Washington and Lee Law Review

Volume 78 | Issue 2 Article 4

Spring 2021

Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments

Sonia M. Gipson Rankin
University of New Mexico School of Law, sgrankin@unm.edu

Recommended Citation
Sonia M. Gipson Rankin, *Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments*, 78 Wash. & Lee L. Rev. 647 (2021).
Available at: https://scholarlycommons.law.wlu.edu/wlulr/vol78/iss2/4

Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments

Sonia M. Gipson Rankin*

Abstract

Issues of racial inequality and violence are front and center today, as are issues surrounding artificial intelligence ("AI"). This Article, written by a law professor who is also a computer scientist, takes a deep dive into understanding how and why hacked and rogue AI creates unlawful and unfair outcomes, particularly for persons of color.

Black Americans are disproportionally featured in criminal justice, and their stories are obfuscated. The seemingly endless back-to-back murders of George Floyd, Breonna Taylor, Ahmaud Arbery, and heartbreakingly countless others have finally shaken the United States from its slumbering journey towards intentional criminal justice reform. Myths about Black crime and criminals are embedded in the data collected by AI and do not tell the truth about race and crime. However, the number of

* Assistant Professor of Law, University of New Mexico School of Law. J.D. 2002, University of Illinois College of Law; B.S. Computer Science 1998, Morgan State University. I would like to thank my husband, Eric Rankin, for introducing this issue to me. Thank you to my colleagues, Nathalie Martin, Alfred Mathewson, Ernest Longa, Joshua Kastenberg, Marsha K. Hardeman, DeForest Soaries, David Gipson, and Naomi Rankin, who provided comments on this paper. I am grateful to my wonderful research assistants, Lauren Jones, Marcos Santamaria, Marissa Basler, and Deanna Warren for their sharp insight and the Interdisciplinary Working Group for Algorithmic Justice for their support. This paper is in honor of my children, Naomi, Sarai, and Isaac, and is dedicated to the memory of George Stinney Jr. Say Their Names.

Black people harmed by hacked and rogue AI will dwarf all historical records, and the gravity of harm is incomprehensible.

The lack of technical transparency and legal accountability leaves wrongfully convicted defendants without legal remedies if they are unlawfully detained based on a cyberattack, faulty or hacked data, or rogue AI. Scholars and engineers acknowledge that the artificial intelligence that is giving recommendations to law enforcement, prosecutors, judges, and parole boards lacks the common sense of an eighteen-month-old child. This Article reviews the ways AI is used in the legal system and the courts' response to this use. It outlines the design schemes of proprietary risk assessment instruments used in the criminal justice system, outlines potential legal theories for victims, and provides recommendations for legal and technical remedies to victims of hacked data in criminal justice risk assessment instruments. It concludes that, with proper oversight, AI can increase fairness in the criminal justice system, but without this oversight, AI-based products will further exacerbate the extinguishment of liberty interests enshrined in the Constitution.

According to anti-lynching advocate, Ida B. Wells-Barnett, "The way to right wrongs is to turn the light of truth upon them." Thus, transparency is vital to safeguarding equity through AI design and must be the first step. The Article seeks ways to provide that transparency, for the benefit of all America, but particularly persons of color who are far more likely to be impacted by AI deficiencies. It also suggests legal reforms that will help plaintiffs recover when AI goes rogue.

Table of Contents

INTRODUCTION 650

I. ARTIFICIAL INTELLIGENCE CAN BE HACKED AND WEAPONIZED...654
   A. *What Is Artificial Intelligence and How Is It Used?*...654
   B. *How Data Is Gathered, Analyzed, and Utilized*..661
   C. *Public Fears and the Reality of Concerns Related to AI*...666

1. Cyberattacks 668
2. Untrustworthy Data 675
3. Disparate Outcomes 678

D. *Prevention Efforts by Scientists Are Insufficient* 681

II. RISK ASSESSMENT INSTRUMENTS IN CRIMINAL JUSTICE ARE NOT SECURE 685

III. LEGAL OPTIONS FOR PEOPLE HARMED BY UNTRUSTWORTHY AI IN CRIMINAL JUSTICE 692

A. *Impact of Data Errors Outside of the Criminal Justice System* 693

B. *Nefarious Actors Provide False Scientific Evidence in Criminal Cases, Infringing on Civil Liberties* 699

C. *Theories of Accountability and Liability from Case Law for Victims of Potentially Untrustworthy AI Used in Criminal Cases* 701

1. Legal Response to Data Requests in Criminal and Civil Cases 702
   - *a.* State v. Loomis 702
   - *b.* Rodgers v. Laura & John Arnold Foundation 704
2. Legal Remedies for Wrongfully Convicted Criminal Defendants Impacted by Hacked Data 706
3. Legal Remedies for Third Parties Injured Due to Flawed Data or Algorithms 710

IV. RECOMMENDATIONS FOR ADDRESSING UNTRUSTWORTHY AI 712

A. *Hacking Will Have a Disproportionate Impact on Black Communities in the Criminal Justice System* 713

B. *There Must Be Stronger Oversight of AI Used in Criminal Justice* 716

C. *Legal Remedies Are Needed for Parties Harmed by Data Hacks in Criminal Justice Risk Assessment Instruments* 721

CONCLUSION 723

INTRODUCTION

In Jordan Peele's 2019 widely praised film, *Us*, protagonist Adelaide meets her doppelgänger, who then proceeds to terrorize Adelaide's family.[1] This look-alike is described as her "Tether," the product of a failed and disbanded United States government experiment, deemed harmless and left to remain below ground.[2] The "Tethered" escape from their confinement intent on replacing the above-ground community.[3] This film captures a growing tension in science and society of artificial intelligence ("AI") today: what happens when AI does not operate as intended? Having AI serve as the foundation of risk assessment instruments, particularly in criminal justice allows instances of benign neglect, nefarious actors, and unintended consequences to change the outcomes in ways that harm society.[4] In other words, AI is as likely to contribute to racism in the law as it is a means to end it.

The courts have yet to address legal issues related to easily hackable AI. Further, the current remedies in place do not offer sufficient recourse for either wrongfully incarcerated defendants or harmed third parties. The criminal justice flowchart is well established. A defendant is arrested for a crime, granted or denied bail, convicted, and sentenced, sometimes with a possibility of being eligible for parole. AI has been used to augment every stage of the criminal justice

[1] US (Universal Pictures 2019); *see* Tasha Robinson, *Jordan Peele's* Us *Turns a Political Statement into Unnerving Horror*, VERGE (Mar. 22, 2019, 10:47 AM), https://perma.cc/TK7F-RA6E.

[2] *See* Tasha Robinson, *Does the Ending of Jordan Peele's* Us *Play Fair With the Audience?*, VERGE (Mar. 25, 2019, 3:32 PM), https://perma.cc/ALM7-5K8M.

[3] *See id.* This theme is also explored in the highly regarded Netflix series, *Stranger Things*. The "upside-down" reflects a distorted version of a small community where the corrupted inhabitants of the mirrored society cross blurred boundaries. *See* Ashley Strickland, *The Weird Upside Down Science Behind '*Stranger Things*'*, CNN (July 4, 2019, 9:55 AM), https://perma.cc/68F6-TAXJ.

[4] *See* United States v. Curry, 965 F.3d 313, 344–46 (4th Cir. 2020) (Thacker, J., concurring) (emphasizing the potential for harm from predictive policing algorithms that may use data that reflects and reinforces racial biases).

decision-making process.[5] Law enforcement uses facial recognition drones that report a high probability that the defendant has committed a crime.[6] Using a pretrial risk assessment instrument, the prosecutor recommends to the court that the defendant not be eligible for bail because the software identifies the defendant as a flight risk or a danger to the community.[7] The judge sentences the defendant after consulting a risk assessment instrument that gives a recommended sentence.[8] The parole board consults another risk assessment instrument that computes its determination of the defendant's release, parole, or probation.[9] Without question, these issues disproportionately affect Black people, Indigenous people, and other communities of color.

Black Americans are disproportionally featured in criminal justice, and their stories are obfuscated. The seemingly endless

[5] From surveillance to pretrial sentencing to probation. For more on surveillance, see Molly Griffard, *A Bias-Free Predictive Policing Tool?: An Evaluation of the NYPD's Patternizr*, 47 FORDHAM URB. L.J. 43, 44 (2019). For more on pretrial sentencing, see generally Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. 439 (2020). For more on probation, see Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2278 (2019); Cade Metz & Adam Satariano, *An Algorithm That Grants Freedom, or Takes It Away*, N.Y. TIMES (Feb. 6, 2020), https://perma.cc/UV2Q-PK3S (last updated Feb. 7, 2020).

[6] *See* Bobby Allyn, *'The Computer Got It Wrong': How Facial Recognition Led to False Arrest of Black Man*, NAT'L PUB. RADIO (June 24, 2020, 8:00 AM) (last updated June 24, 2020, 9:05 PM), https://perma.cc/Q2AS-9LP7 (detailing the false arrest of a Black male following a faulty facial recognition match); Nicholas Bogel-Burroughs, *Baltimore Hopes Surveillance Planes Lower Crime, but Residents Fear Abuse*, N.Y. TIMES (Apr. 9, 2020), https://perma.cc/N5FL-PPU7 (last updated June 3, 2020) (discussing the use of surveillance planes by the Baltimore Police Department).

[7] *See* Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 WL 2429574, at *1 (D.N.J. June 11, 2019) (summarizing a "data-based" risk assessment algorithm which provides quantitative scores and a "decision-making framework" to assist courts in "assess[ing] the risk that [a] criminal defendant will fail to appear for future court appearances or commit additional crimes and/or violent crimes if released").

[8] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 769 (Wis. 2016).

[9] *See Risk Assessment Landscape: Public Safety Risk Assessment Clearinghouse*, BUREAU JUST. ASSISTANCE, https://perma.cc/98EA-QML3 (providing an overview of the various risk assessment tools used across the country at different decision points in the criminal justice system).

back-to-back murders of George Floyd,[10] Breonna Taylor,[11] Ahmaud Arbery,[12] and heartbreakingly countless others have finally shaken the United States from its slumbering journey towards intentional criminal justice reform.[13] The degradation of Black bodies at the hands of law enforcement and domestic terrorists has changed the narrative and halted many widely repeated tropes and excuses of "tough on crime" and "Black-on-Black crime" rhetoric.[14] Myths are embedded in the data collected and do not tell the truth about race and crime in the United States.[15] The number of Black people harmed by manipulated algorithms will dwarf all documented historical records, and the gravity of harm is incomprehensible.

Imagine that a defendant has been convicted and sentenced using hacked software, at any point in the criminal justice process. In one scenario, the defendant receives a longer sentence than what would have been given but for the hacked software. In another, imagine the hacked software erred, recommending a lighter penalty, and the defendant injures a third party while released. Attacks on data are not only possible and probable but have indeed occurred; this can negatively

[10] *See* Tim Arango et al., *Footage of Police Body Cameras Offers Devastating Account of Floyd Killing*, N.Y. TIMES (July 16, 2020), https://perma.cc/335P-B4XK (last updated Aug. 11, 2020).

[11] *See* Rukmini Callimachi, *Breonna Taylor's Family Claims She Was Alive after Shooting but Given No Aid*, N.Y. TIMES (July 6, 2020), https://perma.cc/DJE7-FGMR (last updated Sept. 23, 2020).

[12] *See* Richard Fausset, *Suspects in Ahmaud Arbery's Killing Are Indicted on Murder Charges*, N.Y. TIMES (June 24, 2020), https://perma.cc/GYT2-HR55 (last updated June 26, 2020).

[13] *See* Michael Harriot, *A Timeline of Events That Led to the 2020 'Fed Up'-rising*, ROOT (May 30, 2020, 1:52 PM), https://perma.cc/B5ME-SBZ9.

[14] *See* Hannah Allam, *FBI Announces That Racist Violence Is Now Equal Priority to Foreign Terrorism*, NAT'L PUB. RADIO (Feb. 10, 2020, 4:17 PM), https://perma.cc/LL6F-BACX (covering the announcement of the FBI that the agency made hate-fueled violence a top national security priority, on par with foreign terrorist groups). To understand how pervasive these tropes have been used, see generally Bernard D. Headley, *"Black on Black" Crime: The Myth and the Reality*, 20 CRIME & SOC. JUST. 50 (1983) and Evan Stark, *The Myth of Black Violence*, 38 SOC. WORK 485 (1993).

[15] *See* Mayson, *supra* note 5, at 2227–49 (arguing that the inequality exposed by algorithmic risk assessment should "galvanize a more fundamental rethinking of the way in which the criminal justice system understands and responds to risk").

impact the administration of justice.[16] Inevitable questions arise: What are the legal accountability mechanisms available to address the unknowns of cyberattacks, falsified data, and the unintended consequences created by using unmonitored artificial intelligence? And, does the current rubric of "obstruction of justice," with the fixed statute of limitations, provide enough of a deterrent for would-be bad actors?

The implementation of AI in legal spaces has brought great promise. An array of legal scholars, scientists, and businesses believe that embedding AI into criminal justice reform can lead the United States to a more effective and efficient, bias-free system no longer centered on entrenched historical racism.[17] This Article is not a manifesto against tech and AI in the practice of law. Yet, without transparent safeguards to ensure that data sources have not been manipulated, criminal justice risk assessment instruments must not be used to administer justice. Additionally, this lack of oversight and transparency will have a disproportionate impact on Black people, people of color, and people with low socioeconomic status in the criminal justice system.

Part I of this paper explains the ways AI architecture leaves itself vulnerable to attacks. It describes different cyberattacks such as malware, including computer viruses, worms, and botnets, and Distributed Denial-of-Service attacks. It also outlines the damage by cyberattacks, camouflaged and manipulated data sets, and the unintended outcomes produced by AI. All these and more are the ways that AI can be corrupted to change the integrity of the outcome and undermine a fair criminal justice system. Part II outlines the design schemes of two proprietary risk assessment instruments used in the

[16] *See infra* Part III.A.

[17] *See generally* Katheryn Russell-Brown, *Racial Profiling: A Status Report of the Legal, Legislative, and Empirical Literature*, 3 RUTGERS RACE & L. REV. 61 (2001); Kenneth B. Nunn, *Rights Held Hostage: Race, Ideology and the Peremptory Challenge*, 28 HARV. C.R.-C.L. L. REV. 63 (1993); Paul Butler, *One Hundred Years of Race and Crime*, 100 J. CRIM. L. & CRIMINOLOGY 1043 (2010); Nicole Gonzalez Van Cleve & Lauren Mayes, *Criminal Justice Through "Colorblind" Lenses: A Call to Examine the Mutual Constitution of Race and Criminal Justice*, 40 L. & SOC. REV. 406 (2015); Naomi Murakawa & Katherine Beckett, *The Penology of Racial Innocence: The Erasure of Racism in the Study and Practice of Punishment,* 44 L. & SOC'Y REV. 695 (2010).

criminal justice system, Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS") and the Arnold Tool, and how they hinder justice. Part III documents potential legal theories of accountability and liability available to victims of hacked data in the criminal justice system. Part IV provides recommendations for legal and tech remedies to victims of hacked data in criminal justice risk assessment instruments. Finally, this Article concludes that vigilant oversight is necessary to ensure victims of hacked data are permitted recourse.

I. ARTIFICIAL INTELLIGENCE CAN BE HACKED AND WEAPONIZED

This Part provides technical background on data procurement and how AI depends on secure data. It also outlines a general background on cybersecurity, cyberattacks, and hacking, and how AI can be hacked and otherwise produce unexpected outcomes. An understanding of these matters contextualizes the flaws in the overreliance or blind reliance on data.

A. *What Is Artificial Intelligence and How Is It Used?*

AI is used to make sense of data produced in the legal system, and the legal community must understand hacked data. Failure to do so renders the law unable to appreciate present and future threats from technology.[18] Machine-learning algorithms have fundamentally transformed how life occurs,

[18] As AI becomes more embedded in our shared legal lexicon, more clients will begin expecting attorneys to have strong competence in understanding this field. *See* MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (AM. BAR ASS'N 2012) (requiring a competent lawyer to keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology). Thirty-eight states have adopted the duty of technology competence in some form. *See* Robert Ambrogi, *Tech Competence*, LAWSITES, https://perma.cc/ZJ2C-JW7P (listing states that have formally adopted the duty of technology competence since the ABA formally approved the change to the Model Rules of Professional Conduct in 2012). Florida has even required three out of thirty-three credit hours required every three years must be in approved technology programs. Mark D. Killian, *Court Approves CLE Tech Component*, FLA. BAR (Oct. 15, 2016), https://perma.cc/XCP8-A3EZ.

impacting hiring,[19] access to health care,[20] and criminal justice sentencing.[21] AI is relevant to virtually any intellectual task, and, in the legal community alone, it has affected the practice and administration of law in many ways for the better. From the digitization of court proceedings to the automatic transcription of legal proceedings, there has been a world of innovation that has impacted and resulted in less administrative and repetitive and rote work in law practice.[22] AI plays a role in more strategic legal tasks, including emotional intelligence, advanced problem-solving skills, and creative solutions for improvements to justice and new areas of ethical and legal conundrums.[23] The potential benefit of using AI in the administration of justice is expansive, which ushers in an exciting period as the field evolves.

At present, there is no straightforward way to define what AI means. Intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans, needs to be

[19] *See* Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARV. BUS. REV. (May 6, 2019), https://perma.cc/4F6M-VRMZ (discussing whether hiring algorithms prevent bias or amplify it).

[20] *See* Angela Spatharou et al., *Transforming Healthcare with AI: The Impact on the Workforce and Organizations*, MCKINSEY & CO. (Mar. 10, 2020), https://perma.cc/6VHX-QV7J (noting AI's transformative power on the delivery of health care).

[21] *See* Judge Noel L. Hillman, *The Use of Artificial Intelligence in Gauging the Risk of Recidivism*, AM. BAR ASS'N. (Jan. 1, 2019), https://perma.cc/8EJK-23A4 (noting concern with the judicial use of AI at sentencing to predict a criminal defendant's risk of recidivism and encouraging courts to meet this technological development "with skepticism and close scrutiny").

[22] *See* Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 85 BROOK. L. REV. (forthcoming 2021) ("The principal advantages of artificial intelligence in the administrative context are similar to those in the private sector: accuracy and efficiency.").

[23] What will happen when autopilot arguments can be coded? In what ways will new branches of law and the development of case law and statutes affect or impact? What will improvements that have access to justice look like and feel like in society? These are questions that are raised every time innovation enters society and the law. *See* Garrett & Monahan, *supra* note 5, at 475–91 (detailing the rise of modern risk assessment AI and noting concern).

developed to meet society's needs.[24] In computer science, AI research is defined as the study of "intelligent agents," i.e., any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.[25] These algorithmic processes are intended to rationalize decision-making by minimizing human bias and fallibility.[26] Simply put, AI is to capture not just *what* people think, but more importantly, how people *should* think. The implications run from philosophical to practical. As AI is a newer field in science, math, and engineering, there is no universal definition that defines the work, process, and product.[27] AI is then any technique that authorizes the computer "to mimic human intelligence using logic, if-then rules, decision trees, and machine learning (including deep learning)."[28] AI serves as a

[24] Tech companies like Apple, Amazon, and Google have established themselves as users' personal secretaries, using AI to guide us through the day, where we happily deposit our personal, confidential, and private items, such as calendars, photos, documents, and locations. *See* FRANKLIN FOER, WORLD WITHOUT MIND: THE EXISTENTIAL THREAT OF BIG TECH 2 (2017).

[25] KLAUS SCHWAB, THE FOURTH INDUSTRIAL REVOLUTION 11 (2016).

[26] The architects of AI theorized that the brain itself was a computer, controlled by programs, something possible to replicate back into the machine, though the first ideas for algorithms arose from Gottfried Leibniz, in the late 1600s. *See* FOER, *supra* note 24, at 36, 64. Google founders Larry Page's and Sergey Brin's plan was to create a brain that would not be influenced by human biases, untrustworthy sensory direction, and unexplainable desires that come from physical bodies. *Id.* at 38. The same will be true for Facebook. *Id.* at 64.

[27] Algorithms can be challenging to define, and it is believed that 60 percent of users are still entirely unaware it exists. FOER, *supra* note 24, at 73. Algorithms and the giant umbrella of tech innovations that have emerged in the last five years have and will continue to fundamentally alter the ways we connect, live, and work in the world. *See* SCHWAB, *supra* note 25, at vii. There is no universal definition, and it is currently being treated in the same way that people will use the term "Kleenex" to refer to all tissues or "Xerox" to refer to copying as a genericized trademark. *See* Christoph Henkel & Ruth C. Hauswirth, *What Law Faculty Need to Know about Artificial Intelligence*, ASS'N AM. L. SCHS. (June 12, 2019), https://perma.cc/FCE8-WX4W (noting that there are multiple subfields in addition to there being no universal definition of AI).

[28] Roger Parloff, *Why Deep Learning Is Suddenly Changing Your Life*, FORTUNE (Sept. 28, 2016, 5:00 PM), https://perma.cc/Y643-A7BQ; *see* Meenal Dhande, *What Is the Difference between AI, Machine Learning and Deep Learning?*, GEOSPATIAL WORLD (July 3, 2020), https://perma.cc/F4N7-PUJD

universal term that is used to describe various machine-learning predictive technologies that can be applied to give users a window into large amounts of data, eliminate irrelevant information, and organize data for increased review, efficiency, and accuracy.

Algorithms capture the process for solving a problem within a system, but neither give nor expect a proscribed correct, definitive answer. Algorithms define the process through which a decision is made, and AI uses that data to make decisions. AI has many more legal complexities.[29] The AI can determine what data it thinks is the most valuable and reevaluate importance independent of the developer's intent. Additionally, data is messy. If bad data is input, then the AI will produce inaccurate results.

Since 2015, AI has become faster, less expensive, and more powerful.[30] Because of infinite storage and seemingly unlimited minable data, there are infinite possibilities available through AI.[31] Machine-learning (of which deep learning is a subset) is a type of AI that includes gives computers the ability to learn without being explicitly programmed.[32] The machine-learning

(articulating the interconnected nature of AI, machine-learning, and deep learning).

[29] *See* Bridget Watson, *A Mind of Its Own—Direct Infringement by Users of Artificial Intelligence Systems*, 58 IDEA 65, 70 (2017)

> Part I discusses the evolution of artificial intelligence, the creative thinking capabilities of artificial intelligence systems, how direct infringement by an artificial intelligence system might occur, multiple parties involved in a single artificial intelligence system, and indemnification.

