge about the present and future). "Profiles are patterns obtained from a probabilistic analysis of data; they do not describe reality. Taken to a more abstract level, profiling leads to the identification of patterns in the past, which can develop into a very useful and valuable probabilistic knowledge about non-humans, individuals and groups of humans in the present and in the future." (289)

[123] JOHN DEWEY, EXPERIENCE AND NATURE 167 (Jo Ann Boydston ed. 1987) (originally published in 1925).

[124] DANIEL J. SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR, AND PRIVACY ON THE INTERNET 72-73 (2007).

[125] U.S. DEP'T OF HEALTH, EDUC. & WELFARE, PUB NO. (05) 73–94, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS 112 (1973).

The negative constraining effects of algorithmic predictions are experienced most harmfully by marginalized groups. As Sandra Mayson notes: "Given the nature of prediction, a racially unequal past will necessarily produce racially unequal outputs."[126] Anupam Chander argues that "[a]lgorithms trained or operated on a real-world data set that necessarily reflects existing discrimination may well replicate that discrimination."[127]

Inequality hangs over the past like fog. It lingers for generations. It affects every facet of people's lives, and it seeps into the data about them. The data being guzzled by algorithms of prediction is tainted and sour; it is not fresh. Attempts to cleanse away the effects of historical inequality are thwarted when algorithmic predictions use the dirty data of the past.[128]

Fossilization also works in the opposite direction. It entrenches privilege. Advantages become etched into the future because they exist in the data. With fossilization, the losers keep losing and the winners keep winning.

The fossilization problem doesn't just occur with discrimination and inequality. The problem exists more broadly. As long as algorithmic predictions focus on the past, they perpetuate the status quo. Algorithmic predictions answer the question of "what will happen in the future?' by asking "what happened in the past?" In this way, the past will always cast a shadow on the future.

Although copious quantities of past data fuel algorithmic predictions, the data is a distorted picture of the past. As Oscar Gandy astutely observes, the collection of data is not a neutral process. Data is collected by powerful entities based on their aims and prejudices. Gandy argues: "The exercise of power is seen in the ways that asking some questions, rather than others, will be reflected in the kinds of data that become easier to acquire. We can't ignore the ways in which historical factors have led us to include race, and racial proxies in predictive and explanatory models even where their inclusion made little sense at all."[129]

The data being used by predictive algorithms is not selected at random. Data exists because it was collected or generated, often at the direction of humans and organizations for particular purposes. For example, Gandy notes that police departments have gathered copious data about traffic control and accidents "to satisfy the requirements of insurance companies" but "rarely collected the kinds of data that would support an analysis of racial bias in traffic stops, searches, and arrests."[130] Data is collected to reveal, and it is sometimes not collected to conceal.

[126] Sandra G. Mayson, *Bias In, Bias Out*, 128 Yale L.J. 2218, 2224 (2019).

[127] Anupam Chander, *The Racist Algorithm?*, 115 Mich. L. Rev. 1023, 1036 (2017).

[128] *See* Pauline T. Kim, *Manipulating Opportunity,* 106 Va. L. Rev. 867, 870 (2020) (predictive "systems are likely to distribute information about future opportunities in ways that reflect existing inequalities and may reinforce historical patterns of disadvantage.").

[129] GANDY, CHANCE, *supra* note X, at 63.; *see also* Burk, *Algorithmic Legal Metrics*, *supra* note X, at 1162 ("Numerical transformations are always value-laden, are never deterministic in any objective sense, and always depend upon human judgment.").

[130] GANDY, CHANCE, *supra* note X, at 62.

Predictive algorithms do not just entrench the past, projecting it onto the future, but they are doing so with a particular version of the past. The data that is available is often selected, not simply found, and the data that is *not* available is often the product of deliberate choices. As Elizabeth Joh contends, "Every action—or refusal to act—on the part of a police officer, and every similar decision made by a police department, is also a decision about how and whether to generate data."[131] Thus, the data that algorithms cull from the past is the product of power—and, when used by algorithmic predictions, the data perpetrates the power.

Additionally, there is skewing based on the quantifiable nature of the data being used for algorithmic predictions. Algorithms ignore more qualitative data that isn't reducible to standardized categories. W.H. Auden's poem, *The Unknown Citizen,* written in 1940, brilliantly captures how reductive and incomplete various data points about a person can be. The poem pieces together various facts that the Bureau of Statistics has compiled about a person: he worked in a factory, paid his union dues, "was popular with his mates and liked a drink," owned a "phonograph, a radio, a car and a Frigidaire," was married with five children, and other things. The poem ends:

> Was he free? Was he happy? The question is absurd:
> Had anything been wrong, we should certainly have heard.[132]

The poem's narrator expresses incredulity that anything of significance about the person can't be inferred from the data about him. But readers quickly see the superficiality and hollowness of this depiction of the person; the data fails to capture his personality or anything meaningful about him.

Although modern algorithmic predictions can involve massive quantities of data about people, the data used by these algorithms is not all there is to know about people. Algorithmic predictions use data they can readily digest. Other data is ignored, despite the fact it can reflect key differences and unique attributes about people.[133] As Katrina Geddes aptly notes, the use of algorithmic predictions diminishes the role of "personal narratives" and displaces "embodied and experiential knowledge."[134]

And, as mentioned above, the data used by algorithms is not neutral or naturally-occurring. The "past" that algorithmic predictions rely upon is fabricated. Data exists because someone decides to collect it or build a device to record it. Algorithmic predictions use a crafted past — overtly and subtly constructed with biases, ideologies, and power.

[131] Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 Wm. & Mary Bill Rts. J. 287, 289 (2017).

[132] W.H. Auden, *The Unknown Citizen*, in W.H. AUDEN: COLLECTED SHORTER POEMS 1927-1957 (Edward Mendelson ed. 1976). The poem was written in 1940.

[133] Burk, *Algorithmic Legal Metrics*, *supra* note X, at 1158 (noting how data compiled for algorithms strips away essential details and context).

[134] Katrina Geddes, *The Death of the Legal Subject,* 25 Vand. J. Ent. & Tech. L. 1, 3 (2023).

B. THE UNFALSIFIABILITY PROBLEM

Another problem with algorithmic predictions is the "unfalsifiability problem."[135] Algorithmic predictions can't be established as true or false until some future date – and sometimes never at all.[136] Often, the only way individuals are permitted to challenge predictions is to contest their accuracy. Predictions may assert future "facts," but predictions are neither true nor false because asserted matter has not yet vested.[137]

Algorithmic predictions have great power because people and organizations believe them and rely on them. Some algorithmic predictions eventually turn out to be true, but predictions aren't yet true and may never reach a point where they become true. Predictions can also be false, but we might never know.

Predictions that are unfalsifiable can readily evade accountability. Individuals subjected to algorithmic predictions often lack any meaningful ability to challenge them. In the science fiction movie, *Gattaca,* the protagonist (Vincent Freeman) is deemed to be inferior based on his genetic makeup because he is more likely to have health problems. He is denied the ability to pursue his dream of space flight.[138] There is no way for him to prove that he is fit to do the job. Similarly, a co-worker (Irene Cassini) is restricted from space travel because she has an elevated risk of heart disease. The movie chronicles a dystopian world where people are denied opportunities based on predictions and not afforded any way to contest these denials.

Although the accuracy of predictions that will vest at a certain point can be determined after they vest, decisions are made based on them beforehand. These decisions have consequences at the time of the decision. Waiting to challenge predictions until after they vest will often be too late. Even worse, while some predictions vest in a person's lifetime, other predictions might never vest until they are dead. This type of situation resembles the kind of absurd nightmare that Franz Kafka might have imagined.[139]

C. THE PREEMPTIVE INTERVENTION PROBLEM

In many cases involving algorithmic predictions, decisions to intervene are made. These interventions make it even more difficult to assess the accuracy of a prediction. We call this difficulty the "preemptive intervention problem." Preemptive decisions or interventions circumvent the feedback loop for assessing the accuracy of predictions.

[135] Matsumi, *Rectification Rights*, *supra* note _.

[136] Daniel J. Solove, *Data Mining and the Security-Liberty Debate*, 74 U. Chi. L. Rev. 343, 359 (2008) (noting that "wrongful predictions about whether a person might engage in terrorism at some point in the future are often not ripe for litigation and review.").

[137] Matsumi, *Rectification Rights*, *supra* note _, at 32.

[138] Gattaca (1997).

[139] SOLOVE, THE DIGITAL PERSON, *supra* note X, at X (describing privacy problems with the use of data and computers as analogous to the kind of problems depicted by Franz Kafka's *The Trial).*

Consider an algorithmic system that predicts the likelihood that bus drivers will have a traffic accident.[140] Suppose a bus company decides to create a driver safety program using algorithmic predictions. If a driver's probability for an accident in the next three months exceeds a certain threshold, the driver will be sent to a driving safety training program. An algorithm predicts that Driver X is a high-risk driver, and the company sends Driver X to an extra training program. Subsequently, an accident doesn't occur. Was the prediction wrong? Or was the prediction correct and the training program effectively reduced the risk? The company's intervention makes it difficult to evaluate the accuracy of the prediction. Proponents of the algorithm will proclaim that the algorithm prevented accidents – a compelling narrative that can be hard to refute.

The preemptive intervention problem is closely related to the unfalsifiability problem – but there is an important difference. Although both problems involve the inability to evaluate a prediction's accuracy, the preemptive intervention prevents the prediction from ever vesting, and it can lead to the false narrative that the prediction was correct and that the intervention successfully prevented it from occurring. Suppose there were no intervention for Driver X. The prediction will vest in three months, and it can then be assessed as accurate or inaccurate – feedback that will help improve the predictive model.

Of course, the company could evaluate its overall accidents before and after the use of the algorithm and preemptive intervention. Or it could set up a control group where no intervention is made. These are certainly better ways to proceed, as they enable greater scrutiny and evaluation of the algorithm. Although these more scientific ways of studying the algorithm might be effective for the overall success of the driver safety program, they still do not tell us about each individual case.

Suppose that Driver X believes that the algorithm's predictions about him are incorrect. The company counters that the safety program reduced accidents by more than 60% and therefore is highly effective. "What about me?" the driver says. "The algorithm may work for some, but that doesn't prove that it works for me." The company explains: "It did work for you. You were flagged by the algorithm and trained, and that's why you didn't have an accident." The driver replies: "But how do you know I would have had an accident if you hadn't intervened?" There is no way to know, but there is little the driver can do to challenge the algorithm as it relates to him.

For the company, the fact that the algorithm may be wrong in some individual cases is a small cost that is outweighed by the overall effectiveness of the safety program. A low or even modest error rate in the algorithm is quite acceptable. As long as the algorithm is working well to reduce accidents, it is a success for the company.

But from the standpoint of individual drivers, such as Driver X, they are

[140] Matsumi, *Predictions and Privacy*, *supra* note _, at 159.

thrown under the bus (pardon the pun). Wrong predictions can adversely affect the drivers, subjecting them to unnecessary training and also tarnishing their record by indicating they are high risk drivers. In other situations, preemptive interventions could have far worse effects on individuals, such as being arrested or not being hired.

Often, it is difficult for organizations to resist taking preemptive interventions. Imagine if the company failed to respond to the algorithm's predictions, and Driver X had a crash resulting in the death of many people. Litigation would surely ensue, and the company would look quite bad if they predicted the driver was high risk yet failed to take any action.

There are thus strong incentives for organizations to make preemptive interventions to prevent a situation where they look derelict for not responding to a prediction. Making such preemptive interventions is not necessarily bad, as they can prevent terrible occurrences, but these interventions create problems for individual fairness that must be addressed.

D. THE SELF-FULFILLING PROPHECY PROBLEM

Algorithmic predictions can create a self-fulfilling prophecy effect, shaping the future as forecasted even though it may have been inaccurate.

