FORMULAS FOR TROUBLE:

Why Smart Companies Must Tread Carefully with Algorithms

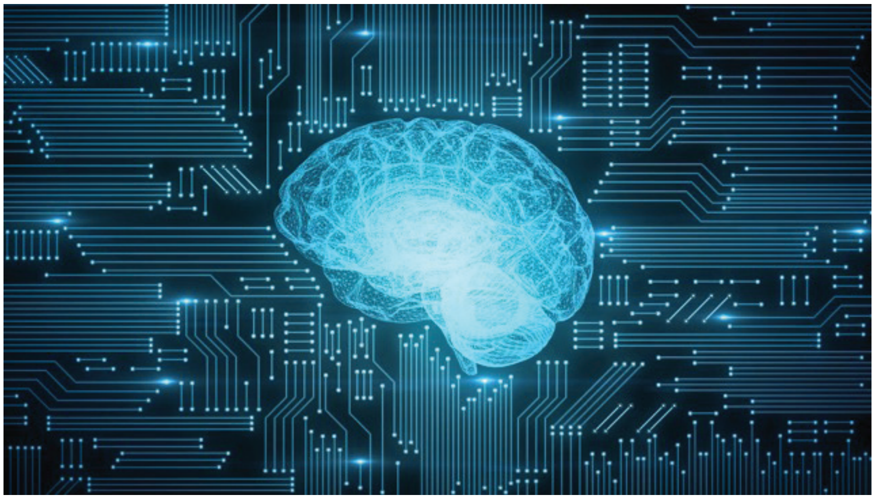
April 2018

OpenMIC

www.openmic.org

TABLE OF CONTENTS

INTRODUCTION



We are living in an age of algorithms and Artificial Intelligence (AI), with rapid advancements in data analytics and cognitive technologies enabling rich and unprecedented insights into 21st century problem solving. As the Internet of Things becomes reality — with digital technologies embedded in almost every aspect of everyday life — these algorithms promise huge breakthroughs across fields such as business, science, medicine, transportation, housing and government.

Artificial intelligence is "one of the most important things that humanity is working on," says Google CEO Sundar Pichai. "It's more profound than … electricity or fire." DJ Patil, former chief data scientist for the U.S. government, says, "We have to remember these algorithms and these techniques are going to be the way we're going to solve cancer. This is how we're going to cure the next form of diseases." Artificial intelligence will also help address global challenges such as climate change and world hunger, experts say.

At the same time, alarms are sounding about possible negative societal impacts of algorithms and artificial intelligence. As more and more companies deploy data-driven technology to target advertisements to potential customers, recruit new employees and offer loans and mortgages, research shows algorithms can bring about unintended, discriminatory outcomes. Journalists, researchers, data scientists and developers are beginning to uncover the dangerously massive scope of potential harm that can be inflicted on people who already face systematic marginalization due to their race, gender, age, class, sexuality, ability, or zip code.

> *AT OPEN MIC, WE BELIEVE THAT COMPANIES EMPLOYING ALGORITHMS AND ARTIFICIAL INTELLIGENCE, AS WITH ALL TECHNOLOGIES, MUST DEVELOP POLICIES AND PRACTICES THAT PROVIDE TRANSPARENT OVERSIGHT, INCLUDING BY ACCOUNTING FOR THE IMPACT OF ALGORITHMS ON PEOPLE'S LIVES AND OUR SOCIAL CONTRACT.*

Responsible businesses must step up to participate in policy dialogues, advocate for values-based industry practices and help set and implement technical standards that promote long-term, sustainable, and equitable economic solutions. Shareholder engagement is essential to ensure they do. Investors can be key players in encouraging companies to assess and address the potential impacts of algorithms before they are deployed. The purpose of this report is to highlight the critical issues and challenges posed by the race to score gains in the marketplace through the unseen hands of algorithms and AI.

Bias is a critical concern. It's without question that big data invites discrimination, but when companies lack transparency and clear insight into the full impact of algorithms, it is often difficult to pinpoint precisely how. Without fully understanding the impact of algorithms, says Kate Crawford, co-founder of the AI Now Institute at New York University and a Principal Researcher at Microsoft Research New York, the danger is that "algorithmic flaws aren't easily discoverable: How would a woman know to apply for a job she never saw advertised? How might a black community learn that it were being overpoliced by software?"

When algorithms learn to replicate the existing patterns of bias in society, companies face significant legal, financial and reputational risk. "Given the potential for such long-term negative implications, it's imperative that algorithmic risks be appropriately and effectively managed," the consulting firm Deloitte warned in a recent report. On the positive side, organizations that employ "a risk-aware mind-set will have an opportunity to use algorithms to lead in the marketplace, better navigate the regulatory environment, and disrupt their industries through innovation."

How can companies and organizations ensure that their use of AI helps build a future that is fair, just and ethical?