[30] *See* Michael Copeland, *What's the Difference Between Artificial Intelligence, Machine Learning and Deep Learning?*, NVIDIA (July 29, 2016), https://perma.cc/FM5J-MGQC (stating that the explosion of AI has a lot "to do with the wide availability of GPUs that make parallel processing ever faster, cheaper, and more powerful . . . [and] practically infinite storage and a flood of data").

[31] *See id.*

[32] "Artificial intelligence" was used from the 1950s to the 1980s to describe the field. "Machine-learning" was used from the 1980s to the 2010s. "Deep learning" has been the best way to describe this work since then. *See* Bella Wilson, *Major Milestones of Artificial Intelligence from 1949 to 2018*, MEDIUM (Apr. 18, 2018), https://perma.cc/27DL-QZXK (listing the most important developments of AI, including the different terms used to refer to AI).

process defines a problem, prepares the data[33] (separating training data from test data),[34] trains the model,[35] and finally deploys the machine-learning (watching for automation bias).[36]

Deep learning is a type of machine-learning where the computer is trained to perform human-like tasks, such as image identification, recognizing speech, and making predictions.[37] However, instead of using predefined equations, deep learning sets up basic parameters about the data and trains the computer to learn independently.[38] Deep learning is made up of algorithms that allow the software to teach itself to perform tasks by exposing multi-layered neural networks to copious amounts of data.[39] One of the best examples of this is AlphaZero, where, in 2017, the computer was able to beat a chess master after learning the game in four hours.[40] Open-source

[33] *See* Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. 192, 192 (2019) (noting concern about the use of "dirty data" from corrupt, racially-biased, or unlawful police practices in algorithmic tools to support predictive policing).

[34] *See* Cassandra Laskowski, *AI Fundamentals for Faculty*, ASS'N AM. L. SCHS. (July 15, 2020), https://perma.cc/J9DC-V8DP (discussing the stages of machine-learning systems).

[35] *See* Anup Bhande, *What Is Underfitting and Overfitting in Machine Learning and How to Deal with It*, MEDIUM (Mar. 11, 2018), https://perma.cc/96CL-3LA7 (explaining that how well a model fits to a data set determines whether the model will yield accurate results); Ali Svoboda, *Applied Regression Analysis: Project 2*, RPUBS (Mar. 19, 2015), https://perma.cc/2STZ-VYKK (experimenting with a dataset of information on houses to see if the house price could be explained by the square footage, age, and features of the home).

[36] "Automation bias refers to a specific class of errors people tend to make in highly automated decision making contexts, when many decisions are handled by automated aids (e.g., computers), and the human actor is largely present to monitor on-going tasks." Linda J. Skitka, *Automation Bias*, UNIV. ILL. CHI. (2011), https://perma.cc/JN4V-BP52.

[37] *See Deep Learning: What It Is and Why It Matters*, SAS, https://perma.cc/FH34-J5ZA.

[38] *See id.*

[39] *See* Dhande, *supra* note 28 (explaining that deep learning "uses some machine learning techniques to solve real-world problems by tapping into neural networks that simulate human decision-making," requiring huge datasets to train itself).

[40] Coding a computer to play chess requires three components: the rules (bishop moves diagonally), strategies (castling), and a goal (claim the king).

machine-learning libraries like Sonnet, released by Google, makes it easier for developers to build neural network components.[41] Another example of how deep learning works is examining the use of Google to have a translation agent that gives correction ideas. Google translation scans millions of books to look for patterns.[42] Usually, "quick brown" is followed by "fox;" thus, the software looks for inferences from patterns in writing.[43] For speech recognition, it had a 90 percent understanding rate.[44] However, that means one in ten words was wrong. The computer had to process an algorithm that

DeepMind, creators of AlphaZero, "said the difference between AlphaZero and its competitors is that its machine-learning approach is given no human input apart from the basic rules of chess. The rest works out by playing itself over and over with self-reinforced knowledge." Samuel Gibbs, *AlphaZero AI Beats Champion Chess Program after Teaching Itself in Four Hours*, GUARDIAN (Dec. 7, 2017, 7:41 AM), https://perma.cc/8BPK-PJUQ. AlphaZero was given no rules, strategies, or goals and had to learn them all by only playing the game. *Id.* DeepMind found that this approach made AlphaZero learn in a "more human-like approach," searching for moves, processing around eighty thousand positions per second in chess. *Id.*; *see* David Silver et al., *A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go Through Self-Play*, SCI. MAG. (Dec. 7, 2018), https://perma.cc/4UXH-WMBQ (detailing the achievements of the AlphaGo Zero program); *see also* FOER, *supra* note 24*,* at 52–53 (discussing deep learning's potential to evolve to do everything).

[41] *See* Mariya Yao, *12 Amazing Deep Learning Breakthroughs of 2017*, FORBES (Feb. 5, 2018, 8:00 AM), https://perma.cc/4U88-VMFM (reflecting on the deep learning breakthroughs of 2017—"The Year of AI").

[42] *See* Quoc V. Le et al., *A Neural Network for Machine Translation, at Production Scale*, GOOGLE AI BLOG (Sept. 27, 2016), https://perma.cc/SHJ7-HRCE (announcing the Google Neural Machine Translation system, which "utilizes state-of-the-art training techniques to achieve the largest improvements to date for machine translation quality").

[43] *See* Tianyi Zhao, *The AI Powers behind Google Translate*, GEO. (May 5, 2019), https://perma.cc/5A7Q-SNDV (discussing Google Translate's automated recognition of patterns and regularities in data); Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES (Dec. 14, 2016), https://perma.cc/2NTF-UZED (explaining Google Translate's ability to rewire itself to reflect patterns from the data it absorbs).

[44] *See* Mike Wheatley, *Google Makes Its Speech-to-Text and Text-to-Speech Services More Accurate and Accessible*, SILICONANGLE (Feb. 21, 2019), https://perma.cc/DL3V-2ZLT.

sorted the words.[45] This self-directed trial-and-error process is deep learning. Applied data science, computer science, and high-tech investment are moving in this direction of such results.[46]

AI is a tool, but it can also be used as a weapon. A multifaceted document, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, is a 2018 report crafted by fourteen institutions from academia, civil society, and industry.[47] The report outlined the economic, political, and human labor expenses related to malicious attacks, new attacks, and changes to the type of character of threats.[48] Similarly, in their article, "How A.I. Could Be Weaponized to Spread Disinformation," Cade Metz and Scott Blumenthal described how weaponized AI could be used to spread disinformation.[49] Another author, Jayshree Pandya, in "The Weaponization of Artificial Intelligence," addressed how AI could be weaponized in cyberwar.[50] Misunderstanding the mathematics of high-dimensional spaces may lead users to false confidence in the ability of deep neural networks to make the right decisions.[51]

[45] *See* Isaac Caswell & Bowen Liang, *Recent Advances in Google Translate*, GOOGLE AI BLOG (June 8, 2020), https://perma.cc/F6EN-WAMU (discussing further advancements in the AI behind Google Translate).

[46] *See* Kevin J. Ryan, *Who's Smartest: Alexa, Siri, and or Google Now?*, INC. (June 3, 2016), https://perma.cc/GA4K-EMK2.

[47] MILES BRUNDAGE ET AL., THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE: FORECASTING, PREVENTION, AND MITIGATION (2018), https://perma.cc/8Y3T-SVUS (PDF).

[48] *Id.* They make four high-level recommendations related to policymaking, reach out to necessary parties about foreseeable harm, identify best practices, and expand stakeholders and domain experts into the conversation. *See id.* at 31 (noting AI is "already being deployed for purposes such as anomaly and malware detection").

[49] Cade Metz & Scott Blumenthal, *How A.I. Could Be Weaponized to Spread Disinformation*, N.Y. TIMES (June 7, 2019), https://perma.cc/3FCM-8W6L.

[50] Jayshree Pandya, *The Weaponization of Artificial Intelligence*, FORBES (Jan. 14, 2019, 12:51 AM), https://perma.cc/V7ZG-8673.

[51] Some other issues include the fact that it can be difficult for networks to converge, and that GANs have yet to converge on large problems. *See* Jonathan Hui, *GAN—Why It Is So Hard to Train Generative Adversarial Networks!*, MEDIUM (June 21, 2018), https://perma.cc/N6WC-KSF7 (explaining

B. *How Data Is Gathered, Analyzed, and Utilized*

The top three spaces where attacks occur are through Big Data,[52] the Internet of Things ("IoT"),[53] and AI, all of which compromise confidentiality, integrity, and availability.[54] These three attack surfaces comprise the primary ways data is gathered, disseminated, and utilized.[55]

Big Data describes types of data and is an attack surface by hackers. Big Data's value comes from four parts: capture, storage, analysis, and action.[56] In creating Big Data, there are distinct ways that data is procured, and tech companies have become the biggest gatekeepers of data. For example, Google provides a hierarchy to information; Facebook uses algorithms to organize social circles; Amazon watches users' purchasing and browsing patterns to recommend further acquisitions.[57] Sets of Big Data are gathered, continually, to be used in AI.[58] The amount of discoverable data is astronomical. The volume of Big Data continues to expand, with 2.4 quintillion new bits of data being created daily.[59] It is anticipated that there would be 44 trillion gigabytes by 2020,[60] including the capture of every red light camera video, Fitbit daily steps, meal trackers, grocery

the difficulty in creating and training generative adversarial networks, which create data as opposed to discriminative models which process data).

[52] *See* Laurel Eckhouse, Opinion, *Big Data May Be Reinforcing Racial Bias in the Criminal Justice System*, WASH. POST (Feb. 10, 2017), https://perma.cc/WQ72-5G2P.

[53] The Internet of Things is defined as the connection via the internet of all computing devices that transmits and receives data. Eric Brown, *Who Needs the Internet of Things*, LINUX.COM (Sept. 13, 2016), https://perma.cc/SP3M-95XG.

[54] *See Artificial Intelligence and Robotics National Institute*, AM. BAR ASS'N 469 (2020), https://perma.cc/Q5AD-932N (PDF) (noting the cyber risks in today's complex world of tech to the Big Data, Internet of Things, and AI).

[55] *See id.* (listing the three primary attack surfaces).

[56] *See Why Big Data Is "The New Natural Resource"*, WASH. POST, https://perma.cc/4YLL-2UV4 [hereinafter *The New Natural Resource*].

[57] *See* FOER, *supra* note 24, at 4–5. One-third of Amazon purchases come from recommendations. *Id.* at 70. Facebook uses upwards of one hundred thousand "signals" when determining data that a user will see. *Id.* at 73.

[58] *See id.*

[59] *See The New Natural Resource*, *supra* note 56.

[60] *See id.*

orders, and podcasts consumed by persons across the world.[61] There is a difference between found data sets and created data sets. For example, a found data set would be basketball statistics, street mapping, traffic data and vehicle locations, and additionally, data from wearable internet which provide biomedical data from wearable internet such as fitness trackers and smartphones.[62] IoT shares house temperatures and air quality.[63] And even legal decisions have joined the world of data points. Through the Caselaw Access Project, all cases in United States history have been digitized and made public and widely available.[64] Each case, date, defendant, and jurisdiction is a data

[61] *See id.* Smart devices, social media, cameras, and sensors will feed the Internet of Things and prepare for harvesting. *See id.*

[62] Secure locations at military centers were compromised because of people wearing their Fitbit trackers. *Fitness App Strava Lights Up Staff at Military Bases*, BBC NEWS (Jan. 29, 2018), https://perma.cc/5MQH-3GYN.

[63] JunHo Jo et al., *Development of an IoT-Based Indoor Air Quality Monitoring Platform*, 2020 J. SENSORS 1, 2, https://perma.cc/2DHG-42BR (PDF).

[64] *See Gallery*, CASELAW ACCESS PROJECT, https://perma.cc/9HHE-85P9 (including only some of what has been digitized). In 2018, the Harvard Law School Library announced its forthcoming project: Caselaw Access Project (CAP) API that published all United States case law for anyone to access for no cost. *About*, CASELAW ACCESS PROJECT, https://perma.cc/AGC4-N8LC. Between 2013 and 2018, the library digitized forty million pages of United States court decisions, encompassing almost 6.5 million individual cases. *Id.* The digitization of 6.5 million cases for the CAP was an audacious undertaking and a true marvel. *Id.* The database purports to have gathered every case from 1658 until 2018, telling the legal story of the United States. *Id.* It tells our legal ancestry and provides the "legal genome" path of whom we are and how we got there through the law. *Id.* The creation of the database was a huge undertaking and will fundamentally influence the understanding of patterns and predictability of cases by jurisdiction, state, and across the nation. Our ability to interpret evidence of bias in judicial decisions will be simple to discern after reviewing and parsing the information based on the information presented. Not only is this audacious collection based on time, but it is also impressive based on breadth. The collection includes all state, federal, tribal, and territorial courts, including American Samoa, Dakota Territory, Guam, Tribal Courts, and the Northern Mariana Islands. *Id.* Not only is the breadth broad, but the level of detail included is extensive. Each volume is broken down to the case-level to include majority and dissenting opinions, with humans reviewing the data for party names, docket numbers, and dates. *Id.* This human-verification process will be vital as we review recent hacks that have been able to fool AI systems. This project's full potential is still being explored/investigated, as users have created Wordclouds, limerick generators,

point for exploration and extrapolation.[65] However, found data does not always tell the full story. One can have incorrect labels[66] or missing or incorrect data.[67]

Created data, on the other hand, consists of data sets from synthesized sources made to mimic real data that can serve as a "sandbox" for developers to make created data.[68] Examples of created data would be Instagram filters that put sunglasses on a face or epidemiology models about the distribution of a vaccine.[69] The assumption that test data will be generally similar to the training data has created its own subfield, Generative Adversarial Networks ("GANs").[70] Algorithms capture the process for solving a problem within a system, but neither give nor expect a proscribed correct, final answer.[71] GANs are created when two neural networks train concurrently to understand that one can serve as a foundation to generate

open-source casebooks, and models to teach people how to write in Python. *Gallery*, CASELAW ACCESS PROJECT, https://perma.cc/9HHE-85P9.

[65] *See About*, CASELAW ACCESS PROJECT, *supra* note 64 (listing key metadata fields within the digitization process).

[66] CLARENCE CHIO & DAVID FREEMAN, MACHINE LEARNING & SECURITY: PROTECTING SYSTEMS WITH DATA AND ALGORITHMS 279 (Courtney Allen ed., 2018) ("This bias causes imperfect data and incorrect labels assigned to samples, affecting the accuracy of the system.").

[67] *See id.* at 280; *see also* Eric Westervelt, *Did a Bail Reform Algorithm Contribute To This San Francisco Man's Murder?*, NAT'L PUB. RADIO (Apr. 18, 2017), https://perma.cc/H2UK-V8P2 ("Judge Reardon followed the recommendation of . . . a computer-generated score that's used . . . to help calculate whether a suspect is a flight risk or likely to return to court. . . . In the case of French, a miscalculation ended in murder.").

[68] *See* Cody Nash, *Create Data from Random Noise with Generative Adversarial Networks*, DEVELOPERS, https://perma.cc/U296-S4JE (explaining how a GAN-Sandbox is used to generate new credit card data).

[69] *See* Lucas Matney, *Instagram's AR Filters Are Getting More Dynamic*, TECHCRUNCH (May 27, 2020, 4:45 PM), https://perma.cc/6QEQ-RZGD (discussing the AR that creates Instagram filters).

[70] *See generally* W. Philip Kegelmeyer, *Adversarial Issues in Machine Learning*, *in* 22 NEXT WAVE 10 (2019), https://perma.cc/AAV6-U9B8 (PDF). For additional ideas on the role of Generative Adversarial Networks, see generally Steven M. Bellovin et al., *Privacy and Synthetic Datasets*, 22 STAN. TECH. L. REV. 1 (2019); Russell Spivak, *Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339 (2019).

[71] *See* FOER, *supra* note 24*,* at 67–69.

possible outcomes for veracity and probability of occurrence.[72] Every day, GANs are improving effectiveness at producing realistic-looking synthetic samples, though they are unable to infer, which remains a complex human trait.[73] The GAN model has shown substantial promise, but the existing errors are significant.[74] The samples produced are quite convincing, but mistakes can include color, style, and in extreme instances, object identity.[75] There are also structural flaws with this model of developing training sets. The discriminator becomes too strong too quickly, and the generator ends up not learning anything.[76] Or the generator only learns the particular weaknesses of the discriminator, and it can take advantage of these to trick the discriminator into classifying generated data as real instead of learning to represent the actual data distribution.[77] The generator can also learn only a minimal subset of the actual data distribution, leading to insufficient variation in the output.[78]

AI is designed to learn from training sets made of found data and created data.[79] Training sets are acquired from every

[72] *See* Barry Chen et al., *Toward a Deep Learning System for Making Sense of Unlabeled Multimodal Data*, *in* 22 NEXT WAVE 1, 4 (2019), https://perma.cc/AAV6-U9B8 (PDF) ("In GANs, two networks compete against each other: The first one learns features effective for generating input data realistic enough to fool the second one.").

[73] *See* Vincent Dumoulin, *Adversarially Learned Inference*, GITHUB, https://perma.cc/BH2T-LDY9.

[74] *See id.* (including photographs to visualize the inaccuracies).

[75] *See id.*

[76] *See* Animesh Karnewar, *V-GAN (Variational Discriminator Bottleneck): An Unfair Fight between Generator and Discriminator*, MEDIUM (Nov. 9, 2018), https://perma.cc/2A7L-2HKX (discussing the problem with a discriminator becoming too powerful). The discriminator classifies generated data as fake so accurately and confidently that there is nothing in the discriminator's back-propagated loss function gradients for the generator to learn. *See id.*

[77] *See id.* ("Generally, if the Discriminator becomes too strong, i.e. it can easily tell the samples apart, it would cease to supply plausible gradients to the Generator for training.").

[78] *See id.*

[79] *See* Alex Moltzau, *Artificial Intelligence and Training Data*, TOWARDS DATA SCI. (Oct. 18, 2019), https://perma.cc/GK7E-UTTX ("The data used to build the final model usually comes from multiple datasets.").

camera, scanner, and device available.[80] Microsoft, Stanford, and Duke have quietly deleted public face recognition data sets.[81] Over ten million images of one hundred thousand individuals who had not been asked for nor given their consent have been removed.[82] Although Microsoft has deleted the database, it is still available to researchers and companies that had previously downloaded it and this data is still being shared on open source websites.[83] Sometimes, there is data that is not meant to be included in AI calculations. The Pentagon has had to remind military troops and defense personnel that in their quest to reach their daily "10,000" steps via their fitness trackers, they were inadvertently transmitting sensitive and secure locations to a third party.[84]

There are many compelling reasons to be pleased with the collection of data, and people have benefited from data sharing across systems through IoT. The productivity increases, quantifiable improvements in quality of life, and lowering costs of regular goods and services provide comfort and excitement

[80] When Google began its process of digitizing all books, it would arrive at libraries, trucking away boxes of books and quickly scanning and returning them. FOER, *supra* note 24*,* at 54. "We are not scanning all those books to be read by people. We are scanning them to be read by an AI." *Id.* at 55 (quoting an unnamed Google engineer).

[81] *See* Madhumita Murgia, *Microsoft Quietly Deletes Largest Public Face Recognition Data Set*, FIN. TIMES (June 6, 2019), https://perma.cc/T2Z2-S78B (explaining that Microsoft, Duke, and Stanford data sets were taken down after a Financial Times report).

[82] *See id.*

[83] *See id.*

> Now it is completely disassociated from any licensing, rules or controls that Microsoft previously had over it. People are posting it on GitHub, hosting the files on Dropbox and Baidu Cloud, so there is no way from stopping them from continuing to post it and use it for their own purposes.

(quoting Alan Harvey).

[84] A twenty-year-old Australian student discovered two years' worth of data uploaded by Strava, a social media platform for athletes with satellite-tracking data for digital fitness devices such as Fitbit. David Martin, *Pentagon Reviews Fitness Tracker Use over Security Concerns*, CBS NEWS (Jan. 29, 2018, 6:51 PM), https://perma.cc/4LAR-7XH5. The tracking outlined military bases and secure sites across the country and mapped their travel patterns by time and location. *Id.*

about the future.[85] With the anticipated trillions of devices such as smartphones, wearable devices, computers, and tablets, generating further trillions of data points, the assistance and surveillance, and ownership of the data will continue to be called into question.[86] While there are many positive applications associated with enjoying a digital presence,[87] those positive applications do not outweigh the real and concerning threats of privacy, surveillance, and stolen information being used to create training sets without the user's permission.[88] And attacks can occur between interconnected devices. A user may be comfortable with the data being used to recommend better running trails, but not to determine if one's insurance company will dictate a minimum number of steps required to be eligible for health insurance.[89] And users may not appreciate that a breach of an individual's streaming service can lead to entry to a person's home security system.[90]

C. *Public Fears and the Reality of Concerns Related to AI*

Doubts about AI are growing. In 2019, the American Bar Association featured an article penned by Judge Nel L. Hillman titled "The Use of Artificial Intelligence in Gauging the Risk of Recidivism."[91] The article outlined three reasons for concern

[85] *See* SCHWAB, *supra* note 25, at 137 (listing the positive impacts of IoT).

[86] *See id.* at 18. It is anticipated that by 2025, there will be one trillion sensors connected to the internet. *Id.* at 26.

[87] Such as an increase in transparency, faster interconnectivity between persons and systems, more space for free speech, faster dissemination of information, and more efficient use of government systems and resources. *See id.* at 123 (listing the positive impacts of digital presence).

[88] *See* Jathan Sadowski, *Companies Are Making Money from Our Personal Data—but at What Cost?*, GUARDIAN (Aug. 31, 2016), https://perma.cc/8ZX9-ZZAR (commenting that the methods and purposes of data collection range from irritating infringements to major intrusions).

[89] *See* Angela Chen, *What Happens When Life Insurance Companies Track Fitness Data?*, VERGE (Sept. 26, 2018, 1:01 PM), https://perma.cc/ZT43-GCTJ (considering the privacy concerns involved with insurance policies that allow customers to share fitness data in exchange for discounts).

[90] *See* Marc Wilczek, *Cybercrime: AI's Growing Threat*, DARK READING (Oct. 4, 2019, 10:00 AM), https://perma.cc/MZ4G-7FK7 (outlining ways IoT can be hacked).

