2020

**AI Report: Humanity Is Doomed. Send Lawyers, Guns, and Money!**

Ashley M. London

James B. Schreiber

Follow this and additional works at: https://dsc.duq.edu/dlr

Part of the Science and Technology Law Commons

**Recommended Citation**


Available at: https://dsc.duq.edu/dlr/vol58/iss1/6

This Article is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Duquesne Law Review by an authorized editor of Duquesne Scholarship Collection.
AI Report: Humanity Is Doomed. Send Lawyers, Guns, and Money!

Ashley M. London, J.D.* and James B. Schreiber, Ph.D. **

I. INTRODUCTION .................................................................98
II. THE MACHINES ARE COMING FOR LAW JOBS .............102
III. INSIDE THE BLACK BOX ....................................................105
   A. The AI Systems Creation Myth.................................106
   B. Mathematical Models Used in Creating AI Systems.................107
   C. Linear/Logistic Models, Tree Based Models, and Neural Networks ..........108
   D. Feeding the Process with Value-Laden Data, an Inherently Biased Process......109
IV. THE PARADE OF HORRIBLES ...............................................111
   A. AI and Hiring ..............................................................113
   B. AI Systems and the First Amendment.....................115
   C. AI, Self-Driving Cars and YouTube’s Pedophile Problem .................116
   D. Data Scraping from Social Networks Tested in Courts ..................121
   E. AI and Legal Research, Education, and Practice........................122
V. A PATH FORWARD TO CREATING TRUSTWORTHY AI ......124
VI. CONCLUSION AND AN ISSUE SPOTTING CHECKLIST ......128

* The idea of this paper came from a presentation delivered by Assistant Professor of Legal Skills Ashley London and Dr. James Schreiber at the Duquesne University School of Law conference held in Pittsburgh, Pennsylvania, on April 26 and 27, 2019, titled Artificial Intelligence: Thinking about Law, Law Practice, and Legal Education. Professor London would like to thank her family, as well as Duquesne University School of Law Professors Ann Schiavone, Maryann Herman, and Katherine Norton for their support. Further, a special thank you to student editor and annotator Erika Dowd.

** Dr. James Schreiber, Professor of Epidemiology and Statistics at the Duquesne University School of Nursing, and Professor London met by chance in 2018 when introduced to collaborate on a statistics project for the law. This paper is designed to explain complex topics in a straightforward manner, so professional issue spotters and problem solvers (i.e., lawyers) will be able to identify areas of concern and be equipped to face the practical and ethical challenges emerging in this era of the rise of the machines.
I. INTRODUCTION

When machines and computers, profit motives and property rights, are considered more important than people, the giant triplets of racism, extreme materialism, and militarism are incapable of being conquered.

-Martin Luther King, Jr.

*Delivered 4 April 1967, Riverside Church, New York City, speaking about the Vietnam War.*

Warren Zevon knew when you are hiding in Honduras and “[t]he sh*t has hit the fan,” it is time to call for the lawyers, guns, and money. Replace “hiding in Honduras” with the real harms caused by Artificial Intelligence (AI) system algorithms, such as enabling systemic workplace gender discrimination, autonomous vehicles striking pedestrians with darker skin tones, and pedophiles being provided with video content of underage children, the refrain sounds more like: “send in the lawyers to sort out the enormous, manmade mess.”

AI systems are powerful technologies being built and implemented by private corporations motivated by profit, not altruism. Change makers, such as attorneys and law students, must therefore be educated on the benefits, detriments, and pitfalls of the rapid spread, and often secret implementation of this technology. The implementation is secret because private corporations place proprietary AI systems inside of black boxes to conceal what is inside. If they did not, the popular myth that AI systems are unbi-

---

1. CLAYBORNE CARSON, A CALL TO CONSCIENCE: THE LANDMARK SPEECHES OF DR. MARTIN LUTHER KING, JR. 158 (2001) (ebook). Civil rights leader Martin Luther King, Jr. delivered this speech on April 4, 1967, at the Riverside Church in New York City. *Id.* He shared the program with other national leaders to condemn the Vietnam War and the arrogance of the wealthy West in its pursuit of profits over the welfare of its people and the people of the war-torn country. *Id.*


3. See MEREDITH WHITTAKER ET AL., AL NOW REPORT 2018, at 4-5 (2018), https://ainow-institute.org/Al_Now_2018_Report.pdf. A ‘black box’ system is one that is not transparent, in other words, what happens inside of that box is not open to scrutiny by anyone other than the creating entity or company. *Id.* Watchdog groups are working diligently to end the ‘black box’ effect in the use of AI systems. See *id.* at 11. It is this secret nature of some AI systems that advocates say violates due process and has given rise to lawsuits that will be discussed later in this paper. See *id.*
ased machines crunching inherently objective data would be revealed as a falsehood. Algorithms created to run AI systems reflect the inherent human categorization process and can, in some respects, become a lazy way to interact with the world because the systems attempt to outsource the unparalleled cognitive skills of a human being into a machine. AI systems can also be extremely dangerous because human categorization processes can be flawed by bias (explicit or implicit), racism, and sexism.

There is a big profit motive in AI system development and implementation. Revenue generated from the direct and indirect application of AI system software is estimated to grow to as much as $36.8 billion by 2025. As a subset, the global legal analytics market alone is expected to reach a staggering value of $1,858 million by 2022. But, as Fei-Fei Li, one of the major developers of these technologies recently argued, "we will hit a moment when it will be impossible to course-correct." What she means is soon it may be impossible to reverse the damage done to vulnerable portions of the population through the widespread use of algorithmic-based systems. Li is a modern voice echoing the prescient statements made by Dr. King in 1967 about the cascade of evils facing society when the human moral compass is outsourced to machines, computers, algorithms, and the profits that flow from their rapid rise and ubiquitous use are prioritized over the condition of humanity. How many mistakes should a machine be allowed to make in the name of developing a deep learning function if those mistakes put marginalized human beings at a further disadvantage, and who is charged with policing this technology when it errs?

In many cases, attorneys are in the best position to monitor, guide, and correct the use of AI systems steeped in the practice of

8. CARSON, supra note 1, at 158.
navigating ethical quagmires and solving problems. After all, the individual licensure of every attorney in this country depends on the ability to abide by the Rules of Professional Conduct. An attorney could be disbarred for violating any one of these rules. In contrast, AI system developers are under no such formalized ethical constraints, and indeed are under very few state or federal rules and regulations governing their conduct or product development.

This article suggests creating widely accepted and enforceable rules of ethics to govern so-called “Trustworthy AI.” This article proposes that the first step in that direction is to introduce attorneys and law students to the basis of AI system development and the ethical guidelines recently promulgated by the European Union Commission (EU). These guidelines suggest the fundamental approach to ensuring AI systems are ethical should be based upon a “[r]espect for fundamental rights, within a framework of democracy and the rule of law.”

When confronted with an AI issue, every attorney and law student should begin by asking the following questions. First, who developed the algorithm and for what purpose? Second, who chose the variables used? Third, who defined success? And forth, who was at the table when the decision points were implemented in the AI development process? Each of these value-laden decision points have an inherent power differential embedded into the decision-making process.

10. See Model Rules of Prof’l Conduct r. 8.4 (AM. BAR ASS’N 1983). This rule provides, in relevant part, that misconduct is: “(c) engag[ing] in conduct involving dishonesty, fraud, deceit or misrepresentation . . . [and] (g) engag[ing] in conduct that the lawyer knows or reasonably should know is harassment or discrimination on the basis of race, sex, religion, national origin, ethnicity, disability, age, sexual orientation, gender identity, marital status or socioeconomic status in conduct related to the practice of law.” Id.