Step one in that process may be to adapt the kinds of risk assessment tools that are already routinely used in most large organizations to gauge and manage conventional business risks. Algorithmic risks often don't show up in these assessments, according to a 2017 report by Deloitte, because of "the complexity, unpredictability, and proprietary nature of algorithms, as well as the lack of standards in this space."

According to Deloitte, corporate managers and boards should ask themselves:

- Does your organization have a handle on where algorithms are deployed?
- Have you evaluated the potential impact if they function improperly?
- Does senior management understand the need to manage algorithmic risks?
- Is there a clearly established governance structure for overseeing the risks from algorithms?
- Is a program in place to manage these risks? If so, are you continuously enhancing the program as technologies and requirements evolve?

In another report, the consulting firm McKinsey & Company likens business users seeking to avoid harmful applications of algorithms to health-conscious consumers who must study literature on nutrition and read labels in order to avoid excess calories, harmful additives, or dangerous allergens.

According to McKinsey, there are three building blocks for controlling algorithmic risk in large organizations:

- Business-based standards for machine-learning approvals
- Professional validation of machine-learning algorithms
- A culture for continuous knowledge development

“Creating a conscious, standards-based system for developing machine-learning algorithms will involve leaders in many judgment-based decisions. For this reason, debiasing techniques should be deployed to maximize outcomes,” McKinsey says. One effective technique is conducting a “pre-mortem” exercise “to pinpoint the limitations of a proposed model and help executives judge the business risks involved in a new algorithm.”

For companies and investors, developing a risk-aware mind-set will require innovative approaches to the challenges raised by AI. Microsoft CEO Brad Smith and Harry Shum, Executive Vice President of Microsoft's Artificial Intelligence and Research Group, believe AI may give rise to new fields of law and to new ethical considerations in the field of computer science:

> *"IN COMPUTER SCIENCE, WILL CONCERNS ABOUT THE IMPACT OF AI MEAN THAT THE STUDY OF ETHICS WILL BECOME A REQUIREMENT FOR COMPUTER PROGRAMMERS AND RESEARCHERS? WE BELIEVE THAT'S A SAFE BET. COULD WE SEE A HIPPOCRATIC OATH FOR CODERS LIKE WE HAVE FOR DOCTORS? THAT COULD MAKE SENSE. WE'LL ALL NEED TO LEARN TOGETHER AND WITH A STRONG COMMITMENT TO BROAD SOCIETAL RESPONSIBILITY. ULTIMATELY THE QUESTION IS NOT ONLY WHAT COMPUTERS CAN DO. IT'S WHAT COMPUTERS SHOULD DO."*

That optimistic vision of life and business in the algorithmic 21st century must be developed around the core principles of fairness, accountability and transparency. Companies in every industry sector — especially their senior executives and boards of directors — will need singular focus and commitment to those principles if the enormous promise of artificial intelligence is to be realized.

KEY TERMS

An **algorithm** is a set of rules that can be used to process information, identify patterns, predict outcomes, and solve tasks. Computer scientists point out that a simple recipe or a list of directions to a friend's house can be an algorithm. For a computer, an algorithm is a set of well-defined instructions that allows the computer to solve a problem set.

Algorithms are capable not only of following instructions, but also of figuring out their own new steps to follow to reach a result or offer solutions. This process is called machine learning. Through **machine learning**, an algorithm can sort massive quantities of data and identify patterns that it uses to develop protocols for making sense of new information. Simply put, an algorithm can decipher from old patterns how to predict the future.

Machine learning algorithms make calculations by attempting to replicate patterns they recognize in existing data. Therefore, the data that engineers feed to machine learning algorithms to teach them to make decisions — the **training data** — is critical to the way that machine learning algorithms will make decisions moving forward, and thus to the outcome of those decisions.

Artificial intelligence — or AI — is a branch of computer science that attempts to build machines capable of intelligent behavior. The "intelligence" of these computers is that they can learn without input from humans. While artificial intelligence and machine learning are often used interchangeably, most data scientists think of machine learning as a subset of artificial intelligence.

One of the most critical challenges when employing algorithmic decision-making is to define the **criteria for success**. What is the intent of the algorithm? What questions are being asked? What principles are being applied? An algorithm will generate outcomes that best align with its understanding of success.

BIAS BAKED IN

Algorithmic tools are being deployed rapidly and across a broad range of industries, without regulatory or voluntary safeguards in place to prevent these tools from replicating systemic bias and discrimination.

“Discrimination and bias are not modern inventions,” says the Center for Democracy & Technology in its comprehensive 2016 report, Digital Decisions. “However, the scope and scale of automated technology makes the impact of a biased process faster, wider spread, and harder to catch or eliminate.”