[91] Hillman, *supra* note 21.

related to AI in the sentencing process: potential violations of due process, the limitations of AI related to unacceptable risks of error and implicit bias, and lastly, that "reliance on AI to predict recidivism improperly cedes the discretionary sentencing power to nonjudicial entities."[92] Judge Hillman is correct on the pulse of significant concerns about the use of AI in the sentencing process. One of the areas that is receiving the least attention is how unstable AI is as a developing science, particularly as it relates to security and the potential for hacking related to the input data and the produced output data.[93]

In a 2017 Pew Research survey poll, 72 percent of U.S. adults reported they were worried that robots and computers would do human jobs.[94] Those adults reported being three times more likely to feel worried as compared to enthusiasm about the role of algorithms in making hiring decisions without human involvement.[95] In the same survey, people were decidedly reluctant to incorporate unique AI practices into their lives.[96] Nevertheless, AI decision-making is happening in every aspect of people's lives, without proper vetting.[97] Cybersecurity is consistently treated as an afterthought instead of being integrated into protocols and principles from the initial design stage.[98] Furthermore, the question is not whether an AI cyberattack has happened yet, but what is the unknown impact

[92] *Id.*

[93] *See id.* (acknowledging that states only recently introduced AI tools in sentencing).

[94] Compared to 33 percent that were enthusiastic. Monica Anderson, *6 Key Findings on How Americans See the Rise of Automation*, PEW RSCH. CTR. (Oct. 4, 2017), https://perma.cc/3UFC-SU63.

[95] *Id.*

[96] *Id.* ("[A]round six-in-ten U.S. adults say they would *not* want to ride in a driverless car (56%) or have a robot caregiver for themselves or a family member (59%).").

[97] *See supra* notes 86–87 and accompanying text.

[98] *See* Barbara Burgess, *Businesses Consider Cybersecurity as an Afterthought despite Growth in Attacks, EY Survey Finds*, EY (Feb. 18, 2020), https://perma.cc/8NDZ-BULY ("Despite the overall growth in cyberattacks, only one-third of organizations say the cybersecurity function is involved at the planning stage of a new business initiative . . . .").

of the cyberattack that has likely already occurred.[99] This Part further describes types of cyberattacks, how data is hacked, and what happens when AI is not monitored for bias.

1. Cyberattacks

Cybersecurity and privacy are areas of concern in AI development. Over the last seventy-five years, scientists have been watching and defending against cybersecurity attacks.[100] The experts warn that it is not a matter of "if" but rather "when" we can anticipate, future breaches and hacks to the software.[101] Hacking is the act of someone or something gaining unauthorized access to a computer device or even an algorithm.[102] Hacking finds weaknesses in the security settings, exploiting them to access confidential information to inject

[99] *See* Wilczek, *supra* note 90.

[100] In 1943, Alan Turing, the "father of Computer Science" led the British effort that developed the first digital machine that could hack German codes, known as the "Enigma" code. *See* Glenn Zorpette, *Breaking the Enemy's Code: British Intelligence Deciphered Germany's Top-Secret Military Communications with Colossus, an Early Vacuum-Tube Computer*, 24 IEEE SPECTRUM 47, 47–51 (1987). In 1982, the United States staged a prototype where they reprogrammed computer equipment intended for a Soviet gas pipeline that caused the pipeline to explode. *See* RICHARD M. NEPHEW, TRANSATLANTIC SANCTIONS POLICY: FROM THE 1982 SOVIET GAS PIPELINE EPISODE TO TODAY 1 (2019), https://perma.cc/89QL-J2ET (PDF). In 1988, the first worm was distributed by the Internet and released in November of that year. *See* Sihan Qing & Weiping Wen, *A Survey and Trends on Internet Worms*, 24 COMPUTS. & SEC. 334 (2005), https://perma.cc/2GPH-XC6F. In 1999, Melissa—the first widespread email worm—and Kak were created and deployed. *See* Nick G., *The Most Telling Cyber Security Statistics in 2020 [Infographic]*, TECHJURY, https://perma.cc/ZZ9H-NYZL (last updated Aug. 20, 2020) [hereinafter Cyber Sec Statistics]. In 2003, the Department of Homeland security began operations creating the national cybersecurity division, and by 2006, NASA had been forced to block emails with attachments before shuttles were launched out of fear of potential hacking schemes. *Id.* In 2009, the Aurora attack hit Google and thirty-three other companies in search of intellectual property and 2010, the Stuxnet attack was uncovered. *Id.* This was considered the first weaponized malware because of its targeted purposely targeted nature and the fact that it is disrupted Iran's nuclear program and the centrifuges used for uranium enrichment. *Id.*

[101] *See Cyber Sec Statistics*, *supra* note 100 ("Cyber security statistics show that this field will only continue to grow commensurately with the demand. Hackers and cyber criminals aren't slowing down . . . .").

[102] *See id.*

harmful data or applications.[103] Between January 2006 and April 2018, there were over eight thousand recorded breaches in identity theft resource centers.[104] Fifty million Facebook accounts were affected by an attack in September 2018 and bar exam students have even found their bar exams hacked.[105]

There are a plethora of types of cyberattacks,[106] and AI has changed the design and broadened the scope of cyberattacks.[107] Malware is defined as malicious software that infects a computer, and this will include computer viruses, worms, Trojan horses, spyware, and adware.[108] Malware aids hackers to gain control over the targeted computer or devices to perform forced

[103] *See id.*

[104] *See Data Breaches*, IDENTITY THEFT RES. CTR., https://perma.cc/NE2R-AFNP (including data breach reports from 2005 to present).

[105] *See* Mike Isaac & Sheera Frenkel, *Facebook Security Breach Exposes Accounts of 50 Million Users*, N.Y. TIMES (Sept. 28, 2018), https://perma.cc/JB7G-LDED (explaining the ramifications of the breach); David Jesse, *Michigan Online Bar Exam Crashes in Middle of Testing; Hacking Attempt Blamed*, DETROIT FREE PRESS (July 28, 2020), https://perma.cc/3GQ3-UYEX (reporting the ramifications of Michigan's hacked bar exam); Chris Opfer, *Florida Scraps Online Bar Exam, Citing Technology Concerns*, BLOOMBERG L. (Aug. 17, 2020, 8:39 AM), https://perma.cc/EC64-PWTW (referring to the hacking of Michigan's exam).

[106] They can also include phishing, ransomware, pharming, viruses, Wi-Fi eavesdropping, and industrial IoT attacks. *Types of Cyber Threats and What They Do*, TIE NAT'L (Jan. 24, 2017), https://perma.cc/5U7N-EC4Q. Ransomware attacks have stymied many local governments throughout the United State in recent years. Ransomware attacks in Atlanta, Georgia in 2018, Baltimore, Maryland in 2019, and Greenville, North Carolina in 2020, removed critical access to data for weeks and cost the cities hundreds of thousands of dollars to restore and protect the systems. *See* Alan Blinder & Nicole Perlroth, *A Cyberattack Hobbles Atlanta, and Security Experts Shudder*, N.Y. TIMES (Mar. 27, 2018), https://perma.cc/B3WK-KFAA; Sean Gallagher, *Baltimore Ransomware Nightmare Could Last Weeks More, with Big Consequences*, ARS TECHNICA (May 20, 2019, 12:47 PM), https://perma.cc/GSZ6-JPLA; Genna Contino, *Greenville Water Phone, Online Payment Systems Restored*, GREENVILLE NEWS (Feb. 4, 2020, 10:30 AM), https://perma.cc/8AVT-EB7N.

[107] *See generally* William Dixon & Nicole Eagan, *3 Ways AI Will Change the Nature of Cyber Attacks*, WORLD ECON. F. (June 13, 2019), https://perma.cc/GF8M-3C7M (pointing to attacks such as impersonation of trusted users, attacks in the background, and faster attacks with more effective consequences).

[108] *See What Is the Difference: Viruses, Worms, Trojans, and Bots?*, CISCO (June 14, 2018), https://perma.cc/L63E-8RBT.

actions and access unauthorized data.[109] Unlike a virus, worms live autonomously in the computer's memory; they do not damage nor alter the hard drive, but instead send themselves to other machines in the network, causing damage by shutting down parts of the network.[110] Botnets create an army of infected computers that act under a hacker's control, and these infected units often function in a way that makes the attack undetectable.[111] Bots are mostly discussed in the media in the context of spreading false information, but bots are also used to attack computers and networks.[112] They can send spam emails, spread malware, or have Distributed Denial-of-Service attacks (DDoS).[113] In a DDoS attack, a botnet army keeps attacking a web server, causing it to fail because of an overload, forcing the web servers to shut down.[114]

[109] *See Cyber Sec Statistics*, *supra* note 100. A solid 99.9 percent of discovered mobile malware was hosted on third party app stores. *Id.* An app store is a type of digital distribution platform for computer software called applications, often in a mobile context, and two of the most popular ones are Google Play Store and Apple App Store. *See* J. Clement, *Number of Apps Available in Leading App Stores 2020*, STATISTA (Nov. 24, 2020), https://perma.cc/L38L-AEFY (providing a chart showing the biggest app stores).

[110] *See* John Markoff, *Worm Infects Millions of Computers Worldwide*, N.Y. TIMES (Jan. 22, 2009), https://perma.cc/U93M-LJSK. In 2009, a worm called Downadup infected almost nine million computers in fourteen days. *See id.*

[111] *See Botnet DDoS Attacks*, IMPERVA, https://perma.cc/8SY2-EF7N (explaining what a botnet is). The number of Windows botnets rose from 29 percent to 34 percent in the first quarter of 2018. Alison DeNisco Rayome, *Major DDoS Attack Lasts 297 Hours, as Botnets Bombard Businesses*, TECHREPUBLIC (Apr. 27, 2018, 5:48 AM), https://perma.cc/97B6-EEHS.

[112] *See* Sam Bocetta, *Has an AI Cyber Attack Happened Yet?*, INFOQ (Mar. 10, 2020), https://perma.cc/5Y3F-R6G9 ("These bots can pretty easily be used for misinformation, like when users marshal them to flood a Twitter thread with false posters to influence an argument. But they can be used to DDos the computers and networks of an enemy.").

[113] *See* Rayome, *supra* note 111.

[114] *What Is a DDoS Attack?*, SUCURI (Aug. 9, 2019), https://perma.cc/D8EJ-3LN5.

Cyberattacks in 2019 and 2020 were rampant.[115] Most companies are unaware of internal and external attacks.[116] Experts believe it takes half a year to detect a data breach and that almost half of all cyberattacks are aimed at small businesses, with 91 percent of attacks launched with a phishing email.[117] Every fourteen seconds, a business operation falls victim to a ransomware attack,[118] and 38 percent of malicious attachments are masked and hidden as Microsoft Office types of files.[119] In 2016, cybercriminals exploited 48 percent of U.S. citizens' credit cards, and the global cost of online crime is anticipated to be $6 trillion by 2021.[120] In the same way, 62 percent of global organizations have admitted that they are not equipped to handle a cyberattack.[121] Between January and September 2019, 7.9 billion data records became vulnerable to data breaches.[122] Additionally, undetected software bugs are another way systems are vulnerable to attack. For example, on January 29, 2019, it was discovered that an iPhone FaceTime

[115] *See Cyber Sec Statistics*, *supra* note 100 ("IoT attacks were up by 600 percent in 2017. In 2019, the attacks reached 2.9 billion events."). In the first half of 2020, thirty-six billion records were exposed, surpassing the total number of records exposed for all of 2019 by a factor of two. *See* RISKBASED SECURITY, 2020 Q3 REPORT: DATA BREAK QUICKVIEW 10 (2020), https://perma.cc/W74U-7WAK (PDF).

[116] *See id.* ("Perhaps the more concerning side to cyber security statistics, in general is the number of incidents that have gone unreported. Speculation would lead one to believe that the figure of 31% is significantly lower than reality.").

[117] *Id.*

[118] *Id.*

[119] *See id.* Experts studying breaches in 2020 found that 84 percent were financially motivated. *See* VERIZON, 2020 DATA BREACH INVESTIGATIONS REPORT 10 (2020), https://perma.cc/6K3A-LXAY (PDF).

[120] *See* Steve Morgan, *Cybercrime to Cost the World $10.5 Trillion Annually by 2025*, CYBERCRIME MAG. (Nov. 13, 2020), https://perma.cc/XNM2-K4WL. Some experts see damages related to cyberattacks are expected to exceed $5 trillion by 2024. Wilczek, *supra* note 90. Thirty-one percent of organizations had experienced a cyberattack at the operational infrastructure level, and the most concerning point is that a number of these incidents have gone unreported. *Cyber Sec Statistics*, *supra* note 100.

[121] *See Cyber Sec Statistics*, *supra* note 100.

[122] *See* INGA GODDIJN & JAKE KOUNS, DATA BREACH QUICKVIEW REPORT (2019), https://perma.cc/B9CU-FMJL (PDF).

bug would allow a person to eavesdrop on any iPhone user.[123] One attack against Instagram locked users out of their social profiles in August 2019,[124] and a bug in the code led to a data breach in November 2019.[125] AI cyberattacks are on the rise because of vulnerable storage services and an increase in data that people give to companies that are then sold to third-party companies.[126] These are examples of types of attacks that have been reported; however, because of companies' private nature, it is not clear how many companies, software distributors, and organizations have had data breaches.

Attacks against government entities are particularly pernicious as they directly impact a necessary part of people's daily lives in the United States. A reported 7.3 percent of data breaches in the United States, from 2014 to 2018, occurred on government or military entities.[127] For instance, a recent attack in April 2018, where 3.75 million Social Security numbers and bank account details were data-mined,[128] can lead to public

[123] A major new bug was introduced in iOS 12.1 that allowed anyone to FaceTime groups to listen in on the audio and potentially YouTube video of anyone else using iOS. It worked ranging from new iPhone XS to iPhone 5S, as long as it was using iOS 12.1 and above. *See* Jake Swearingen, *Major iPhone FaceTime Bug Lets You Eavesdrop on Any iPhone User*, INTELLIGENCER (Jan. 28, 2019), https://perma.cc/WXT9-5YPT (explaining the details of the FaceTime bug). Apple anticipated needing a week to fix the bug, advising users to disable FaceTime on their phones. However, they quickly changed this policy and disabled group FaceTime from the system level. *See* Tom Warren, *Apple Disables Group FaceTime Following Major Security Flaw*, VERGE (Jan. 29, 2019, 12:04 AM), https://perma.cc/3N2C-PRRX.

[124] *See* Bocetta, *supra* note 112.

[125] *See id.*; Sarah Kuranda & Reed Albergotti, *New Instagram Bug Raises Security Questions*, INFO. (Nov. 16, 2018, 3:32 PM), https://perma.cc/BR3F-ZLJQ (probing the dangers of Instagram's recent security flaw).

[126] *See* Bocetta, *supra* note 112 ("[E]veryday people are giving more data to companies than ever before, particularly through device or app usage or through subscription services.").

[127] *See* EXPERIAN, DATA BREACH RESPONSE GUIDE 4 (2018), https://perma.cc/K9V5-TZ93 (PDF) (reporting that 46.3 percent of breaches were through business, 27.1 percent in medical and health care industries, 12.6 percent in banking and credit financial industries, and 6.7 percent in education).

[128] *See* Rex Hammock, *Taskrabbit, IKEA's Gig-Economy Home Service Marketplace, Gets Hit by Hackers*, SMALLBUSINESS.COM (Apr. 17, 2018), https://perma.cc/6F5D-QXUP. The attack was made by a botnet that used

mistrust in the government's ability to safeguard private data. While it may be encouraging that it is not a more significant percentage of identified data breaches, the very governmental entity responsible for securing people's data is in reality denying people of their liberty interest, particularly as it relates to incarceration. Hackers can develop machine algorithm hacking methods effortlessly or use botnets to spread an attack.[129] The scope of the impact of hacked AI is unquantifiable.[130] Also, every time a bot system makes an attack, it becomes better when it attempts to attack again.[131]

These pale in comparison to the largest cyberattack against the United States federal agencies that was detected in December 2020. It is believed that malicious code was snuck into an update on the software Orion (made by SolarWinds, a network-monitoring company) in March 2020.[132] Several government organizations, Departments of State, Defense, Homeland Security, Treasury, and Commerce, and the National Institutes of Health reported their networks being breached.[133]

dependent machines (called in the industry "slave machines") to perform a DDoS attack on the servers. Bocetta, *supra* note 112.

[129] *See* Bocetta, *supra* note 112 (noting the ease with which AI-assisted attacks and algorithms are created).

[130] Consumers who have had personal information exposed through hacking, theft, or negligence have with increasing frequency brought actions, often class actions, against the business that held such information in its computer system. This Article collects and discusses cases that have addressed the liability of private businesses to governments and consumers for a breach of data security for consumers' information when such breach has occurred in the course of the private business in question. *See* Eric C. Surette, Annotation, *Liability of Businesses to Governments and Consumers for Breach of Data Security for Consumers' Information*, 1 A.L.R.7th Art. 2 (2021).

[131] *See* Bocetta, *supra* note 112.

[132] *See* Kari Paul & Lois Beckett, *What We Know—and Still Don't — About the Worst-Ever US Government Cyber-Attack*, GUARDIAN (Dec. 19, 2020), https://perma.cc/U758-65MX (detailing the malware attack on Orion that allowed hackers to steal information from several U.S. government departments).

[133] *See* Ellen Nakashima & Craig Timberg, *DHS, State and NIH Join List of Federal Agencies—Now Five—Hacked in Major Russian Cyberespionage Campaign*, WASH. POST (Dec. 14, 2020, 10:20 PM), https://perma.cc/8UDH-NH3J ("The list of victims of the cyberespionage, which already included the Treasury and Commerce departments, is expected to grow and to include more federal agencies and numerous private

The Energy Department and the National Nuclear Security Administration, responsible for maintaining the United States nuclear stockpile, reported their business networks being comprised in the attack.[134] Security teams realized that their initial relief that they had not used the compromised systems turned to panic when they realized which third-party applications had also been compromised.[135] The attackers used stealth and several tactics to fly under the radar of detection. They used United States based internet addresses, timed their intrusions during working hours, and other careful acts to avoid raising alarms.[136] It is believed that this is the gravest cyberattack against the United States in years and it can take months to determine which technology supply chains and networks were compromised.[137]

companies . . . ."); David E. Sanger & Nicole Perlroth, *More Hacking Attacks Found as Officials Warn of "Grave Risk" to U.S. Government*, N.Y. TIMES (Dec. 17, 2020), https://perma.cc/2VY3-GV67 (last updated Jan. 5, 2021) [hereinafter *More Hacking*] (noting that the attack appeared to extend "beyond nuclear laboratories and the Pentagon, Treasury, and Commerce Department systems").

[134] *See More Hacking*, *supra* note 133.

[135] *See id.*

[136] *See* David E. Sanger et al., *Billions Spent on U.S. Defenses Failed to Detect Giant Russian Hack*, N.Y. TIMES (Dec. 16, 2020), https://perma.cc/Z2WN-TUF5 (last updated Jan. 2, 2021) [hereinafter *Billions*]. The hacked company at the heart of this, SolarWinds, has had unstable security measures as employees' passwords were leaked last year. *See* Raphael Satter et al., *Hackers Used SolarWinds' Dominance against It in Sprawling Spy Campaign*, REUTERS (Dec. 15, 2020, 8:05 PM), https://perma.cc/L939-URPC (emphasizing the vulnerability of SolarWinds' security prior to the attack). SolarWinds' update server password was "solarwinds123." *Id.*

[137] *See More Hacking*, *supra* note 133. This will not be limited to only the United States government as banks and Fortune 500 companies also use the network management tool from Orion. *See Billions*, *supra* note 136. This is not the first major hack against a United States federal organization. The United States Office of Personnel Management was hacked in 2014 and security-clearance files on 22.5 million Americans, and 5.6 million sets of fingerprints, were taken without detection. *See* David Alexander, *5.6 Million Fingerprints Stolen in U.S. Personnel Data Hack: Government*, REUTERS (Sept. 23, 2015, 10:50 AM), https://perma.cc/65LH-BP3U; Ellen Nakashima, *Hacks of OPM Databases Compromised 22.1 Million People, Federal Authorities Say*, WASH. POST (July 9, 2015, 7:33 PM), https://perma.cc/9EFT-7UB5 (noting that

2. Untrustworthy Data

Attacks on machine-learning is an acute worry of cybersecurity experts.[138] The technology can manufacture alternative data sets (images, music, speech, and even dialogue) that can be merged with actual, real-world data in legal systems with little to no oversight.[139] This false data should leave attorneys, judges, and juries uneasy. Scientists have discussed adversarial issues in machine-learning,[140] but an adversary's goals are not always clear. There are three broad attack categories: quality, confidence, and evasion attacks.[141] Quality attacks are used to drive down the effectiveness of machine-learning from the training data, to likely convince the code not to execute an effective analytic or cause it to waste time trying to improve itself.[142] Confidence attacks are used to decrease the effectiveness, without impacting the accuracy of the training data.[143] Evasion attacks are designed to concoct specific outcomes for future test samples.[144] Machine-learning is vulnerable to attack because of two necessary assumptions, that the test data is the same as the training data, and that the "ground-truth" labels used in the training data are accurate.[145]

hackers exposed sensitive information of over twenty-two million between two major breaches of U.S. government databases in 2014).

[138] *See* Harold Kilpatrick, *The Malicious Use of Artificial Intelligence in Cybersecurity*, SECUREAGE (Aug. 24, 2018), https://perma.cc/6WQW-N6TE (recognizing machine-learning as an area of particular concern in cybersecurity).

[139] *See id.* (defining machine-learning).

[140] *See* Kegelmeyer, *supra* note 70, at 12 (citing Amir Globerson & Sam Roweis, *Nightmare at Test Time: Robust Learning by Feature Deletion*, ICML '06: PROC. 23RD INT'L CONF. ON MACHINE LEARNING 353–60 (2006)) (asserting that adversarial aspects of machine-learning have been discussed for over a decade).

[141] *See id.* An "adversary" refers to a malicious entity whose aim is to prevent the security measure. *See* Jeremiah Blocki, *Adversary Attacks*, CARNEGIE MELLON UNIV. SCH. COMPUT. SCI., https://perma.cc/THB4-KPJH. It is presented in the idea of a "game" between the user and the system trying to attack it. *See id.*

[142] *See* Kegelmeyer, *supra* note 70, at 12.

[143] *See id.*

[144] *See id.*

[145] *See id.* at 12–13.

Machine-learning is compromised when the data pool is poisoned, inserting fraudulent data or code, generating false positives or false negatives.[146] There are adversarial examples of deep neural networks that can be attacked on pre-existing audio and visual recognition models.[147] Authors Nicolas Carlini and David Wagner discovered that there was a way that they could insert adversarial examples into the speech-to-text systems through slight distortions.[148] These authors argue that targeted adversarial examples exist in the audio domain by attacking a state-of-the-art speech-to-text transcription neural network (DeepSpeech).[149] To do this, the authors conducted an experiment where they embedded speech into audio that typically should not be recognizable as speech.[150] By choosing silence as the "target," Carlini and Wagner were able to hide audio from a speech-to-text system.[151] With a 100 percent success rate, Carlini and Wagner were able to turn any audio

[146] *See* Kilpatrick, *supra* note 138 (listing different malicious uses for machine-learning).