Critics of the Dropout Early Warning System (DEWS) used by many schools in the U.S. point to findings that it is "negatively influencing how educators perceive students, particularly students of color.[141] In what is referred to as the "Pygmalion Effect," people perform better when expectations are higher.[142] In a pioneering study in 1968, Robert Rosenthal and Lenore Jacobson told teachers that certain students selected randomly were expected to be "intellectual bloomers." Based on IQ testing of the students before and after the study, the researchers found that the "intellectual bloomers" showed significant gains on the test.[143] When algorithms predict certain students will excel and others will flounder, the Pygmalion Effect can make the students more likely to perform as expected.

Scholarship allocation algorithms commonly used in higher education can cause self-fulfilling prophecy effects. One study found that the awarding of a scholarship and the amount increased the likelihood of graduation.[144] Additionally, "[s]cholarships can influence a student's attitude and level of commitment to college."[145]

[141] Todd Feathers, *False Alarm: How Wisconsin Uses Race and Income to Label Students 'High Risk',* The Markup (Apr. 27, 2023), https://themarkup.org/machine-learning/2023/04/27/false-alarm-how-wisconsin-uses-race-and-income-to-label-students-high-risk.

[142] *Why Do We Perform Better When Someone Has High Expectations of Us? The Pygmalion Effect Explained*, The Decision Lab, https://thedecisionlab.com/biases/the-pygmalion-effect.

[143] ROBERT ROSENTHAL & LENORE JACOBSON, PYGMALION IN THE CLASSROOM (1968).

[144144] Alex Engler, *Enrollment Algorithms are Contributing to the Crisis of Higher Education,* Brookings (Sept. 14, 2021).

[145] *Id.*

Consider a case where a predictive software at a university predicts which students are likely to be unsuccessful and drop out of school. Suppose the algorithm predicts that a student is in the high-risk category. A school official reaches out to her, explains the situation, and tells her that the school is willing to help. However, the student, upon learning about her "probable" future, decides that continuing on at the university is pointless for her. She had been struggling to catch up with her studies because she was also working part-time to support her mother, who has been working hard to pay her tuition. She thus decides that continuing on at the university is not worth the continued sacrifice because she is predicted to fail. The prediction becomes a self-fulfilling prophecy.

People make decisions based on all sorts of information, some of it reliable and some not. For some people, it can turn out to be fortunate to know the probable outcome in advance and make decisions accordingly. At the same time, the life stories of successful people prove that success does not come without accompanying failures and that success often defies the odds.[146] Had these people stopped trying because their forecasted future didn't look successful, their innovations and accomplishments would never have happened. Of course, people persevere even when they know that the odds are against them, but algorithmic predictions might make their perseverance seem more foolhardy and make them more inclined to quit.

Often, though, the decisions to act on algorithmic predictions aren't made by the individuals subjected to them. Decisions are made by organizations, and individuals aren't given the chance to prove the predictions wrong.

The self-fulfilling prophecy problem does not just affect individuals; it also affects groups as well as populations of people with similar characteristics. When algorithmic predictions are unleashed at a large scale, they start to have cumulative effects. In each individual case, these effects may be small, but when aggregated and played out over time, the effects add up to something much larger.

For example, predictions about people who are part of a marginalized group might indicate that they will not have promising future prospects or not be a good credit risk or have weak earning potential. When made at large scale, decisions based on these predictions turn these predictions into a self-fulfilling prophecy.[147] People from that group will be denied loans, jobs, or other opportunities that weaken their finances and otherwise constrain their ability to prosper. The data from these cases adds up and gets fed back into the algorithms, further strengthening the correlations and reinforcing the

[146] Consider, for example, life stories of Oprah Winfrey, Michael Jordan, Lionel Messi, Lady Gaga, The Beatles, Walt Disney, Steven Spielberg, Stephan King, Steve Jobs, Thomas Edison, Abraham Lincoln, and Albert Einstein.

[147] Geddes, *Legal Subject, supra* note X, at 24 ("A poor credit score can trap individuals in cycles of financial precarity that affirm the score's prediction, as where, for example, more punitive credit terms for a 'high-risk' debtor increases the debtor's risk of default."); Burk, *Algorithmic Legal Metrics*, *supra* note X, at 1163-70 (explaining how "credit scores do not merely predict default, but actually facilitate default").

predictions.

The self-fulfilling prophecy problem occurs frequently in the criminal justice context. As Oscar Gandy observes, “If police base their models or expectations on arrest data, and if racial profiling increases the racial disparity in the numbers of blacks who are stopped, searched, arrested, and convicted, then the statistics reflecting this racial disparity will serve as evidence in support of the appropriateness of the technique.” [148] Marginalized communities are disproportionately subjected to surveillance.[149] Generally, the more the watchful eye looks, the more legal infractions it will find, justifying even more surveillance—a self-perpetuating spiral.

Cathy O’Neil argues that in many cases, insufficient data is gathered about instances when algorithms make mistakes.[150] Not only does this skew the picture of the accuracy of algorithms, but also it leads to the self-fulfilling prophecy. The lie can sometimes become truth. As O’Neil puts it, algorithms can “generate their own reality.”[151]

E. Racing Towards the Prediction Society

Algorithmic predictions are problematic not only because each problem perpetuates inequalities, infringes fairness, or exacerbates existing problems, such as discrimination and disproportionate impact. They are problematic because, as a whole, they distort or deprive our power to choose and create our own future. The problems we have discussed above raise an overarching and fundamental issue: *Who controls our future?*

1. Creating the Future

As the Nobel-prize winning physicist Dennis Gabor said, “the best way to predict the future is to create it.” [152] This is what decisions based on algorithmic predictions often do.

Algorithmic predictions are actually not predictions as many think them to be – they should actually be understood as *creations*. The metaphor of predictions as peering into the future is inapt. Algorithmic predictions don’t

[148] Gandy, Chance, *supra* note X, at 124-25.

[149] Scott Skinner-Thompson, Privacy at the Margins 16 (2021); *see also* Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor 6 (2015) (“People of color, migrants, unpopular religious groups, sexual minorities, the poor, and other oppressed and exploited populations bear a much higher burden of monitoring and tracking than advantaged groups.”).

[150] Cathy O’Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy 133 (2016).

[151] *Id.* at 133.

[152] Mireille Hildebrandt, *New Animism in Policing: Re-animating the Rule of Law?*, *in* The SAGE Handbook of Global Policing 407 (Ben Bradford et al. eds., 2016), According to Quote Investigator, this quote is attributed to various individuals, including Abraham Lincoln and Ilya Prigogine. See Quote Investigator, *We Cannot Predict the Future, But We Can Invent It*, https://quoteinvestigator.com/2012/09/27/invent-the-future/ (last visited 7th May 2023).

actually show the future like a vision in a crystal ball. They are not clairvoyant. In fact, they are not really predictions of the future; they are past correlations. The prediction emerges from the users of these algorithms who assume that correlations in the past will repeat in the future. The algorithms themselves are thus not actually predicting; it is the users of the algorithms who are making predictions based on the strength of this underlying assumption.

The metaphor of prediction as a vision of the future makes predictions seem more passive and neutral than they are. *Algorithmic predictions not only forecast the future; they also create it.* When used to make decisions about people, algorithmic predictions are an exercise of power over them in an effort to control the future. Algorithmic predictions mine the past to help powerful entities make decisions in the present in order to shape the future.

Because algorithmic predictions forecast the future by projecting the past onto the future, individuals' ability to choose and create their future is impacted by the past. If they have performed poorly in the past, they will likely be forecasted to perform poorly in the future. Consequently, they are less likely to be given opportunities in the future.

Because predictions are virtually impossible to falsify, individuals have little recourse. They can either decry the predictions and suffer the consequences. Or, they can try to play the game and do actions that might influence the algorithms.[153] Instead of challenging the predictions, they might focus on trying to achieve better scores. In this way, predictions can become tyrannous; they can force people to play along.

As with surveillance, algorithmic predictions can stifle individual uniqueness, creativity, and expression. Julie Cohen argues that people's identity is formed through self-authorship as well as by society and the web of relationships people have with others.[154] Privacy is essential for flourishing; it enables "spaces for the play and the work of self-making."[155] Surveillance chills and ultimately destroys this process. Surveillance "fosters a kind of passivity" and makes it hard for people to pursue and express their differences.[156]

Algorithmic predictions work in a similar, yet distinct way. They don't inhibit uniqueness; they either ignore it or penalize it. They standardize and eliminate diversity. They close off opportunities for deviations and innovations. The algorithmically-predicted future affords no room for people to take paths less chosen. It won't allow people to defy the odds. Decisions

[153] Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 Notre Dame L. Rev. 1, 3-4 (2018) (discussing how and why individuals game algorithms and the countermeasures designers of algorithms use); Geddes, *Legal Subject, supra* note X, at 24 (discussing how individuals can "feel compelled to perform a set of alien practices" to generate data in order to make credit scoring algorithms deem them more creditworthy).

[154] Julie E. Cohen, *What Privacy Is For,* 126 Harv. L. Rev. 1904, 1910 (2013).

[155] *Id.* at 1911.

[156] JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 140, 141 -52 (2012).

based on algorithmic predictions can amount to a ruthless denial of agency and selfhood.

When people's freedom and ability to choose and create their own future is curtailed and diminished due to proliferation of predictions, organizations developing and deploying predictive software are gaining greater power over the future and are shaping it more.

When properly understood, algorithmic predictions are a very problematic exercise of power that has dramatic effects for individuals and society.[157] Of course, algorithmic predictions can also have many good uses and effects. Many times, actions taken on predictions are done for good intentions, and on many occasions, these actions actually result in good outcomes. Crime can be reduced. Accidents can be prevented. Health can be improved. School graduation rates can be raised.

Architecting the greater social good can be noble undertaking. But acting on predictions to shape the future is a profound exercise of power, and the potency of this power is often not fully appreciated. The use of algorithmic predictions can have bad side-effects; they can create problems despite the best of intentions. But society is rushing forward with algorithmic predictions. A multitude of private and public sector entities are using these predictions recklessly, without forethought. There is often little scrutiny or accountability. This is a dangerous and irresponsible way to create the future.

2. The Powerlessness of the Predicted

As a result of decisions based on algorithmic predictions, individuals' power to choose and create their future is gradually distorted and deprived. Instead, the power is transferred to entities that develop and deploy predictive software. These entities are often not intentionally and collusively trying to take away control from individuals. They have many aims, such as maximizing profit or trying to do something good such as reduce crime or accidents. But as predictions are increasingly used, people become governed by predictions.

Algorithmic predictions cannot be fully understood without looking at the entities that create and use them. These entities have particular goals and aims; they exercise power in particular ways. For many entities, their primary goal is not accuracy – as long as the predictions are somewhat better than chance, they will suffice. Nor is their primary goal to make the world a better place. In many cases, the aim is efficiency – saving time and money. Automated or partially-automated decisions are faster. In the modern world, large organizations must make a multitude of decisions about vast numbers of people. It is no surprise they turn to automation for help.

[157] *See* Mireille Hildebrandt, *Profiling and the Identity of the European Citizen, in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES 303, 308 (Mirielle Hildebrandt & Seth Gutwirth eds. 2008) (arguing that profiling is an exercise of power; the people profiled "lack the feedback regarding what happens to their data and how the knowledge inferred from them may be put to use").

For individuals, on the other hand, these goals provide little benefit. Individuals suffer the costs. It might become difficult or impossible for people to escape from a prediction, regardless of whether it is accurate or inaccurate.