11. In 2017, a non-profit group, known as Future of Life Institute (FIL), established a set of guidelines that form an AI code of ethics known as the Asilomar AI Principles. See Asilomar AI Principles, FUTURE LIFE INST., https://futureoflife.org/ai-principles/ (last visited Jan. 24, 2020). This code includes suggestions such as: recommending a healthy exchange between AI researchers and policy makers; when applying AI to personal data, a person’s real or perceived liberty must not be unreasonably curtailed; and humans should choose how and whether to delegate decisions to AI systems to accomplish human-chosen objectives. Id. Unlike the Model Rules of Professional Conduct for attorneys, these are guidelines without a mechanism for enforcement. See id. They have, however, been endorsed by California in August 2018, as well as by AI researchers at Google DeepMind, Facebook, Apple, and more. FLI Team, State of California Endorses Asilomar AI Principles, FUTURE LIFE INST. (Aug. 31, 2018), https://futureoflife.org/2018/08/31/state-of-california-endorse-asilomar-ai-principles/.


13. Id. at 9.
apparatus.\textsuperscript{14} Even armed with the best intentions, a developer cannot account for all potential sources of bias, including implicit or unconscious bias.\textsuperscript{15} It becomes especially important then to ask the questions posed by Meredith Whittaker, Executive Director of the AI Now Institute, "[w]hat assumptions about worth, ability and potential do these systems reflect and reproduce? Who was at the table when these assumptions were encoded?"\textsuperscript{16} The majority of people at the table developing these technologies are white, and they are male.\textsuperscript{17} There is a crisis of diversity at the heart of the AI sector.\textsuperscript{18} At Facebook for example, only 15% of all AI researchers are female.\textsuperscript{19} At Google, that number shrinks to 10\%.\textsuperscript{20} For African American workers, those numbers are even smaller.\textsuperscript{21} At Google, 2.5\% of its full-time workforce is black, while at Microsoft and Facebook that number increases to 4\%.\textsuperscript{22} Current data on the state of


\textsuperscript{15} See About Us, PROJECT IMPLICIT, https://implicit.harvard.edu/implicit/aboutus.html (last visited Jan. 22, 2020). Founded in 1998, Project Implicit is a non-profit collaboration between researchers from Harvard University, the University of Washington, and the University of Virginia. Id. The goal of the organization is to educate members of the public about hidden, or "implicit" biases. Id. The group developed the Implicit Association Test (IAT) that has generated data and research regarding implicit racial attitudes across the country. See Preliminary Information, PROJECT IMPLICIT, https://implicit.harvard.edu/implicit/takeatest.html (last visited Jan. 22, 2020). Implicit bias is defined as "[t]he attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. [T]hese are [a]ctivated involuntarily, without awareness or intentional control. [T]hey [c]an be either positive or negative." Cheryl Staats et al., State of the Science: Implicit Bias Review 10 (2017), http://kirwaninstitute.osu.edu/researchandstrategicinitiatives/implicit-bias-review/. Implicit biases are formed as a result of mental associations that have formed from direct and indirect messages we receive from the world, and people, around us. See id.


\textsuperscript{20} Id.

\textsuperscript{21} Id.

\textsuperscript{22} West et al., supra note 18, at 11.
gender and racial diversity in the field of AI is decidedly grim, both in the corporate industrial sector and in academia, where 80% of all AI professors are men. When AI tools such as facial recognition systems mistakenly categorize a black person’s face as a gorilla, or when Uber’s application suspends transgender drivers due to an oversight in its programming, these problematic outputs are a sign of flawed algorithmic input affecting human beings in discriminatorily, socially, and legally unacceptable ways. An urgent re-evaluation is in order, along with systemic design process changes, because the development of AI is not just about profits, it is about power.

II. THE MACHINES ARE COMING FOR LAW JOBS

For a profession operating in a system based upon the principle of stare decisis, there exists a strong bias supporting the rapid development and application of so-called machine learning in law and the legal field. In a 2018 American Bar Association (ABA) study, attorneys reported saving time and increasing efficiency were the biggest advantages of adopting of AI systems in law firms. Companies seeking to sell AI systems to law firms say firms need to adopt this technology as of yesterday to “[s]tay in the [g]ame.” AI systems allegedly help firms maximize their budgets by increasing speed in areas such as contract review, mechanizing repetitious

27. Stare decisis means to stand by things decided and not to disturb settled points of law. Stare decisis, BLACK’S LAW DICTIONARY (8th ed. 1992). Stare decisis is the doctrine of precedent under which it is necessary for a court to follow earlier judicial decisions when the same points arise again in litigation. Id.
tasks, and increasing a firm’s ability to scale services to both new and old clients to turn higher profits.  

Promoters of the disruption of technology in the legal field say that in the “short run” AI systems will lead to “greater legal transparency, more efficient dispute resolution, improved access to justice . . . . [And] lawyers will be empowered to work more efficiently . . . .”31 For example, the average human attorney can review a contract in ninety-two minutes, or approximately fifteen billable increments of six minutes each, while an AI system can perform the task in twenty-six seconds.32 Big law firms, government and public-interest organizations, and law schools are all being asked to do more with less money, the idea that machines can become as powerful as an expensively trained advocate in the law is seductive. For example, an AI system dubbed Lex Machina (Latin for “law machine”) acquired by LexisNexis in 2015, is on the thirteenth expansion of its legal analytics platform that began with a focus on Intellectual Property (IP) cases.33 The product mines litigation data to provide attorneys with information such as the average duration of a legal matter, damage awards, resolution, opposing counsel litigation history, and historic rulings from judges on motions and other decisions.34 The company’s website says its programming is powered by proprietary algorithms that are “new,” “unorthodox,” and “extremely valuable.”35 Such enthusiastic promotion belies the fact that the development and implementation of AI systems is complex, multi-faceted, and potentially fraught with issues. Attorneys will therefore be called upon to course correct when the offspring of these projects go awry.

If profit is one of the biggest motives spurring law to look for ways to embrace AI systems, there are other identifiable factors at play.

30. See id.
34. See Hichman, supra note 29.
in the sharp rise of legal technology applications, including: a reduction in entry-level law jobs across the country,\textsuperscript{36} a recent slump in law school admissions figures,\textsuperscript{37} the expense of civil litigation, and the need to try to close the ever-growing justice gap for low-income families.\textsuperscript{38} Lawsuits are expensive and so are the large white-collar law firms that appear to be the fastest adopters of AI system technology.\textsuperscript{39} The average civil lawsuit in America today costs between $43,000 and $122,000 from complaint to verdict.\textsuperscript{40} It is little wonder then that, according to the ABA’s 2018 Legal Technology Survey Report, AI system usage is greatest at law firms with over one hundred attorneys.\textsuperscript{41} At least one large law firm, the prominent international Big Law firm O’Melveney & Myers LLP, based in Los Angeles, California, made headlines when it announced it was pioneering the introduction and use of AI in its recruiting and hiring process for associates to improve diversity.\textsuperscript{42} While this may not lower costs to its clients, a move to increase diversity is certainly a good public relations for a large law firm. Black attorneys make up approximately 3.3% of lawyers in Big Law, and women continue to be underrepresented in leadership roles.\textsuperscript{43} Fortune 500 companies are looking to spend their legal dollars with more diverse law firms, so applying AI systems in this context serves both altruistic


\textsuperscript{38} Lewis Creekmore et al., \textit{The Justice Gap: Measuring the Unmet Civil Legal Needs of Low-Income Americans} 6 (2017), https://www.lsc.gov/sites/default/files/images/TheJusticeGap-FullReport.pdf (reporting 86% of civil legal problems reported by low-income Americans in the past year received inadequate or no legal help).


\textsuperscript{40} Paula Hannaford-Agor & Nicole L. Waters, \textit{Estimating the Cost of Civil Litigation} 7 (2013), http://www.courtstatistics.org/~media/Microsites/Files/CSP/DATA%20PDF/CSPH\_online2.ashx.