[147] *See generally* Nicolas Papernot & Patrick McDaniel, *Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning*, PA. STATE UNIV. DEP'T OF COMPUT. SCI. & ENG'G (Mar. 13, 2018), https://perma.cc/X4MM-7YXV (PDF) (outlining how the lack of robustness in adversarial settings leads to complications in the AI predictions). Also, Mainuddin Ahmad Jonas and David Evans presented how adversarial attacks on image classifications can be hidden in layers, leading to adversarial attacks. Mainuddin Ahmad Jonas & David Evans, *Poster: Enhancing Adversarial Example Defenses Using Internal Layers*, INST. ELEC. & ELECS. ENG'RS SYMP. ON SEC. & PRIV. (2018), https://perma.cc/TR7F-JC7P (PDF) [hereinafter *Enhancing Adversarial Example Defenses*].

[148] *See* Chris Edwards, *Hidden Messages Fool AI*, 62 COMMC'NS ACM 13, 13–14 (Jan. 2019) (citing Nicholas Carlini & David Wagner, *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, INST. ELEC. & ELECS. ENG'RS SYMP. ON SEC. & PRIV. (2018), https://perma.cc/7HK4-5YFB (PDF)) (explaining Carlini and Wagner's discovery of how to translate adversarial examples into speech-to-text systems).

[149] *See id.* (describing Carlini and Wagner's attack on the DeepSpeech engine published as open-source code).

[150] *See id.* ("Rather than using noise to confuse the system, [Carlini] had found the engine was susceptible to slightly modified recordings of normal speech or music.").

[151] *See id.*

waveform into any target transcription.[152] The authors also found that it was possible to hide speech inside audio by adding adversarial noise to cause the speech-to-text transcription neural network to transcribe nothing; music can be transcribed as arbitrary speech, and audio can transcribe up to fifty characters per second, hiding false data.[153]

Google scientist, Christian Szegedy, studied deep neural networks and discovered intrinsic blind spots in the deep neural networks that were present despite a robust design.[154] There were contradictions in the neural network's capability to perform accurately. Despite the neural net's high ability to perform and seemingly robust design, it had difficulty performing when given adversarial examples that had been generated by another neural network.[155]

The speech could be hidden so that a deep neural network, trained through adversarial machine-learning, might convert it to a false data source, and none would be the wiser. The hack happening at the creation of training sets that are used to train the generator and fool the discriminator, could lead to output data made from hacked means and then wrongly deciding algorithms. Moreover, these methods of hacks are the types that are know right now, and do not account for the other methods that could be in use. The complexity of generative adversarial networks is that the data breaches happened in such a discreet, purposeful, undetectable manner that it takes multiple layers of examination to discover precisely why the attack was initiated, where the attack occurred, and who perpetrated the attack.

[152] *See id.* ("The attacks buried subtle glitches and clicks in the speech or music at a level that makes it hard for a human hearing the playback to detect.").

[153] *See id.* For more ideas on how these adversarial examples work, Mainuddin Ahmad Jonas and David Evans outlined how internal layer information in deep neural networks can help us gain insights into the nature of adversarial examples, to provide insight into improving defenses. *See Enhancing Adversarial Example Defenses*, *supra* note 147.

[154] Christian Szegedy et al., *Intriguing Properties of Neural Networks*, INT'L CONF. ON LEARNING REPRESENTATIONS 2 (2014), https://perma.cc/G5XF-5SC8 (PDF).

[155] *See id.* (explaining that if one neural net is used to generate a set of adversarial examples, those examples are challenging for another neural network to perform).

3. Disparate Outcomes

Data is only as good as the people who produce it and the security that protects it. Bad actors can manufacture results that do not serve their intended goals or targets but have a second-order effect[156] that will cause others to suffer either as a result of a backlash or policy shift. This technology is not as advanced as is purported. In 2018, the Pentagon's Defense Advanced Research Projects Agency (DARPA) announced a $2 billion commitment to the development of "third-wave" AI tools that would show reasoning and contextual awareness—common sense.[157] Technology that lacks the common sense of an eighteen-month-old[158] should not determine people's access to due process. AI has recently had two very public incidents of disparate outcomes in the last year. In 2019, Apple Credit Card's reliance on their technology led to disparate credit limits being issued to husbands and wives with the same credit history, sometimes, with a variable differential of 20:1 limit compared to the other.[159] In the same year, United Health was

[156] *See* RADEK SILHAVY, ARTIFICIAL INTELLIGENCE METHODS IN INTELLIGENT ALGORITHMS 302 (2019) (detailing the harmful effects that information and communications technology can have on the environment).

[157] *See* Mark Jones, *DARPA Wants to Give AI Common Sense Using Child Psychology*, TECHHQ (Nov. 2, 2018), https://perma.cc/UM3M-6WCP (exploring DARPA's multi-year commitment to developing common sense in AI).

[158] At this time, machine-learning does not have the "common sense" of an eighteen-month-old human child. *See* Alison Gopnik, *The Ultimate Learning Machines*, WALL ST. J. (Oct. 11, 2019, 11:00 AM), https://perma.cc/VYV5-57N3 ("Babies seem to learn much more general and powerful kinds of knowledge than AIs do, from much less and much messier data."); Jack Corrigan, *DARPA Wants to Build Computers with "Common Sense,"* NEXTGOV (Oct. 26, 2018), https://perma.cc/K8PJ-G8E6 (noting that DARPA will compare AI tools against the cognitive abilities of children ages zero to eighteen months); Melanie Mitchell, *AI Can Pass Standardized Tests—but It Would Fail Preschool*, WIRED (Sept. 10, 2019, 6:00 AM), https://perma.cc/FS5M-F7RT (arguing that DARPA seems "quite far" from developing an AI system with the common sense of an eighteen-month-old).

[159] *See* Neil Vigdor, *Apple Card Investigated after Gender Discrimination Complaints*, N.Y. TIMES (Nov. 10, 2019), https://perma.cc/SG8L-64MJ (writing about an upset Twitter user whose Apple Card spending limit was twenty times higher than his wife's, despite her higher credit score). Goldman Sachs said it would reevaluate credit limits on a case-by-case basis. *See* Kif Leswing, *Goldman Sachs Will Reevaluate Apple Card Credit Limits After Bias*

accused of using an algorithm that led to sicker Black Americans receiving differential medical treatment than less ill White persons,[160] even though a robust body of law prohibits this type of unconstitutional outcome.[161] These examples show that the implementation of the technology requires vigilant and constant oversight before it is allowed to be used on the public.

One of the problems of AI is that false data results and racist data—implicit and intentional—may also be produced, leading to disparate outcomes. "Tay" was an AI bot designed and released by Microsoft Corporation in 2016.[162] It was designed to mirror the language patterns of a nineteen-year-old girl and programmed to learn all human interactions by interacting with people via Twitter.[163] Within sixteen hours, Tay had to be shut down due to several racist, anti-Semitic, and sexually charged messages in response to other Twitter users.[164] Microsoft had not been able to determine if Tay's racist responses were based on a "repeat after me" capability (which may or may not be a

Allegations, CNBC (Nov. 11, 2019, 7:26 PM), https://perma.cc/3MZF-KMTP. Some experts are arguing that the disparate outcomes are not based on the algorithm coding errors. *See, e.g.*, Diane Harris, *Apple Card Gender Bias? Don't Assume Its Discrimination, Experts Warn*, NEWSWEEK (Nov. 12, 2019, 6:18 AM), https://perma.cc/U49N-P75J (considering various possible causes of the disparities).

[160] Melanie Evans & Anna Wilde Mathews, *New York Regulator Probes UnitedHealth Algorithm for Racial Bias*, WALL ST. J. (Oct. 26, 2019, 7:00 AM), https://perma.cc/8MAL-6QSM ("New York's insurance regulator said it is launching an investigation into a UnitedHealth Group Inc. algorithm that a study found prioritized care for healthier white patients over sicker black patients.").

[161] A letter to UnitedHealth by the Department of Financial Services and Department of Health outlined that "the N.Y. Insurance Law, N.Y. Human Rights Law, N.Y. General Business Law, and federal Civil Rights Act [all] protect against discrimination for certain classes of individuals." Letter from Linda Lacewell, Superintendent, N.Y. State Dep't of Fin. Servs. & Howard Zucker, Comm'r, N.Y. State Dep't of Health, to David Wichmann, Chief Exec. Officer, UnitedHealth Grp. Inc. (Oct. 25, 2019), https://perma.cc/N3KS-KYFB (PDF) [hereinafter Letter from Linda Lacewell].

[162] *See* Peter Bright, *Tay, the Neo-Nazi Millennial Chatbot, Gets Autopsied*, ARS TECHNICA (Mar. 25, 2016, 7:15 PM), https://perma.cc/FR6N-JVVU (stating that Microsoft created Tay in 2016 to replicate a similar Chinese bot).

[163] *Id.*

[164] *See id.* (listing examples of the tweets).

built-in feature), whether it was a learned response, or if it was an example of complex behavior that cannot be traced to code.[165] Users told Tay that the Holocaust did not occur.[166] Tay did not know what the Holocaust was (only that it was a proper noun), and since users told Tay it did not happen, Tay then proceeded to operate and evolve with that base understanding.[167]

Peter Lee, Microsoft's corporate vice president of Microsoft Research, apologized and acknowledged that the testing done did not accurately prepare for the fact that the public would actively seek to destabilize and attack the bot.[168] Caroline Sinders, a machine-learning designer and Fellow of the Digital Harvard Kennedy School, documented concerns about machine-learning and bots and how they learn.[169] She noted that AI must be trained using a body of data, and this corpus must be sorted through knowledge trees that direct a question or type of question to a pre-formed answer.[170] However, Microsoft did not restrict specific pre-imagined queries from being directed to certain outcomes.[171] While they did code for sensitive topics such as Eric Garner's murder by law enforcement, it did not have hard-coded responses to particular terms such as rape, domestic violence, or Holocaust denials.[172] This lack of pre-thought is a fundamental design flaw, as things like racism, privacy, and danger continue to be embedded in AI.[173] The level of cyberattacks that can happen (not in the public eye, as with Tay, but deep within the design of the AI) is

[165] *Id.*

[166] *Id.*

[167] *See id.* (explaining that Tay was incapable of recognizing why users would lie to her about the Holocaust).

[168] *Id.*

[169] Caroline Sinders, *Microsoft's Tay Is an Example of Bad Design*, MEDIUM (Mar. 24, 2016), https://perma.cc/Y4Y6-X6WX.

[170] *Id.*

[171] *Id.*

[172] *Id.*

[173] Facebook deals with this as it relates to requests to remove false information from its social media platform. *See Working to Stop Misinformation and False News*, FACEBOOK (2017), https://perma.cc/WY3S-PHJY (outlining the steps Facebook is taking to combat false news).

uncertain, and the most concerning emerging issue in cybersecurity today.[174]

Additionally, there is growing alarm within the legal community about the unintended outcomes from relying on AI created outcomes. For instance, "Poverty Lawgorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities" is a guide authored by Michele Gilman that assists lawyers who advocate for low-socioeconomic clients about the hidden ways AI has been imbedded in family law, housing, workers' rights, immigration surveillance, public benefits, schools and education, and consumer law.[175] These efforts by attorneys[176] remind all that the legal community must be vigilant in reviewing all instances of third-party software and applications used by United States systems.

D. *Prevention Efforts by Scientists Are Insufficient*

Bad data—whether found or created—affects the outputs, and the AI will learn inaccurate results, producing disastrous outcomes.[177] Software designers are continually working to ensure efficiency, accuracy, transparency, and accountability.[178] It is humorous to read stories of how AI mistakes a cat for

[174] *See* Kegelmeyer, *supra* note 70, at 14 (acknowledging that adversarial machine-learning is a new and developing field but contending that the dangers of cyberattacks require continued use of machine-learning methods).

[175] Michele Gilman, *Poverty Lawgorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities*, DATA & SOC'Y (Sept. 15, 2020), https://perma.cc/4H4K-LR9G (PDF).

[176] *See* Karen Hao, *The Coming War on the Hidden Algorithms That Trap People in Poverty*, MIT TECH. REV. (Dec. 4, 2020), https://perma.cc/4ECR-QDXB (explaining how attorneys are fighting the automated systems that deny the poor housing, jobs and basic services).

[177] *See* Thomas C. Redman, *If Your Data Is Bad, Your Machine Learning Tools Are Useless*, HARV. BUS. REV. (Apr. 2, 2018), https://perma.cc/3PHW-LAD2 ("The quality demands of machine learning are steep, and bad data can rear its ugly head twice both in the historical data used to train the predictive model and in the new data used by that model to make future decisions.").

[178] *See* Darrell M. West & John R. Allen, *How Artificial Intelligence Is Transforming the World*, BROOKINGS INST. (Apr. 24, 2018), https://perma.cc/J4AC-AESD (showing how software designers can anticipate problems and analyze specific issues within AI systems).

guacamole[179] or a turtle for a rifle.[180] However, the same AI can make life-threatening mistakes. It can wrongly decide that a patient does not need medical care based on faulty data sets.[181] It can wrongly decide that a wife will be a higher credit risk than her husband based on faulty data sets.[182] It can wrongly determine that a defendant has a higher likelihood of recidivism based on faulty data sets.[183] Left unsupervised, unregulated AI directly impacts access to life, liberty, the pursuit of happiness, and fairness in the legal system.[184]

[179] *See* Jonathan Zittrain, *The Hidden Costs of Automated Thinking*, NEW YORKER (July 23, 2019), https://perma.cc/W55L-MZT4 (describing a machine-learning model that became 99.99 percent sure it was given a photograph of guacamole, even though the photograph was a cat to human eyes).

[180] *See* Anish Athalye et al., *Synthesizing Robust Adversarial Examples*, PROC. 35TH INT'L CONF. ON MACH. LEARNING (June 7, 2018), https://perma.cc/N5FQ-PXPG (PDF) (describing a study in which an algorithm consistently classified poses of a 3D-printed turtle as a rifle); Kim Martineau, *Why Did My Classifier Just Mistake a Turtle for a Rifle?*, MIT NEWS (July 31, 2019), https://perma.cc/P4S2-KBP5.

[181] United Health was accused of using an algorithm that led to sicker Black Americans receiving medical treatment less often than less ill White persons. *See* Evans & Mathews, *supra* note 160. "New York's insurance regulator said it is launching an investigation into a UnitedHealth Group Inc. algorithm that a study found prioritized care for healthier white patients over sicker black patients." *Id.*

[182] Apple Credit Card led to disparate credit limits being issued to husbands and wives with the same credit history, sometimes at the magnitude of a twenty times difference. Vigdor, *supra* note 159. Goldman Sachs said it would reevaluate credit limits on a case-by-case basis. *See* Leswing, *supra* note 159. Some experts are arguing that the disparate outcomes are not based on the algorithm coding errors. *See* Harris, *supra* note 159.

[183] *See* Cynthia Rudin et al., *The Age of Secrecy and Unfairness in Recidivism Prediction*, 2.1 HARV. DATA SCI. REV. 1, 2–4 (2020) (discussing the lack of transparency and data inconsistencies in predictive modeling in criminal justice databases).

[184] Amazon has created the leading facial recognition software called "Rekognition," which Amazon has advertised and promoted to police agencies for use in criminal investigations. *See* John Warner, *If You're Worried Artificial Intelligence Is Coming for You, Read Melanie Mitchell's New Book*, CHI. TRIB. (Nov. 4, 2019, 7:00 AM), https://perma.cc/8KZZ-L4N9. The American Civil Liberties Union (ACLU) investigated and discovered that this software has been leading to the misidentification of people across the nation, as the software incorrectly connected New England professional athletes to mugshot databases. *Facial Recognition Technology Falsely Identifies Famous Athletes*, ACLU MASS. (Oct. 21, 2019, 2:00 PM), https://perma.cc/RC77-HJL9.

Programmers and scientists, and cybersecurity experts are continually working on methods to protect data and systems.[185] Furthermore, scholars are studying how to use AI to defeat nefarious AI.[186] However, AI developers are taught to let efficiency be a driving force.[187] Though efficiency allows for a smooth system,[188] it can flatten human thought, producing what one expects and wants to see, rather than what should be produced.[189] Unexpectedly, technology has crossed into an area that coders did not adequately anticipate in the design phase: the implications of a self-determining code devoid of human, subconscious norms. As Professor Jon Kleinberg, a computer scientist from Cornell University, explains:

We have, perhaps, for the first time, built machines we do not understand. . . . [B]ecause they act like us, it would be reasonable to imagine that they think like us too. But the reality is that they do not think like us at all; at some deep level, we

Nearly 17 percent of the athletes were falsely identified, and an independent computer science expert verified the results. *Id.* A similar test completed by the ACLU of California, found the misidentification of twenty-eight sitting members of Congress with a disproportionate number of the false matches being people of color. Steven Melendez, *Amazon's Face-Recognition Tool Falsely Matched California Lawmakers to Mugshot*s, *ACLU Says*, FAST CO. (Aug. 14, 2019), https://perma.cc/U46W-A6BA.

[185] For examples, see Bocetta, *supra* note 112; *The Threat of AI-Powered Cyberattacks Looms Large*, AI BUS. (Sept. 25, 2019), https://perma.cc/86CL-2WKU; Ramsés Gallego, *AI and Security: Machine Learning Is a Threat Detection Game-Changer*, TECHBEACON (July 5, 2020), https://perma.cc/XJE8-ZG3K.

[186] *See generally* John Leyden, *AI-Powered Honeypots: Machine Learning May Help Improve Intrusion Detection*, DAILY SWIG (May 11, 2020), https://perma.cc/PE22-VWTX.

[187] *See* FOER, *supra* note 24, at 71 ("When programmers are taught algorithmic thinking, they are told to venerate efficiency as a paramount consideration.").

[188] *See id.*

[189] *See id.* at 70 (arguing that algorithms remove humans from the whole process of inquiry). "Data, like victims of torture, tell its interrogator what it wants to hear." *Id.* at 71. Technology and culture writer Nicholas Carr stated, "The more time we spend immersed in digital waters, the shallower our cognitive capabilities become due to the fact we ceased exercising control over our attention: 'The Net is by design an interruption system, a machine geared for dividing attention.'" SCHWAB, *supra* note 25, at 101–02 (citing NICHOLAS CARR, THE SHALLOWS: WHAT THE INTERNET IS DOING TO OUR BRAINS (2011)).

don't even really understand how they're producing the behavior we observe. This is the essence of their incomprehensibility.[190]

There are efforts for AI to be more directed in their creation by expanding the types of networks created or in new fields of study, as at Stanford[191] and the Massachusetts Institute of Technology.[192]

However, scientists understand they are barely at the level of even knowing that an attack did occur. Cybersecurity has become less dependable,[193] and scientists patch the breaches, as opposed to fortifying all points of entry for a hack.[194] Scholars examining this work understand there must be a sense of "watchful paranoia."[195] But, this hacked AI can slip into technology used in the criminal justice system. Moreover, even the most expeditious committees in the U.S. House of Representatives will be alarmingly too late, considering that this technology is already being used to decide people's liberty interests in the criminal justice system.[196]

[190] Jon Klienberg & Sendhil Mullainathan, *We Built Them, But We Don't Understand Them*, EDGE (Jan. 21, 2015), https://perma.cc/MB9W-MCFY.

[191] *See* Ethan Baron, *Stanford Unveils New AI Institute, Built to Create 'A Better Future for All Humanity'*, MERCURY NEWS (Mar. 18, 2019, 5:07 PM), https://perma.cc/CBA7-5R9M (describing a new institute dedicated to using AI to build the best-possible future).

[192] *See* Terri Park, *Advancing Artificial Intelligence Research*, MIT NEWS (Nov. 18, 2020), https://perma.cc/492M-B4HL (discussing a new collaboration that awards funding to projects that target the advancement of trustworthy AI, enhancing human cognition in complex environments, and AI for everyone).

[193] *See* Jack Wallen, *10 Cybersecurity Stories in 2019 That Make Us Feel Less Secure*, TECHREPUBLIC (Dec. 15, 2019, 1:10 PM), https://perma.cc/JDM7-K3RF (discussing the memorable security threats in 2019).

[194] *See id.* (explaining that for many threats the initial point of entry needs serious vetting and security which includes a level of risk many businesses are not willing to take).

[195] Kegelmeyer, *supra* note 70, at 14.

[196] House Resolution 153 was referred to committee on February 27, 2019, to develop guidelines for the ethical development of artificial intelligence. H.R. Res. 153, 116th Cong. (2019).

II. RISK ASSESSMENT INSTRUMENTS IN CRIMINAL JUSTICE ARE NOT SECURE

The Sentencing Reform Act of 1984,[197] part of the Comprehensive Crime Control Act of 1984,[198] had a seemingly pure motive: to increase consistency in U.S. federal sentencing by decreasing recidivism.[199] Unfortunately, these reforms were made in a haphazard method that was not evidence-based nor vetted against existing data.[200] Stopping crime with technology has become a lucrative industry.[201] And since then, courts and correction departments have been using algorithms to determine a defendant's "risk" of not appearing for court appearances. These algorithms have been used in determining bail, sentencing, and parole.[202] Jurisdictions are beginning to analyze the code's efficacy and accuracy. [203]

[197] Pub. L. No. 98-473, 98 Stat. 1987 (codified as amended in scattered sections of 18 and 28 U.S.C.).

[198] Pub. L. No. 98-473, 98 Stat. 1976 (codified as amended in scattered sections of 18 U.S.C.).

[199] *See* Charles Summers & Tim Willis, *Pretrial Risk Assessment Research Summary*, BUREAU JUST. ASSISTANCE 1 (Oct. 18, 2010), https://perma.cc/G6AQ-3RRX (PDF) [hereinafter *Pretrial Risk Assessment Research Summary*].

[200] Some have studied that arguments made in murder and violent crime decreases were causally linked to higher mandatory minimums put into effect in the 1980s. *See* Doris Layton Mackenzie et al., *Sentencing and Corrections in the 21st Century: Setting the Stage for the Future*, NAT. CRIM. JUST. REFERENCE SERV. (July 2001), https://perma.cc/5SJH-MCUK (PDF).

[201] *See* Griffard, *supra* note 5, at 48 (noting how predictive policing has developed into a "multi-million dollar business" (citing Andrew G. Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1131–32 (2017)).

[202] *See* Alex Chohlas-Wood, *Understanding Risk Assessment Instruments in Criminal Justice*, BROOKINGS INST. (June 19, 2020), https://perma.cc/8CW2-PA4F (discussing algorithmic tools designed to predict the risk that the defendant will fail to appear in court).