Suppose an algorithm correctly identifies people who will steal from their employer 80% of the time. In the 20% of cases where it is wrong, it fails to identify a future thief half the time (false negatives) and incorrectly identifies a non-thief as a thief half the time (false positives). Its creators view it as a smashing success and unleash the algorithm upon the world. Suppose the algorithm makes predictions about 100 million people. Assuming the algorithm performs as well as it tested, it will still be wrong 20% of the time. This error rate is far from minimal or trivial, as it will affect 20 million people, and 10 million adversely, as they would be falsely designated as a future thief.

Ultimately, decisions and actions based on algorithmic predictions are often made before the prediction is determined to be true or false. Organizations might rejoice in the 80% accuracy rate and write off the 10% false positives as a small cost. They might justify such action on the fact that not relying on the algorithm will yield even worse results, as human predictions might be significantly more inaccurate.

At first glance, the above scenario appears to justify the use of the algorithm. If the algorithm can be used to make a prediction that appears to be more accurate than ordinary human judgment, why not use it? Human decisions are fraught with error, bias, cognitive limitations, and a legion of other flaws that algorithms can potentially avoid.

The more accurate algorithmic predictions appear to be, the more confidence their creators have in them. Humility with human prediction is replaced by arrogance with algorithmic prediction. Algorithmic predictions are trusted more and used more.

The unfortunate people who are false positives will find it hard to escape the negative consequences of the prediction. Rational employers will not hire anyone whom the algorithm predicts will be a thief. The odds are so high that the algorithm is right that it would be folly not to follow it. In practice, although the algorithm is right only 80% of the time, it will likely be relied upon 100% of the time.

The problem of unfalsifiability makes this situation even more terrible for individuals subjected to a wrong prediction. In many cases, the prediction will never be verified. For particular individuals, in the example above, if the person isn't hired, we will never know if she would have stolen or not.

Algorithmic predictions can be implemented at a much greater scale and ease than human predictions, increasing the number of people who lose out because of them. The more widely such predictions are used, the more the losers will suffer. The data giving rise to the correlation of future thievery might also be used by other algorithms to make other adverse predictions about these people. Meanwhile the organizations using the algorithmic predictions reap all the benefits; the costs are mostly (if not entirely) borne by the individuals who are false positives.

Moreover, decisionmakers can readily mistake an algorithm's stated probability with its accuracy. An algorithm's creators might claim a high rate of probability, making people trust the algorithm more. But just because an algorithm's predictions are claimed to be at a high likelihood doesn't mean that the claim is correct.

As the use of algorithmic predictions continues to escalate, individuals might find themselves systematically disadvantaged by these predictions and unable to escape.

Katrina Geddes makes an important related point – predictive decisions about an individual based on statistical data "no longer require the input of the underlying individual."[158] The individual previously "represented a privileged source of information about their intentions, motivations, and moral capabilities." Algorithmic predictions often exclude this information, resulting in what Geddes terms "the death of the legal subject."[159] She observes that algorithmic predictions treat "data subjects not as unique individuals, but as patterns of behavior."[160]

3. A Future of Predictions

Imagine the trajectory–predict if you will–a future in which many more decisions about our lives are made by algorithmic predictions. A world in which algorithmic predictions are legion is a terrifying dystopia.

There has been extensive discussion about "surveillance creep," where surveillance increases little by little, each change feeling small and incremental, until we find ourselves in a world like George Orwell's *Nineteen Eighty-Four.* With predictions, we are not in a creep but a sprint.

Because algorithmic predictions assume that past patterns repeat in the future, individuals are bound by the past and their ability and opportunities to create their future would be limited.

We are increasingly living in a "scored society" where people are assessed and ranked in countless dimensions of their lives.[161] Even if it is not a "social scoring system" operated by governments, various fragmented scoring systems, each with their own specific purposes, can be meshed and woven into our society, affecting more decisions and opportunities, closing in the walls around our zone of freedom.

[158] Geddes, *Legal Subject, supra* note X, at 5.

[159] *Id.*

[160] *Id.*

[161] Citron & Pasquale, *Scored Society, supra* note X, at 2-4.

The future is at stake. Our power to shape our own futures – or self-determination – is under grave threat. Opportunities will increasingly be denied based on algorithmic predictions. People will be subject to greater amounts of surveillance and scrutiny based on algorithmic predictions, leading to a self-fulfilling prophecy.

Prediction problems deprive and distort our ability and opportunities to choose and create our own future. Instead, companies and data brokers are contouring our future. As Dan Burk aptly notes, "current practices mark a shift from quantification of social statistics in order to *describe* and *predict* relationships to quantification of social relationships in order to *monitor* and *control* them."[162]

Of course, humans have always made predictions. But human predictions are not systematic. Algorithmic predictions are different – and they are increasing and threaten to become pervasive. Occasional use of predictions, whether human or algorithmic, do not pose as great of a threat as the pervasive use of algorithms. With the occasional and limited use of predictions, there are avenues for escape and large pockets of freedom. But as algorithmic predictions spread more widely, they become more enveloping and oppressive. As Carissa Véliz contends that "by making forecasts about human behavior just like we make forecasts about the weather, we are treating people like things. Part of what it means to treat a person with respect is to acknowledge their agency and ability to change themselves and their circumstances."[163]

Recall our discussion of the movie *Minority Report,* where people can be convicted based on a prediction of a crime they will commit in the future. The predictions are highly accurate—far more accurate than a criminal trial would be. Yet, even so, punishing people for crimes they haven't yet committed crosses an ethical line and is fundamentally at odds with basic concepts of fairness. It is hard to imagine scenarios where it would be ethically acceptable to punish a person for future predicted wrongdoing.

In other contexts, decisions based on predictions can be acceptable, such as admissions decisions based on predictions of a student's likelihood of success at college or employer hiring decisions based on predictions of a job candidate's likelihood of being a productive employee. Although they can be acceptable, this doesn't mean that they are free of all ethical concerns. Decisions based on predictions about people still exist in an ethical gray zone. Algorithms, however, make these situations more ethically troubling because they are more mechanical, systematic, and consistent than human predictions. The problem grows worse as more algorithmic predictions are used in more dimensions of people's lives. People will have agency in theory, but in practice, decisions about them will be made as if their agency didn't

[162] Dan L. Burk, *Algorithmic Legal Metrics*, 96 Notre Dame L. Rev. 1147, 1154 (2021).

[163] Carissa Véliz, *If AI Is Predicting Your Future, Are You Still Free?* Wired (Dec. 27, 2021); *see also* Barbara Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment,* 88 Yale L.J. 1408, 1414 (1979) ("The attempt to predict an individual's behavior seems to reduce him to a predictable object rather than treating him as an autonomous person.").

matter.

In *Discipline and Punish,* Michel Foucault argued that Jeremy Bentham's ruthlessly efficient design for a prison—called the Panopticon—is an apt metaphor for modern times.[164] The Panopticon is designed with prison cells arrayed in a circle around a central observation tower. At all times, the prisoners in the cells can be watched. The observers in the tower can be obscured so that the prisoners cannot see if they are being watched at any particular moment. This type of ubiquitous surveillance leads to the prisoners internalizing discipline and conforming; they become "docile."[165] Foucault noted that the rise of surveillance technologies was turning the entire world into a Panopticon.

Predictions create a different kind of prison, not a Panopticon but a *Predicticon*. In addition to being constantly watched, people's every action, every click, and every twitch is being recorded and analyzed for patterns and then resulting in fortune or ruin. The Panopticon chills outliers; it aims to induce them to conform. The Predicticon casts outliers out of consideration. It doesn't care if they conform. Even worse, the Predicticon usurps people's stories and writes a clichéd ending. Instead of a grand Borgesian library with an infinitude of tales, the imaginatively-stultified Predicticon allows only the same predictable stories. The Predicticon is a prison of predictions, where our future paths are closed off, where we live in a cell that continually constricts.

Even when predictions are highly accurate, it is important to preserve the possibility of change. In Fyodor Dostoevsky's *Notes from Underground* (1864)*,* the underground man laments a future day when "all human actions" will be "calculated . . . like a table of logarithms."[166] He complains that "everything will be so precisely calculated and designated that there will no longer be any actions or adventures in the world."[167] Although Dostoevsky's protagonist would surely be surprised at the sophistication of today's algorithms, they are still far from 100% accurate. Entities, however, are making decisions based on algorithmic predictions that result in people being treated as if their agency were irrelevant. There is a social value in preserving space for agency, for not surrendering to the tyranny of the predictions.

Decisions based on algorithmic predictions often penalize people based on the actions of others. For example, because many people with similar behaviors or characteristics to a person did something wrong, the algorithm might predict that the person will do the same offense. Penalizing a person on this basis can be unethical in many circumstances. As Katrina Geddes contends, "algorithmic prediction effectively punishes the underlying

[164] MICHEL FOUCAULT, DISCIPLINE AND PUNISH (originally published 1975).

[165] *Id.* at 138.

[166] FYODOR DOSTOEVSKY, NOTES FROM UNDERGROUND (Richard Pevear & Larissa Volokhonsky trans. 1993) (originally published in 1864).

[167] *Id.*

individual for membership of a statistical group."[168] People should be judged based on their own actions and choices, not those of others. Although we do not take the position that judging people in this way is inherently immoral, it raises ethical concerns. As algorithmic predictions become pervasive, the concerns grow more troubling.

* * *

The use of algorithmic predictions for matters involving humans is fraught with problems. On the surface, these predictions seem beguiling. They gleam with the promise higher accuracy and less bias than human predictions. But algorithmic predictions are, in reality, power dressed up with math. They are used not to see the future but to shape it. They draw from a particular construction of the past and aim to give it an iron grip on the future. The entities using algorithmic predictions are not predicting the future to understand it but to control it.

When entities gain control over people's future, people lose control. Organizations make decisions and choices based on predictions, and people's own decisions and choices no longer matter.

III. ALGORITHMIC PREDICTIONS AND THE LAW

The use of algorithms and the generation of inferences creates headaches for the law. Algorithmic predictions, however, cause migraines. A key reason for the law's failure is that the law has failed to single out predictions and treat them differently from other inferences. Any meaningful improvement in privacy regulation for predictions depends upon a recognition that predictions are unlike other types of inferences. The law thus far lacks the focus as well as the tools to deal with algorithmic predictions. In this Part, we discuss several ways in which the law currently struggles to address algorithmic predictions.

A. LACK OF A TEMPORAL DIMENSION

The GDPR and several other data protection and privacy laws provide special protection for automation and profiling. But the current approach of these laws falls short in addressing prediction problems because they lack a temporal dimension and do not do enough to address the problems of predicting the future.

The GDPR's approach is embodied in several privacy laws in other jurisdictions around the world. However, these rules and rights related to profiling and inferences, including right not to be subject to solely automated decision-making or profiling, are not sufficient to address prediction problems. Certainly, the GDPR does a lot to address automated decision-

[168] Geddes, *Legal Subject, supra* note X, at 31.

making—more than most other laws—but as we discuss throughout this Part, there are several limitations that prevent it from adequately addressing the problems we discussed in Part II.

Predictions fall under the definition of profiling because the GDPR defines "profiling" as "any form of automated processing of personal data . . . to evaluate certain personal aspects" of an individual, particularly "to analyse or *predict* aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."[169] But the GDPR's definition turns on automation and profiling, and it treats inferences of the past, present, and future the same. This lack of temporal dimension impacts whether and how rules on automated decision-making or profiling can be used to tackle the problems we discussed in the previous Part.

When it comes to automated decision-making and profiling, the GDPR starts by treating it with the general rules for the processing of personal and sensitive data. If the data processing is solely automated with legal or similarly significant effects for individuals, then the GDPR provides special additional protections.[170] Processing that involves human involvement (that is more than perfunctory) is not protected by these special protections[171] According to EU interpretative guidance, human involvement "must ensure that any oversight of the decision is meaningful, rather than just a token gesture" and be "carried out by someone who has the authority and competence to change the decision."[172] If a "human being reviews and takes account of other factors in making the final decision," then it is not solely automated.[173] Because general provisions apply to all profiling and automated decision-making, various data protection principles in the GDPR apply to automated decision-making or profiling so long as the output (the profile) is personal data.[174] Also, data subjects can exercise various individual rights provided for in the GDPR.