\textsuperscript{41} See \textit{1 American Bar Association Legal Technology Survey Report}, supra note 39, at 21.


and profitability goals.\textsuperscript{44} Since 2012, legal technology startups have raised $757 million in capital to develop new AI systems technology.\textsuperscript{45}

In 2017, the McKinsey Global Institute found that while nearly half of all legal tasks could be automated by current technology, only 5\% of all jobs could be entirely automated.\textsuperscript{46} Applying its current definition of technology—widely available or being tested in a lab—McKinsey estimates 23\% of a lawyer’s job can be automated.\textsuperscript{47} If lawyers and law students are not aware of the trends, the primarily privately-held technology companies alone will set the pace and tone of the adoption of AI systems. Legal professionals will increasingly be called upon to predict where conflicts will arise and how humans program personal bias and potential illegalities into these algorithmic models as well as to police offending AI systems.

III. INSIDE THE BLACK BOX

AI systems are neither intelligent nor “artificial intelligence.” It is more like, “artificial artificial intelligence.”\textsuperscript{48} AI is humans helping machines help humans perform tasks better, faster, more economically, and even predictively.\textsuperscript{49} In seeking to assign that uniquely human characteristic of intelligence to computers, the risk of potential ethical issues increases in proportion to one’s reliance on a machine with no independent moral compass, conscience, or rich background of experience (schemas and scripts) to draw upon to make nuanced distinctions.

There is also a fundamental difference in the application of automated or predictive analytics services and those that purport to use AI systems. For example, a credit card loyalty program might use predictive analytics to determine whether it could increase reward redemption by spending more money marketing to specific credit card holders.\textsuperscript{50} Predictive analytics systems review data from the

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{44} Id.
\item \textsuperscript{47} Lohr, supra note 45.
\item \textsuperscript{48} See CATHY O’NEIL & RACHEL SHUTT, DOING DATA SCIENCE 169 (2014).
\item \textsuperscript{49} See id.
\end{enumerate}
\end{footnotesize}
past to spot patterns, which allow human users to make predictions, test certain assumptions, and take action.\textsuperscript{51}

On the other hand, an AI system is a term used by the European Union to describe the next step on the predictive analytics continuum, which is also sometimes referred to as "machine learning."\textsuperscript{52} AI systems make assumptions, reassess the models, and reevaluate all of the data inputted into them without the intervention of a human "operator."\textsuperscript{53} Taking this a step further, the term "deep learning" is used primarily within neural network systems by AI systems to complete complex tasks, like classifying large data sets, or operating a self-driving car where the machine must be prepared to interact with a variety of variables at lightning speed.\textsuperscript{54} The system, essentially, begins training itself by making mistakes.

A. The AI Systems Creation Myth

Humans process a massive amount of data every day, and one way the brain manages to do that quickly and efficiently is by using its almost unparalleled ability to categorize everything—fellow humans, laws, social situations, and even recognizing everyday objects. "The need for effective retrieval from this vast storehouse of information has prompted humans to develop a storage strategy based on semantic coding and organization of input information."\textsuperscript{55}

In short, a process. In fact, scientists note human intelligence is based upon abilities that are superior to anything yet conceived and built by a human, i.e., "intelligent machines."\textsuperscript{56} Statistical models can be used as one lens to understand and represent reality.\textsuperscript{57} The models, though, are artificial constructions where assumptions are made, extraneous details are removed, and others are left as abstractions.\textsuperscript{58} Each one of those assumptions, removals, and abstractions are decision points. Thus, one must not only examine what was included but focus also on what was not included and the processes that led to those decisions. AI systems will always be first

\begin{itemize}
  \item \textsuperscript{51} Id.
  \item \textsuperscript{52} Eur. Comm’n Guidelines, supra note 12, at 36. "Machine learning" is a term that will be used throughout this paper to signify "artificial intelligence systems" that employ machine learning to make assumptions, learn, and provide predictions on larger scale. Reavie, supra note 50.
  \item \textsuperscript{53} Reavie, supra note 50.
  \item \textsuperscript{54} Marr, supra note 9.
  \item \textsuperscript{55} Uday A. Athavankar, Categorization . . . Natural Language and Design, Design Issues, Spring 1989, at 100, 100.
  \item \textsuperscript{56} Id.
  \item \textsuperscript{57} See O’Neil & Shutt, supra note 48, at 28.
  \item \textsuperscript{58} Id.
\end{itemize}
and foremost, a human endeavor complete with very human deficiencies, blind spots, and occasional flashes of brilliance. Acknowledging the human bias encoded into an algorithm is the first step in exploring the false creation narrative that AI systems are born completely unbiased, operate as flawless science drones, and do not make the same mistakes as humans.

Unlike machines, humans experience the world in real time and there are many moments of curiosity or grey areas of doubt. These moments generally create a desire to understand what is going on or what has happened. “The fuzziness of [those] boundaries, [is] an important characteristic of the human categorization process. . . .”59 To better understand the world, the mind does not finely discriminate between highly similar concepts.60 Instead, “the mind automatically selects the cognitively economical option of neglecting the infinite differences among objects to behaviorally and cognitively usable proportions.”61 While this process might not matter so much when deciding upon whether or not to define a coffee-drinking receptacle as a Tervis Tumbler, a Starbucks travel mug, or a Styrofoam cup, it becomes problematic when humans engage in social categorization. Decades of research have demonstrated that categorizing people in terms of their social identities can lead to stereotyping and prejudice.62

B. Mathematical Models Used in Creating AI Systems

AI systems are the results of some mathematical model or algorithm. Algorithms are nothing more than a set of rules that a computer can follow. Models are mathematical expressions linking variables of interest to other variables of interest.63 When discussing AI systems, terms like algorithms and machine learning are used interchangeably. While the terms may have different meanings according to the context in which they appear, the end goal is to predict and classify a set of data using programmer-driven decision points. Prediction is where the goal is to forecast something like the price of a car, house, or salary request. This is a numeric based prediction. In classification, on the other hand, the goal is to accurately place, a person for example, in a pre-defined category, such

59. Athavankar, supra note 55, at 104.
60. Id. at 102.
61. Id.
63. Model and algorithm are used interchangeably in this proposal for brevity.
as a yes or no. A third and slightly different goal is to create clusters that may be used later to predict or classify.

C. Linear/Logistic Models, Tree Based Models, and Neural Networks

Three common prediction methods are used to reach the goal of building an accurate model, or an algorithm: linear/logistic models, tree based models, and neural networks. The linear/logistic model involves the creation of a best fit line through a set of data points. This mathematical procedure is used for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets of the points from the curve. “The proof uses calculus and linear algebra” to find a relationship between variables. It is useful because it is simple to use when predicting a continuous outcome, such as the price of a house. In applied statistics, an outcome variable is “predicted” as an equation of variables of interest, otherwise known as independent variables. Logistic regression is used when the outcome variable is categorical such as “yes” or “no.” These tools work well if the predictor variables are not overly related to each other, but they can also miss complex relationships between variables.