[203] *See* Kim Steven Hunt & Robert Dumville, *Recidivism Among Federal Offenders: A Comprehensive Overview*, U.S. SENT'G COMM'N (Mar. 2016), https://perma.cc/QN5M-CLEM (PDF). The federal government acknowledged it still needed to complete a study on the scores produced by PRAIs. Eric Holder, U.S. Att'y Gen., Speech Presented at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference, Philadelphia, PA (Aug. 1, 2014).

The objective of predictive policing tools is to reduce criminal activity in a community.[204] Various tools can outline where crime has occurred but also predict potential crime. Pretrial Risk Assessment Instruments (PRAIs) purport to assist courts in predicting future behavior of defendants related to recidivism risks and failure to appear at trial; PRAIs are used in almost every state.[205] Over sixty risk-assessment tools are being used in the criminal justice system, combining variables such as demographics, family background, and additional factors related to criminal history and psychological and sociological considerations.[206] Risk-assessment tools generally outperform expert opinion by about 10 percent[207] and are seen as not substantially distinguishable from the human error rate of judges and parole boards.[208] However, the use of the tool

[204] *See* Tim Lau, *Predictive Policing Explained*, BRENNAN CTR. FOR JUST. (Apr. 1. 2020), https://perma.cc/N6SC-RT72 (stating the predictive policing is designed to identify where to deploy police or to identify people who are more likely to commit a crime).

[205] *See Algorithms in the Criminal Justice System: Risk Assessment Tools*, ELEC. PRIV. INFO. CTR., https://perma.cc/RX7W-ZS6S (last updated Feb. 2020).

[206] *See* Alyssa M. Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303, 309 (2017). For a sampling of state statutes regulating risk assessment instruments in criminal justice, see IDAHO CODE ANN. § 19-1910 (2021); OHIO REV. CODE ANN. § 5120.111 (LexisNexis 2021); N.J. STAT. ANN. § 2A:58C-2 (West 2021); OHIO REV. CODE ANN. § 5120.115 (LexisNexis 2021); ALASKA STAT. ANN. § 33.07.020 (2021); ARIZ. REV. STAT. ANN. § 5-201 (2021); CAL. PENAL CODE § 1320.25 (West 2021); DEL. CODE ANN. tit. 11, § 2104 (2021); 725 ILL. COMP. STAT. 5/110-6.4 (2021); MD. CODE ANN. MD. RULES 4-216.1 (LexisNexis 2021); N.J. STAT. ANN. § 2A:162-25 (West 2021); UTAH CODE ANN. § 78A-6-124 (LexisNexis 2021); UTAH CODE JUD. ADMIN. 3-116 (LexisNexis 2021); VT. STAT. ANN. tit. 13, § 7554C (2021); W. VA. CODE ANN. § 5A-5-7 (LexisNexis 2021); COLO. REV. STAT. ANN. § 17-22.5-404 (West 2021); GA. CODE ANN. § 42-9-45 (2021); HAW. REV. STAT. ANN. § 706-670 (LexisNexis 2021); N.D. CENT. CODE ANN. § 12.1-01-04 (West 2021); OKLA. STAT. ANN. tit. 57, § 332.21 (West 2021); 42 PA. CONS. STAT. § 2154.7 (2021); VA. CODE ANN. § 17.1-803 (2021).

[207] Kia Rahnama, *Science and Ethics of Algorithms in the Courtroom*, 5 U. ILL. J.L. TECH. & POL'Y 169, 175 (2019) (citing Anna Maria Barry-Jester et al., *The New Science of Sentencing*, MARSHALL PROJECT (Aug. 4, 2015, 7:15 AM), https://perma.cc/Z988-NTHJ).

[208] *See id.* at 175–76 (noting that "the underlying truth [is] that algorithms will . . . be designed and created by people who inevitably hold value-laden presumptions and intuitions is in escapable").

varies across most of the United States jurisdictions with court approval.[209]

The first goal of PRAIs is to ensure that pretrial decisions are more consistent across jurisdictions.[210] There are no standardized metrics for risk assessments, so defendants are categorized based on subjective judgments of pretrial officers, which can result in inconsistent, disparate, and potentially arbitrary recommendations in contrast to the intent of the Bail Reform Act of 1984[211] and pretrial recommendations from the American Bar Association,[212] the National Association of Pretrial Services Agencies[213] and the National Institute of Justice.[214] The second goal is to maximize the number of successful pretrial decisions.[215] This goal is achieved "by maximizing the number of defendants who are released before they are tried, without negatively affecting appearances and court rates or public safety."[216] The factors that were associated most with termination for pretrial risk were related to the nature of the charges pending at the time of the arrest, the history of criminal arrests and convictions and active community supervision at the time of the arrest, history of failure to appear, history of violence, residential stability, employment civility, community ties, and substance abuse.[217] For instance, the PRAI used in New Orleans was based on the most extensive, most diverse set of pretrial records ever assembled—750,000 cases from nearly three hundred

[209] *See* State v. Headley, 926 N.W.2d 545, 552–53 (Iowa 2019) (holding that it was within the discretion of the trial court to consider risk assessment tools on their face if it was used in a presentence investigation report and its use did not violate the defendant's due process rights).

[210] *See Pretrial Risk Assessment Research Summary*, *supra* note 199, at 1.

[211] 18 U.S.C §§ 3141–3156

[212] *See Pretrial Risk Assessment Research Summary*, *supra* note 199, at 1.

[213] *See id.*

[214] *See id.*

[215] *Id.* at 2.

[216] *Id.*

[217] *See id.* (citing Marie VanNostrand & Kenneth J. Rose, *Pretrial Risk Assessment in Virginia*, VA. DEP'T CRIM. JUST. SERVS. (May 1, 2009), https://perma.cc/32R4-FF9R (PDF)).

jurisdictions.[218] Research shows that actuarial risk assessment instruments can provide some predictive benefits for pretrial decisions.[219] It is noteworthy that although PRAIs state they do not explicitly rely on factors such as race, ethnicity, or geography,[220] variables such as "risk" can be a proxy for race.[221]

State legislation related to the regulation of risk assessment instruments is varied and, in some instances, vague. States use a variety of terms such as "risk assessment instruments" ("RAI") and "risk assessment tools."[222] Very few of these statutes explicitly state that the instrument they are using is digital. Some describe it as a "worksheet" while others simply do not address the question.[223] Some states only list specific departments using their RAI (i.e., juvenile detention, probation office), while others only list information at a local district level.[224] Two commonly used products are COMPAS and the Arnold Tool. This section analyzes these two products.

[218] Shelbi Flynn et al., *Pretrial Risk Assessment: The Use of Evidence-Based Assessment Tools During Bond Setting*, CITY OF NEW ORLEANS, https://perma.cc/3WCY-ECVS.

[219] *See Pretrial Risk Assessment Research Summary*, *supra* note 199, at 4.

[220] MATTHEW DEMICHELE ET AL., THE INTUITIVE-OVERRIDE MODEL: NUDGING JUDGES TOWARD PRETRIAL RISK ASSESSMENT INSTRUMENTS 24 (2018), https://perma.cc/7KET-8854 (PDF).

[221] *See* Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237, 238 (2015).

[222] *See* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1780–81 (2018) (describing nationwide adoption of such tools).

[223] *See* Brian Netter, *Using Group Statistics to Sentence Individual Criminals: An Ethical and Statistical Critique of the Virginia Risk Assessment Program*, 97 J. CRIM. L. & CRIMINOLOGY 699, 701 (2007) (explaining that the Virginia system relies upon "simple worksheets that tally demerits for past crimes with additional penalties for demographic characteristics found to be correlated with the commission of crime"). I opted not to include any statutes that only say something to the effect of "we use an RAI" with no further detail. Just because a state is not represented does not necessarily mean they are not using some form of AI.

[224] The few statutes that go into detail about their standards require checking every five years for accuracy. Indiana's statute notes explicitly that rules will be adopted for RAI standards "before January 2020." IND. CODE § 35-33-8-0.5 (2021).

COMPAS is an acronym for Correctional Offender Management Profiling for Alternative Sanctions.[225] COMPAS was developed by Equivant (formerly Northpoint)[226] and is used throughout the United States to determine pretrial detention, sentencing, or probation and parole.[227] The COMPAS scale is an algorithmically determined assessment that claims to predict an individual's risk of recidivism.[228] The COMPAS assessment roughly estimates the risk of recidivism using several variables. Combining data from interviews with the offender,[229] information derived from the offender's criminal history and observations of the person,[230] and other unknown factors, COMPAS derives a score to represent a defendant's likelihood of recidivism or potential behavior while incarcerated.[231] COMPAS provides users with a pretrial release risk scale,[232] general recidivism scale,[233] and a violent recidivism scale.[234]

The Arnold Tool is another PRAI that has gained national attention. The Arnold Tool was created by Arnold Ventures (formerly The Laura and John Arnold Foundation) and has been implemented in more than forty jurisdictions.[235] The Arnold Tool is implemented differently depending on the jurisdiction. For instance, the New Jersey Constitution states that the Arnold Tool is to calculate the defendant's dangerousness, history of failure to appear, and obstruction of the criminal

[225] *See* Ed Yong, *A Popular Algorithm Is No Better at Predicting Crimes than Random People*, ATLANTIC (Jan. 17, 2018), https://perma.cc/BD4A-MHP8.

[226] *See id.*

[227] *See id.*

[228] *See* NORTHPOINTE, PRACTITIONER'S GUIDE TO COMPAS CORE 27 (2015).

[229] For a sample of the COMPAS Risk Assessment questions, see Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://perma.cc/925Z-W289.

[230] Such as criminal involvement, relationships and lifestyle, personality and attitudes, family, and social exclusion. *See Algorithms in the Criminal Justice System: Risk Assessment Tools*, *supra* note 205, at 26.

[231] *See* NORTHPOINTE, *supra* note 228, at 27.

[232] *See id.*

[233] *See id.* at 26.

[234] *See id.* at 28.

[235] *See, e.g.*, *Public Safety Assessment for Pretrial Release and Detention*, N.M. COURTS, https://perma.cc/9ULH-2NTA.

justice.[236] By comparison, New Mexico has a constitutionally mandated implementation that does not require a calculation of a defendant's failure to appear or obstruct justice.[237] Jurisdictions also differ in the steps the courts must follow before rendering a verdict, such as four additional steps in Arizona, five steps in Santa Cruz County, California, and in New Jersey, ten additional steps that must be taken.[238]

Nearly 120 criminal justice organizations have called for a halt to the use of all PRAIs,[239] and district attorney associations and criminal defense organizations agree that risk assessment tools are opening unintended consequences that leave defendants and victims without recourse.[240] Civil rights organizations note six areas of concern regarding the use of algorithmic decision-making tools: lack of transparency, lack of accuracy, failure to provide the necessary information, the perpetuation of racial bias and discrimination, subjective interpretation by decision-makers, and measurement of group risk instead of individual risk.[241] COMPAS came to the public's

[236] *See* RAUL TORREZ & DIANNA LUCE, MINORITY REPORT: AD HOC COMMITTEE TO REVIEW PRETRIAL RELEASE AND DETENTION PROCEDURES 4 (2020), https://perma.cc/FZ4G-YUV9 (PDF) [hereinafter MINORITY REPORT] (citing N.J. STAT. ANN. § 2A:162-18 (West 2020)).

[237] *See id.* (citing N.M. CONST. art. II, § 13).

[238] *See id.* at 5 (comparing jurisdictions). New Mexico requires no additional steps. *See id.* The University of New Mexico Institute for Social Research published a response to the Minority Report outlining points of agreement and dissension about the interpretation of the data gathered. UNIV. N.M., INST. FOR SOC. RSCH., RESPONSE TO THE MINORITY REPORT (2020), https://perma.cc/7F7Z-V44E.

[239] *See* LEADERSHIP CONF. EDUC. FUND, THE USE OF PRETRIAL "RISK ASSESSMENT" INSTRUMENTS: A SHARED STATEMENT OF CIVIL RIGHTS CONCERNS 10, https://perma.cc/WB7C-XH7D (PDF) [hereinafter USE OF PRETRIAL "RISK ASSESSMENT" INSTRUMENTS] (listing signatory organizations).

[240] *See* Eric W. Siddall, *The Real World and the Failure of "Bail Assessment Tools"*, ASS'N OF DEPUTY DIST. ATT'YS, https://perma.cc/K2FW-Q58N ("The Arnold tool has led to the massive release of violent criminals and tragic results."); MINORITY REPORT, *supra* note 236, at 3 (noting that 23 percent of defendants released before trial committed new crimes during the pretrial period).

[241] *See* USE OF PRETRIAL "RISK ASSESSMENT" INSTRUMENTS, *supra* note 239, at 2–4 (recommending principles that "provide tools and guidance for reducing the harm that these assessments can impose").

attention because of a prominent ProPublica article[242] which outlined concerning outcomes from the product and a Wisconsin Supreme Court case, *State v. Loomis*,[243] that also defined glaring concerns of the product. The *ProPublica* article noted that COMPAS performed worse on one measure of performance (false positive rates) for Black individuals than White individuals.[244] Other researchers counter that the disparity can be explained by differences in the underlying offense rates for each race without a biased model.[245] However, the Wisconsin Supreme Court in *State v. Loomis* addressed this issue. Mr. Loomis wanted clarity and redress due to the proprietary nature of the software, the inability to identify high-risk persons because of the way the data is gathered, lack of cross-validation with the Wisconsin specific population, concerns about disproportionately classified minority offenders, and that the software was designed for post-sentencing determinations.[246] Scientist Kristin Lum has also discovered concerning outcomes

[242] *See* Angwin et al., *supra* note 229 (arguing that the risk assessment tool was unreliable in predicting violent crime and produced racially disparate risk scores).

[243] 881 N.W.2d 749 (Wis. 2016).

[244] *See* Angwin et al., *supra* note 229; *see also* Alex Chohlas-Wood, *Understanding Risk Assessment Instruments in Criminal Justice*, BROOKINGS INST. (June 19, 2020), https://perma.cc/EX3H-WD3D (characterizing the *ProPublica* findings as the "most notable claim" of discrimination made against a risk assessment tool).

[245] *See id.* (citing Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, CORNELL UNIV. (Aug. 14, 2018), https://perma.cc/2BTB-4CD7 (PDF)). After applying a traditional measure of model fairness, researchers noted that evidence of racial discrimination faded. *See* Cholas-Wood, *supra* note 244 (citing Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear.*, WASH. POST (Oct. 17, 2016, 5:00 AM), https://perma.cc/3KTN-VK3V).

[246] *See Loomis*, 881 N.W.2d at 769; Angwin et al., *supra* note 229 (explaining Loomis' arguments against the use of COMPAS in his sentencing decision). For details on how ProPublica analyzed the COMPAS Recidivism Algorithm, see Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://perma.cc/S669-RR9N. A group of Stanford researchers determined it was virtually impossible to create a predictive model for all races that did not protect disparities in those who suffer the harm of incorrect predictions, though it has been contested. *See* Corbett-Davies et al., *supra* note 245 ("[T]here is a mathematical limit to how fair any algorithm—or human decision-maker—can ever be.").

from the use of the Arnold Tool. Booking charges that did "not result in a conviction (i.e., charges that are dropped or end in an acquittal) increased the recommended level of pretrial supervision in around 27 percent of cases evaluated by the tool."[247] Mr. Loomis, the *ProPublica* article, and Dr. Lum note that a lack of transparency in the algorithms leaves neither camp with clarity.[248] AI products used in the criminal justice system are just as vulnerable to attack as other industries.[249] As AI is used in policing, pretrial detention, sentencing, and probation, these products must increase transparency to ward against cyberattacks and bias due to various factors.

III. LEGAL OPTIONS FOR PEOPLE HARMED BY UNTRUSTWORTHY AI IN CRIMINAL JUSTICE

The Sixth Amendment[250] affords defendants the right to face their accusers and to review the evidence against them.[251]

[247] Kristian Lum et al., *The Impact of Overbooking on a Pre-Trial Risk Assessment Tool*, *in* FAT* '20: PROCS. OF THE 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 482, 482 (2020), https://perma.cc/HD6R-X2NP (PDF).

[248] *See, e.g.*, *Loomis*, 881 N.W.2d at 757 (noting Mr. Loomis' assertion that using the COMPAS risk assessment tool at sentencing "violates a defendant's right to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevents him from assessing its accuracy").

[249] *See, e.g.*, Andy Greenberg, *Hack Brief: Anonymous Stole and Leaked a Megatrove of Police Documents*, WIRED (June 22, 2020, 12:48 PM), https://perma.cc/7TRF-563K (reporting the leak of more than a million files from more than two hundred state, local, and federal law enforcement agencies).

[250] "In all criminal prosecutions, the accused shall enjoy the right . . . to be informed of the nature and cause of the accusation; to be confronted with the witnesses against him . . . ." U.S. CONST. amend. VI.

[251] *Brady v. Maryland* provides the United States' standard of "discovery" as outlined in the Sixth Amendment; this standard introduces the production of evidence by the prosecution that would be favorable to the accused. *See* 373 U.S. 83, 89–90 (1963); MARK J. MAHONEY, THE RIGHT TO PRESENT A DEFENSE 12 (1994) (noting that *Brady*'s discussion of discovery rights was based in the Compulsory Process Clause of the Sixth Amendment). Mahoney argues that because of the lack of a common law right to discovery before trial and restrictions to certain types of materials, discovery practices in the United States do not reflect all of the ways compulsory process requirements should be interpreted. MAHONEY, *supra*, at 14. And in the

Though the evidence procurement framework is established in *Brady v. Maryland*,[252] it does not incorporate the reality of how AI works. Other paths must be created for recourse. This section will discuss how data breaches are dealt with in the law, outside of the criminal justice system, and through case law from Wisconsin and New Jersey. It also presents potential theories of liability for wrongfully convicted defendants and third-party victims based on AI-based software.

A. *Impact of Data Errors Outside of the Criminal Justice System*

In light of the threat landscape and the technology capabilities, it is vital to think of cybersecurity through statutes such as the California Consumer Privacy Act of 2018[253] and the European Union General Data Protection Regulation of 2016.[254] Data protection and privacy law provide an underpinning for why secured data is so necessary to justice. Since May 2018, the European Union General Data Protection Regulation (EU GDPR) has provided data protection and privacy, giving individuals control over their personal data.[255] The EU GDPR creates a presumption that applying algorithms to personal data is unlawful, barring certain circumstances.[256] This has led to

absence of an express right to discover algorithms, a person must be permitted a method to confront their accuser, as outlined in the Confrontation Clause of the Sixth Amendment. U.S. CONST. amend. VI.

[252] 373 U.S. 83 (1963).

[253] CAL. CIV. CODE § 1798.100 (West 2021).

[254] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[255] *See* Sahar Bhaimia, *The General Data Protection Regulation: The Next Generation of EU Data Protection*, 18 LEGAL INFO. MGMT. 21, 21–22 (2018) (explaining that the EU GDPR updated existing EU data protection law but retained its core principles and values).

[256] *See* Andrew Burt & Stuart Shirrell, *AI Is Rising, and Governments Are Starting to React*, LAW.COM (Feb. 28, 2018), https://perma.cc/RFR2-8FU9. The regulation also offers several considerable rights, such as the right to receive a type of explanation when an algorithm makes a decision that has a particular kind of impact. *See id.*

recourse for harmed individuals. For instance, the French Data Protection Authority fined Google €50 million ($57 million) for EU GDPR violations connected to the unauthorized harvesting and use of personal data.[257] Modeled in part on the EU GDPR, the California Consumer Privacy Act (CCPA), which took effect in January of 2020, protects California residents regarding their personal data and its use.[258] Both laws provide the right to be informed, the right of portability, the right to access, and the

[257] *See* Mathieu Rosemain, *France Fines Google $57 Million for European Privacy Rule Breach*, REUTERS (Jan. 21, 2019, 11:31 AM), https://perma.cc/8TYJ-4XJZ (explaining that the EU GDPR gives European regulators the ability to levy "fines of up to 4 percent of global revenue for violations"). Microsoft quietly took down their facial recognition data after a Financial Times report. *See* Murgia, *supra* note 81 (citing Madhumita Murgia, *Who's Using Your Face? The Ugly Truth About Facial Recognition*, FIN. TIMES (Sept. 18, 2019), https://perma.cc/7YJS-5GKE). Technology policy researcher, Michael Veale suggests Microsoft realized they violated Article 9 of the EU GDPR. *See id.* ("There is reason to believe that the people in data set cannot be considered to expressly and clearly have made their faces public." (quoting Michael Veale)).

[258] *See* CAL. CIV. CODE § 1798.100. The CCPA is the first comprehensive data privacy law in the United States. *See* LAURA JEHL & ALAN FRIEL, CCPA AND GDPR COMPARISON CHART 1 (2018), https://perma.cc/3DUD-4CN3 (PDF). It was signed into law on June 28, 2018, taking effect on January 1, 2020. *See id.* It is designed to give consumers control over the collection, use, and even the sale of their personal data. *See id.* A space of interest to U.S. privacy advocates is how the CCPA aligns with the EU general data protection regulation in terms of enforcement and allows consumers to control their personal information even when held by third parties. Protections related to the right to notice of data collection, the right to access data collected and request deletion, and the right to opt out of the sale of personal information, could potentially fundamentally change the nature of data collection. CAL. CIV. CODE §§ 1798.100(a)–(b), (d), 120(a). As many tech companies are in California, this can have unusually far-reaching implications. Right now, these are protections for natural persons who are California residents, but it will be important to watch if any natural persons outside of California will attempt to access these protections. *See id.* § 1798.100(g) (defining "consumer"). "The Right to be Forgotten" in the EU GDPR and CCPA can potentially afford individuals a route to private rights of actions. *See* GDPR, art. 17 (establishing a "right to erasure"); CAL. CIV. CODE § 1798.105 (providing consumers with some rights regarding the "deletion of personal information collected by businesses"). For a comparison of GDPR and CCPA, see JEHL & FRIEL, *supra*. Further study of CCPA can provide an additional route for recourse for wrongfully convicted defendants and third parties, depending on how the data was gathered.

right to erasure.[259] The EU GDPR creates a "privacy by default" legal framework, requiring companies to have a legal basis for processing personal data in the EU, meaning prior consent.[260] In contrast, the CCPA creates transparency in California's data economy and rights to consumers, providing an opt-out mechanism.[261] This distinction of prior consent, unique to the EU GDPR, makes the difference, providing a legal strategy for privacy first through user control and second, by providing a foundation for examining data and the law.[262]

Hacked data is an issue that has arisen in several civil cases. In *Beyer v. Symantec Corp.*,[263] a class action was filed targeting the cybersecurity firm Symantec Corp. over software flaws that allegedly rendered consumers' devices vulnerable to hackers.[264] In another California case, *Smith v. Adobe Systems, Inc.*,[265] Ms. Smith asserted a claim based on "strict liability because there were allegedly design defects in Adobe's products and services—more specifically, 'security flaws,'—which purportedly allowed a computer hacker to 'hijack control of [her] browser, files, [and] web content' and 'to silently reprogram [her] hardware [and] user settings.'"[266] In *National Election Defense*

[259] *See* JEHL & FRIEL, *supra* note 258 (comparing the EU GDPR and CCPA). Called the "right to erasure" under GDPR, but "right to deletion" under CCPA. *See CCPA vs GDPR: Compliance with Cookiebot*, COOKIEBOT, https://perma.cc/8QTR-ALKA (last updated Nov. 30, 2020) [hereinafter *CCPA vs GDPR*] (noting minor differences between these rights).