There is special protection for "solely" automated individual decision-making or profiling. Individuals have a "right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[175] However, there are exceptions, such as for contracts, when authorized by law, or with consent. Individuals have the "at least the right to obtain human intervention . . . to express his or her point of view and to contest the decision."[176] Data controllers must provide information about

[169] GDPR, art. 4(4) (emphasis added).

[170] GDPR, art. 22.

[171] Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation* 2016/679, (2018), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

[172] *Guidelines on Automated Individual Decision-Making, supra* note X, at 21.

[173] *Id.* at 20.

[174] GDPR art. 5(1)(a) - (e); art. 6(1)(a) - 6(1)(f).

[175] GDPR, art. 22(1).

[176] Id. at art 22(3).

"the existence of automated decision-making, including profiling" as well as "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[177]

These rules certainly have benefits and virtues, especially for problems which they are intended to solve. However, they are not designed to address prediction problems. For example, fossilization problem cannot be solved by simply requiring input data and output (i.e., profile) to be accurate; such a requirement can worsen fossilization.

More broadly, these provisions in the GDPR are very vaguely sketched, and it remains unclear exactly how they can apply meaningfully in practice, both in profiling involving past or present as well as predictions involving future.[178] Being provided with "meaningful information about the logic involved" may solve problems associated with profiling. But such logic is hardly of help for the prediction problems.

The GDPR views problems involving profiling as caused in part by automation and curable by inserting humans into the process or by giving individuals a voice in the process. But the GDPR focuses too myopically on automation as the problem. Automation certainly exacerbates problems and could even be said to be a cause of some problems, but this does not mean that it *is* the problem. In a superb and comprehensive analysis, Rebecca Crootof, Margot Kaminski, and Nicholson Price contend that adding a "human in the loop" does not cleanse away problematic decisions and can make them worse.[179] The GDPR (and other laws) are far too vague about the roles humans should play, how they should work "in tandem with a machine."[180]

Even with greater guidance about human involvement in algorithmic predictions, numerous empirical studies provide reason for skepticism that a human in the loop will be helpful. In a great overview and synthesis of the research, Ben Green notes that studies demonstrate that people are overly deferential to automated systems, ignore errors in such systems, and override algorithmic decisions at the wrong times.[181] Additionally, "people cannot reliably balance an algorithm's advice with other factors, as they often overrely on automated advice and place greater weight on the factors that algorithms emphasize."[182]

Over in the United States, in a different context, the involvement of a human in the loop has also been viewed as a cure for any problems with algorithmic

[177] GDPR, art. 13(2)(f) and 14(2)(g).

[178] Halefom Abraha, *Regulating Algorithmic Employment Decisions Through Data Protection Law*, European Labour L.J. 1, 10 (2023).

[179] Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 Vand. L. Rev. 429, 482 (2023).

[180] *Id.* at 437.

[181] Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 Computer Law & Security Rev. 1, 7 (2022).

[182] *Id.* at 9.

predictions. In *State v. Loomis,* a defendant challenged a trial court's use of a risk assessment algorithm to determine his sentence.[183] The defendant argued that the algorithmic prediction violated his Constitutional right to due process because he didn't receive an "individualized sentenced" based on his "unique character."[184] The Wisconsin Supreme Court rejected the defendant's due process challenge because the trial judge had the ultimate discretion to decide the defendant's sentence.[185] This holding indicates that algorithmic predictions are fine as long as a human is involved in the decision; such predictions can, in essence, be human-washed.

Many algorithmic predictions are employed to aid humans in making decisions; they are not just deployed to decide solely on their own. The involvement of a human doesn't magically cure the problems with algorithmic predictions. For human involvement to be the answer, the law must set forth exactly how humans would ameliorate the problems with algorithmic predictions in particular cases. Instead, the law just points to a human and says: "Hey, there's a human, so all is fine" even though it remains unclear what the human is to do.[186]

The *Loomis* court required some procedural protections, such as greater transparency about the use of algorithmic predictions,[187] but this thin veneer of process doesn't address the problems we discussed earlier on. As Alicia Solow-Niederman argues, the *Loomis* processes are an "algorithmic grey hole" – they provide an "appearance of legality" but are in fact empty.[188] She contends that human involvement is meaningless because "Judges cannot make rational, informed choices about how to weight the tool's input if they do not evaluate how the tool operates and then calibrate their own decision-making protocol accordingly."[189]

The California Consumer Privacy Act (CCPA), arguably one of the strongest privacy laws in the United States, tries valiantly to account for inferences, and it succeeds in some parts but also fails in others. The CCPA recognizes inferences as personal data.[190] The CCPA defines "personal information" to include "inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior,

[183] Wisconsin v. Loomis, 881 N.W.2d 749 (Wis. 2016).

[184] *Id.* at 764.

[185] *Id.* at 765-67.

[186] Reuben Binns & Michael Veale, *Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR*, 11 International Data Privacy Law 319 (2021), https://doi.org/10.1093/idpl/ipab020; SCHUFA Case.

[187] *Id.*

[188] Alicia G. Solow-Niederman, *Algorithmic Grey Holes,* 5 J. Law & Innovation 116, 124 (2023).

[189] *Id.* at 125.

[190] Jordan M. Blanke, *Protection for "Inferences Drawn:" A Comparison between the General Data Protection Rule and the California Consumer Privacy Act*, 2 GLOBAL PRIVACY L. REV. 81 (2020); Jordan Blanke, *The CCPA, 'Inferences Drawn,' and Federal Preemption*, 29 Richmond J. L. & Tech. 53 (2022).

attitudes, intelligence, abilities, and aptitudes."[191] The California Office of the Attorney General has declared that "inferences appear to be at the heart of the problems that the CCPA seeks to address."[192]

Nevertheless, the CCPA merely considers inferences to be personal data and regulates them as such. What is missing is a focus on the quality and use of inferences – these are the issues involved with predictions. Labeling inferences as personal data is certainly an important step, but it is far from enough.

The California Office of the Attorney General explained that inference is "essentially a *characteristic* deduced about a consumer (such as 'married,' 'homeowner,' 'online shopper,' or 'likely voter') that is based on other information a business has collected (such as online transactions, social network posts, or public records)."[193] The word "characteristic" does not contain temporal dimension.

The GDPR, CCPA, and countless other laws lump all inferences together and treat them largely the same, but the temporal dimension matters. Privacy law's failure to distinguish predictions from other inferences is a significant flaw and oversight.

Even if humans can be brought in to improve an algorithmic prediction, many of the problems with these predictions remain. They are still predictions even if they have the human touch. Even if a human were to override the algorithm and make a prediction based on the human's experience or hunch, the prediction would still be unfalsifiable. A human might be able to look at something else beyond the past data that the algorithm used, so there is the potential that the human could provide an escape from the fossilization problem. But what would the human look at? Most likely the human's past experience, which is just past data in a less voluminous and systematic form. Humans would likely not be able to cure the preemptive intervention or self-fulfilling prophecy problems.

The problems we discussed emerge from the practice of prediction; algorithms exacerbate these problems through their automation. Tempering algorithms with humans merely focuses on the automation dimension, but the problems with prediction still remain. Ultimately, the law should better address predictions about people, whether algorithmic or human or hybrid.

B. THE TRUE-FALSE DICHOTOMY

Privacy law is built around a true-false dichotomy about facts. Privacy law often protects against the disclosure of true facts, and it also protects against false information through defamation law, rights to correct data in records,

[191] CCPA, Cal. Civ. Code § 1798.140(v)(1)(K).

[192] California Office of the Attorney General, Opinion of Bob Bonta and Susan Duncan Lee, No. 20-303 (Mar. 10, 2022), at 13.

[193] Opinion of Rob Bonta, Attorney General, and Susan Duncan Lee, Deputy Attorney General, (March 10, 2022) (emphasis added).

and duties of data accuracy. However, privacy law struggles to handle anything that falls in between the binary poles of truth and falsity.

Predictions do not fit well in existing privacy law under true-false dichotomy approach because a matter asserted in a prediction is often not true or false. A prediction is a guess until the future state of affairs being predicted comes to pass. Certainly, some predictions can be proven false at the start if based on false data or faulty logic. But many predictions are not based on either. The truth or falsity of these predictions can only be determined in the future – if at all.

The public disclosure facts tort, which is recognized in most states, protects against the widespread disclosure of true information about a person that is highly offensive to a reasonable person.[194] A prediction, which is neither true nor false, does not fit with this tort.

Other torts address false information. Defamation law—the torts of libel and slander—protects against the dissemination of false information that causes reputational harm.[195] Similarly, the tort of false light requires false information.[196] To prevail under these torts, plaintiffs must prove falsity, but a prediction isn't false. A prediction is, to some extent, an opinion about the future. Under defamation law, there can be liability for statements of fact but generally not for statements of opinion. According to the U.S. Supreme Court, a "statement of opinion relating to matters of public concern which does not contain a provably false factual connotation will receive full constitutional protection."[197]

Ultimately, at least in the United States, plaintiffs bear the burden of proof in a defamation case and must prove falsity. Because predictions are neither true or false, plaintiffs are heading down a dead-end road.

C. The Failure of Rectification Rights

The struggle to handle predictions between binary poles of true and false is accentuated in exercising rights to rectify or correct. Countless privacy statutes provide people with rights of rectification or correction. Under the EU's GDPR, data subjects have the right to have errors in their personal data corrected"[198] and to have "incomplete personal data completed."[199] In the U.S., under the CCPA, individuals have a right to "request a business that maintains inaccurate personal information about the consumer to correct that inaccurate personal information."[200] Rights to correct are quite common in privacy laws in the U.S. and around the world.

[194] Restatement (Second) of Torts § 652D (1977).

[195] *See* Restatement (Second) of Torts § 558 (1977) (requiring a "false and defamatory statement concerning another").

[196] Restatement (Second) of Torts § 652E (1977).

[197] Milkovich v. Lorain Journal Co., 497 U.S. 1, 20 (1990).

[198] GDPR art. 16.

[199] GDPR art. 16.

[200] CCPA, 1798.106(a).

These rights, however, are ill-suited for algorithmic predictions. A right to correct enables people to correct false data in their records, but it does not provide any redress or protection against predictions – even ones that are dubious or harmful.[201]

Many privacy statutes also have duties to maintain "data quality" – that personal data be accurate, complete, and up-to-date. For example, the GDPR provides for the principle of accuracy, requiring that personal data be ”accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay."[202] But this duty turns on the truth-falsity dichotomy. Unverifiable predictions are not inaccurate. The real issue is whether predictions are just, fair, and not causing unwarranted harm. These considerations don't fit into the true-false binary.

The EU's *Guidelines on Automated Individual Decision-Making and Profiling*[203] by the Article 29 Data Protection Working Party (now the European Data Protection Board - EDPB) highlight how the right to rectification fails in prediction cases:

> Individuals may wish to challenge the accuracy of the data used and any grouping or category that has been applied to them. The rights to rectification and erasure apply to both the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person).[204]

Challenges for accuracy won't work for predictions. Regarding predictions, the *Guidelines* provide the following example:

> A local surgery's computer system places an individual into a group that is most likely to get heart disease. This 'profile' is not necessarily inaccurate even if he or she never suffers from heart disease. The profile merely states that he or she is more likely to get it. That may be factually correct as a matter of statistics.[205]

The *Guidelines* recognize that the prediction is not necessarily inaccurate even if it never becomes true because it may have been "factually correct as a matter of statistics." However, the GDPR would never accept statistical "correctness" for a past or present inference, such as age, past locations, or current medical condition. Imagine if as a matter of statistics, a profile indicates that a person presently has heart disease – an inference that is false. Even if as a matter of statistics, the person likely has heart disease,

[201] Matsumi, *Rectification Rights*, *supra* note _, at _.