This isn’t a problem for the future — machine-learning already dictates our world today, and algorithmic tools are impacting the global marketplace at a massive scale. According to International Data Corporation (IDC), worldwide revenues for “cognitive/AI solutions” will mushroom from nearly $8 billion in 2016 to $47 billion in 2020. Fastest growth is expected in banking, securities investment services, manufacturing, retail, and healthcare. “Cognitive/AI systems are quickly becoming a key part of IT infrastructure and all enterprises need to understand and plan for the adoption and use of these technologies in their organizations,” says IDC.

> *THE MYTH OF ALGORITHMIC OBJECTIVITY IS A CRITICAL BARRIER TO UNCOVERING AND ADDRESSING THE WAYS IN WHICH ALGORITHMIC DECISION-MAKING CAN BE UNFAIR. EXAMPLES OF ALGORITHMIC BIAS ABOUND.*

One problem, says mathematician, data scientist and former hedge fund quant Cathy O’Neil, who authored *Weapons of Math Destruction*, is the perception that algorithms are neutral because they rely on math — and people generally trust math. The myth of algorithmic objectivity is a critical barrier to uncovering and addressing the ways in which algorithmic decision-making can be unfair.

Examples of algorithmic bias abound.

Facebook’s algorithms, for example, have enabled dozens of brand-name companies — such as Amazon, Verizon, UPS — to exclude older workers from job ads, according to an investigative report from ProPublica and The New York Times. One employment expert called this practice “blatantly unlawful.” As recently as November 2017, Facebook’s “ethnic affinity” advertising algorithm allowed housing advertisers to exclude potential home buyers who were “African Americans, mothers of high school kids, people interested in wheelchair ramps, Jews, expats from Argentina and Spanish speakers,” ProPublica found.

In 2015, researchers at Carnegie Mellon found that Google’s online advertising system showed ads for higher-income executive jobs nearly six times more often to men than to women.

The inherent risks of AI bias are now readily conceded by tech industry leaders, who often highlight the need for humans to analyze and take responsibility for what they do with artificial intelligence. A 151-page report by Microsoft, published in January 2018, provides multiple examples of how algorithmic decision-making can go wrong. Among them:

> “An AI system could also be unfair if people do not understand the limitations of the system, especially if they assume technical systems are more accurate and precise than people, and therefore more authoritative. In many cases, the output of an AI system is actually a prediction. One example might be ‘there is a 70 percent likelihood that the applicant will default on the loan.’ The AI system may be highly accurate, meaning that if the bank extends credit every time to people with the 70 percent ‘risk of default,’ 70 percent of those people will, in fact, default. Such a system may be unfair in application, however, if loan officers incorrectly interpret ‘70 percent risk of default’ to simply mean ‘bad credit risk’ and decline to extend credit to everyone with that score — even though nearly a third of those applicants are predicted to be a good credit risk. It will be essential to train people to understand the meaning and implications of AI results to supplement their decision-making with sound human judgment.”

Algorithms’ flaws carry significant reputational risk for companies, which greater transparency can help address. Addressing the problem requires businesses to distinguish between the intention of an algorithmic tool and its real-world impact. According to the non-profit Data & Society Research Institute, “the designer of an algorithm may have no intentions of producing discriminatory results. For example, algorithmically inferring race with a high degree of accuracy without actually knowing race is relatively easy. Unless an analyst is testing to make sure that race is not a factor, the correlates that enable such discrimination to occur can often go unnoticed.”

Case in point: a 2012 Wall Street Journal investigation into pricing on the web site of Staples, the office supply company, found that users were offered different prices for the same product depending on their location. An algorithm’s decision was driven by competition: “ZIP Codes whose center was farther than 20 miles from a Staples competitor saw higher prices 67 percent of the time. By contrast, ZIP Codes within 20 miles of a rival saw the high price least often, only 12 percent of the time.” The unintended consequence of this formula — which was hidden from consumers — was that customers in poorer communities were more likely to be charged higher prices for the same product since Staples had fewer nearby competitors in these areas.

> *ONLY BY MEASURING AND EXAMINING THE OUTCOMES OF AN ALGORITHM'S DECISION — INCLUDING UNINTENDED CONSEQUENCES — CAN INVESTORS AND THE PUBLIC TRULY UNDERSTAND WHETHER ADVANTAGES GAINED BY BUSINESSES COME AT THE EXPENSE OF BROADER HARMS TO SOCIETY.*

Importantly, even when algorithms are intentionally programmed to counteract bias, discrimination can still occur, especially when cost-effectiveness is prioritized. In 2016, a study by marketing professors at the MIT Sloan School and the London School of Business found that social media advertising for technology and science jobs that was "explicitly intended to be gender-neutral in its delivery" resulted in far fewer women seeing the ad — even though women who saw the ad were more likely to click on it. How could this be? The researchers concluded that "women aged 18-35 are a prized demographic and as a consequence are more expensive to show ads to. This means that an ad algorithm which simply optimizes ad delivery to be cost-effective, can deliver ads which are intended to be gender-neutral in what appears to be a discriminatory way." In order to mitigate this discriminatory impact, the study advised advertisers "set different budgets for female and male advertising campaigns, but also further separate out bidding strategies by age as well as gender, to ensure that they do reach younger women."