Tree based models come in three general types: decision trees, random forest, and gradient boosting. Decision trees are easy to understand and visually appealing. These are generally yes or no rules based on the data and all possible outcomes that can be seen through the branches of the tree and are used for classification and regression. A random forest, a “collection of decision trees,” is used as an ensemble whose results are aggregated, and a random forest uses many decision trees based on rules created from subsamples. The combination of these trees increases the performance level of

---

68. See Hosmer & Lemeshow, supra note 65, at 1.
the model overall. By aggregating many smaller decision trees, this method limits overfitting as well as errors due to bias.\textsuperscript{72} Gradient boosting uses a weaker decision tree than a random forest and creates a group of decision trees to create a high performing model. One issue in gradient boosting is small changes in the data set can create radical changes in the model along with difficulty in explaining the predictions.\textsuperscript{73}

Neural networks use a hidden layer, commonly termed interconnected neurons, which send messages to each other.\textsuperscript{74} A neural network is also known as “deep machine learning” and is a new name for an approach to AI systems in existence since 1944.\textsuperscript{75} Neural nets are modeled loosely on the concept of the human brain and consist of deeply interconnected processing nodes.\textsuperscript{76} Deep learning models use several of these layers stacked on top of each other to create results or decisions.\textsuperscript{77} The significant difference between neural networks and the other methods is the ability to handle extremely complicated tasks, e.g., image recognition; but neural networks can be slow to develop.\textsuperscript{78} In order to produce results or predictions, the nodes must be trained using weighted data sets.\textsuperscript{79} Theorists find the level of opacity in the training and feeding of these neural nets to be problematic in terms of being able to identify problematic decision points being used to produce data.\textsuperscript{80} Because of this, neural nets have cycled in and out of favor with developers since their inception.\textsuperscript{81}

D. Feeding the Process with Value-Laden Data, an Inherently Biased Process

No matter what AI system is used, each one must be fed massive quantities of data to begin its process and each one employs value-laden assumptions. The data is typically messy when it is first collected. Therefore, the first step in any process requires that the

\textsuperscript{72} See Liberman, supra note 70.
\textsuperscript{73} Kelley, supra note 64.
\textsuperscript{74} KEVIN GURNEY, AN INTRODUCTION TO NEURAL NETWORKS 13 (1997).
\textsuperscript{76} Id.
\textsuperscript{77} See EUR. COMM’N GUIDELINES, supra note 12, at 36.
\textsuperscript{78} See generally Mengye Ren et. al., Learning to Reweight Examples for Robust Deep Learning, in 80 PROCEEDINGS OF MACHINE LEARNING RESEARCH, THIRTY-FIFTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING 4334 (Jennifer Dy & Andreas Krause eds., 2019).
\textsuperscript{79} Id.
\textsuperscript{80} Hardesty, supra note 75.
\textsuperscript{81} Id.
data is organized and cleaned prior to running through the model in an attempt to achieve the desired result. The desired result of any model must be defined as its “success,” or what it hopes to achieve by running the data through the given set of decision points. Examples of successful endpoints for AI systems can include determining who is more likely to pass the bar exam, which consumers are more likely to purchase a promoted product, and who is more likely to be a recidivist in the criminal law context. If the data set is large enough, it is split into random parts to create an algorithm and then to verify or validate that model with the other data split or splits. Once this work is completed, new data typically arrives, and there will be a move to an optimization phase based on the current model and what was defined as success.

While this process may sound objective and very scientific, it is the opposite. Every system is laden with the values programmed into them by the human developers, and each programmed value will have an inherent power differential. Even the data that exists was not pulled together by a strictly objective decision-making system. Value laden means the person who developed the algorithm chose the variables included, the definition of success, and the optimization process of that success definition. That is a great deal of power. The most widely-reported issues are that of racial bias and sexism, but it would be a mistake to think that only those “hot-button” social issues are implicated. An individual who is subject to the application of any given algorithm could potentially be categorized in any number of ways separate and distinct from gender or skin color. “[H]umans are likely among the richly multidimensional stimuli” and many distinct categories may be applied simultaneously such as occupation, religion, sexual orientation, socio-economic status, and education.

Dr. Cathy O’Neil, a data scientist, formerly working with Wall Street is at the forefront of ringing the alarm about the dangers of the sudden overarching influence of AI systems. Dr. O’Neil’s research has demonstrated that mathematical models are not unbiased, and that the unregulated use of big data reinforces discrimination. Dr. O’Neil continues to call on the modelers of algorithms to take responsibility for the use of black box algorithms and charges policy makers to regulate their use. Power differentials

82. O’NEIL & SHUTT, supra note 48, at 41.
83. Bodenhausen et al., supra note 62, at 125.
84. See generally CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION (2016).
focus on who is making the decision, how many people were involved in that decision, who is accountable for that decision, and who holds the decision makers accountable.

Decision points are human opinions embedded in a mathematical model. Often, these systems are based on the opinions of the person who has access to the data and thinks these variables will work best or the variables that appear to work best. This personal conflict is reflected in an individual’s selection of products and environment and also in the selection of variables plugged into an algorithm’s model. “One seeks assurance and psychological comfort that come from predictable responses expected from the category and also looks for deviations representing personal identity.”

However, if a human programmer’s personal identity (explicit or implicit) is that of a racist, then high levels of racial prejudice are almost inevitably going to become part of any machine-driven categorization scheme and the AI system will perpetuate a bias toward stereotypically expected behavior.

IV. THE PARADE OF HORRIBLES

Without an increase in oversight, big data algorithms can magnify and replicate the biases that exist in our society at large, leading to bigger issues that have already begun to appear in the court systems. So, the fact that human beings create AI should give society pause because humans are fallible. The algorithmic systems that turn data into information and predictions rely on imperfect input, logic, probability, and those who design them.

Under former President Barack Obama, the White House released several key reports on big data to advance the conversation about the use of such systems and to ensure that these systems do not become barriers to entry for certain groups of people. In addition, one of the reports sought to ensure that the output of these systems was

86. Athavankar, supra note 57, at 107.
87. Id.
88. Bodenhausen et al., supra note 62, at 127.
91. Id.
not rooted in hidden stereotypes that could “hardwire discrimination, reinforce bias, and mask opportunity.”

One critical area is the increasingly problematic use of algorithms in the criminal justice system. New policies in states such as California, New Jersey, and New York, are rolling out so-called “risk assessment” algorithms that recommend to judges whether a person who has been arrested should be released. In Broward County, Florida, a risk assessment scoring system called COMPAS, used on more than 7,000 people in 2013-2014, was shown to be biased against black suspects. ProPublica obtained the risk scores and checked to see how many of the people classified by the AI system were charged with new crimes over the next two years and found that only 20% of the people predicted to commit violent crimes actually did. It also found that the algorithm being used was only slightly better than a coin flip. The program was also more likely to falsely flag black defendants as future criminals, at twice the rate as white defendants. While on the other hand, white defendants were mislabeled as low risk more often than black defendants.

Michelle Alexander wrote in a New York Times opinion article about the problems on machine learning risk assessment algorithms, e-carceration, and the down-stream effects of those algorithms. The down-stream effects of these algorithms are not getting nearly enough attention—especially the risk that entire communities of people could become trapped in digital prisons that lock them out of opportunity. E-carceration is a relatively new term of art used to describe the use of technology to deprive people of their liberty, specifically the use of algorithms that purport to appear color-blind and unbiased. It is important for attorneys and law students to remember that these “products” are being created by private corporate interests and sold to states for shareholder profit. Even if the algorithms, programs, and GPS-enabled electronic monitoring devices that the algorithms control are employed by government entities subject to judicial oversight, the private corporations

92. Id.
93. Alexander, supra note 85.
95. Id.
96. Id.
97. Id.
98. Id.
99. Alexander, supra note 85.
100. Id.
that produce these are not held to similar standards of transparency or accountability.\textsuperscript{101}

The following are examples illustrating the how, why, and what can go wrong with the output of the application of AI systems in a civil context, or the downside risk. The companies employing these AI systems are largely protected by existing laws designed to keep corporate trade secrets concealed from public scrutiny and protected from litigation and recovery by damaged plaintiffs.