[202] GDPR art. 5(1)(d).

[203] Article 29 Working Party, *Guidelines on Automated Individual Decision-Making and Profiling* [hereinafter *Profiling Guidelines*].

[204] *Profiling Guidelines*, at 17.

[205] *Id.* at 18.

what matters would be the actual truth or falsity of this inference. According to the *Guidelines,* in prediction cases, however, false inferences are considered not necessarily inaccurate if they are statistically "correct."

This inconsistency happens because there are at least two approaches on how to assess accuracy of inferences -- (1) substantive accuracy and (2) procedural accuracy -- and they are often commingled and used without being clearly distinguished.[206]

Substantive accuracy involves determining whether the prediction becomes true.[207] Suppose that an algorithm predicts there is an 80% chance that person X will watch movie Y. Person X in fact watches movie Y. Here, the prediction was substantively accurate, according to this view.

Procedural accuracy focuses on the procedure or method of making the prediction.[208] Taking the above example, the system that predicts there is an 80% chance that person X will watch movie Y is *procedurally accurate* if the method for making the prediction is valid and that the probability of 80% was accurate. This approach looks at whether the algorithm relied on accurate and sufficient data as well as generally accepted computational methods. Note, however, that the procedural accuracy approach is not fully a measure of the accuracy of a prediction. The soundness of the underlying data and method doesn't guarantee that the prediction is accurate.

Rectification thus becomes very difficult (and often impossible) in prediction cases. The *Profiling Guidelines,* however, desperately try to make it seem as though rectification rights can work. In the example of the prediction of future heart disease, the *Guidelines* note that a data subject can exercise the rectification right by presenting more data by using more advanced computer and statistical model:

> Nevertheless, the data subject has the right, taking into account the purpose of the processing, to provide a supplementary statement. In the above scenario, this could be based, for example, on a more advanced medical computer system (and statistical model) factoring in additional data and carrying out more detailed examinations than the one at the local surgery with more limited capabilities.[209]

While this approach may initially appear to be reasonable, it is problematic in several ways.[210] Data subjects are being asked to do a task far beyond their capabilities. The *Profiling Guidelines* note that data subjects could produce "a more advanced computer system (and statistical model)" than the one they are challenging. Few data subjects are able to become experts in

[206] Matsumi, *Rectification*, *supra* note _, at 34; Matsumi, *Predictions and Privacy*, *supra* note _, at 197.

[207] *Id.*

[208] Matsumi, *Predictions and Privacy*, *supra* note _, at 197.

[209] *Id.*

[210] For detailed discussion, see Matsumi, *Rectification*, *supra* note X, at 44.

mathematics and computer science, and even fewer have the time and resources to develop better alternative algorithmic models. This suggestion by the *Guidelines* does not empower data subjects; it just burdens them with a task they will likely find to be impossible.

Importantly, the problems with algorithmic predictions transcend truth and falsity. Even a highly accurate algorithm that is created with the best process can still result in the problems we discussed in the previous Part. Algorithmic predictions can't be addressed with a simplistic truth-falsity binary. The GDPR and nearly all other privacy laws lack the tools to address the problems. The *Guidelines* valiantly attempt to finesse the issue, but ultimately end up with a rather absurd position.

Despite the difficulty (and in many cases, impossibility) in trying to evaluate predictions on a true-false basis, laws still keep trying to do so. For example, the U.S. Fair Credit Reporting Act (FCRA) is built around preventing inaccuracies. The FCRA states that its purpose is to promote "fair and accurate credit reporting. Inaccurate credit reports directly impair the efficiency of the banking system, and unfair credit reporting methods undermine the public confidence which is essential to the continued functioning of the banking system."[211]

The FCRA focuses far too heavily on the truth-falsity dichotomy and has scant protections against unfair and harmful predictions. Although the ingredients to calculate and generate credit scores—such as credit history and credit information, including any "item of information contained in a consumer's file"—can be disputed by individuals, the credit score itself is not subject to such dispute.[212]

The credit score is a prediction. It is the consumer reporting agency's determination about the risk of extending credit to a person. The risk relates to a future occurrence – the likelihood a person will pay back a debt in the future. All of the data used to determine the score may be accurate, but it doesn't mean that the score is an accurate reflection of a person's credit risk – or a fair or ethical one.

D. THE LIMITATIONS OF INDIVIDUAL PRIVACY RIGHTS

Many privacy and data protection laws provide individuals with rights to transparency as well as rights to object, contest, or opt out.[213] These rights often fail to adequately address prediction problems.

Regarding automated decision-making, the GDPR takes this approach, requiring data controllers to provide information about "the existence of automated decision-making, including profiling" as well as "meaningful information about the logic involved, as well as the significance and the

[211] FCRA, 15 U.S. Code § 1681(a)(1).

[212] Matsumi, *Rectification*, *supra* note _, at 19.

[213] Margot E. Kaminski & Jennifer R. Urban, *The Right to Contest AI,* 121 Colum. L. Rev. 1957 (2021).

envisaged consequences of such processing for the data subject."[214] The onus then shifts to the data subjects to exercise their right to object to the data processing.

Providing individuals with rights is a hollow and incomplete way to protect privacy.[215] Rights place the burden on individuals to manage their own privacy, a task individuals are ill-equipped to do. With modern machine learning algorithms, it is nearly impossible for individuals to understand the particular risks to them that an algorithm might create. Knowing the "logic involved" in an algorithm is not enough – one needs to know the data that the algorithm is trained on. This data can amount to personal data on millions of people – sometimes billions of people – and for privacy reasons, it cannot be disclosed to data subjects or to the public at large.[216] As Mireille Hildebrandt argues, EU data protection law is too focused on personal data and fails to "deal with *patterns of correlated data.*"[217]

Moreover, the data that machine learning algorithms use to make predictions is not static. The algorithms constantly guzzle data and evolve. To keep up, individuals would have to monitor each algorithm constantly.

Thus, the predominant approach in privacy laws to inform individuals and leave it to them to manage their own privacy fails miserably.[218] There is no easy way to provide individuals with the information they need to understand the algorithms used to make decisions about them. The algorithms are changing based on endless streams of data being piped into them. Even the creators of the algorithms struggle to understand exactly why the algorithms make certain decisions.

Ultimately, the regulation of algorithmic predictions will fail if merely aimed at specific individuals. As Salomé Viljoen astutely observes, "individualist claims subject to individualist remedies . . . are structurally incapable of representing the interests and effects of data production's population-level aims."[219] Algorithmic predictions must be addressed at the societal level.

E. THE CHALLENGES OF ANTI-DISCRIMINATION LAW

Anti-discrimination law is another body of law that could possibly help regulate algorithmic predictions, but it also fails to account for some of the special challenges posed by algorithmic predictions mainly because prediction problems are not necessarily anti-discrimination problems.

[214] GDPR, art. 13(2)(f) and 14(2)(g).

[215] Daniel J. Solove, *The Limitations of Privacy Rights,* 98 Notre Dame L. Rev. 975 (2023).

[216] Ronald Leenes, *Reply: Addressing the Obscurity of Data Clouds, in* PROFILING THE EUROPEAN CITIZEN, *supra* note X, at 293, 300 (noting limits of transparency to protect individuals from profiling).

[217] Mireille Hildebrandt, *Profiling and the Identity of the European Citizen, in* PROFILING THE EUROPEAN CITIZEN, *supra* note X, at 303, 321.

[218] Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma,* 126 HARV. L. REV 1879 (2013).

[219] Salomé Viljoen, *A Relational Theory of Data Governance,* 131 Yale L.J. 573, 578 (2021).

In general, anti-discrimination involves unfair or prejudicial treatment of people and groups based on unalterable characteristics, such as race, gender, age, or sexual orientation. In the U.S., for example, treating individuals unequally based on race, color, national origin, religion, sex including gender identity and sexual orientation, familial status, or disability in housing is prohibited by the Fair Housing Act. Unequal treatment on the basis of race, color, sex, ethnic origin, age, or disabilities in employment violates the Civil Rights Act, the Age Discrimination in Employment Act, Pregnancy Discrimination Act, or the Americans with Disabilities Act.[220]

In the EU, the Charter of Fundamental Rights prohibits "[a]ny discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation."[221] The Employment Equality Framework Directive mandates equal treatment in employment and the Racial Equality Directive broadly restricts racial or ethnic origin discrimination.[222]

Undergirding many of the problems with algorithmic predictions are their discriminatory effects. Discrimination through the use of algorithms is particularly troublesome because of its systematic nature. As Solon Barocas and Andrew Selbst contend, automation can transform bias into a "a formalized rule" that would have systematic effects.[223]

However, anti-discrimination law struggles to address algorithmic predictions. As Talia Gillis aptly argues, anti-discrimination law often makes what she calls "the input fallacy" by focusing too much on biased inputs.[224] Trying to exclude biased input data will fail to prevent algorithms from producing biased outputs. "The input fallacy creates an algorithmic myth of colorblindness by fostering the false hope that input exclusion can create non-discriminatory algorithms." [225] As Gillis observes, "a protected characteristic can be 'known' to an algorithm even when it is formally excluded." [226] Gillis argues that the law's focus should be on anti-discriminatory outputs rather than inputs.

Another problem with discrimination is the myth that automation is neutral and free of human bias. As Ifeoma Ajunwa observes, "the human hand remains present in all automated decision-making."[227] She invokes the

[220] Title VII of the Civil Rights Act, 42 U.S.C. 2000e et seq. (1964); Age Discrimination in Employment Act, 29 U.S.C. 621-634; Pregnancy Discrimination Act of 1978 (amending the Civil Rights Act); Americans with Disabilities Act, 42 U.S.C. 12101-12213.

[221] EU Charter of Fundamental Rights art. 21.

[222] Employment Equality Framework Directive, Council Directive 2000/78/EC (Nov. 27, 2000); Racial Equality Directive, Council Directive 2000/43/EC (June 29, 2000).

[223] Barocas & Selbst, *Disparate Impact*, *supra* note X, at 682.

[224] Talia B. Gillis, *The Input Fallacy,* 106 Minn. L. Rev. 1175, 1180 (2022).

[225] *Id.* at 1180-81.

[226] *Id.* at 1183.

[227] Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 40 Cardozo L.

famous example of the Mechanical Turk, a machine in the 18th century that could play chess. The machine was a fraud – a human was inside.[228] Humans play a role in the data that algorithms are trained on.[229] As Kate Crawford argues "AI is neither *artificial* nor *intelligent.* Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications."[230] The use of algorithmic predictions can thus result in discrimination that is cloaked behind the veil of "neutral" mathematics.

Algorithmic predictions can conceal discrimination by relying on proxy data. This might not be intentional. For example, location can be used as a proxy for race, as can other types of data that correlate. Suppose that there is prior discrimination against people of a certain race that results in their being viewed as less productive in the workplace and more likely to be fired. The algorithm can be programmed to avoid using the correlation of race to determine whether the person will be a successful hire. But the algorithm might find a correlation between a person's home location and the same negative career data because a high percentage of people of that race live in a particular area. Ultimately, location then becomes a proxy that performs the same invidious predictive function as race, though now the prejudice is "algorithm-washed" to make it appear neutral.

Anti-discrimination law is based on existing categories of discrimination, but algorithmic predictions can create new categories of discrimination based on new characteristics.[231] For example, if an algorithm predicts that people who are short are less successful in their careers, it might systematically target taller people. The algorithm might be picking up on existing bias against short people, but because height is not a protected category in anti-discrimination law, the law will do little to stop the use of height in the equation. The result might be an increase in height discrimination.