U.S. law recognizes a theory of "disparate impact," which occurs when a seemingly neutral policy produces a disproportionate adverse effect or impact on a protected class of people. Some data scientists argue that "finding a solution to big data's disparate impact will require more than best efforts to stamp out prejudice and bias; it will require a wholesale reexamination of the meanings of 'discrimination' and 'fairness.'"

Most companies treat their algorithms as trade secrets, a corporate asset to be protected in a metaphorical "black box," far from public scrutiny. But many critics agree with Marc Rotenberg, executive director of the Electronic Privacy Information Center, who says the problem with "black boxes" is that "even the developers and operators do not fully understand how outputs are produced."

"We need to confront the reality that power and authority are moving from people to machines," Rotenberg says. "That is why #AlgorithmicTransparency is one of the great challenges of our era."

Algorithmic transparency is not about exposing trade secrets; it is about disclosing the impact of an algorithm's decision on consumers, and particularly the most marginalized groups. Only by measuring and examining the outcomes of an algorithm's decision — including unintended consequences — can investors and the public truly understand whether advantages gained by businesses come at the expense of broader harms to society.

Case Study: Predictive Hiring

The hiring process provides a good example of the purported benefits and inherent risks of algorithms. Companies are increasingly using automated hiring tools that include personality tests, skills tests, questionnaires and other exams to make judgements about candidates' employability. By one estimate, the HR software market is expected to top $10 billion by 2022.

Some of the biggest marketers of algorithmic tools in the human resources field include Oracle, whose Taleo software "enables companies to easily source, recruit, develop, and retain top talent with an engaging, social, and data-rich talent management software suite"; SAP, whose human resource software offers to help employers "find the right talent, develop future leaders, and engage all employees with automated, transparent processes, and a digital HR experience"; and IBM, which acquired Kenexa, an HR software developer, in 2012 for $1.3 billion.

Many companies purchase automated human resources tools to *reduce* bias in the hiring process. Without a doubt, implementing equitable hiring practices is a critical and urgent endeavor for economic equity.

But how effective are automated hiring tools at reducing bias?

It's true that HR software companies promote the supposed neutrality of their tools as an advantage over traditional hiring methods. For example, HireVue, one of the more popular big data-driven human resources software companies, writes, "Not only does AI cut costs and speed up onboarding, it doesn't care about an applicant's club membership or what their favorite sports team is." Goldman Sachs and Unilever have reportedly used technology from HireVue that analyzes the facial expressions and voice of job candidates to advise hiring managers.

Similarly, the banking giant Citigroup uses Koru, predictive hiring software that says it "identifies the drivers of performance in your company, increases high quality hires, and reduces bias." Koru claims to measure each job candidate for seven "impact skills," including "Grit, Rigor, Impact, Teamwork, Curiosity, Ownership, and Polish." According to Fortune, "the software uses algorithms that search for signs of grit in past behavior. It's less a matter of any individual sign than the accumulation of them. Maybe a candidate was on the volleyball team. But what really matters is how long the person persisted — while, say, holding down a full-time job — as well as the leadership role she attained and the solo projects she completed. The software can suggest follow-up interview questions that let employers dig deeper."

While predictive software companies claim their products help reduce bias, many data scientists say these algorithms can behave in dangerously biased ways.

According to the Center for Democracy & Technology, "if training data for an employment eligibility algorithm consists only of all past hires for a company, no matter how the target variable is defined, the algorithms may reproduce past prejudice, defeating efforts to diversify by race, gender, educational background, skills, or other characteristics."

Kelly Trindel, chief analyst for the Equal Employment Opportunity Commission, has testified:

> "...if the training phase for a big data algorithm happened to identify a greater pattern of absences for a group of people with disabilities, it might cluster the relevant people together in what's called a high absenteeism risk profile. This profile need not be tagged as disability; rather, it might appear to be based upon some common group of financial, consumer or social media behaviors. It may not be obvious to the employer, or even to the data scientist who created the algorithm, that subsequent employment decisions based on this model could discriminate against people with disabilities.
>
> Similarly, if most previously successful employees at a firm happen to be young, white or Asian American men, then the model will codify success in this way. If married women of a particular age are more likely to churn, then the model will codify this in proxies and subsequently predict lower success rates for similar women. All of this can happen without informing the model of disability status, age, race or gender, and all while giving the appearance that the machine is working just as it should, to increase worker ROI."

Speaking before the same EEOC hearing in 2016, Dr. Kathleen Lundquist, CEO of APTMetrics explained: “[A]lgorithms are trained to predict outcomes which are themselves the result of previous discrimination.” This not only has the potential to magnify bias, but also appears to transfer the responsibility of fairness from human to computer. When job applicants’ tests are “scored” by algorithms, company hiring managers and applicants’ alike lack an understanding of how the decision (or score) was determined.