[260] *See CCPA vs GDPR*, *supra* note 259.

[261] *See What Is GDPR?*, COOKIEBOT, https://perma.cc/HD8A-8TBV (detailing the GDPR's extensive requirements); *California Consumer Privacy Act (CCPA)*, COOKIEBOT, https://perma.cc/XU74-WGK3 (last updated Nov. 30, 2020) (explaining that consumers can opt out of having their data sold and request the deletion of collected data). The main rights of both laws are the right to be informed, the right of portability, and the right to access. *See* JEHL & FRIEL, *supra* note 258, at 3–5 (providing a side-by-side comparison of the elements of these rights in each jurisdiction). For additional comparisons, see DATAGUIDANCE & FUTURE OF PRIV. F., COMPARING PRIVACY LAW: GDPR V. CCPA 26 – 35 (2018), https://perma.cc/JXG3-MPWY (PDF).

[262] *See CCPA vs GDPR*, *supra* note 259 (noting that the CCPA does not require prior consent).

[263] No. 18-cv-02006, 2019 WL 935135 (N.D. Cal. Feb. 26, 2019).

[264] *Id.* at *1.

[265] No. C-11-1480, 2011 WL 4404152 (N.D. Cal. Sept. 21, 2011).

[266] *Id.* at *1.

Coalition v. Boockvar,[267] the plaintiff alleged that there was a much-known vulnerability of the voting system, but could not prove that they were hacked.[268] Without defined harm to defined plaintiffs, there are no legal remedies.[269] These plaintiffs could not recover because they had no way to either access the algorithms to prove harm had occurred or would occur without correction. This can lead to years of lost opportunities for wrongfully convicted defendants and loss of revenues and resources.[270] But, some plaintiffs have found redress.

Flawed AI used by states has amounted to millions of dollars in lawsuits and fines. The publication, *The Markup*, noted that "even an error rate of 1 percent could upend the lives of hundreds of thousands of people."[271] These instances are

[267] No. 674 M.D. 2019 (Pa. Commw. Ct. Jan. 24, 2020).

[268] *Id.* In the complaint, the plaintiff stated:

> Yet the Commonwealth has chosen to endorse a new voting system, the ExpressVote XL, which fails at every one of these core functions and violates the plain requirements set forth in the law to guarantee them. Moreover, there are continued and credible complaints that the system is neither secure nor reliable, and is capable of being hacked.

Brief for Petitioner at 1, Nat'l Election Def. Coal. v. Boockvar, No. 674 M.D. 2019 (Pa. Commw, Ct. Jan. 24, 2020).

[269] *See* Clapper v. Amnesty Int'l USA, 568 U.S. 398, 409 (2013) ("To establish Article III standing, an injury must be 'concrete, particularized, and actual or imminent; fairly traceable to the challenged action; and redressable by a favorable ruling.'" (quoting Monsanto Co. v. Geertson Seed Farms, 561 U.S. 139, 149 (2010))).

[270] *See, e.g.*, Stephanie Wykstra, *Government's Use of Algorithm Serves Up False Fraud Charges*, UNDARK (June 1, 2020), https://perma.cc/5GVD-GSAR (reporting that a Michigan agency falsely charged more than forty thousand people with unemployment fraud, causing many of them to lose homes and job opportunities).

[271] Lauren Kirchner & Matthew Goldstein, *Access Denied: Faulty Automated Background Checks Freeze Out Renters*, MARKUP (May 28, 2020, 5:00 PM), https://perma.cc/G5L8-M3UH.

occurring in housing,[272] privacy,[273] and other categories.[274] Lack of government oversight into AI can lead to devastating outcomes for individuals. For example, from October 2013 to September 2016, the Michigan Data Automated System falsely accused thousands of people of committing fraud and took millions of dollars from them.[275] The software, the Michigan Integrated Data Automated System ("MiDAS"), was supposed to detect fraud and automatically charge people with

[272] Davone Jackson was denied low-income housing because of misidentification, leading to almost a yearlong of housing insecurity. *See id.* (noting that Jackson's suit was settled for an undisclosed sum). Glen Patrick Thompson Sr. and his son were left homeless because it connected him to another person who had been evicted. *See id.* (noting that Patrick's suit was also settled for an undisclosed sum). William Hall Jr. was misidentified as a sexual offender and could not get approved for a duplex. *See id.* (noting that Hall's suit is pending).

[273] *See* Davey Alba, *A.C.L.U. Accuses Clearview AI of Privacy 'Nightmare Scenario'*, N.Y. TIMES (May 28, 2020), https://perma.cc/BEK2-2ZBZ (last updated June 3, 2020) (reporting that the ACLU sued Clearview for allegedly "violat[ing] a state law that forbids companies from using a resident's fingerprints or face scans without consent"). The company created a database of over three billion photos across the internet, including Facebook, YouTube, Twitter, and Venmo. *See id.* "Clearview has set out to do what many companies have intentionally avoided out of ethical concerns: create a mass database of billions of faceprints of people, including millions of Illinoisans, entirely unbeknownst to those people, and offer paid access to that database to private and governmental actors worldwide." Complaint at 19, ACLU v. Clearview AI, Inc., No. 9337839 (Ill. Cir. Ct. May 28, 2020). Facebook "agreed to pay $550 million to settle a class-action lawsuit over its use of facial recognition technology in Illinois. . . ." Natasha Singer & Mike Isaac, *Facebook to Pay $550 Million to Settle Facial Recognition Suit*, N.Y. TIMES (Jan. 29, 2020), https://perma.cc/MRJ4-54EN; *see also In re* Facebook Biometric Info. Priv. Litig., 326 F.R.D. 535, 540 (N.D. Cal. 2018) (allowing class certification for privacy action against Facebook).

[274] *See, e.g.*, Maddy Varner & Aaron Sankin, *Suckers List: How Allstate's Secret Auto Insurance Algorithm Squeezes Big Spenders*, MARKUP (Feb. 25, 2020, 5:00 PM), https://perma.cc/J2KH-UAKM (finding that Allstate's new risk algorithm charged more to customers who were already paying the highest premiums and "denied meaningful decreases to thousands of Allstate customers who the company's new risk profile showed were paying too much").

[275] *See* Robert N. Charette, *Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold*, IEEE SPECTRUM (Jan. 24, 2018, 5:00 PM), https://perma.cc/3YMQ-5PX7 (reporting that the Michigan Unemployment Insurance Agency collected about $66 million in one year due to the false charges); Wykstra, *supra* note 270 (noting that the agency failed to repay millions of dollars in erroneous charges for years).

misrepresentation and demand repayment to the state, garnishing wages.[276] After two years of using MiDAS, the agency charged forty thousand people, billing them at five times the original benefits at a rate of 400 percent plus interest.[277] It was determined that 93 percent of the charges were erroneous.[278] Algorithms in Arkansas and Idaho erroneously cut Medicaid benefits.[279] In 2019, a Dutch court found that an algorithm that detected welfare fraud violated human rights, ordering the government to stop using it.[280] If a person committed negligence to this degree, it would be prosecuted as criminal negligence. However, AI developers are protected by outdated mechanisms.

[276] *See* Wykstra, *supra* note 270 ("[C]lass actions lawsuits allege that the system searched unemployment datasets and used flawed assumptions to flag people for fraud . . . .").

[277] *See id.*

[278] *See id. Bauserman v. Unemployment Ins. Agency*, No. 333181, 2017 WL 3044120 (Mich. Ct. App. July 18, 2017) and *Cahoo v. SAS Analytics Inc.*, 912 F.3d 887 (6th Cir. 2019), are two class-action suits brought forward by people who were impacted by the software in place. The legal director of the University of Michigan Law School's Workers' Rights Clinic testified before the Michigan Senate Oversight Committee that he believed close to twenty thousand people were being actively pursued and were having their wages garnished. *See* Wykstra, *supra* note 270 (noting that the director testified in March 2020); *Oversight Committee Hearing*, MICH. SENATE (Mar. 10, 2020), https://perma.cc/JAP3-NBJ3 (providing video of the testimony). It would take almost six years before some people would have the charges dismissed. Wykstra, *supra* note 270.

[279] *See* Colin Lecher, *What Happens When an Algorithm Cuts Your Health Care*, VERGE (Mar. 21, 2018, 9:00 AM), https://perma.cc/VJ4F-CUB3 (describing problems resulting from the implementation of algorithms to allocate home health care hours in Arkansas and Idaho); Michele Gilman, *AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead*, CONVERSATION (Feb. 14, 2020, 8:45 AM), https://perma.cc/KV7C-RRRQ ("Program-wide algorithmic errors have . . . plagued Medicare eligibility determinations in states such as Indiana, Arkansas, Idaho and Oregon.").

[280] *See* Jon Henley & Robert Booth, *Welfare Surveillance System Violates Human Rights, Dutch Court Rules*, GUARDIAN (Feb. 5, 2020, 8:18 AM), https://perma.cc/2HKD-FTQJ ("This is one of the first times a court anywhere has stopped the use of digital technologies and abundant digital information by welfare authorities on human rights grounds." (quoting the UN special rapporteur on extreme poverty and human rights)).

B. *Nefarious Actors Provide False Scientific Evidence in Criminal Cases, Infringing on Civil Liberties*

Nefarious actors provide false evidence in the criminal justice system, infringing on constitutional protections. In 2012, Annie Dookhan was arrested for allegedly faking drug results, forging documentation, and mixing samples at a state police lab.[281] A recorded 1,140 inmates were convicted based on her potentially tainted evidence.[282] Malicious behavior by individuals is one morass; but, benign neglect by states exacerbates the harm. The Public Safety Crime Lab in Houston promoted Jonathan Salvador despite evidence that his practice of "dry-labbing" samples (where forensic analysts report results of tests that they never executed) had put in jeopardy close to five thousand drug cases.[283] This is just as intentionally

[281] *See* CRIMESIDER STAFF, *Annie Dookhan, Chemist at Mass. Crime Lab, Arrested for Allegedly Mishandling Over 60,000 Samples*, CBS NEWS (Sept. 28, 2012, 6:45 PM), https://perma.cc/MVH9-RRYU ("[Dookhan] tested more than 60,000 drug samples involving 34,000 defendants during her nine years at the lab.").

[282] *See id.* It was initially believed to be ninety samples. Commonwealth v. Scott, 5 N.E.3d 530, 536 (Mass. 2014). Dookhan provided false scientific credentials and drew no concerns from her supervisors for years. *See* COMMONWEALTH OF MASS., OFF. OF THE INSPECTOR GEN., INVESTIGATION OF THE DRUG LABORATORY AT THE WILLIAM A. HINTON STATE LABORATORY INSTITUTE 113–15 (2012), https://perma.cc/XB9G-YXJ7 (PDF) ("The most glaring factor that led to the Dookhan crisis was the failure of management."). She intentionally turned negative samples into positive and acknowledged she could not identify which cases were impacted. *See Scott*, 5 N.E.3d at 536 (describing Dookhan's admissions to state police). Defendants in cases where Dookhan served as the primary or secondary chemist were "entitled to a conclusive presumption that [the chemist's] misconduct occurred . . . that it was egregious, and that it was attributable to the Commonwealth." Commonwealth v. Gardner, 5 N.E.3d 552, 556–57 (Mass. 2014).

[283] See the dissent in *State v. Lui*, 315 P.3d 493, 521 (Wash. 2014), outlining several works concerning the laboratory misconduct in Houston. *See also* James Pinkerton & Brian Rogers, *Crime Lab Analyst Kept on Job Despite Shoddy Work*, HOUS. CHRON. (Apr. 5, 2013, 10:25 PM), https://perma.cc/BX7T-PFJD (discussing the retention of a forensic analyst despite high error rate). Other examples abound. *See* Melendez-Diaz v. Massachusetts, 557 U.S. 305, 318–19 (2009) (describing incidents); Thomas J. Lueck, *After Falsified Test Results, Kelly Orders Forensic Shakeup*, N.Y. TIMES (Apr. 20, 2007), https://perma.cc/BAM7-3NUY (reporting that two police crime lab analysts falsely reported results of drug tests). In 2013, New York City's medical examiner's office had to review more than eight hundred rape cases where a

malicious and nefarious as a rogue analyst or department,[284] as it shows a careless attitude towards integrity.

Internationally, countries are already having to backtrack after software errors led to cases being inappropriately determined. In 2019, Denmark reviewed over ten thousand court cases that may have been improperly decided because of a software bug in a cell phone tracking technology.[285] Two problems occurred in the Denmark situation. One was that during the conversion process of raw data into evidence, the system omitted some of the data creating fewer clear images of where the cell phone had been.[286] Second, some of the cell phone tracking data connected phones to the incorrect cell phone towers, potentially linking innocent people to crime scenes.[287] It is estimated that these impacted cases go back to 2012; it will require extensive work to see if any of this data proved to be decisive in verdicts against defendants.[288] There have been many cases that have been brought by plaintiffs who have attempted redress through wrongful convictions based on

lab technician had mishandled DNA evidence for over ten years. Mark Hansen, *Crime Labs under the Microscope after a String of Shoddy, Suspect and Fraudulent Results*, AM. BAR ASS'N J. (Sept. 1, 2013, 10:20 AM), https://perma.cc/F3LH-DNEH. Similar instances in St. Paul, Minnesota, West Virginia, and Oklahoma City have been documented and many occurred over an extended period with little to no oversight by leadership. *See id.* (arguing for increased regulation of crime labs).

[284] After information was unearthed about serious negligence and misconduct by the ASCLD/LAB in North Carolina, an audit was conducted that unpacked over 230 cases of "SBI [State Bureau of Investigation] agents with[holding] exculpatory evidence or distort[ing] evidence . . . over a 16-year period." Radley Balko, *North Carolina's Corrupted Crime Lab*, REASON FOUND. (Aug. 23, 2010, 4:30 PM), https://perma.cc/LBT4-LP87; *see* Craig Jarvis, *Report Criticized SBI Crime Lab's Lack of Documentation*, NEWS & OBSERVER (Aug. 20, 2016, 3:43 PM), https://perma.cc/W6V3-7FSW (last updated Aug. 21, 2016) (discussing the findings of an independent investigation of the SBI crime lab). *See generally* Joseph R. John, *ASCLD/LAB Interim Inspection Report*, AM. SOC'Y OF CRIME LAB'Y DIRS. (2010), https://perma.cc/ZM7V-8ASQ (PDF).

[285] *See* Martin Selsoe Sorensen, *Flaws in Cellphone Evidence Prompt Review of 10,000 Verdicts in Denmark*, N.Y. TIMES (Aug. 20, 2019), https://perma.cc/9542-KQER.

[286] *See id.*

[287] *See id.*

[288] *See id.*

intentional or negligent handling of evidence.[289] If it is already shown that state contracted labs provided false evidence and testimony, and it is already clear that technology can be drastically wrong, there must be recourse available to people harmed in the United States criminal justice system by potentially unregulated AI.

C. *Theories of Accountability and Liability from Case Law for Victims of Potentially Untrustworthy AI Used in Criminal Cases*

This subpart will outline criminal and civil case law where risk assessment instruments were used. It will also describe legal theories for accountability and liability for victims of potentially untrustworthy data used in criminal cases. Potential plaintiffs must be able to explain the harm caused by likely hacked and unregulated AI.[290]

[289] *See generally, e.g.*, Creach v. Dookhan, No. 20-10714, 2020 WL 3256890 (D. Mass. June 16, 2020); Penate v. Hanchett, 944 F.3d 358 (1st Cir. 2019); Green v. N.C. State Bureau of Investigation Crime Lab, No. 11-cv-69, 2018 WL 4356778 (W.D.N.C. Sept. 12, 2018); Spencer v. Dookhan, No. 16-cv-12080, 2017 WL 2785423 (D. Mass. June 27, 2017); Spencer v. Dookhan, No. 13-11431, 2014 WL 6904377 (D. Mass. Dec. 5, 2014); Jones v. Han, 993 F. Supp. 2d 57 (D. Mass. 2014); Solomon v. Dookhan, No. 13-10208, 2014 WL 317202 (D. Mass. Jan. 7, 2014); Cage v. City of Chicago, 979 F. Supp. 2d 787 (N.D. Ill. 2013); Jimenez v. City of Chicago, 830 F. Supp. 2d 432 (N.D. Ill. 2011); Jimenez v. City of Chicago, 877 F. Supp. 2d 649 (N.D. Ill. 2012); Jimenez v. City of Chicago., 732 F.3d 710 (7th Cir. 2013); McCarty v. Gilchrist, 646 F.3d 1281 (10th Cir. 2011); Lincoln v. City of Greenville, No. 4:10-CV-21, 2011 WL 285231 (E.D.N.C. Jan. 25, 2011); Bryson v. Macy, 611 F. Supp. 2d 1234 (W.D. Okla. 2009); Bryson v. Macy, No. CIV-05-1150, 2009 WL 10672213 (W.D. Okla. June 17, 2009); Bryson v. City of Okla. City, 627 F.3d 784 (10th Cir. 2010); Holmes v. Pierce, No. 04 C 8311, 2009 WL 57460 (N.D. Ill. Jan. 7, 2009); Holmes v. Hardy, 608 F.3d 963 (7th Cir. 2010); Washington v. Wilmore, 407 F.3d 274 (4th Cir. 2005); Pierce v. Gilchrist, 359 F.3d 1279 (10th Cir. 2004); *In re* W. Va. State Police Crime Lab., Serology Div., 445 S.E.2d 165 (W. Va. 1994); *In re* Investigation of the W. Va. State Police Crime Lab., Serology Div., 438 S.E.2d 501 (W. Va. 1993).

[290] By way of background, Charles H. Moellenberg Jr. et al., addressed methods for companies to monitor how the legislatures and courts are shaping tort law to apply to products, components, and software incorporating AI and ways to use contractual warranties, indemnities, and limitations to control liability risks. Charles H. Moellenberg, Jr. et al., *United States: Mitigating Product Liability for Artificial Intelligence*, MONDAQ (Mar. 22, 2018), https://perma.cc/XGQ6-8N3A.

1. Legal Response to Data Requests in Criminal and Civil Cases

a. State v. Loomis

The risk assessment instrument, COMPAS, was at the center of the *State v. Loomis* decision.[291] Loomis was barred from reviewing the algorithms in the software and challenged how the proprietary algorithm calculated his risk when determining sentencing.[292]

In February 2013, Eric Loomis was charged with five criminal counts related to a drive-by shooting in La Crosse, Wisconsin.[293] He denied being involved in the shooting but did admit that later that evening, he had driven the same car involved in the shooting.[294] He was arrested and pleaded guilty to two lesser charges of eluding an officer and no contest to operating a vehicle without its owner's consent.[295] The judge sentenced Loomis within the limits of the two charges where he entered a plea.[296] Loomis filed a motion for post-conviction relief.[297] However, the Wisconsin Supreme Court ruled that the use of their risk assessment tool at sentencing did not violate the defendant's due process right to be sentenced based on

[291] *See* State v. Loomis, 881 N.W.2d 749, 753 (Wis. 2016) (focusing on whether using COMPAS while sentencing defendants violates their due process rights).

[292] *Id.* at 761.

[293] *Id.* at 754.

[294] *Id.*

[295] *Id.*; *see* Brief of Plaintiff-Respondent at 2, State v. Loomis, 881 N.W.2d 749 (Wis. 2016) (No. 2015AP157-CR), 2016 WL 485419, at *2 (mentioning that Loomis pled no contest to two charges, including operating a vehicle without its owner's consent).

[296] *Loomis*, 881 N.W.2d at 756. During intake, the Wisconsin Department of Corrections interviewed Loomis, gathered information from his criminal file and entered it into COMPAS. *See id.* "On the attempting to flee an officer charge, the circuit court sentenced Loomis to four years, with initial confinement of two years and extended supervision of two years." *Id.* at 756 n.18. "For operating a vehicle without the owner's consent, the circuit court sentenced Loomis to seven years, with four years of initial confinement and three years of extended supervision, to be served consecutively with the prior sentence." *Id.*

[297] *Id.* at 756.

accurate information, nor did the use of risk assessment tool at sentencing violate a defendant's due process right to an individualized sentence.[298] The Supreme Court of Wisconsin noted that the trial court judge said she based her determination not solely on the COMPAS score but on several additional factors and noted that risk scores may not be used "to determine whether an offender is incarcerated" or "to determine the severity of the sentence."[299] The court did not go to the extent of saying that COMPAS could not be used but did find that there should be five written warnings for judges that they are to review before assessing the pretrial score assigned by COMPAS. The five warnings were: noting the proprietary nature of the software, noting the inability to identify high-risk persons because of the way the data is gathered, noting the lack of cross-validation with the Wisconsin specific population, noting the concerns about disproportionately classified minority offenders, and noting that the software was designed to be used only for post-sentencing determinations.[300]

[298] *Id.* at 757, 792. The risk assessment tool's consideration of defendant's gender did not violate defendant's due process rights. *See id.* at 766–67.

[299] *Id.* at 769. There is continuing research into what judges explain about their thought process in how much weight they give to a risk assessment instrument in deciding their opinion. *See* Joy Wang, *UNM Legal Experts Break Down Judge's Decision to Hold Alleged Rapist in Pretrial Detention*, KOB4 (Jan. 30, 2020, 1:13 PM), https://perma.cc/W6YU-Y37E (considering a judge's use of factors such as severity of a new charge in addition to the algorithm).