Beyond legally protected characteristics and unprotected characteristics that are often connected to bias, there are other characteristics that are not connected to any bias. Algorithms can identify correlations between seemingly random innocuous characteristics and bad outcomes. The algorithm will start to systematically disfavor people with these characteristics.

The result is a new set of winners and losers. People with characteristics that correlate to bad outcomes will be disfavored. Some of these characteristics might be things that people can change, such as a prediction that people who wear pants rather than shorts in the summer are more likely to be productive workers. But people might never find out that their decision to wear pants or shorts affects them in this way, and this could lead to a new type of

Rev. 1671, 1681 (2020).

[228] *Id.* at 1705.

[229] *Id.* at 1707.

[230] KATE CRAWFORD, ATLAS OF AI 8 (2021).

[231] Tal Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 Seton Hall L. Rev. 995, 1012 (2017).

discrimination based on certain arbitrary choices.

There are other characteristics that are more troubling than the choice of clothing. Predictions based on lifestyle preferences, consumption of ideas and entertainment, and so on might disfavor people based on their beliefs or free speech. If buying more than three books by John Steinbeck correlates to a higher risk of defaulting on a loan, then the algorithm can lead to a new form of discrimination against people who enjoy reading Steinbeck's books.

Predictions based on characteristics that people can't change are also quite troubling. Suppose that there is a correlation between people with foot size and a greater risk of defaulting on a loan. Algorithmic predictions might result in such people being systematically disfavored for loans. Over time, a self-fulfilling prophecy can emerge. Such people might have greater financial struggles because of the difficulties in obtaining loans, which could strengthen the prediction that they are bad loan risks. Additionally, it could lead to new predictions about other things, such as being a greater security risk or a less desirable hire.

Existing anti-discrimination law is currently not up to the task of addressing these problems.

IV. REGULATING ALGORITHMIC PREDICTIONS

Algorithmic predictions pose several unique, vexing, and devastating problems. The law has thus far failed to grapple with these problems. Indeed, the law has failed to take even the first steps.

How should algorithmic predictions be regulated? The answer is immensely complicated. There is no one-size-fit-all answer; we can't solve it with a silver bullet. There are, however, issues the law must grapple with and goals the law should strive for, and we discuss them in this Part.

A. RECOGNIZING ALGORITHMIC PREDICTIONS

First and foremost, policymakers and thought leaders must be aware of the problems with algorithmic predictions. The law must recognize and differentiate the unique risks and challenges associated with predictions from algorithmic profiling.

Algorithmic predictions should be defined under existing data protection or privacy laws. These laws can lay down additional rules including how algorithmic predictions can be generated as well as when and for what purposes they can and can't be used.

What duties should apply to forecasters when they make predictions? When should decisions based on predictions be restricted? How should the societal effects of algorithmic predictions be addressed? The law cannot effectively grapple with these issues if predictions are lumped in with all other inferences.

B. Addressing Prediction Problems

At the broadest level, algorithmic predictions are challenging to regulate because they demand both a scientific approach and a humanities approach. The scientific approach involves scientific values such as transparency, scrutiny, and continual testing. But applying a more rigorous scientific method to algorithmic predictions isn't enough. The humanities approach calls for a broad-ranging examination and oversight of the ethical implications of algorithmic predictions. We propose the following recommendations as a starting point.

1. A Scientific Approach

Algorithmic predictions should be regulated with the scientific method, much as the Food & Drug Administration (FDA) requires pharmaceuticals to be safe and effective. Without adequate testing and scrutiny, algorithmic predictions are no better than a form of pseudo-science. The law should demand scientific rigor for algorithmic predictions.[232]

To adequately put algorithmic predictions under control of science, the law must not only focus on input data and information, but also on output data and effects as well as how outputs are generated.[233] Existing privacy laws underscore on various requirements on input date, but less so on output date and rarely on the means to generate output. Assuring accuracy of input data does not assure accuracy of output data.[234]

First, the law must lay down minimum standards to generate predictions. The higher the stakes of decisions on people's lives, the more rigorously algorithmic predictions must be scrutinized. Today, the allure of algorithmic predictions often leads to their being embraced with too much enthusiasm and too little skepticism. The story of phrenology provides useful lessons. Originally concocted by Franz Joseph Gall in the late 1700s, phrenology was the idea that the size and shape of people's skulls were correlated to intelligence and personality traits.[235] Phrenology ironically became popular after it was dismissed as "utterly destitute" by the *Edinburgh Review* in 1815.[236] In one attempt to predict criminality based on physical characteristics, Cesare Lombroso, once referred to as "father of criminology," argued that it was possible to identify criminals by looking at facial features.[237] Police used phrenology to "typify criminals and arrest them, even

[232] Andrew Tutt, An FDA for Algorithms, 69 Administrative Law Review 83 (2017).

[233] *See, e.g.,* Elizabeth M. Renieris, *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse* (2023).

[234] Gillis, *supra* note X, at 1180.

[235] Pierre Schlag, Law and Phrenology, 110 HARV. L. REV. 877, 877-86 (1997).

[236] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 143.

[237] Cesare Lombroso, *Criminal Man, According to the Classification of Cesare Lombroso* (1911) [hereinafter *Criminal Man*]. The original Italian book, entitled *L'Uomo delinquente*,

in the absence of any evidence a crime had been committed."[238] By the 1950s, the phrenology craze had burned out.[239] Despite the red flags, phrenology was pursued with zeal. It was cloaked in the vestments of "science" without being subject to the normal methods and rigors of science. This situation is especially toxic, and it is a grave danger that can still be present with modern algorithmic predictions.

Although phrenology has long been recognized as a pseudo-science, similar claims are resurfacing.[240] For example, a press release by Harrisburg University claimed that "[a] group of Harrisburg University professors and a Ph.D. student have developed automated computer facial recognition software capable of predicting whether someone is likely going to be a criminal."[241] They claimed their prediction model was "80 percent" accurate and had "no racial bias."[242] A coalition of more than 2,400 experts in a variety of fields released a public letter condemning the paper, and said "[their] claims are based on unsound scientific premises, research, and methods, which numerous studies spanning our respective disciplines have debunked over the years."[243]

The group from Harrisburg University isn't the only one.[244] In 2016, researchers from Shanghai Jiao Tong University claimed that their algorithm could predict criminality using face images.[245] Engineers refuted the paper's claims, calling this approach physiognomy.[246]

These examples demonstrate that questionable attempts to make predictions continue to occur – even old and discredited predictive approaches take on new life today. The studies discussed above were performed in the academic community, where scientists and mathematicians scrutinize each others' work. But many creators and users of algorithmic predictions exist outside the academic community and are not subject to the same culture of scientific rigor. Their algorithms are hidden, often wrapped in corporate trade secrets, and not scrutinized by independent

was published in 1876, and the first English translation was published in 1911. See Matsumi, *Predictions and Data Protection*, *supra* note _, at 6.

[238] AJUNWA, QUANTIFIED WORKER, *supra* note X, at 145.

[239] *Id.*

[240] *Id.*

[241] Harrisburg University, *HU Facial Recognition Software Predicts Criminality.* This press release is removed from Harrisburg University's website, but can be accessed at http://archive.is/N1HVe (last visited 2023-04-28).

[242] *Id.*

[243] Coalition for Critical Technology, *Abolish the #TechToPrisonPipeline* (June 22, 2020), https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16 (last visited 2023-04-28).

[244] Matsumi, *Predictions and Data Protection*, *supra* note _, at 6. See WIRED, *An Algorithm That 'Predicts' Criminality Based on a Face Sparks a Furor* (2020-06-24), https://www.wired.com/story/algorithm-predicts-criminality-based-face-sparksfuror/ (last visited 2023-04-28)

[245] Xiaolin Wu and Xi Zhang, *Automated Inference on Criminality using Face Images* (2016), https://arxiv.org/pdf/1611.04135v2.pdf.

[246] Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov, *Physiognomy's New Clothes*, https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a ().

experts.

The law must mandate a scientific approach for all algorithmic predictions before they are used to make decisions having meaningful effects on people's lives. A rare example of a law that takes this approach is the Federal Data Protection Act of Germany.[247] Under the Act, "the use of a probability value regarding specific future behaviour" of an individual "for the purpose of [scoring]" is prohibited unless it meets certain conditions, including that the calculation must be based on a "scientifically recognised mathematical statistical method."[248]

Unlike the GDPR, the burden must not be placed on individuals to prove a better alternative. The burden must be on the creators and users of the algorithmic predictions. If the accuracy of an algorithmic prediction cannot be determined by testing, then it should not be permitted to be used for any decisions of any consequence in people's lives.

Although algorithmic predictions have the potential to be better in some ways than human ones, auditing and testing is essential. As Orly Lobel argues, "[w]ith AI, there is something tangible to scrutinize, unlike the black box of the human decision-maker."[249] But these benefits depend upon actually subjecting algorithmic predictions to rigorous scrutiny. Algorithmic predictions must be viewed with humility and skepticism because they are not infallible.

To address the preemptive intervention problem as well as self-fulfilling prophecy problem, organizations should be required to conduct controlled experiments on their algorithms to demonstrate that their predictions are producing accurate results at an appropriate level for the use. The higher the stakes of the decision for individuals' lives and well-being, the higher the level of accuracy should be.

In science, controls are used in experiments to eliminate alternate explanations. For example, when a company claims its software can predict the risk of recidivism, the company should be required to conduct controlled experiments to demonstrate that their predictions meet a certain level of accuracy and are reliable.

As Frank Pasquale aptly argues, we must eradicate "black boxes" that shield algorithms from scrutiny and accountability,[250] But prediction problems

[247] Bundesdatenschutzgesetz ("BDSG") in German. The English translation is available at https://www.gesetze-im-internet.de/englisch_bdsg/.

[248] BDSG art 31(1)2 ("data used to calculate the probability value are demonstrably essential for calculating the probability of the action on the basis of a scientifically recognized mathematic-statistical procedure").

[249] ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 71 (2022).

[250] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 17-18 (2015); *see also* Tal Zarsky, *Transparent Predictions,* 2013 U. Ill. L. Rev. 1503 (2013); Devin R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law,* 31 Harv. J. L. & Tech. 1 (2017); James Grimmelmann & Daniel Westreich,

create another layer of complexity to the issue: What kind of transparency should be required for predictions? The GDPR provides for an individual right to have an explanation, but this is not the most effective way to ensure for adequate review. As Lilian Edwards and Michael Veale aptly contend, machine learning algorithms are quite difficult to explain to individuals, who are often ill-equipped to comprehend them.[251] The law should focus less on providing explanations to individuals and more on ensuring that algorithmic predictions are properly evaluated by expert regulators.[252] As Talia Gillis and Josh Simons argue, "accountability is the foundational goal" and the nature of transparency and algorithmic explanation requirements should be designed to serve this goal.[253]

There are limits to evaluating algorithms. Input data can be difficult to examine, as it is constantly being fed into the algorithm. Machine learning algorithms are constantly evolving. Moreover, although algorithmic predictions can be tested over many cases, individual cases can never be fully tested because they only occur once. Thus, focusing on particular individual cases will often not be sufficient and productive. The law must ensure that predictive algorithms are evaluated by experts in a scientifically appropriate way.[254]

No decisions of any significant import on people's lives should be made unless the algorithm can be properly evaluated. Thus, this approach would strongly reject the Wisconsin Supreme Court's decision in *Loomis* to deny the defendant information about how a sentencing algorithm worked to protect the algorithm creator Northpointe's trade secrets. [255] If the algorithm's creator won't lay bare all relevant information about how the

Incomprehensible Discrimination, 7 Cal. L. Rev. Online 164 (2017).