Because hiring managers only encounter the qualified candidates the algorithms detect, there is no way of knowing how often the algorithm is making unfair rejections. Data scientist Cathy O’Neil notes: “If you think about it, if you filter people out from even getting an interview, you never see them again....There’s no way to see, to learn that you made a mistake on that — that person would have been a good employee — because they’re gone.”

“TARGET, TRACK AND PUNISH”

Artificial Intelligence will contribute as much as $15.7 trillion to the world economy by 2030, according to a June 2017 PwC report. Unfortunately, most of that won’t benefit people who need it most.

Technologist Anil Dash recently told Pew: “The best parts of algorithmic influence will make life better for many people, but the worst excesses will truly harm the most marginalized in unpredictable ways. We’ll need both industry reform within the technology companies creating these systems and far more savvy regulatory regimes to handle the complex challenges that arise.”

Data scientists generally agree that those most threatened by algorithms are those traditionally disenfranchised, including Black people, Native people, people of color, women, people with disabilities, older people, and people in the criminal justice system.

Algorithms can reproduce bias even in attempts to measure our individual physical or behavioral characteristics — our “biometrics.” For example, Apple has provided refunds to iPhone X customers who have repeatedly discovered that the phone’s face recognition software can have trouble accurately recognizing people of color, raising questions about how Apple tests its software and how — or based on whom — it determines “effectiveness.”

A New York Times article, “Facial Recognition Is Accurate, if You’re a White Guy”, reviews a study by M.I.T. researcher Joy Buolamwini measuring the performance of face recognition systems created by Microsoft, IBM and Chinese company Megvii. Buolamwini found that the algorithms delivered 99% accuracy rates when scanning the faces of white men, with disparate results for people of color: “the darker the skin, the more errors arise — up to nearly 35 percent for images of darker skinned women.”

Yet, companies are using face recognition algorithms to verify the identity of consumers in online payment systems and banking; for photo identification and marketing on social media; in lending processes; and even to assess behavior of job candidates. Some companies are experimenting with the possibility to personalize an advertisement based on the perceived “mood” of the face in front of it.

As government agencies deploy new forms of technology under the guise of national security, airlines and other companies face new risks. Over four years, U.S. Customs and Border Protection (CBP) seeks to carry out a $1 billion "biometric exit" face scanning program; already, face scanning technology is in place at 11 airports across the nation, for passengers traveling on some flights with American Airlines, British Airways, Delta, Emirates, JetBlue and United Airlines.

> *SOME NEW AI PRODUCTS SEEM BORN FROM SCI-FI FILM SCENARIOS…ONE COMPANY ADVERTISES "FACIAL PERSONALITY ANALYTICS" THAT ALLOW "SECURITY COMPANIES AND AGENCIES TO BE MORE EFFECTIVE IN DETECTING ANONYMOUS PERSONS OF INTEREST."*

In recent testimony to a Congressional subcommittee, Laura Moy, Deputy Director of the Center on Privacy & Technology at Georgetown Law, noted that "hiring algorithms have been accused of unfairly discriminating against people with mental illness. Sentencing algorithms — intended to make sentencing fairer by diminishing the role of potentially biased human judges — may actually discriminate against Black people. Search algorithms may be more likely to surface advertisements for arrest records — regardless of whether such records exist — when presented with characteristically Black names."

In a powerful new book, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, Virginia Eubanks investigates the impact of algorithms and predictive risk models on poor and working class Americans, presenting three case studies: an attempt to automate eligibility processes for the State of Indiana's welfare system; an electronic registry of the unhoused in Los Angeles; and a risk model that promised to predict which children will be future victims of abuse or neglect in Allegheny County, Pennsylvania.

When algorithms were employed in Indiana's public assistance program, Eubanks reports, "The assumption that automated decision-making tools were infallible meant that computerized decisions trumped procedures intended to provide applicants with procedural fairness. The result was a million benefit denials."

Eubanks argues that algorithmic decision-making creates a "digital poorhouse" which "uses integrated databases to target, track, and punish." She says engineers and data scientists concerned about the economic and social implications of their designs should ask themselves two questions: "Does the tool increase the self-determination and agency of the poor? Would the tool be tolerated if it was targeted at non-poor people?"

Data researcher Mimi Onuoha warns about the danger of what she calls "algorithmic violence" brought about by "the availability of huge datasets, advances in computational power, leaps in fields like artificial intelligence and machine learning, and the subsequent incorporation and leveraging of all these things into a hierarchical and unequal society."

Some new AI products seem born from sci-fi film scenarios. Faception, an Israeli company, advertises "facial personality analytics" employing computer-vision and machine learning which "automatically reveals personalities in real-time." The technology purportedly allows "security companies and agencies to be more effective in detecting anonymous persons of interest."