A. AI and Hiring

As early as the 1990s, online job applications such as Monster.com allowed employers to advertise employment opportunities for a lower price than if the employer placed a help wanted ad in the classified section of the local newspaper.\textsuperscript{102} Soon, employers began to accept applications via online platforms, which led to the need to find ways to track, sort, identify, and process the sheer volume of applications received in order to find a candidate that best suited the employers' needs.\textsuperscript{103} Seeing an opportunity to generate revenue, technology vendors began making increasingly complicated programs that employed algorithms with lofty goals such as increasing diversity or forecasting future outcomes in the form of scores or rankings of candidates and using the incredible amounts of data being submitted via these online platforms from both job seekers and employers alike.\textsuperscript{104} In 2018, a staggering 60\% of technology companies reportedly plan to invest in AI software to facilitate recruitment because companies perceive that using machines instead of human capital saves time and money.\textsuperscript{105}

Employers seeking workers have three basic goals: reduce time to hire, reduce cost per hire, and maximize the quality of a hire such that qualified (and that word alone is loaded with human-specified definitions of what it means to be “qualified”) candidates will stay longer with the company to benefit the business.\textsuperscript{106} Turnover in terms of time, money, and manpower is costly, and since it takes an

\begin{itemize}
\item \textsuperscript{101} See generally id.
\item \textsuperscript{103} Id.
\item \textsuperscript{104} Id.
\item \textsuperscript{106} BOGEN & RIEKE, supra note 102, at 6.
\end{itemize}
average of six weeks to fill a job opening, employers and their recruiters want to get it right the first time.

There are hiring tools on the market that purport to assist employers in these goals. For example, Amazon, the automation reliant e-commerce giant, began using hiring tools in 2014 to ramp up its hiring process. Using resumes submitted to the company over ten years (the data set), the algorithm used to sort through these resumes and penalized those that included words such as, “women” or “women’s,” and downgraded graduates of two all-female colleges. On the other hand, it privileged resumes featuring strong, masculine, words such as “executed” and “captured.” Amazon abandoned its machine learning system for hiring because the system did not like women. Given that Amazon’s workforce is about 60% male, this is not shocking. The company reportedly created 500 computer models and taught them to recognize 50,000 terms that showed up on candidates’ resumes, but still ended up with biased results against gender and randomly promoted underqualified candidates.

While Amazon admitted its mistake and said it was killing that particular machine learning project, the fact remains that companies around the world are implementing or have implemented technologies like this to recruit candidates for employment. Giant global companies such as Goldman Sachs Group Inc., Unilever, and Wal-Mart Stores Inc. are reportedly using algorithms to diversify candidate pools and to fast-track employees to management positions. According to Unilever, the company’s AI can filter between

107. Id.
110. Id.; Reuters, supra note 89.
111. Id.; Reuters, supra note 113.
112. Id.
113. Id.
60% and 80% of candidates resulting in 80% of applicants who are interviewed by a human in the company’s Human Resources Department actually being hired.\footnote{115} However, the input of those algorithms and the results of its application are uniformly kept in the dark. As Goodman, a staff attorney at the American Civil Liberties Union (ACLU) Racial Justice Program points out, “these tools are not eliminating human bias—they are merely laundering it through software.”\footnote{116}

\section*{B. AI Systems and the First Amendment}

Activists and watchdogs will tell you that the biggest concern regarding the proliferation of AI systems remains transparency. On April 2, 2018, a federal judge allowed attorneys with the ACLU to proceed with a First Amendment case\footnote{117} challenging the federal Computer Fraud and Abuse Act, which appears to prevent studies on the discriminatory use of algorithms by making it a crime to violate a website’s terms of service.\footnote{118} Terms of service, contained in the fine print, often include rules against creating multiple tester accounts, providing inaccurate contact information, or using automated methods to record publicly available data like search results and ads.\footnote{119} Those terms are set by individual sites and can change at any time.\footnote{120} Researchers use practices like setting up dummy accounts to test whether sites are more likely to show higher interest rate loan ads to people of color or to show higher paying jobs to men who search employment listings.\footnote{121}

The case was filed on June 29, 2016 by the ACLU on behalf of plaintiffs Christian W. Sandvig, Kyaratso Karahalios, Alan Mislove, Christopher Wilson, and First Look Media Works, Inc.; two of those plaintiffs were Associate and Assistant Professors of Computer Science at Northeastern University, who designed a study to test whether the ranking algorithms on major online hiring websites

\begin{thebibliography}{99}
\item\label{115} Id.
\item\label{116} Goodman, supra note 89.
\item\label{117} Sandvig v. Sessions, 315 F. Supp. 3d 1 (D.D.C. 2018). The attorneys on the case are Esha Bhandari and Rachel Goodman of the American Civil Liberties Union Foundation and Arthur B. Spitzer and Scott Michelman of the American Civil Liberties Union of the District of Columbia, Washington, D.C. Id.
\item\label{119} Id.
\item\label{120} Id.
\item\label{121} Id.
\end{thebibliography}
produce discriminatory results. The study tested whether women or people of color were adversely affected by the use of these algorithms. The complaint states that without the ability to conduct online audit testing, “policymakers and the American public will have no way to ensure that the civil rights laws continue to protect individuals from discrimination in the twenty-first century.” The court’s most recent decision permits Professors Mislove and Wilson to proceed with their claims that their research activity—which requires providing false information to websites as part of their tester profiles—is protected under the First Amendment. The case continues to work its way through the court system in the United States District Court for the District of Columbia.

C. **AI, Self-Driving Cars and YouTube’s Pedophile Problem**

Algorithmic output is ever-present, whether one is using a crosswalk as a pedestrian in a city that is testing autonomous vehicles or watching a video on YouTube. Self-driving cars are more likely to hit pedestrians of color regardless of the time of day, according to a February 2019 study of the object detection systems currently used in autonomous vehicles. Touted as the modern solution for a reduction in transit costs that translate to better goods pricing for consumers, self-driving cars are also sold as a planet-saving solution to reduce our individual reliance on cars and, thus, reduce the consumption of fossil fuels and reduce emissions. However, it was not until the Department of Defense sponsored a series of challenges between the years 2004-2007 that Google, Inc. began seriously investing in the technology to the point of testing autonomous vehicles in Pittsburgh, Pennsylvania. As a result of the extensive testing being done in Pittsburgh, Mayor Bill Peduto signed an executive order outlining objectives and expectations for autonomous

---

123. Id.
124. Id. at 2.
125. Id. at 35.
vehicle testing in March 2019.\textsuperscript{130} Pittsburgh is one of the first cities to pass such legislation, which calls for transparency and knowledge of autonomous vehicle testing occurring on public streets.\textsuperscript{131} Lawmakers were quick to note that these limitations and expectations did not apply to the technology’s commercialization, and no provisions on enforcement or penalties were created for companies who fail to meet these standards.\textsuperscript{132}

Today’s autonomous cars are powered by predictive algorithms that rely on large sets of data that must perform tasks such as: recognizing road signs; obeying the applicable speed limit; and, perhaps most importantly, knowing when to apply the brake system to avoid hitting objects like human pedestrians.\textsuperscript{133} It takes an enormous number of robust data sets being inputted into the algorithms by engineers for the machine learning mechanisms to begin accurately predicting the variables that these self-driving cars will encounter in real life.\textsuperscript{134} The problem goes back to who is inputting this data and creating the programs. A team of researchers at Georgia Institute of Technology in Atlanta, Georgia, recently published their findings suggesting that the standard object detection used by autonomous vehicles has a higher predictive accuracy for pedestrians who score lower on the Fitzpatrick Scale of skin types.\textsuperscript{135} First developed in 1972 by Harvard researcher and dermatologist Dr. Thomas B. Fitzpatrick, as part of a study on the effects of sunscreen and skin types, this scale characterizes the color of a person’s skin based on its reactive categories, i.e., color.\textsuperscript{136} This classification system was adopted by the Food and Drug Administration (FDA) in 1972 for the evaluation of sun protection factor (SPF) values of sunscreen.\textsuperscript{137} Generally speaking, categories one through three correspond to lighter skin tones than categories four through six.\textsuperscript{138}