[251] Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 Duke L. & Tech. Rev. 18, 67 (2017). For a further critique of the right to explanation, *see* Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 Int'l Data Privacy L. 76, 97 (2017) (noting limited scope of the explanation). For a different position in the debate, *see* Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation,* 71 Int'l Data Privacy L. 233 (2017). For further explication of the GDPR's right to explanation, see Margot E. Kaminski, *The Right to Explanation Explained,* 34 Berkeley Teach. L.J. 189 (2019).

[252] *See* Devin R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 Harv. J.L. & Tech. 1, 64 (2017) (arguing that transparency alone is insufficient to ensure that algorithms are operating lawfully; there must be oversight and evaluation by competent experts); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085, 1088 (2018) ("Explanations of technical systems are necessary but not sufficient to achieve law and policy goals, most of which are concerned not with explanation for its own sake, but with ensuring that there is a way to evaluate the basis of decision-making against broader normative constraints such as antidiscrimination or due process.").

[253] Talia B. Gillis & Josh Simons, *Explanation < Justification: GDPR and the Perils of Privacy,* 2 J. L. & Innovation 71, 81, 94 (2019).

[254] Andrew Tutt contends that an expert agency is needed to evaluate the "safety and efficacy" of algorithms. Andrew Tutt, *An FDA for Algorithms,* 69 Admin. L. Rev. 83 (2017).

[255] Wisconsin v. Loomis, 881 N.W.2d 749 (Wis. 2016) Rebecca Wexler argues that trade secrets should not be privileged when used in sentencing and other criminal proceedings. *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System,* 70 Stan. L. Rev. 1343 (2017).

algorithm works, then it shouldn't be used.

A danger with subjecting algorithmic predictions to the scientific method is that such an approach could devolve into shoddy science. A weak form of review could provide algorithmic predictions with a sheen of legitimacy that is ultimately unwarranted. In its heyday, phrenology was cloaked in the vestments of science; while many phrenologists were charlatans, some attempted to pursue it scientifically.[256] Phrenologists wanted to believe; they became ensconced in intricacies and lost sight of the forest for the trees; they were too quick to accept any supporting evidence and reject conflicting evidence.[257]

Our call for a scientific approach to algorithmic predictions is thus a call for a rigorous one. We must be careful to avoid lending too much scientific legitimacy to algorithmic predictions, as many are still quite underdeveloped. Some are even crude. Moreover, the very premise that future human behavior can be predicted based on past data is dubious at best. Algorithmic predictions should be viewed with a heathy dose of skepticism.

2. A Humanities Approach

Evaluating the accuracy of algorithmic predictions with a scientific approach is necessary but not sufficient. The law must look beyond accuracy in assessing algorithmic predictions. The review of algorithmic predictions must account for all individual and social harms and risks. We thus recommend that in addition to a scientific approach to the review of algorithmic predictions, there must also be a humanities approach. This humanistic analysis cannot be secondary to the scientific one. Even a scientifically sound algorithm can cause significant harm or be misused. The humanistic analysis is just as important as the scientific one — and perhaps even more so.

In some cases, machine predictions are likely to be more accurate than human predictions. Barbara Underwood notes that studies demonstrate that humans "rely primarily on information about the case at hand, paying relatively little attention to background information about other cases."[258] Algorithmic predictions address this problem, taking into account a wide body of data. If accuracy were the only goal, predictive algorithms would be preferable, but predictive algorithms have the problematic side-effects that we discussed earlier. Accuracy can't be assessed in many cases; and in other cases, accuracy is beside the point.

Many scholars propose productive process-based measures to address the use of algorithms and AI generally – audit trails, education, testing, transparency, algorithmic impact assessments, technological tools, humans

[256] Schlag, *Phrenology, supra* note X, at 890-91.

[257] *Id.* at 889-95.

[258] Barbara Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment,* 88 Yale L.J. 1408, 1428 (1979).

in the loop, and others.[259] We agree that many of these measures are useful, but as Ari Waldman aptly notes, process-based solutions alone are insufficient. [260] Waldman argues that "algorithmic decision-making empowers engineers to make policy decisions, embedding their ingrained commitment to efficiency and their indifference to privacy and other social values in society."[261]Algorithmic predictions must be subjected to a wide-ranging substantive ethical inquiry.

For the review of algorithmic predictions, broad standards such as "unfairness" can be useful because of the multifarious and evolving issues involved with algorithmic predictions.[262] Such a standard would allow courts or regulatory agencies to develop a body of guiding cases. The Federal Trade Commission (FTC) already has a legal mandate to protect against "unfair" acts and practices under the FTC Act.[263] Section 5 bars "unfair or deceptive acts or practices in or affecting commerce."[264] An "unfair" act or practice is 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."[265]

In a case-by-case common-law style manner, the FTC has developed the meaning of its unfairness standard.[266] The FTC can readily use its authority

[259] See Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. Rev. 54, 107–28 (2019) (algorithmic impact assessments); Margot E. Kaminski, *The Right to Explanation, Explained*, 34 Berkeley Tech. L.J. 199 (2019) (right to explanation of the logic behind an algorithm); Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249, 1305-13 (2007) (audit trails, transparency); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 Ga. L. Rev. 109,113–15 (2017) (right to explanation of an algorithm's development); Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood*, 47 Soc. Stud. Sci. 216, 217 (2017) (humans in the loop); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633 (2017) (technological tools rather than transparency can ensure accountable algorithms); Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. Pa. L. Rev. Online 189, 202 (2017) (technological tools are insufficient; auditing is essential).

[260] Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 Fordham L. Rev. 613 (2019).

[261] *Id.* at 627.

[262] *See* Dennis D. Hirsch, *From Individual Control to Social Protection: New Paradigms for Privacy Law in the Age of Predictive Analytics*, 79 Md. L. Rev. 439, 447 (2020) (arguing that "the FTC should use its "unfairness authority" to draw lines between big data practices that are socially appropriate and those that are not"); Andrew D. Selbst & Solon Borocas, *Unfair Artificial Intelligence: How FTC Intervention can Overcome the Limitations of Discrimination Law,* 171 U. Pa. L. Rev. _ (forthcoming 2023) (arguing that FTC unfairness can be used to address algorithmic discrimination).

[263] CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY (2016); Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014).

[264] 15 U.S.C. § 45.

[265] 15 U.S.C. § 45(n); Federal Trade Commission, *Policy Statement on Unfairness* (1980), Appended to International Harvester Co., 104 F.T.C. 949, 1070 (1984), http://www.ftc.gov/bcp/policystmt/ad-unfair.htm.

[266] Solove & Hartzog, *FTC and the New Common Law, supra* note X. In a related, but different approach, Sandra Wachter and Brent Mittelstadt argue for a right to reasonable inferences under the GDPR. Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 Colum. Bus. L. Rev. 494 (2019).

to restrict unfair acts and practices to address the problems with predictions that we have described in this Article.

However, there is more to consider in evaluating algorithmic predictions than fairness alone. Determining fairness with regard to predictions is quite complicated as there can be competing conceptions of fairness. For example, in the context of insurance, insurers might argue that it is fair to charge different rates based on different risk groups.[267] Insurers could simply spread risks equally, offering insurance to all at the same premium. Some might argue that such an approach is unfair to people in lower-risk groups as they are paying more than the prediction indicates they ought to pay for their level of risk. Others might argue that highly-individualized premium decisions should be made so that people pay exactly what the predictions indicate for their risk level.

Both of these positions involve conceptions of what insurance should aim to do (spread risk broadly or individualize risk more granularly) as well as which of these conceptions is most fair. Fairness might demand ignoring actuarial data and charging the same rates to promote anti-discrimination or equity. It might be fair to not use genetic information since people's genetic predisposition to certain illnesses are beyond their control. But it might be fair to use other factors (such as age) to differentiate rates, even though age is a category of discrimination.

There are societal implications for how these decisions are made, such as what limits ought to be placed on individualization of rates and which factors should be used and which should be barred. Ultimately, these decisions about predictions are ones that should not be left to the unfettered discretion of companies to decide. Decisions based on algorithmic predictions can have vast rippling societal effects. With algorithmic predictions, we are not dealing with a few isolated decisions, but much more systematic decision-making that has profound effects on the opportunities of individuals as well as entire groups of people.

Algorithmic predictions do not just affect the particular individuals being subjected to them. Because algorithms reify past data about certain characteristics, they can perpetuate stereotypes and past discrimination for all individuals in certain marginalized groups. Privacy law must break from its rather individualistic focus.

The effect of algorithmic predictions on human agency must also be considered. Decisions based on algorithmic predictions might not directly interfere with a person's autonomy. A person might remain fully autonomous, without their choices being influenced or coerced. Instead, decisions based on algorithmic predictions treat people as if they lack free will. Certain decisions based on algorithmic predictions deny people opportunities based on an assumption that they are likely to make choices that they haven't yet made. This type of action based on people's predicted future choices ignores their freedom to choose. People are still free to choose,

[267] GANDY, COMING TO TERMS WITH CHANCE, *supra* note X, at 116-17.

but no matter what they choose or would have chosen, their agency is not accounted for by the decisionmakers; it is just noise to be ignored. Thus, the law must evaluate how decisions based on predictions interfere with an ethical commitment to respect agency.

C. Regulatory Challenges

Algorithmic predictions present other regulatory challenges that require considerable thought and attention.

Focal Points. The law must have the right focal point. Regulatory intervention must be prudential. Focusing on the mere existence of a prediction is too broad, as too many decisions involve predictions to regulate them all. Even focusing on the existence of an algorithmic prediction can be far too broad, as so many predictions involve probabilities. Focusing on whether a prediction is made by a computer via a machine learning algorithm is too narrow, as cruder forms of predictions based on probabilities can create problems. Focusing on whether a decision is based on algorithmic predictions are solely automated is also too narrow of a focus, as predictions made with human involvement can be problematic.

The focal point must be on the problems. Automation itself is not inherently bad or good. Adding a human in the loop doesn't make the problems with algorithmic predictions go away.

Regulation should avoid trigger points such as algorithms or automation, as this could exclude some problematic predictions. Instead, the law should directly confront the problems we discussed.

Context. We can't regulate all predictions the same, and it's impractical to regulate all predictions. People make predictions all the time. When parents hire a babysitter, they are making a prediction that the babysitter will be responsible and safe. When people choose friends, get married, or select a job, they are basing their decisions on predictions. There are simply too many predictions to regulate them all.

The law should consider the following questions: What the prediction is used for? Who is the beneficiary of the predictive system? Who is making the prediction? A government entity? A private company? A lone individual? For what purpose?

There are some situations where the law should completely bar the use of algorithmic predictions – even when highly accurate. For example, suppose that an algorithm can predict with near perfect accuracy that a person will commit a crime in the future. The fact that the prediction is accurate is far from the most important consideration. Even if the prediction would be more accurate than a trial, a decision to convict the person preemptively would undermine the longstanding value not to punish people for something they haven't done yet.

Predictions involving decisions of greater consequence should be more strictly regulated. Not all predictions warrant the same treatment under the law. The law should look to the harms and risks created by various predictions and more stringently regulate those that cause the most harm or risk.

Data Inputs and the Development of Algorithms. The data used by an algorithmic prediction creates complicated issues. Consider credit scoring. New types of data are being gathered about people for credit scoring, and the growing trove of data can pose privacy concerns. However, this new data can provide benefits. Talia Gillis notes that existing credit scoring relies on factors that "are a product of pre-existing disadvantage or discrimination." [268] In many cases, the use of "nontraditional" data to determine creditworthiness "may allow for the expansion of credit to populations that have traditionally been excluded from credit markets."[269] More data isn't necessarily better or worse, and a lot must be considered when evaluating the types of data used to make an algorithmic prediction.