[300] *Loomis*, 881 N.W.2d at 769. The "proprietary nature of COMPAS" prevents the disclosure of how risk scores are calculated;

1. COMPAS scores are unable to identify specific high-risk individuals because these scores rely on group data;

2. although COMPAS relies on a national data sample, there has been "no cross-validation study for a Wisconsin population";

3. studies "have raised questions about whether [COMPAS scores] disproportionately classify minority offenders as having a higher risk of recidivism"; and

4. COMPAS was explicitly developed to assist the Department of Corrections in making *post*-sentencing determinations.

b. Rodgers v. Laura & John Arnold Foundation

Though errors using the Arnold Tool have occurred,[301] to date there has been only one case of the Arnold Tool potentially harming a third party not involved in a criminal matter. On April 5, 2017, Jules Black was arrested by the New Jersey State Police and charged for being a felon in possession of a firearm.[302] The Arnold Tool assigned Black a low Public Safety Assessment (PSA) score, and he was released without a determination of a need for bail.[303] Three days later, he allegedly killed Christian Rodgers.[304] Rodgers was survived by his mother, June Rodgers, who brought a suit both individually and on behalf of her son against the Arnold Foundation.[305]

This was the first time the Arnold Foundation had been a named defendant in a tort suit, and it proved challenging to establish a cause of action. Ms. Rodgers framed the Arnold Tools algorithm within the product liability failure under the New Jersey Products Liability Act (PLA),[306] focusing on the fact that the tool was designed in a defective manner.[307] The court noted

[301] In an additional case, Lamonte Mims allegedly murdered Edward French two weeks after he was released by a judge who relied on the Arnold Tool, despite Mr. Mims having been charged with possession of two guns and being on probation for burglary. Westervelt, *supra* note 67. The Pretrial Division Project of San Francisco Sheriff's Office acknowledged that her staff mistakenly entered incorrect data for Mr. Mims, leading the product to give an incorrect score that the judge relied on for making a decision. *See id.*

[302] Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 WL 2429574, at *1 (D.N.J. June 11, 2019).

[303] *See id.*

[304] *See id.* For more details on Christian Rodgers' death, see Joe Hernandez, *Mother of Slain N.J. Man Blames Computer Program for His Shooting*, WHYY.ORG (Mar. 23, 2018), https://perma.cc/4Q57-NJTH.

[305] *See Rodgers*, 2019 WL 2429574, at *1. "After the murder, the state Judiciary updated the algorithm to recommend preventative detention for anyone charged with serious gun crimes." Hernandez, *supra* note 304.

[306] N.J. STAT. ANN. § 2A:58C-2 (West 2020).

[307] The New Jersey statute states in part, that the product caused a harm not reasonably fit, suitable or safe for its intended purpose because it:

> a. deviated from the design specifications, formulae, or performance standards of the manufacturer or from otherwise identical units manufactured to the same manufacturing specifications or formulae, or
>
> b. failed to contain adequate warnings or instructions, or

the Restatement (Third) of Torts defined the presentencing assessment outside the term "product" under the New Jersey PLA, as the PSA was neither considered a tangible product distributed commercially for use or consumption nor a non-tangible "other item."[308] The PSA was seen as "information, guidance, ideas, and recommendations as to how to consider the risk a given criminal defendant presents."[309] The court also noted that under the First Amendment, information and guidance are not subject to tort liability as they are seen as speech instead of a product.[310] In conclusion, the court also found a failure to establish proximate causation and the fact that the PSA omitted risk indicators of firearm possession and sex-crimes as the PSA is one of many pieces of different information that a judge takes into account.[311]

Although in different positions, both Mr. Loomis and Ms. Rodgers found that their inability to access the algorithm and data used restricted their opportunities to be heard.[312] As Part II of this paper described how AI is untrustworthy in leading decisions, there must be ways to protect a plaintiff's potential tort remedies when there is no transparency of the algorithm. Eric Surette outlined in the article, "Liability of Businesses to Governments and Consumers for Breach of Data Security for

c. was designed in a defective manner.

Rodgers, 2019 WL 2429574, at *2 (citing N.J. STAT. ANN. § 2A:58C-2 (West 2020)).

[308] *See id.* The Restatement (Third) of Torts would consider "non-tangible items such as 'other items,'" to include services, human blood, and human tissues. *Id.* at *2 (citing RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 19 (AM. L. INST. 1998)).

[309] *Id.* at *3.

[310] *Id.*

[311] *Id.* The district court's opinion was affirmed by Rodgers v. Christie, 795 F. App'x 878 (3d Cir. 2020).

[312] *See* State v. Loomis, 881 N.W.2d 749, 760–64 (Wis. 2016) (finding the claim lacking because neither the inability to challenge "the scientific validity of the risk assessment" nor the inability to ensure that sentencing is "based on accurate information" violated due process); *Rodgers*, 2019 WL 2429574, at *2–3 (rendering Public Safety Assessments outside the scope of product liability, making it impossible for plaintiff to bring a claim).

Consumers' Information,"[313] data breach security concerns,[314] and this could provide a roadmap to connect data breaches in criminal justice software. Once the algorithms in the AI are made transparent, and if a data breach can be proven, plaintiffs can begin to prepare plausible causes of action.[315]

2. Legal Remedies for Wrongfully Convicted Criminal Defendants Impacted by Hacked Data

Persons wrongfully convicted based on problematic AI should be permitted to use legal remedies similar to plaintiffs in tort cases who have been convicted by falsified data from human actors. The question is to determine if hacked data may be used to challenge a conviction or sentence. Cary Coglianese and David Lehr, in "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era," provide a strong rationale for why machine-learning should be used in administrative agencies.[316] First, this is because delegating decision-making to machines is likely not prohibited by Congress/statutes.[317] Coglianese and Lehr support machine-learning in administrative agencies because it is possible to use machine-learning without violating due process under the Fifth Amendment.[318] They also note that machine-learning within administrative agencies will not be discriminatory, so long as those implementing it have employed

[313] Surette, *supra* note 130.

[314] *Id.*

[315] *See id.*; *In re* Zappos.com, Inc., No. 3:12-cv-00325, 2013 WL 4830497 (D. Nev. Sept. 9, 2013) (ordering that plaintiffs had standing for a data breach claim where they alleged that a data breach occurred and the information to prove this would be accessible to them).

[316] Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1154 (2017).

[317] *See id.* at 1178–84

> A Congress that deliberately contemplated and authorized an agency to use machine learning would presumably also understand the need to provide guidance about the necessary objective function for algorithms to optimize, and it would be more likely than usual to articulate a sufficiently clear set of goals that would pass the intelligibility muster.

[318] *Id.* at 1184–91.

machine-learning in good faith.[319] However, hackers do not act in good faith and are a well-known problem in the tech community.[320]

The fact that a harmed person has no evidentiary mechanisms to question algorithmic accusers violates the intent of the Sixth Amendment's Confrontation Clause.[321] Adversarial machine-learning, code that has written itself, modified itself, and improved itself, with no indication where there might have been nefarious attacks to the code, withholds from defendants an opportunity to review the testimonial evidence against the accused. Andrea Roth has outlined the history of not allowing machines to testify.[322] She warns against the unknown of "black box dangers" where human error and machine error can cause a machine to produce faulty outcomes.[323] Potentially hacked data makes machine testimony even more of a necessity.[324] The information needed is withheld because of a protected cloud server.[325] The lack of privity between the harmed party (wrongfully convicted defendant or a third party) and the software company allows the state to flee from any

[319] *Id.* at 1191–93.

[320] Operating systems and software are full of undisclosed and undetermined vulnerabilities, whether because of software-hardware incompatibilities, leaving the products used open to cyberattacks and hacks. *See supra* Part I.C; Steve Symanovich, *5 Reasons Why General Software Updates and Patches Are Important*, NORTON (Jan. 7, 2019), https://perma.cc/XEW7-HJUP.

[321] U.S. CONST. amend. VI. Professor Roth outlines the history of the Confrontation Clause and its changing role with the rise of technology. *See* Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2040–41 (2017). Coglianese observes that machine-learning could violate due process under the Sixth Amendment's right to be confronted with the witnesses against him. *See* Coglianese & Lehr, *supra* note 316, at 1184–85.

[322] Roth, *supra* note 321, at 2043.

[323] Professor Roth outlines the history of the radar detector. *Id.* at 2015–19.

[324] *See* Zittrain, *supra* note 179 (discussing how AI's inability to explain its reasoning makes it so "there's no easy way to predict how it might fail when presented with specially crafted or corrupted data").

[325] *See* Coglianese & Lehr, *supra* note 316, at 1220 ("Agencies must properly and securely store these data to minimize threats to privacy intrusions, especially when many administrative applications of machine learning will require inter-agency sharing through the cloud.").

responsibility.[326] The federal government, the District of Columbia, and thirty-five states have compensation statutes for people who have been wrongfully convicted[327] and this must be expanded to include victims who are wrongfully convicted based on faulty AI.

Legal solutions using available tort remedies are currently limited as algorithms are classified as intangible property, like the information in books or media.[328] This property is protected as trade secrets and confidential information, not products, so it does not permit persons harmed by AI-based products in the criminal justice system to access products liability.[329] It is also challenging to access defamation, invasion of property, or breach of duty because of a lack of privity between the third-party software company and individuals impacted by the software, as the privity of contract is between the company and the state corrections or court system and potent indemnification clauses shield the company.[330]

Once the plaintiff can show the AI generated a decision recommendation based on flawed data or a malfunctioning AI that violates Equal Protection, the plaintiff can file a cause of action against the software manufacturer or the state.[331]

[326] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 760–64 (Wis. 2016) (outlining why a defendant is unable to see the information contained inside the algorithm).

[327] *See Compensating the Wrongly Convicted*, INNOCENCE PROJECT, https://perma.cc/6DRC-2MKX (PDF). The following fifteen states do not: Alaska, Arizona, Arkansas, Delaware, Georgia, Idaho, Kentucky, New Mexico, North Dakota, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, and Wyoming. *See id.*

[328] *See* Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 WL 2429574, at *3 (D.N.J. June 11, 2019) (explaining why no remedy exists in tort law).

[329] *See* Roth, *supra* note 321, at 2028 ("Creators of proprietary algorithms typically argue that the source code is a trade secret or that it is unnecessary to prepare a defense . . . . But it is not clear that trade secret doctrine would protect the source code of an algorithm used to convict or impose liability.").

[330] *See* Moellenberg et al., *supra* note 290 (recommending such contracts and indemnity clauses specifically to control the liability risk that comes along with using AI).

[331] *See* Coglianese & Lehr, *supra* note 316, at 1191–205 (discussing, in the context of federal agencies, how an AI-generated decision could violate Equal Protection).

Scientists have studied the adversarial aspects of machine-learning for over a decade,[332] so any software developer and customer should be aware of the potential risk in using the product. Where the flaw occurs should have no bearing on the plaintiff's right of recovery, as the software developer sent a defective product into the industry. If the software company warned the municipality or court that it could not guarantee protected data, remedies might be arguably limited. The state should bear the burden of this error, not a wrongfully convicted person or wronged third party.[333] Additionally, because of the way AI works,[334] it is not reasonable for an indemnification clause in a contract to shield a company from civil liability, as it takes time to determine if the software has an error, typically at the risk of people serving prison sentences.[335] Criminal defendants must be seen as customers of the court, as they, too, are members of the general public and deserve a route to bring forth a suit if their constitutional protections have been violated.

Plaintiffs wrongfully convicted under flawed AI should also be able to access defamation and invasion of privacy causes of action depending on the extent of the conviction and on whether the conviction led to the termination of employment, impacted child custody decisions, or led to the loss of property or standing in the community. The CCPA permits private civil actions for

[332] Kegelmeyer, *supra* note 70, at 10, 12. *See generally* Amir Globerson & Sam Roweis, *Nightmare at Test Time: Robust Learning by Feature Deletion*, *in* ICML 2006: PROCEEDINGS OF THE 23RD INTERNATIONAL CONFERENCE ON MACHINE LEARNING 353–60 (2006), https://perma.cc/EJ8D-3CVQ.

[333] Many scholars have focused on how companies can protect themselves from this potential lawsuit through more carefully crafted contracts to shield them from civil liability whether they design the AI or have incorporated another product into their own. *See generally* Emily Garrison et al., *Artificial Intelligence: The New Frontier for Assessing Insurance Coverage*, POLICYHOLDER PERSP. (June 6, 2019), https://perma.cc/TX7B-TLYF; *Artificial Intelligence Liability—Don't Overlook Your Risk*, HUB INT'L (Oct. 11, 2018), https://perma.cc/9PBE-96C9.

[334] The INTRODUCTION discusses the multiple parties involved in a single artificial intelligence system. *See* Watson, *supra* note 29, at 80.

[335] Once the AI software is up and running, the developer will insist on the end-user essentially signing a contract to say the software meets the specifications set out at the beginning of development. *See* Michael Carson & Greg McEwen, *Artificial Intelligence Misdiagnosis: Who Is to Blame?*, LAW. MONTHLY, https://perma.cc/E3RM-P7R3 (last updated July 3, 2019).

data breaches only; these are statutory damages for not less than $100 to $750 per consumer per incident or actual damages if the company failed to implement reasonable data security.[336]

One of the latest examples of a securities class action lawsuit arising out of a data breach or other cybersecurity and AI incident, on October 24, 2019, a plaintiff shareholder filed a securities class action lawsuit against California-based software company Zendesk after fifteen thousand Zendesk Support and Chat accounts had been accessed without authorized permission.[337] Additionally, it is not unheard of for civil actions to be brought forward in this manner. In *In re Adobe Systems, Inc. Privacy Litigation*,[338] the plaintiffs were customers of a software retailer whose computer systems were hacked, resulting in the exposure of the customers' personal information.[339] The customers alleged that the software retailer did not maintain "reasonable security practices" to protect customer data, in violation of California Civil Code Section 1798.81.5(b).[340]

3. Legal Remedies for Third Parties Injured Due to Flawed Data or Algorithms

A victim harmed by a person wrongfully released due to flawed data in the criminal justice system should also have legal remedies available, either through the state or the software

[336] Kristin Madigan, *Data Privacy: California, GDPR*, *in* ARTIFICIAL INTELLIGENCE AND ROBOTICS NATIONAL INSTITUTE 475 (2020), https://perma.cc/DYC8-B8ZB (PDF); CAL. CIV. CODE §§ 1798.150(a)(1), (a)(1)(A) (West 2021).

[337] *See* Kevin LaCroix, *Zendesk Hit with Data Breach-Related Securities Suit*, D&O DIARY (Oct. 27, 2019), https://perma.cc/3T5V-TA3X; *see also* Beyer v. Symantec Corp., No. 18-cv-02006, 2019 WL 935135, at *1 (N.D. Cal. Feb. 26, 2019) (claiming that Symantec failed to update an open source code "for at least seven years, resulting in critical vulnerabilities" in the products that plaintiffs had purchased); Diaz v. Intuit Inc., No. 15-cv-01778, 2017 WL 4355075, at *1 (N.D. Cal. Sept. 29, 2017) (claiming Intuit knowingly allowed fraudsters to file returns by maintaining lax security measures); Diaz v. Intuit, Inc., No. 15-cv-01778, 2018 WL 2215790, at *1 (N.D. Cal. May 15, 2018) (same).

[338] 66 F. Supp. 3d 1197 (N.D. Cal. 2014).

[339] *Id.* at 1207.

[340] *Id.* at 1210.

company. One cause of action could be negligence. The first step would have to assess if the state owed a duty to protect the victim.[341] The standard of reasonable care extends to foreseeable plaintiffs for foreseeable harm.[342] As for foreseeable plaintiffs, an argument could be made that a person who has been released from prison could foreseeably harm another. The state and the software developer should foresee that hacked or hackable AI in a product used in criminal courts could lead to a dangerous defendant being released and harm to the general public. This is what happened to Christian Rodgers when Jules Black was released and allegedly killed Rodgers.[343] However, this is a very fact-specific analysis. It can be based on biased rationales, harkening to historical myths of people of color being more likely to engage in criminal behavior.[344]

It is foreseeable that software can be hacked. The threat of hacked AI or insecure data is so concerning to corporations that there has been a concerted effort to reinforce legal protections through indemnification clauses between the developer, the manufacturer, and the customer (the state).[345] The question arises regarding what type of data error (leading to AI error) would lead to potential liability. If the data is flawed (as in found data sets), there may be no liability, as this may not be considered foreseeable within the duty of care, or it may be considered a causal break in a factual causation analysis, or could be seen as too far removed in a proximate causation

[341] *See* Dimick v. Hopkinson, 422 P.3d 512, 521 (Wyo. 2018) ("The elements of negligence are well known: '(1) the defendant owed the plaintiff a duty to conform to a specified standard of care; (2) the defendant breached the duty of care; (3) the breach proximately caused injury to the plaintiff; and (4) the injury is compensable by money damages.'" (quoting Brown v. Big Horn Cty. Sch. Dist. No. 3,388 P.3d 542, 546–47 (Wyo. 2017))).

[342] *See* Johnson v. A/S Ivarans Rederi, 613 F.2d 334, 351 (1st Cir. 1980) (labeling an action as negligent where it was foreseeable that an employee would use the dangerous area in the course of their work).

[343] *See* Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 WL 2429574, at *4 (D.N.J. June 11, 2019) (noting that Black's generated Public Safety Assessment score influenced his release).

[344] *See generally* Phillip Atiba Goff et al., *Not Yet Human: Implicit Knowledge, Historical Dehumanization, and Contemporary Consequences*, 94 J. PERSONALITY & SOC. PSYCH. 292 (2008).

[345] *See* Moellenberg et al., *supra* note 290 (recommending the use of "contractual warranties, indemnities, and limitations to control liability risk").

analysis. However, there might be a different analysis to know if hacked data could create a path to liability.

Another path a plaintiff might pursue is product liability. Typically, a plaintiff would have access to sue a software company if the software proved defective.[346] However, the courts and the criminal justice systems are the entities that have privity of contract.[347] There is no privity of contract between the potential civil plaintiffs who are also the criminal defendants. Whether products liability claims can be pursued and by whom will depend on the terms of the contract, which will often provide indemnification to the software company for logical fallacies in the code.[348] Also, the lack of privity between the software company and the third party can bar claims.[349] While the software company will have a contract with the court system, there are often indemnification clauses in the contract that will protect the software company from any liability that a third party might face.[350] Equity and fairness require that third parties be able to recover when harmed by hacked algorithms used in court.

IV. RECOMMENDATIONS FOR ADDRESSING UNTRUSTWORTHY AI

Hacking by nefarious actors is a true threat to criminal justice reform. Without revolutionary efforts, there will be no justice. Lack of scrutiny will cause further disparities in the Black community, communities of color, and low-socioeconomic individuals.[351] There must also be hyper-vigilance in monitoring

[346] *See, e.g.*, Wendorf v. JLG Indus., Inc., 683 F. Supp. 2d 537, 540 (E.D. Mich. 2010) (alleging a software defect associated with a machine's controls).

[347] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 754 (Wis. 2016) (discussing how a private company designed COMPAS to support the Department of Corrections, a part of the state, "when making placement decisions, managing offenders, and planning treatment").

[348] *See* Moellenberg et al., *supra* note 290 (arguing for more indemnification clauses and greater protection for software companies).

[349] *See, e.g.*, Flory v. Silvercrest Indus., Inc., 633 P.2d 383, 387 (Ariz. 1981) (holding that lack of privity will preclude recovery in the case of some warranties).

[350] *See* Moellenberg, *supra* note 290.

[351] *See* Barry Friedman, Opinion, *The Worrisome Future of Policing Technology*, N.Y. TIMES (June 22, 2018), https://perma.cc/F78C-RUBP

the creation, distribution, and manipulation of training sets to guard against cyberattacks. Finally, it is time for machine-learning algorithms to be redefined as tangible property so plaintiffs can access product liability causes of action due to faulty product design.

A. *Hacking Will Have a Disproportionate Impact on Black Communities in the Criminal Justice System*

Black communities are disproportionately represented in the criminal justice system for reasons that have little to do with crime. Every jurisdiction is permitted by statute, regulation, political will, or common practice to maintain records on policing as they see fit.[352] There is no nationwide database on death at the hands of the police, records of arrests, detention, length of detention, plea bargains, sentencing, incarceration lengths, or the demographics of any of the deceased, defendants, or victims.[353] At best, studies rely on estimates, and those numbers paint a grim picture.[354] The Bureau of Justice Statistics found that from June 2015 through March 2016, on average, there was an arrest-related death rate of four-per-day (the 1,348 deaths BJS acknowledged did not include the deaths under federal or tribal law enforcement).[355] The number of Black people harmed by manipulated algorithms will dwarf all historical records, and the gravity of harm is incomprehensible.

("Whether written by humans or a product of machine learning, algorithms take past facts and magnify them into future police actions. Much of street policing in recent years . . . has been deployed disproportionately against minorities and in poor neighborhoods.").

[352] *See, e.g.*, *National Use-of-Force Data Collection*, FED. BUREAU INVESTIGATION, https://perma.cc/7WJK-KMX4 (explaining that participation in this collection of use of force data is open to any jurisdiction but still voluntary).

[353] Rob Picheta & Henrik Pettersson, *American Police Shoot, Kill and Imprison More People than Other Developed Countries. Here's the Data*, CNN (June 8, 2020, 7:13 AM), https://perma.cc/7MV6-9QPU.

[354] *See id.* ("If every US state were counted as a country, the 31 countries with the highest incarceration rates in the world would all be US states, according to the Prison Policy Initiative.").

[355] *See id.*

Charlon McIlwain, author of *Black Software: The Internet & Racial Justice, From the AfroNet to Black Lives Matter*,[356] tells the story of three civil rights-era leaders.[357] Those leaders foresaw the complexities of race, racism, and technology.[358] McIlwain describes how A. Philip Randolph ("the philosopher"), Bayard Rustin ("the planner"), and Roy Wilkins ("the visionary") traced how computing and automation could handily further mask inequalities with shallow capitulation by the people.[359] These men saw the future. Race, racial intent, and racial disparities are so rooted in the code, so pervasive, and so fraught with foregone conclusions, that it is poor design not to have unpacked the racialized exacerbation that occurs from code design.[360] A model following a "burden-shifting test" can help mitigate the harms.[361]

Questions to be raised are several: "Is the model fair? Does the model have a valid business justification? Are there alternative models that are fairer, but maintain reasonable predictive ability?"[362] Intentional design will help stop further disparities.

[356] CHARLTON MCILWAIN, BLACK SOFTWARE: THE INTERNET & RACIAL JUSTICE, FROM THE AFRONET TO BLACK LIVES MATTER (2019).

[357] Charlton McIlwain, *The Three Civil Rights-Era Leaders Who Warned of Computers and Racism*, SLATE (Jan. 30, 2020, 5:50 AM), https://perma.cc/D3ZY-NGXL.

[358] *See generally id.*

[359] Dr. McIlwain calls A. Phillip Randolph the chief ethicist who believed public interest should direct technological creation, Bayard Rustin, the social engineer, who outlined technological governance and a need for the people to be trained for employment, and Roy Wilkins, the visionary, who foresaw a future where computers would be trained to catalog racist ideologies. *Id.*

[360] "Algorithms don't have to look at race to be racist." Friedman, *supra* note 351. PredPol, another algorithm was without a way to correct racial bias, leading to data that did no more than intensify bias. *See* Griffard, *supra* note 5, at 51 (citing Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 16–17 (2016)).