The researchers noted that earlier studies, which showed issues with facial recognition software regarding the proper identification of both women and those with Fitzpatrick skin types four through

\textsuperscript{131} \textit{Id.}
\textsuperscript{132} \textit{Id.}
\textsuperscript{134} \textit{Id.}
\textsuperscript{135} \textit{Wilson et al., supra note 126, at 1.}
\textsuperscript{136} Silonie Sachdeva, \textit{Fitzpatrick Skin Typing: Applications in Dermatology}, 75 \textit{Indian J. Dermatology Venerology & Leprology} 93, 93 (2009).
\textsuperscript{137} \textit{Id.}
six, compared to groups with a higher degree of facial recognition accuracy, i.e., white men, inspired them to employ the scale to categorize pedestrians for purposes of the study.\footnote{Wilson et al., supra note 126.} They also cite the ACLU report that found Amazon’s facial recognition system incorrectly matched a number of darker-skinned members of Congress to mugshots from arrests across the country.\footnote{See Jacob Snow, Amazon’s Face Recognition Falsely Matched 28 Members of Congress with Mugshots, ACLU (July 26, 2018, 8:00 AM), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.}

The Georgia Tech researchers concluded that standard models for object detection, trained on standard data sets, appear to exhibit a higher rate of precision in regard to people lower on the Fitzpatrick skin type scale.\footnote{Wilson et al., supra note 126.} In plain language, this means that autonomous vehicles avoid hitting lighter skinned people at a higher rate than darker skinned people. The researchers also showed that some changes during the algorithm’s “learning” phase—the time when it is beginning to crunch data to come to conclusions and make predictions that can be replicated over time with greater accuracy—can partially mitigate this disparity if the source of capture bias is not considered before the models are deployed.\footnote{Id.}

The study, which has not yet been peer reviewed, is not without its critics who say that the Georgia Tech researchers did not use the same datasets (i.e., the photos, images of pedestrians, and street conditions, for example) as the developers of the autonomous vehicles.\footnote{Bill Howard, Cameras, AI on Self-Driving Cars May Miss Darker-Skinned Faces, ExtremeTech (Mar. 7, 2019, 12:42 PM), https://www.extremetech.com/extreme/287152-cameras-ai-on-self-driving-cars-may-miss-darker-skinned-faces.} If nothing else, this groundbreaking study offers critical insight into the risks of algorithmic bias, especially for those human beings with darker skin tones, and challenges developers to consider the diversity of data required to protect all drivers and pedestrians.

 Algorithms used as part of online platforms can be just as dangerous if they are not programmed, employed, and monitored properly. For example, YouTube’s Digital Playground, an automated recommendation system that connects viewers to content powered by AI technology, has come under fire in June 2019 for suggesting home videos of children to pedophiles.\footnote{K.G. Orphanides, On YouTube, a Network of Pedophiles Is Hiding in Plain Sight, Wired (Feb. 20, 2019), https://www.wired.co.uk/article/youtube-pedophile-videos-advertising.} Videos of children playing in their own backyards, wearing bathing suits, doing
gymnastics, or just getting dressed have racked up more than 400,000 views per video due to the automated algorithm that prompts users to view other video content through a progression of recommendations based on prior views. YouTube’s algorithm specifically suggests videos that are seemingly popular with other pedophiles, most of which have hundreds of thousands of views and feature disturbingly inappropriate comments. While YouTube, which has billions of users worldwide, began disabling some of the comments when the matter was brought to its attention, the algorithm itself is still in use and drives 70% of views on the platform. The company shrouds the details of how the system formulates these choices in secrecy. Jonas Kaiser, a researcher at Harvard’s Berkman Klein Center for Internet and Society, first stumbled upon the videos while researching a project focusing on YouTube in Brazil. He does not believe YouTube designed the program to serve the prurient interests of pedophiles, but the effect of a “disturbingly on point” algorithm is to connect these viewers with both innocent and sexually-charged video content driven by the expressed preferences of its users.

YouTube has not discontinued the use of its Digital Playground algorithm because it is a lucrative business for the San Bruno, California based company purchased by Google in 2006 for $1.65 billion and now operating as a subsidiary of the tech giant. The company continues to monetize the algorithm by selling advertisement space to major corporations who pay to place their content in streams of highly-popular videos. In February 2019, Wired published an article in its United Kingdom online edition that showed one video of two young girls doing yoga was accompanied by pre-roll advertising from L’Oréal and had almost two millions views. The magazine alerted other advertisers who began questioning YouTube’s policies and pulling advertisement deals. Official company policies promulgated in 2017 state that YouTube will disable comments on videos

146. See Orphanides, supra note 144.
147. Fisher & Taub, supra note 145.
148. Id.
149. Id.
151. Orphanides, supra note 144.
152. Id.
153. Id.
where users say "inappropriate" things, "provid[e] guidance for creators who make family-friendly content," "engag[e] and learn[] from experts," and "doubl[e] the number of Trusted Flaggers" to heighten efforts to protect families and kids using the platform. These vague policies should prompt any attorney or law student who has read cases such as New York v. Ferber, Jacobellis v. Ohio, and United States v. Williams to ask what those cases' principles really mean in practice. Because even when a questionable or inappropriate comment is disabled on these YouTube videos—typically of children acting innocently—the algorithm continues to promote these videos and allow viewers to continue to watch and share them, meaning that the cycle continues.

---

155. 458 U.S. 747 (1982). The Supreme Court reached a unanimous decision in this case out of New York that challenged a New York child pornography statute that prohibited persons from knowingly promoting a sexual performance by a child under the age of sixteen. Id. at 749. The statute gave an extremely wide definition of sexual conduct. Id. at 751. The Supreme Court held that the statute did not violate the First Amendment because of the state's overwhelming interest in "safeguarding the physical and psychological well-being of a minor" and protecting children from being exploited to produce pornographic materials. Id. at 756-57 (quoting Globe Newspaper Co. v. Superior Court, 457 U.S. 596, 607 (1982)). "The prevention of sexual exploitation and abuse of children constitutes a government objective of surpassing importance." Id. at 757.
156. 378 U.S. 184 (1964). The Supreme Court reversed the decision by the Supreme Court of Ohio that resulted in an upheld conviction against a local movie theater manager for possessing and exhibiting an allegedly obscene film. Id. at 195-96. While the court reversed the conviction of the theater manager upon judging the material in the movie not to be obscene, it reaffirmed that a state's interest in protecting children from obscene material was a "legitimate and indeed exigent interest of the States and localities throughout the Nation in preventing the dissemination of material deemed harmful to children." Id. at 195. The differentiating factor in this case is that the movie at issue contained a love scene between a male and a female, which the court concluded was not aimed at a child audience but instead the public at large. Id. at 195-96.
157. 553 U.S. 285 (2008). In this case, the Supreme Court decided that a federal statute aimed at criminalizing the possession and distribution of material described as child pornography (whether or not it actually depicted underage participants) was not overbroad and therefore not violative of the First Amendment. Id. at 288. The Court noted that the statute at issue "tracks the material held constitutionally proscribable in Ferber and Miller: obscene material depicting . . . children engaged in explicit conduct." Id. at 292-93. Further, Supreme Court precedent holds the First Amendment does not protect child pornography. Id. In this case, the statute required that the defendant holds material out to be real child pornography or that the defendant leads others to believe the material being offered is real child pornography; therefore, the issues are questions of fact and not vague or indeterminate. Id. at 306. "Child pornography harms and debases the most defenseless of our citizens. Both the State and Federal Governments have sought to suppress it for many years, only to find it proliferating through the new medium of the Internet." Id. at 307.
158. Orphanides, supra note 144.
D. Data Scraping from Social Networks Tested in Courts