> *INDEED, ONE OF THE "HOTTEST" CURRENT GROWTH AREAS FOR ALGORITHMS IN THE U.S. IS IN THE DELIVERY OF GOVERNMENT SERVICES, INCLUDING ADMINISTRATION OF THE CRIMINAL JUSTICE SYSTEM, AS CITIES AND STATES DEPLOY NEW DATA-DRIVEN TOOLS IN "PREDICTIVE POLICING," BAIL DETERMINATION, CRIMINAL SENTENCING, PROBATION, PAROLE, AND JUVENILE DETENTION.*

For example, predictive policing software — which analyzes data on past *arrests* to gain insight into where and when future crime is most likely to occur — is being increasingly deployed by major police departments in the U.S., including New York City, Los Angeles, Atlanta and Chicago. Leading marketers of predictive policing software include IBM, Hitachi, Motorola (a division of Lenovo), LexisNexis (a division of the RELX Group) and the privately-owned Palantir.

While researchers such as Jennifer Bachner, a government expert at Johns Hopkins University, contend that "predictive policing offers an opportunity to make significant advances toward a safer and more just society," critics such as Jennifer Lynch, an attorney with the Electronic Frontier Foundation, say predictive policing will "only further focus police surveillance on communities that already are overpoliced and could threaten our constitutional protections and fundamental human rights."

The RAND Corporation studied the Chicago Police Department's use of predictive software to decrease that city's notoriously high murder rate. In 2016, RAND concluded: "To make a long story short: It didn't work."

A major focus is pre-trial risk assessment: developing tools to help judges decide where defendants will await trial — at home or in jail. By law, this decision hinges on the judge's prediction of what the defendant would do if released. In many cities and states, new "risk assessment algorithms" use data about a defendant's age, location, family history, and a host of other factors to make a recommendation to a judge regarding the defendant's "likelihood" of re-arrest or failure to appear at the next court date.

Civil rights and racial justice groups have questioned the effectiveness of risk assessment tools because they have been found to replicate racism. In 2016, ProPublica found that a software used across the country to assess risk — COMPAS — was biased against Black people; in Florida, the researchers found that the algorithm was more likely to falsely flag Black defendants as future criminals, "wrongly labeling them this way at almost twice the rate as white defendants" while white defendants were also more likely to be mislabeled as low risk. (A Dartmouth College computer science study recently found that COMPAS is no better at predicting an individual's risk of recidivism than choices made by random volunteers recruited from the internet.)

And a January 2018 paper by data science researchers John Logan Koepke and David Robinson concludes that today's risk assessment tools make "zombie predictions," with predictive models that are trained on data from older bail regimes, and "are blind to the risk-reducing benefits of recent bail reforms." Koepke and Robinson argue that these new tools "risk giving an imprimatur of scientific objectivity to ill-defined concepts of 'dangerousness'" and "pave the way for a possible increase in preventive detention."

Some experts believe that such pre-trial risk assessment tools could offer great benefit, if the algorithm focuses on "predicting judges' decisions, rather than defendant behavior." A recent paper by a team led by Cornell University's Jon Kleinberg created a simulation with an algorithm, trained through machine learning, "to improve and understand" judges' decision-making about defendants — "a concrete prediction task for which there is a large volume of data available". The study found that, by integrating algorithms into an economic framework (in this case, using different econometric strategies like quasi-random assignment of cases to judges) and "being clear about the link between predictions and decisions", pre-trial risk assessment tools could result in "reductions in all categories of crime, including violent ones. Importantly, such gains can be had while also significantly reducing the percentage of African-Americans and Hispanics in jail." Importantly, for the algorithm to effectively reduce crime and incarceration rates, it focused on human decision-making by judges, rather than attempting to predict defendants' future behavior.

As these algorithmic tools are increasingly deployed in the criminal justice system, "we should have the right not just to see and understand what the risk assessment tools say, but to independently audit the results that come from their introduction in the system where they are used," says Hannah Sassaman, a policy director at the Philadelphia-based Media Mobilizing Project and a current Soros Justice Fellow focusing on community oversight of risk assessment algorithms. Sassaman adds: "This could mean including independent data scientists, focused on the rights of communities, on robust community advisory boards governing child welfare, criminal justice, or other contexts using algorithmic risk assessment. It could also mean halting the use of these algorithms if they are not producing the results we want."

A 2017 report by the AI Now Institute at New York University, concluded that "core" public agencies responsible for areas such as criminal justice, healthcare, welfare, and education should no longer use "black box" AI and algorithmic systems, including both AI systems licensed from third party vendors and algorithmic processes created in-house. "The use of such systems by public agencies raises serious due process concerns, and at a minimum they should be available for public auditing, testing, and review, and subject to accountability standards," the report said.