[361] *See* Nicholas Schmidt, *Ethical Algorithms & How Attorneys Can Save Us from Biased AI*, CONSILIENCEML (Feb. 19, 2020), https://perma.cc/VP4P-XN5U (PDF).

[362] *Id.*; *see* Nicholas Schmidt, *Ethical Algorithms: Fixing Discriminatory Machine Learning and Biased AI*, *in* MASTERCLASS: UNDERSTANDING MACHINE LEARNING 62, 72 (2020), https://perma.cc/RAW6-S7TD (PDF); Lum, *supra* note 247, at 8 ("[A] pretrial risk assessment instrument must be developed with community input, revalidated regularly by independent data scientists with

Appreciating that the machines are designed to replicate human behaviors reminds us that the biases in human nature are embedded in the algorithms. It is also foreseeable to anticipate that nefarious actors will use adversarial machine-learning to entwine falsehoods into actual data and outputs. The New York Police Department uses software it expects to produce models that obfuscate race and gender data and other potential proxies for sensitive data.[363] However, even this product has raised serious civil rights and civil liberties issues because of how the software can exacerbate disparities.[364]

If a designer can create AI, then a designer can create fair AI.[365] Teaching and developing an explicitly anti-racist code of conduct for AI designers that also monitors for anti-Blackness in the code will provide a start to addressing the issue of hacking. The first complexity is defining fairness. Lack of conditional parity across the entire group defines discrimination in the code, and machine-learning experts and practitioners must continue to implement fairness in the code and during quality control checks.[366] "Explainable AI," designed in the last five years, gives a model for making the software transparent.[367] Understanding the reasons hackers attack systems will help

that input in mind, and subjected to regular, meaningful oversight by the community.").

[363] *See* Griffard, *supra* note 5, at 45 ("'Patternizr is a new, effective, and fair recommendation engine . . . [that] when used properly, encourage[s] precision policing approaches instead of widespread, heavy-handed enforcement techniques.'").

[364] *See id.* ("[C]onsider whether the developers' goal to build a bias-free predictive policing tool is actually achievable given the limitations of its inputs—racially-biased historic criminal justice data—and its users humans with the potential for errors and cognitive biases."). For additional ideas, see generally Nicole Turner Lee et al., *The Role of AI in the Criminal Justice System* (Living with AI: The Human Impacts of AI Symposium, June 11, 2020), https://perma.cc/DAK5-9G6T [hereinafter *Role of AI in the Criminal Justice System*].

[365] *See* Henk Griffioen, *Fairness in Machine Learning with PyTorch*, GO DATA DRIVEN (May 22, 2018), https://perma.cc/3J6Y-BGPY.

[366] *See id.*

[367] *See* Schmidt, *supra* note 362; Richard Tomsett, *Explainable AI (XAI)*, *in* MASTERCLASS: UNDERSTANDING MACHINE LEARNING 33–61 (2020), https://perma.cc/RAW6-S7TD (PDF).

train developers to create ways to combat criminal justice algorithm assaults better.

B. *There Must Be Stronger Oversight of AI Used in Criminal Justice*

Regulation ensures that the use, distribution, and adoption of innovative technologies are serving society's best interests.[368] Security and privacy of the internet network, data rights, and protection must be carefully monitored. Data gathering has been used as a tool to disempower people.[369] As users will be profiled, analyzed, and considered a quantifiable output before they can act, processes must ensure that an algorithmic proposed output is not considered determinative.[370] Scientists have also determined that algorithms that are more transparent and straightforward perform with the same accuracy as the COMPAS algorithm.[371] If there are higher accuracy and

[368] *See* Schwab, *supra* note 25, at 69–78. It is even anticipated that something as routine as census gathering can be accomplished through Big Data sources. *Id.* at 144. Schwab outlines the governance principles that should be followed during an era of market disruption. Though security and privacy are listed, they serve as the foundation as none of the disruption can be considered forward-moving if it is not correct, transparent, and accountable. *Id.* at 72.

[369] *See* Tim Elfrink, *Once-Secret Files from Gerrymandering Strategist Show GOP Misled Court, Watchdog Group Claims*, WASH. POST (June 7, 2019, 5:01 AM), https://perma.cc/B7XU-55MU (discussing the impact of racial data on gerrymandering).

[370] The question has been asked, "How do we maintain our individuality, the source of our diversity and democracy, in the digital age?" Schwab, *supra* note 25, at 100. This question is being unpacked by philosophers, attorneys, theologians, and coders across the world.

[371] *See* Yong, *supra* note 225 ("[T]his training-wheels algorithm could perform just as well as COMPAS, with an accuracy of 67 percent, even when using just two pieces of data—a defendant's age, and their number of previous convictions."). However, transparent algorithms may not lead to an explanation that will assist a plaintiff in court. *See generally* 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 (2018); Elaine Angelino et al., *Learning Certifiably Optimal Rule Lists for Categorical Data*, J. MACH. LEARNING RSCH. 18(234) (2018), https://perma.cc/WHN5-3XK7; Robin A. Smith, *Opening the Lid on Criminal Sentencing Software*, DUKE TODAY (July 19, 2017), https://perma.cc/MX4Z-V6NE.

transparent products available, they should be used. Scientists, states, and Congress will then have a better model to follow in creating new tech.

Companies must produce a framework for every stage in this process to ensure accountability, including audits, and impact statements. Scientists and developers must continue to train datasets to measure intended variables, quantify and mitigate bias in statistical models, and not conflate multiple distinct predictions.[372] Social scientists and lawyers must be incorporated in the software design process to ensure public policy goals are reflected in the tools and are reproducible for court challenges.[373] There can be the creation of source code on places like GitHub,[374] designed using adversarial machine-learning to check for equal protection violations in code. For example, Bnh.ai, a law firm dedicated to legal issues related to AI and analytics,[375] has outlined ten questions an organization should answer in gauging liability related to the use of AI.[376] Questions such as "how are your organization's models audited for security or privacy vulnerabilities" and "does your company have response plans in place to address AI/ML incidents" of attacks or failures, can guide government systems that decide what types of AI to use.[377] It can also help map whether they have correctly and thoroughly thought through all implications of deploying the software.[378]

Every state and jurisdiction must do a complete and thorough inventory of algorithms that are in use in their

[372] *See* PARTNERSHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 16, 18, 22 (2019), https://perma.cc/C9K3-FR56 (PDF).

[373] *See Role of AI in the Criminal Justice System*, *supra* note 364; PARTNERSHIP ON AI, *supra* note 372, at 27–28, 30.

[374] *See* Georgios Gousios et al., *Lean GHTorrent: GitHub Data on Demand*, MSR 2014: PROCEEDINGS OF THE 11TH WORKING CONFERENCE ON MINING SOFTWARE REPOSITORIES 384 (2014), https://perma.cc/HH8X-GZSZ (PDF).

[375] Seth Colander, *Bnh.ai Is a New Law Firm Focused Only on AI*, VENTURE BEAT (Mar. 19, 2020, 5:00 AM), https://perma.cc/KR7C-VJK7.

[376] BNH.AI, TEN QUESTIONS ON AI RISK, https://perma.cc/Q2MJ-5Y2J (PDF).

[377] *Id.*

[378] *Id.*

criminal justice system, cataloging companies, software, and data sources. If a state does not make its algorithms transparent, it must monitor and process PRAIs used. States should ensure that found data is accurate, show due diligence in reviewing the data, see if created data has been manipulated, and provide the public with a clear understanding of the limits of the technology, beyond even the five warnings presented in *Loomis*. AI developers, states, corrections departments, and judges should query software and algorithms used in the criminal system, principally looking for the ways AI perpetuates bias, the properties of algorithms necessary to mitigate AI bias, and the five places to review for bias in the algorithmic guideline process.[379] If it is evident that AI outcomes further exacerbate racial disparities, it must immediately be removed from the decision-making process by law enforcement, district attorneys, judges, and parole boards.

Federal studies and reform efforts must be unrestrained and directed towards genuine reform. Congressional AI studies must be completed with the input of privacy advocates, scholars, and scientists to counter the outsized influence of tech lobbyists.[380] According to quarterly reports filed with Congress, Carnegie Mellon University was the only organization to

[379] They must know who collects the data, how the algorithm is trained, how it works, how it will be used, how it will be used to understand the feedback loop and how new outcomes influence the next phase of the software. *See generally* Steven D. Pearson et al., *Is Consensus Reproducible? A Study of an Algorithmic Guidelines Development Process*, 33 MED. CARE 643 (1995).

[380] *See* Growing Artificial Intelligence Through Research Act, H.R. 2202, 116th Cong. (2019) ("[R]equir[ing] certain federal activities related to artificial intelligence, including implementation by the President of a National Artificial Intelligence Initiative."); Countering Online Harms Act, H.R. 6937, 116th Cong. (2019) (requiring the Federal Trade Commission to conduct a study on artificial intelligence, and other purposes); GAINS Act, H.R. 6950, 116th Cong. (2019) (requiring the Secretary of Commerce and the Federal Trade Commission to conduct a study on artificial intelligence, and other purposes); AI JOBS Act of 2019, H.R. 827, 116th Cong. (2019) (promoting a 21st-century artificial intelligence workforce); Financial Transparency Act of 2019, H.R. 4476, 116th Cong. (2019) ("[R]equir[ing] federal financial regulatory agencies to adopt specified data standards with respect to format, searchability, and transparency."); *see also* David McCabe, *How Tech's Lobbyists Are Using the Pandemic to Make Gains*, N.Y. TIMES (Apr. 3, 2020), https://perma.cc/YZ3Q-UXR4 (discussing the tech industry's efforts to lobby the federal government to adopt more cloud-based services).

disclose "artificial intelligence" as a federal lobbying issue in 2015.[381] Product liability and other liability issues will soon be subject to more lawsuits, and there should be a federal answer to protect people harmed by AI.[382] On November 17, 2020, the U.S. Office of Management released its final guidance on the regulation of AI, following a February 2019 executive order.[383] Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence,"[384] provides a risk-based approach that prioritizes stronger protections for AI systems that show a potential for higher risk with a focus on economic growth, but is seen as not balanced with an understanding of AI harms.[385] The White House AI Regulatory guidance recognizes a distinction between private sector AI regulation and governments deploying AI systems,[386] but without a clear acknowledgement that private sector third-party software used in government

[381] Gopal Ratnam & Kate Ackley, *Artificial Intelligence Is Coming. Will Congress Be Ready?*, ROLL CALL (June 10, 2019, 7:00 AM), https://perma.cc/8NWK-VHHE.

[382] *See id.* ("[O]paque automated advertising systems driven by algorithms could perpetuate discrimination and avoid scrutiny . . . . Tech-savvy lawmakers say Congress must be better educated before passing legislation addressing artificial intelligence to avoid repeating the failures made with earlier internet technologies."); Chris Opfer, *AI Hiring Could Mean Robot Discrimination Will Head to Courts*, BLOOMBERG L. (Nov. 12, 2019, 6:01 AM), https://perma.cc/XU88-3DKZ ("The Equal Employment Opportunity Commission is already investigating at least two cases involving claims that algorithms unlawfully excluded certain groups of workers during the recruitment process, and seven attorneys told Bloomberg Law it's just a matter of time until courts are asked to weigh in on similar arguments.").

[383] *See* Alex Engler, *New White House Guidance Downplays Important AI Harms*, BROOKINGS INST. (Dec. 8, 2020), https://perma.cc/44GL-58W8 (citing Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 11, 2019)). This executive order directed the federal government to develop five branches for furthering AI: "(1) invest in AI research and development (R&D), (2) unleash AI resources, (3) remove barriers to AI innovation, (4) train an AI-ready workforce, and (5) promote an international environment that is supportive of American AI innovation and its responsible use." *Artificial Intelligence for the American People*, WHITE HOUSE, https://perma.cc/MV2V-5CVJ.

[384] Russel Vought, *Guidance for Regulation of Artificial Intelligence Applications*, WHITE HOUSE (Nov. 17, 2020), https://perma.cc/84WQ-JHNJ.

[385] *See* Engler, *supra* note 383 ("[T]here is a real risk that this document becomes a force for maintaining the status quo, as opposed to addressing serious AI harms.").

[386] *Id.*

systems is happening, this will not achieve the effect of proper regulation. Additionally, the guidance states "Agencies should consider new regulation only . . . in light of the foregoing section . . . that Federal Regulation is necessary."[387] Engler notes that without modernized enforcement processes, the current system allows mechanisms to circumvent the law using algorithms.[388] This was attempted recently by Housing and Urban Development when it attempted to implement a new rule that made it impossible for a plaintiff to prove they were discriminated against by an algorithm.[389] It is unknown what will change with the Biden administration's policies on AI.[390]

[387] *See id.* (quoting VOUGHT, *supra* note 384, at 2).

[388] *Id.*

[389] *See* Elizabeth Fernandez, *Will Machine Learning Algorithms Erase the Progress of the Fair Housing Act?*, FORBES (Nov. 17, 2019), https://perma.cc/YBX7-9B2Z ("[T]he proposal drastically limits the recourse of those who feel that they have been discriminated against—so much so that it may be impossible to show discrimination existed."). The Interdisciplinary Working Group on Algorithmic Justice—a group of ten computer scientists, legal scholars, and social scientists from the Santa Fe Institute and the University of New Mexico—submitted a formal response to this proposal that articulated how algorithms in housing applications may be inherently biased against certain groups of people. Letter from Sonia Gipson Rankin et al., The Interdisciplinary Working Grp. on Algorithmic Just., to Off. of the Gen. Couns., Rules Docket Clerk, Dep't of Hous. and Urb. Dev. (Oct. 18, 2019), https://perma.cc/EH7Q-PBD3 (PDF). On October 25, 2020, the United States District Court of Massachusetts issued a preliminary injunction against enforcement of the proposed HUD rule. Memorandum and Order Regarding Plaintiffs' Motion for Preliminary Injunction Under 5 U.S.C. § 705 to Postpone the Effective Date of HUD'S Unlawful New Rule, Mass. Fair Hous. Ctr. v. U.S. Dep't of Hous. & Urb. Dev., (No. 20-11765) (D. Mass. Oct. 25, 2020), 2020 WL 6390143.

[390] *See* Engler, *supra* note 383 ("It is hard to imagine that changing this guidance is going to be a leading priority of the Biden White House, given all its other pressing problems."). The National Security Commission on Artificial Intelligence—which includes executives from Microsoft, Amazon Web, and Google—submitted a 756-page report to President Biden and Congress laying out their vision for "winning the AI era." NAT'L SEC. COMM'N ON ARTIFICIAL INTEL., FINAL REPORT (2021), https://perma.cc/7GSU-Y3T2 (PDF).

C. *Legal Remedies Are Needed for Parties Harmed by Data Hacks in Criminal Justice Risk Assessment Instruments*

The use of AI through criminal justice risk assessment instruments cements harm to defendants, victims, and stifles the administration of justice. What Apple and United Health did by failing to protect women and Black Americans is already contrary to established Equal Protection under the Fourteenth Amendment.[391] These constitutional violations beg the question: why did the algorithm not scan for violations against protected classes when deployed? And if a judge consults a risk assessment instrument and the tool is also violating the Equal Protection, the decision must be allowed review.

Tort remedies also provide the traditional means of shifting all or part of the economic and non-economic loss from one entity to another due to harm caused by misconduct, deliberately or through inattention.[392] The functions and goals of negligence law are to deter unsafe activities, compensate injured victims, encourage economic growth and progress, and improve effectiveness, efficiency in legal administration, and fairness.[393] Scholars have noted the limitations because the limits of tort law remain undefined, and potentially hacked criminal justice data is a legitimate concern every plaintiff should raise. In *Bauserman v. Unemployment Insurance Agency*,[394] the court noted that the case was remanded so that a cognizable constitutional tort claim could be identified.[395] Resident Fellow

[391] The superintendent of the New York Department of Financial Services, Linda Lacewell said, "Any algorithm, that intentionally or not results in discriminatory treatment of women or any other protected class of people violates New York law." Sridhar Natarajan & Shahien Nasiripour, *Viral Tweet About Apple Card Leads to Goldman Sachs Probe,* BLOOMBERG BUS. (Nov. 9, 2019), https://perma.cc/53TS-B6E2. A letter to UnitedHealth by the Department of Financial Services and Department of Health outlined that "New York Insurance Law, the New York Human Rights Law, the New York General Business Law, and the federal Civil Rights Act all protect against discrimination for protected classes of individuals." Letter from Linda Lacewell, *supra* note 161.

[392] *See* DOMINICK VETRI ET AL., TORT LAW AND PRACTICE 4 (2020).

[393] *See id.* at 13.

[394] 950 N.W.2d 446 (Mich. Ct. App. 2019).

[395] *Id.*

of Yale's Information Society Project, Anat Lior, author of the article, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*,[396] advocates for the adoption and application of a strict liability regime on current and future AI accidents, by delving into and exploring the realm of legal analogies in the AI context, thereby promoting the agency analogy, and subsequently, the *respondeat superior* doctrine.[397] And in *The AI Accident Network: Artificial Intelligence Liability Meets Network Theory*,[398] Lior argues for a way to integrate network theory into the field of AI tort law presenting a new methodology about the appropriate liability regime that should apply when AI causes damages.[399] *In Civil Liability for Artificial Intelligence: What Should its Basis Be?*,[400] scholar Jean-Sébastien Borghetti outlines that AI used in different fields can be addressed through particular liability regimes, whether strict liability or general liability.[401] To protect liberty and fairness, the definition of product liability must identify AI and algorithms as falling within the definition of a

[396] Anat Lior, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*, 46 MITCHELL HAMLINE L. REV. 1043, 1043 (2020)

> This article explains why the agency analogy is the best-suited one in contrast to other analogies that have been suggested in the context of AI liability (e.g., products, animals, electronic persons, and even slaves). As a result, the intuitive application of the respondeat superior doctrine provides the AI industry with a much-needed underlying liability regime that will enable it to continue to evolve in the years to come, and its victim to receive remedy once accidents occur.

[397] *Id.*

[398] Anat Lior, *The AI Accident Network: Artificial Intelligence Liability Meets Network Theory*, 95 TUL. L. REV. (forthcoming 2021).

[399] *Id.*

[400] Jean-Sébastien Borghetti, *Civil Liability for Artificial Intelligence: What Should Its Basis Be?*, 17 LA REVUE DES JURISTES DE SCIENCES PO 94 (2019).

[401] *Id.* The harm caused by autonomous vehicles, for example, is probably better addressed through strict liability regimes for traffic accidents through a general liability for the AI regime. *See* Charikleia Bertsia, *Legal Liability of Artificial Intelligence-Driven Systems (AI)*, INT'L HELLENIC UNIV. (2019), https://perma.cc/KE59-2U5U (PDF) (analyzing the Product Liability regime in the European Union to determine whether it suitably addresses issues raised by increasing AI usage).

product and not just the final software. The Third Restatement of Torts must also define AI and machine-learning algorithms as a product. As it took time to define harm by any product, harm by hacking and cybercrimes, though complex, must be given space for redress by victims. These steps can provide wrongly convicted persons and other victims recourse under the law.

CONCLUSION

If there is oversight, AI can increase fairness in the criminal justice system. Else, failure to ensure the validity of AI-based products will lead to extinguishing liberty interests enshrined in the Constitution. United States Supreme Court Justice Oliver Wendell Holmes, Jr., set the nation on course with his understanding of this principle, in his oft-cited observation: "The life of the law has not been logic; it has been experience."[402] AI is how we will rewrite society's rules, and how we will explain, defend, and refine the Constitution. It is time for the law to be forward-thinking in protecting people from potential harms of AI, whether it has been weaponized or has entered the public sphere without proper scrutiny. Achieving fairness is a lofty and necessary goal, but cementing disparate outcomes will decelerate justice's evolution.

Ernest Rutherford, the "father of nuclear physics," said, "[y]ou should never bet against anything in science at odds of more than about 10-12 to 1 against."[403] Most Americans support policies that restrict the scope of autonomous technologies. Yet society is beyond this moment and cannot, nor should it, go backward. One day, AI will mimic only the most optimal principles and protocols of human nature. It will show that humans strive to be filled with compassion, justice, and fairness. It will not show greed and bias, nor will it exacerbate superficial

[402] OLIVER WENDELL HOLMES, THE COMMON LAW 1 (1881). "The felt necessities of the time, the prevalent moral and political theories, intuitions of public policy, avowed or unconscious, even the prejudices which judges share with their fellow-men, have had a good deal more to do than the syllogism in determining the rules by which men should be governed." *Id.*

[403] RICHARD J. LIPTON, THE P=NP QUESTION AND GÖDEL'S LOST LETTER, viii (2010).

or historical errors and harms. This aim propels researchers forward to augment human decision points with AI. There is a need for the systems to protect privacy, offer transparency, and purposely require software developers and states that use this technology to ensure it is serving the public. These public policy concerns will rebut many established torts, contracts, and patent laws and principles. The technology illuminates flaws and inequities that have always been in the system. Because of the exponential speed of technology, legal systems have not kept up with the rate of change. New principles must ensure at a minimum that there are no further inequities created in the system. Social justice must ensure that rational thought is not being manipulated to perform as the machine would.[404] If not careful, mindful, and vigilant, history will find humankind responding as an *algorithm* would rather than ensuring the human spirit and capacity to improve is captured in the technology. [405] Now is the time to use AI to devise a society without the historic human errors of bias.

Noted anti-lynching advocate, Ida B. Wells-Barnett, posited, "The way to right wrongs is to turn the light of truth upon them."[406] Transparency is vital to safeguarding equity through design and must be the first step. If an algorithm is used in the criminal justice system and has been hacked, it is a defective product that harms everyone. Tort law must rapidly adapt to allow plaintiffs theories of accountability and liability through tort reform under state and federal law. The goal is to understand our biases and work with them, rather than hide from them. AI in criminal justice will need law and tech to reduce biases, improve justice, and achieve fairness.

[404] *See* FOER, *supra* note 24, at 77 ("That's why Facebook has so few qualms about performing rampant experiments on its users. The whole effort is to make human beings predictable—to anticipate their behavior, which makes them easier to manipulate."); *id.* at 220 ("Machines are increasingly suggesting the most popular topics for human inquiry, and humans are increasingly obeying.").

[405] Oscar Wilde wrote in his 1889 essay, *The Decay of Lying* that "Life imitates Art far more than Art imitates Life." OSCAR WILDE, THE DECAY OF LYING 10 (1891).

[406] LORI AMBER ROSSENER ET AL., POLITICAL PIONEER OF THE PRESS 117 (2018).