The legal application of “scraping” data from social networks without the network’s consent has been tested by the courts. The latest in a series of high-profile cases out of the United States Court of Appeals for the Ninth Circuit was decided on September 9, 2019 in hiQ Labs, Inc. v. LinkedIn Corp. The court of appeals affirmed the district court’s grant of a preliminary injunction in favor of hiQ Labs, Inc. In effect, the court of appeals ruled that LinkedIn could not deny a data analytics company access to publicly available member profiles, a move allowing the controversial practice of data scraping to continue and placing the business interests of a company over the privacy concerns raised by LinkedIn Corp. The court found that hiQ established a likelihood of irreparable harm to its business should the preliminary injunction be allowed to stand. It noted that hiQ raised serious questions about whether its stated causes of action were preempted by the Computer Fraud and Abuse Act (CFAA). Ultimately, the court decided the CFAA’s prohibition on accessing a computer “without authorization” is only violated when the person attempts to “circumvent” a computer’s generally applicable access rules; not when a data scraping company like hiQ is accessing data made publicly available, like the LinkedIn user profiles. The court also left open potential state law remedies to victims of data scraping such as: trespass to chattels, copyright infringement, misappropriation, unjust enrichment, conversion, breach of contract, and breach of privacy.

The CFAA is the government’s attempt to criminalize hacking, or the unauthorized access to computers and networks. The CFAA provides a civil remedy that provides for a fine or imprisonment. Academics and researchers may now use this recent ruling to justify the use of data-scraping bots to conduct research into the discriminatory effects of algorithms. The analysis is different in the realm of profit-seeking companies such as hiQ who need access to

159. 938 F.3d 985 (9th Cir. 2019).
160. Id. at 1005.
161. Id.
162. Id. at 993.
163. Id. at 999.
164. Id. at 999-1002.
165. Id. at 1001-02.
166. Id. at 1004.
data to survive. Litigation regarding the scope of the CFAA as to the legal and illegal harvesting and use of data will continue to be used to test the boundaries of what it means to be a public website and who can access and copy information scraped from a so-called public website. The key CFAA language, “without authorization” may one day appear in front of the Supreme Court as courts across the country subject the federal statute to conflicting interpretations.

E. AI and Legal Research, Education, and Practice

Providing lawyers and law students with access to courses on legal analytics or data science will become an increasingly critical part of the modern legal practice and law school experience. Law schools that do not offer such courses in the design, development, implementation, use, and legal ramifications of big data will need to move in this direction or find themselves left behind. In fact, “technology competence” has been on the ABA’s radar since the approval of an amendment to comment 8 of Model Rule 1.1 in 2012.

So far, thirty-eight states have adopted the revised comment, including the Commonwealth of Pennsylvania. The revised comment reads as follows:

[to maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject.]

At least two jurisdictions, Florida and North Carolina, have recently adopted mandates, which state that all licensed attorneys in

---


171. MODEL RULES OF PROF’L CONDUCT r. 1.1 cmt. 8 (AM. BAR ASS’N 1983).


173. MODEL RULES OF PROF’L CONDUCT r. 1.1 cmt. 8 (emphasis added).
those states must complete continuing legal education (CLE) credits devoted to technology training.\textsuperscript{174} But as this article has illustrated, it will not be enough for attorneys and law students to simply know how to use the newest technologies, but rather they must also understand what is going on inside of the black boxes of AI systems to maintain competency.

One of the places attorneys and law students encounter AI systems every day is in the legal research systems used by popular legal databases such as LexisNexis, Westlaw, Ravel, Casetext, and Fastcase. Type in a case name or a search term, and voila, an unseen algorithm is generating the corresponding results. One might even expect each of the systems to return similar results when faced with similar or identical inquiries, but that is not the case. A 2017 study conducted by Susan Nevelow Mart, an Associate Professor and Director of the Law Library at the University of Colorado School of Law, showed there was very little overlap in the cases that appeared in the top ten results returned by each of the databases she examined.\textsuperscript{175} An average of 40\% of the cases were unique to one database, and only about 7\% of the cases were returned in search results in all six databases, which demonstrates that each database is somehow privileging information, or using different decision points, to get to the results generated.\textsuperscript{176} If a researcher knew what a search algorithm was privileging, then better or more accurate results could be obtained for clients in a business where time really is money.\textsuperscript{177} Simply answering inquiries is not where the AI application to legal research will stop. Legal research providers such as LexisNexis are rolling out the beta versions of analytics products now. For example, LexisNexis is releasing a product called, Context. This language analytics program supposedly will allow legal professionals to build arguments designed to sway judges in favor of their clients.

Machine intelligence is predicted to be one of the greatest disruptors of the role of lawyers in the history of the legal profession—most specifically in the areas of discovery, legal research, document generation, and predicting outcomes. Regulatory issues will con-


\textsuperscript{176} Id.

\textsuperscript{177} Id. at 389.
continue to arise, as will issues in the area of professional responsibility, or legal ethics.\textsuperscript{178} But the question remains: who or what is watching the computers and programmers responsible for creating these AI systems that we have shown touch on many aspects of our modern lives?

V. A PATH FORWARD TO CREATING TRUSTWORTHY AI

Created by humans, employed by humans, affecting humans, and profiting humans, AI systems should be developed and governed in a way to maximize trust in both their creation and output. Without aligning to ethical norms, AI systems cannot be trustworthy.\textsuperscript{179} Lawyers and law students are charged with protecting fundamental human rights with a high degree of ethical responsibility. The first step to understanding what it means to create ethical computer systems is examining the only set of well-developed guidelines for the ethical implementation and use of AI systems in the world, those promulgated by the EU.

In 2018, the European Union Commission, which is a politically independent executive arm of the European Union,\textsuperscript{180} produced the first report of its kind on the development of ethical guidelines for trustworthy AI in 2019.\textsuperscript{181} The report and its guidelines attempt to set forth three pillars to substantiate its goal of supporting "ethical, secure and cutting-edge AI made in Europe."\textsuperscript{182} The first two focus on the economics of AI development, but the third focuses on "ensuring an appropriate ethical and legal framework to strengthen European values."\textsuperscript{183} The report was designed to be delivered to AI stakeholders, those people and corporations designing, developing, deploying, implementing, using, or being affected by AI.\textsuperscript{184} Compliance with the guidelines is discretionary, but AI systems do not operate in a lawless world.\textsuperscript{185}

The report outlined that trustworthy AI has three components, it should be lawful, ethical, and robust.\textsuperscript{186} The focus within the report is on the ethics and robust components, as the legal component will

\textsuperscript{178} ALARIE ET AL., supra note 31, at 13.
\textsuperscript{179} EUR. COMM’N GUIDELINES, supra note 12, at 4-5.
\textsuperscript{181} EUR. COMM’N GUIDELINES, supra note 12, at 1.
\textsuperscript{182} Id. at 4 (citations omitted). The first two pillars are: "i) increasing public and private investments in AI to boost its uptake[] and ii) preparing for socio-economic changes." Id.
\textsuperscript{183} Id.
\textsuperscript{184} Id. at 5.
\textsuperscript{185} Id. at 6.
\textsuperscript{186} Id. at 5.
vary from country to country. The ethics argument is founded on the fundamental rights established within the European Union Treaties and European Union Charter with a common component of human dignity. Human dignity is the idea that every human being has intrinsic worth. Additionally, every human is a moral subject and not an object, and thus, AI systems must be developed in a manner that "respects, serves and protects humans' physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs." This statement presents a high bar conceptually, without details of what each of this means pragmatically. For example, what are the concrete red lines that would clearly infringe on mental integrity and, therefore, should not be crossed? As noted by the Committee, the focus of what should be done versus what can be done becomes another central focal point in the ethical discussions of an AI system.