Given widespread concern and growing evidence about the discriminatory nature of AI tools, the major companies producing these products and selling them to government agencies should be held to demonstrate the fairness of these tools before — not after — they are deployed. **Investors can be key players in engaging companies to audit algorithmic impact at the outset.**

REGULATING ALGORITHMS

Calls to regulate algorithmic tools and artificial intelligence are not uncommon. But that doesn't mean wide-scale regulation is imminent - or, if regulations are adopted, that they will be easy to enforce.

New York City in December 2017 enacted a law requiring a task force to examine the city's "automated decision systems" now used by multiple agencies to target services from high school placements to firehouses to criminal justice decisions. Passed unanimously by the City Council and signed by Mayor Bill de Blasio — who has pledged to make New York "the fairest big city in America" — the measure designed to tackle algorithmic discrimination is the first of its kind in the United States. The task force will study how city agencies use algorithms to make decisions that affect New Yorkers' lives, and whether any of the systems appear to discriminate against people based on age, race, religion, gender, sexual orientation or citizenship status. To help accomplish that, advocates have urged that the fledgling effort consider a framework structured around Algorithmic Impact Assessments (AIAs).

However, the New York City law doesn't extend to businesses or other organizations that use algorithms. And it's not clear how many other U.S. cities and states will follow New York's example.

In fact, the U.S. does not have an overarching federal privacy law that would regulate algorithms across the board. Rather, personal information in the United States is regulated by sector, with protections under a number of federal laws, including the Fair Credit Reporting Act, the Health Insurance Portability and Accountability Act or other federal equal opportunity laws. Some states have their own privacy laws.

Much of the pressure to regulate algorithms comes from Europe, where a new European Union regulation — the General Data Protection Regulation (GDPR) — takes effect May 25, 2018. The GDPR requires government or private entities to obtain an individual's explicit, well-informed, and affirmative consent before using that individual's personal information.

The GDPR also provides a basic "right to explanation" to individuals affected by "automated decision-making that has legal or similarly significant effects on them." Companies are required to "give individuals information about the processing; introduce simple ways for them to request human intervention or challenge a decision; carry out regular checks to make sure that your systems are working as intended."

Importantly, the GDPR applies if an algorithm collects or processes personal information of any EU citizen, regardless of where the company deploying the algorithm is located. Global companies such as Facebook, Google and Amazon have already implemented major changes to their privacy policies. Julie Brill, a corporate vice president and deputy general counsel at Microsoft, told the *New York Times*, "We embrace G.D.P.R. because it sets a strong standard for privacy and data protection rights, which is at the core of our business."

DEFINING FAIRNESS AND ACCEPTING RESPONSIBILITY

“Will the net overall effect of algorithms be positive for individuals and society or negative for individuals and society?”

That’s the question put recently to 1,300 technology experts, scholars, corporate practitioners and government leaders in a survey by Pew Research Center and Elon University’s Imagining the Internet Center. The results? Experts are evenly divided: 38 percent of respondents predicted that the positive impacts of algorithms will outweigh negatives for individuals and society in general, while 37 percent said negatives will outweigh positives; 25 percent said the overall impact of algorithms will be evenly balanced.

Meanwhile, there’s significant concern among corporate leaders:

- A 2017 survey of more than 1,000 large companies by Accenture found that fewer than a third have a high degree of confidence in the fairness and auditability of their AI systems, and fewer than half have similar confidence in the safety of those systems. “Clearly, those statistics indicate fundamental issues that need to be resolved for the continued usage of AI technologies,” the researchers concluded.
- Seventy-seven percent of CEOs in a 2018 PwC survey said AI and automation will increase vulnerability and disruption to the way they do business. Other concerns cited in the CEO survey were potential for biases and lack of transparency (76 percent), ensuring governance and rules to control AI (73 percent), risk to stakeholders’ trust and moral dilemmas (73 percent), potential to disrupt society (67 percent), and lack of adequate regulation (64 percent).

Legal, regulatory and financial risks for companies loom large. In addition to the problems at tech platforms like Facebook and Google, such major online companies as Airbnb, Uber and Lyft have faced legal and reputational risk for allowing discrimination to occur on their platforms. Algorithms enable or exacerbate the problem.

Civil rights groups, data scientists, and civic technology organizations have begun to organize efforts to protect civil rights in the era of big data around three core principles: fairness, accountability, and transparency. They advocate for responsibility, explainability, accuracy, auditability, and fairness. But the conversation about how to achieve all that is only beginning to emerge. A technical paper on “Accountable Algorithms,” by a group of academics from Princeton and Fordham, and the non-profit Upturn, argues that transparency — and scrutiny of “black box” algorithms — isn’t sufficient to establish safeguards for accountability in automated decisions. Ensuring accountability will require sustained and close collaboration among computer scientists, policymakers, and lawmakers.