Rigorous scrutiny must be applied to the way that data used for algorithmic predictions is gathered and shaped. As Kate Crawford observes, "Data and data sets are not objective; they are creations of human design."[270] Consider a prediction that a criminal will commit another crime based on recidivism data. At first glance, recidivism data appears to be rather clear, but upon closer examination, it's highly malleable and depends upon how "recidivism" is defined.[271] Recurrence of the same crime? The same category of crime? Any crime? How long after release from prison can a new offense take place to count as recidivism? Any time, even if more than 10 years afterwards? Should only convictions count for recidivism? Or arrests too? Recidivism statistics thus depend upon how recidivism is defined and a multitude of subjective judgments made by the human designers of algorithms. Moreover, the recidivism statistics can be skewed by disproportionate policing. As the vast majority of crimes are resolved by plea bargaining, the convictions might not reflect the actual crimes committed or even whether defendants are guilty or innocent since many take the plea rather than risk trial.

Beyond the data algorithms use, the development process for algorithms must be rigorously examined. As Jessica Eaglin aptly argues, "Tools constructed to estimate recidivism risk reflect numerous normative choices. There is no such thing as a 'value-free' tool."[272] Eaglin chronicles how human choices occur at multiple stages in the process. [273] For adequate accountability and review, the human reasoning and choices in the development of algorithmic prediction systems should be documented. The review of algorithmic predictions should avoid merely looking at the

[268] Gillis, *supra* note X, at 1192.

[269] Gillis, *supra* note X, at 1208.

[270] Kate Crawford, *The Hidden Biases in Big Data*, Harv. Bus. Rev. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data.

[271] Jessica M. Eaglin, Constructing Recidivism Risk, 67 Emory 59, 72 (2017)

[272] *Id.* at 88.

[273] *Id.* at 72-88.

algorithms; the provenance and creation process of the algorithms must also be scrutinized.[274]

Escape and Expiration. When a prediction is unverifiable, there must be a possibility of escape from its effects. Unverifiable algorithmic predictions place people in a kind of permanent purgatory. There is no way for people to challenge the predictions. People should have a right not to live in an unverifiable limbo forever.

Certain high-consequence algorithmic predictions should have an expiration date or be specific in terms of temporal validity. For example, a prediction that a person is likely to perpetrate a terrorist attack should not lead to the person being on a suspected flyer list or no fly list indefinitely. There must be a reasonable point at which a prediction expires.

Algorithmic predictions should be revisited when practical. Certain decisions might be impractical, such as hiring decisions or school admissions decisions, as the parties have moved on. But sentencing decisions can involve people who remain incarcerated when better methods to make the decisions might be developed. Except for certain situations involving early release or parole, the basic sentencing decisions are often made once-and-for-all. Generally, the law in the U.S. is rather poor about reopening matters even when new facts are discovered. But a robust commitment to the scientific method would demand that decisions based on discredited algorithmic predictions be revisited.

Individual Redress and Recourse vs. Actions by Regulators. When should the law provide rights to individuals to challenge algorithmic predictions? When should regulators be the ones to take on this role? Some combination of individual rights and regulatory enforcement is necessary, but the optimal mix is quite complicated to determine.

For individual redress, when should individuals be given a right to challenge an algorithmic prediction? On what basis should individuals be able to raise the challenge? Accuracy is one basis the law allows for challenges, but this is far too narrow.

At the same time, it is impractical to allow individuals to challenge any decision made by an algorithmic prediction. These decisions are made en masse at an enormous scale; allowing individuals to challenge these decisions at any time for any reason could create an endless flood of litigation. Individuals should certainly have some recourse, but this might involve a regulator reviewing the algorithm and deciding one way or the other for all similarly-situated individuals.

Regardless of whether algorithms are reviewed via individual litigation or regulatory review, when algorithms involve high-stakes decisions about

[274] Gillis & Simons, *Explanation, supra* note X, at 92 (a "technical explanation of a machine learning model" alone is insufficient because the "choices made by humans in its design and implementation" are key to understanding the model).

individuals, the burden should be on the creators and users of the algorithms to establish their accuracy and viability.

Societal Level Perspective. The law must also focus on algorithmic predictions at the societal level, just only at the individual level. Each individual algorithm could be fine, but the totality of too many algorithmic predictions could have serious implications for people's freedom. The overall extent of use of predictions matters as well, as we don't want a dystopia where decisions about nearly every facet of our lives are made through algorithmic predictions.

The law should avoid merely addressing one algorithm at a time. The problems involve the totality of algorithmic predictions and their effect on society, human flourishing, and respect for human agency. Algorithmic predictions must be addressed individually as well as collectively.

Room for Uncertainty. Algorithmic predictions purport to be a kind of ultimate weapon in the longstanding war against uncertainty. Humans have always been at the mercy of the uncertain future. Being subjected to chance and luck makes us feel powerless, vulnerable to the arbitrary whims of an indifferent universe. Probabilities and statistics offer a way out of this existential malaise, a more scientific way to cope with the future beyond hope and superstition.

With ever-improving technologies to predict, why would we ever want to go back to the brutal world of chance? Yet, chance has virtues. It brings unexpected things, some of which are wonderful and better than the status quo. We might discover our truths are wrong; we might be pleasantly surprised at learning new things. The person who seemed like a questionable hire might turn out to be the very best employee. The student who seemed like a failure might one day win the Nobel Prize.

Nassim Nicholas Taleb explains the significance of what he calls "black swans" – unpredictable occurrences that have an enormous impact. Taleb tells the story of how Europeans thought all swans were white until explorers landed in Australia, where they discovered to their astonishment that black swans existed.[275] Taleb argues that "we are demonstrably arrogant about what we think we know" and "[o]ur predictions may be good at predicting the ordinary, but not the irregular, and this is where they ultimately fail."[276] "*The inability to predict outliers,*" Taleb observes, "*implies the inability to predict the course of history.*"[277] We thus have way too much arrogance in the ability to predict using algorithms. Far from lamenting our lack of predictive powers, we should accept the surprises and the new challenges and opportunities they bring.

We must avoid falling victim to the tyranny of predictions. We must find a

[275] NASSIM NICHOLAS TALEB, THE BLACK SWAN: THE IMPACT OF THE HIGHLY IMPROBABLE xvii (2007).

[276] *Id.* at 138, 149.

[277] *Id.*at xx.

way to restrain ourselves. The complexity and mysteries of human life can't be translated solely into standardized and quantifiable data. As Mireille Hildebrandt argues, it is essential to respect the "foundational incomputability of human identity" and to "*protect what counts but cannot be counted:* the fragile but robust, indeterminate but sustainable, ecological and irreducibly subjective self."[278] The urge to predict will remain strong. We will want to use every tool we can to predict. But sometimes, like Odysseus, we must tie ourselves to the mast or wax our ears and exercise restraint. Just because we can use an algorithmic prediction doesn't mean we should. Clearly, algorithmic predictions are going to be used, and widely. But we must be humble and not let the entities zealously trying to control the future curtail the unexpected.

* * *

The issues discussed above have no easy answers. Right now, however, they are rarely being addressed by the law, and the first step is to recognize the need to address these issues.

Consider the following example of how the law addressed problems with algorithmic predictions from more than a century ago. The example comes from Dan Bouk's illuminating book on the history of life insurance, *How Our Days Became Numbered.* In the 1880s and 1890s, the life insurance industry charged African Americans higher premiums and offered lower payouts based on statistics that African Americans had a lower life expectancy.[279] Between 1887-1894, several states passed laws to ban the practice. As Bouk writes, the "insurers pressed for recognition that past and present differences necessitated discrimination. . . Antidiscrimination forces nitpicked the opposition's statistics, but mostly conceded the past to insurers. Yet they kept winning because they made a better case for the future, a case that statistics could not touch."[280]

Ironically and sadly, the insurers responded by finding ways to avoid insuring African Americans. Most insurers "stopped soliciting blacks." One large company continued to market insurance to African Americans but subjected them to "the most strenuous medical scrutiny."[281]

This example illustrates an attempt by the law to address the fossilization problem. Although the statistics showed that African Americans had a lower life expectancy than whites, the laws viewed preventing fossilization as more important than individualizing rates based on race. Although the actuarial data about life expectancy was correct, the laws were passed with an aspirational goal – if living conditions improved for blacks under a more

[278] Mireille Hildebrandt, *Privacy and Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning,* 20 Theoretical Inquiries in L. 83, 86, 96 (2019) (emphasis in the original).

[279] DAN BOUK, HOW OUR DAYS BECAME NUMBERED: RISK AND THE RISE OF THE STATISTICAL INDIVIDUAL 34-35 (2015).

[280] *Id.*at 45.

[281] *Id.*at 51.

equitable society, they'd live longer. The laws strived to base insurance decisions on how the world should and can be not on how it is.[282]

Despite their failure, the laws banning disparate treatment in insurance rates were a positive measure. The laws might have fared better had they addressed the runarounds. Although the laws failed in execution, they represent an important step – they recognized the problem of fossilization and they turned away from the law's typical focus on accuracy alone. The law intervened to shape the future rather than cede control of the future to companies and the rigid hand of past data.

Recently, the EU has been developing an AI Act.[283] The AI Act takes risk-based approach and differentiates between uses of AI that create: (1) an unacceptable risk, (2) a high risk, or (3) low or minimal risk. The Act prohibits AI systems that exploit "vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm."[284] AI systems that provide "social scoring of natural persons for general purpose by public authorities" is also prohibited as it "may lead to discriminatory outcomes and the exclusion of certain groups."[285]

Although the AI Act does not provide distinct regulation for algorithmic predictions, the AI Act's risk-based approach has promise because it is focused on problems. But a lot still remains undetermined, as the efficacy of the Act depends upon *how* the risks are identified and addressed.[286] As demonstrated by the example involving laws forbidding insurance companies from charging higher life insurance premiums based on race, laws can fail for many reasons, such as loopholes or poor enforcement. At this juncture, we can only attempt to predict the efficacy of the EU's AI Act, and we remain humble about making predictions.

We should also look beyond algorithms to all forms of prediction. We must evaluate when it is appropriate to use predictions about people's future behavior no matter how these predictions are made. Making predictions is likely unavoidable in some cases, but we should evaluate how extensively a decision should rest on a prediction. For example, consider algorithmic predictions of recidivism at sentencing. The focus is often on whether the algorithmic output is accurate or biased. But there are broader issues that also must be explored: Should future recidivism be a factor in sentencing? If so, what weight should it have in the overall sentencing determination?

[282] *Id.* at 41-53.

[283] Proposed AI Act, https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence.

[284] Proposed AI Act, art. 5(1)(b); 5.2.2 of the Explanatory Memorandum.

[285] Proposed AI Act, art 5(1)(c); Proposed AI Act, recital 17.

[286] For an excellent discussion of how various privacy laws, including the AI Act, address risk, see Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. Rev. (forthcoming 2023) draft on file with the authors.

Prediction is fraught with peril and problems. Modern algorithmic predictions are beguiling because they appear to be an advance from older cruder forms of prediction. But the algorithms do not make predictions any less problematic; in fact, they make the problems worse.

CONCLUSION

We are awash in algorithmic inferences about many facets of our lives. Algorithmic predictions warrant special attention. They are different from inferences about the past and the present. They cause a unique set of problems that the law does not adequately address.

Humans have long wanted to minimize chance and risk. The world is precarious, and life is fraught with uncertainty. The desire to predict is quite understandable. Enthusiasts for algorithmic predictions see a bright future of sophisticated and accurate forecasting. Yet, algorithmic predictions are being embraced far too quickly and without adequate attention to their problems and sufficient concern about their limitations.

As algorithmic predictions proliferate, they threaten to change society. Such predictions shift control over people's future, taking it away from individuals and giving the power to entities to dictate what people's future will be. It is essential that we turn our focus to algorithmic predictions and regulate them rigorously. Our future depends on it.