The second fundamental right is the "freedom of the individual," which includes freedom to make life decisions for oneself and freedom from sovereign intrusion. But, there is a clear acknowledgement that at times intervention must occur at the government level to ensure equal access the benefits and opportunities of using AI systems. Additionally, AI systems must not have "(in)direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation." Thus, the focus must be on how to improve individual life, freedoms, and positive engagement in society and not for power or manipulation. The result is to improve individual and collective well-being. Related, the report authors also argue that AI systems must be based on a respect of democracy, justice, and the rule of law and that the systems should serve to maintain and foster democratic processes. Included in this argument is the commitment to the rule of law and to ensure due process and equality before the law. The final fundamental rights are equality, non-discrimination, solidarity, and citizens' rights. The AI system

---

187. Id. at 9-10.
188. Id. at 10.
189. Id.
191. EUR. COMM’N GUIDELINES, supra note 12, at 11-12.
192. Id. at 10.
193. Id.
194. See id. at 11.
195. Id.
196. Id.
should not generate unfairly biased decisions, which obviously includes respecting vulnerable populations. AI systems have the potential to improve the function of government yet could negatively impact individuals and infringe on their rights; thus, safeguards must be built into the systems.

After the fundamental rights, the report states that there are four ethical principles to guide AI systems. They are: respect for human autonomy, prevention of harm, fairness, and explicability. Respect for human autonomy in this context is the ability of individuals to have full and effective self-determination over themselves. Again, the goal is to improve human experiences and is best accomplished with human oversight of the processes in the AI systems. Prevention of harm is met through the principle that AI systems should not adversely affect human beings. This relates back to human dignity along with mental and physical integrity. Technical robustness requires that it is not open to malicious use. Thus, vulnerable populations should receive more attention and be included in the development and implementation of these systems.

The Committee also created a “non-exhaustive” list of seven non-hierarchical interacting areas of concern that should be a focus during development, implementation, and the life cycle of the AI system:

1. **Human agency and oversight**: Including fundamental rights, human agency and human oversight

2. **Technical robustness and safety**: Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility

3. **Privacy and data governance**: Including respect for privacy, quality and integrity of data, and access to data

4. **Transparency**: Including traceability, explainability and communication

---

197. Id. at 12.
198. See id.
199. Id.
200. Id.
201. Id.
202. Id.
5. **Diversity, non-discrimination and fairness[•]** Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation

6. **Societal and environmental wellbeing[•]** Including sustainability and environmental friendliness, social impact, society and democracy

7. **Accountability[•]** Including auditability, minimisation and reporting of negative impact, trade-offs and redress.\(^2^{03}\)

Of the seven areas of concern, transparency needs further discussion due to the key concepts of traceability, explainability, and communication.\(^2^{04}\) These are related to explicability above but warrant more information. As the impact on people’s lives increases, there must be a path for explaining the system’s decision-making process. As for human decisions, the focus must also be on how the use of the system is shaping the decision-making process from its design to rationale to implementation. Finally, for communication, humans have the right to know that they are interacting with an AI system. There must also be a mechanism that allows humans to decide not to engage with the system.

The final version of the report is not without critics. Committee Member Dr. Thomas Metzinger wrote an editorial in Der Tagesspiegel, that the report is an ethics whitewashing and a marketing sales narrative.\(^2^{05}\) More importantly, he writes that trustworthy AI is conceptual nonsense because machines cannot be trustworthy.\(^2^{06}\) But Dr. Metzinger also noted the EU guidelines are currently the best thing that is out there at this time.\(^2^{07}\)

In sharp contrast, the United States first introduced its “American AI Initiative” through an Executive Order issued by President Donald Trump in February 2019.\(^2^{08}\) The order, titled “Executive Order on Maintaining American Leadership in Artificial Intelligence,” lists five principles that drive the initiative and can be summarized as follows: (1) the United States must drive technological breakthroughs in AI systems; (2) the United States must drive development of technical standards to reduce barriers to testing and

\(^{203}\) Id. at 14.

\(^{204}\) See id.

\(^{205}\) Metzinger, supra note 190.

\(^{206}\) Id.

\(^{207}\) Id.

deployment of AI systems; (3) the United States must train American workers to develop and apply AI system technologies; (4) the United States “must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people;” and (5) the United States must promote an international environment to support “American AI research and innovation and open[] markets for American AI industries.”

The word “ethics” does not appear even once in the order. However, making a path for profitability and support of research and development for the creation and growth of the AI systems industry is front and center. In fact, President Trump specifically names artificial intelligence as a research and development priority in his 2019 Fiscal Year Budget, and he calls it a key area of focus. The budget requests more than $84 billion in research, engineering, and prototyping activities to maintain “technical superiority.” The Executive Order calls on the National Science and Technology Council (NSTC) Select Committee on Artificial Intelligence to coordinate this American AI Initiative.

In the meantime, lawyers and law students in the United States should consider using the EU Committee’s framework to spark a discussion about the development of our own set of ethical guidelines for the development of so-called Trustworthy AI, especially as it inexorably assumes a role of dominance. As Ronald Regan said, restating a maxim first introduced by rabbinic sage Hillel the Elder, “[i]f not us, who? And if not now, when?”

VI. CONCLUSION AND AN ISSUE SPOTTING CHECKLIST

AI systems are only as good as the human creators behind the algorithms. AI systems can help close the justice gap for low-income families or help connect pedophiles to view video content featuring young children. AI systems can promote or disadvantage women and minority job candidates. AI systems can serve our virtues or our vices. As Dr. Martin Luther King, Jr. foretold, when

209. Id.
211. Id.
these machines powered by algorithms built for profit become more important than basic human dignity, then the destructive forces of our changeable human nature—the “giant triplets of racism, extreme materialism, and militarism”—are given free rein. Have we hit the time when it is “impossible to course-correct” as Fei-Fei Li warns? Or does humanity still have time to address the real issues caused by the proliferation of AI systems, without proper checks and balances before the very computers humans build independently decide how this all ends? It is time to send in the lawyers and the money. Maybe not the guns, oh, wait . . .

**AI ATTORNEY ISSUE-SPOTTING CHECKLIST**

The following is a list of basic questions any attorney or law student should ask when working with AI systems, in addition to consulting the list of seven non-hierarchical interacting areas of concern listed in the EU report and discussed above. These are where the potential ethical issues may arise in the creation and application of any AI system.

- What is the goal of this algorithm?
- What data is being inputted?
- Who is in charge of inputting the data?
- What are the algorithm’s decision points?
- Who decided on those decision points?

---


216. See Ben Tarnoff, *Weaponised AI is Coming. Are Algorithmic Forever Wars Our Future?*, _GUARDIAN_ (Oct. 11, 2018, 5:00 PM), https://www.theguardian.com/commentisfree/2018/oct/11/war-jedi-algorithmic-warfare-us-military; see also _Contracts for Oct. 25, 2019, U.S. DEPT DEFENSE_, https://www.defense.gov/Newsroom/Contracts/Contract/Article/1996639/ (last visited Apr. 14, 2020). On Oct. 25, 2019, the U.S. Department of Defense announced that Microsoft Corporation had been awarded the $10 billion, ten year contract to create the Joint Enterprise Defense Infrastructure Program (JEDI) Cloud missile defense system. _Id._ This is a cloud computing system that weaponizes artificial intelligence and includes the use of unmanned drones that can be programmed to locate targets in real time, essentially making it less time consuming to find people to kill in war zones. _Id._ The system will be designed to serve United States forces all over the world. _Id._
• Were potential issues of bias accounted for in constructing those decision points and how?

• Do you have an ethicist on the development team? Do you have a true critical outsider providing input?