CONSULTANTS AT ACCENTURE PREDICT THAT ONE OF THE NEWEST JOBS IN THE FIELD OF ARTIFICIAL INTELLIGENCE WILL BE AN "ETHICS COMPLIANCE MANAGER" ACTING AS WATCHDOG AND OMBUDSPERSON FOR "UPHOLDING NORMS OF HUMAN VALUES AND MORALS."

One such collaborative effort is the Partnership on AI, launched in 2016 by founding corporate partners Amazon, DeepMind, Facebook, Google, IBM, and Microsoft. The organization now has more than 50 members, including NGO partners such as ACLU, AI Now Institute, Amnesty International, Center for Democracy & Technology and the Data & Society Research Institute. The partnership says it "works to study and formulate best practices on artificial intelligence technologies, advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society."

A May 2016 White House report on big data and algorithmic systems suggested that to "avoid exacerbating biases by encoding them into technological systems, we need to develop a principle of 'equal opportunity by design' — designing data systems that promote fairness and safeguard against discrimination from the first step of the engineering process and continuing throughout their lifespan."

Consultants at Accenture predict that one of the newest jobs in the field of artificial intelligence will be an "ethics compliance manager" acting as watchdog and ombudsperson for "upholding norms of human values and morals — intervening if, for example, an AI system for credit approval was discriminating against people in certain professions or specific geographic areas." Accenture suggests: "The ethics compliance manager could work with an algorithm forensics analyst to uncover the underlying reasons for such results and then implement the appropriate fixes."

But all this will take corporate willingness to acknowledge the risks and a determination to tackle algorithmic accountability. Skeptics abound. Frank Pasquale, a University of Maryland law professor and leading expert in data science, has said that perhaps most difficult part of addressing the issue of algorithmic transparency is getting tech companies to "accept some kind of ethical and social responsibility for the discriminatory impacts of what they're doing."

The fundamental barrier to algorithmic accountability, according to Pasquale, is convincing companies "to invest serious money" and empower staff to ensure both legal compliance and broader ethical compliance. Without that, "we're not really going to get anywhere."

RESOURCES

A sampling of organizations that explore the business and societal implications of algorithms and artificial intelligence:

Ethics and Governance of Artificial Intelligence
This collaboration between The Berkman Klein Center for Internet & Society at Harvard University and the MIT Media Lab conducts evidence-based research to guide decision-makers in the private and public sectors and undertakes pilots "to bolster the use of AI for the public good."

The AI Now Institute at New York University
AI Now is an interdisciplinary research center that seeks to understand the social implications of artificial intelligence in four areas: Rights & Liberties; Labor & Automation; Bias & Inclusion; Safety & Critical Infrastructure.

Data & Society
A research institute focused on social and cultural issues arising from data-centric technological development.

Center for Democracy and Technology
CDT's Privacy and Data Project explores the changing role of technology in daily life and its impact on individuals, communities, and public policy.

The Partnership on AI
Launched in 2016, this 50-member alliance of corporate and NGO heavyweights "works to study and formulate best practices on artificial intelligence technologies, advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society."

Algorithmic Justice League
Led by Joy Buolamwini, the League is a collective of activists, coders, artists, academics, companies, citizens, legislators and regulators who aim to highlight algorithmic bias through media, art and science; provide space for people to voice concerns and experiences with coded bias; and develop practices for accountability during the design, development, and deployment of coded systems.

The Center for the Fourth Industrial Revolution
Organized by The World Economic Forum, the Center serves as “a hub for global, multi-stakeholder cooperation to develop policy frameworks and advance collaborations that accelerate the benefits of science and technology.”

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
IEEE is the world’s largest technical professional society. This initiative is designed to ensure that those designing and developing AI systems are “educated, trained, and empowered to prioritize ethical considerations … for the benefit of humanity.”

Fairness, Accountability, and Transparency in Machine Learning
An annual event that brings together researchers and practitioners “concerned with fairness, accountability, and transparency in machine learning.”

The Future of Life Institute
Organizational mission: “Most benefits of civilization stem from intelligence, so how can we enhance these benefits with artificial intelligence without being replaced on the job market and perhaps altogether?”

DeepMind Ethics & Society
DeepMind is a for-profit company specializing in AI that Google acquired in 2014 and is now part of the Alphabet group. It espouses the belief that “all AI applications should remain under meaningful human control, and be used for socially beneficial purposes” The Ethics & Society Unit seeks to help technologists put ethics into practice and help society anticipate and direct the impact of AI.

ITI Artificial Intelligence Policy Principles
The Information Technology Industry Council (ITI) represents large tech companies and calls itself “the global voice of the tech sector.”

Electronic Privacy Information Research Center
A public interest research center in Washington, DC.

Data Science for Social Good
The Data Science for Social Good Fellowship is a University of Chicago summer program to train aspiring data scientists to work on data mining, machine learning, big data, and data science projects with social impact.