

## Perspective

## High Tech, High Risk: Tech Ethics Lessons for the COVID-19 Pandemic Response

Emanuel Moss<sup>1,2,\*</sup> and Jacob Metcalf<sup>1,\*</sup><sup>1</sup>Data & Society Research Institute, New York, NY 10010, USA<sup>2</sup>CUNY Graduate Center, New York, NY 10016, USA\*Correspondence: [emoss@gradcenter.cuny.edu](mailto:emoss@gradcenter.cuny.edu) (E.M.), [jake.metcalf@datasociety.net](mailto:jake.metcalf@datasociety.net) (J.M.)<https://doi.org/10.1016/j.patter.2020.100102>

**THE BIGGER PICTURE** This paper describes the coupling of machine learning and the social production of risk in general, with specific illustrations drawn from machine learning applications in response to the COVID-19 pandemic. As the COVID-19 pandemic has drastically reshaped society around the world, many have looked to machine learning as a technology capable of addressing large problems at scale, and machine learning applications have been seen as desirable interventions in mitigating the risks of the pandemic disease. However, machine learning, like many tools of technocratic governance, is deeply implicated in the social production and distribution of risk. Therefore, the role of machine learning in the production of risk must be considered as engineers and other technologists develop tools for the current crisis. The paper concludes by describing the role of risk management in the effort to institutionalize ethics in the technology industry, and how such efforts can benefit from understanding the social production of risk through machine learning.



**Concept:** Basic principles of a new data science output observed and reported

## SUMMARY

The COVID-19 pandemic has, in a matter of a few short months, drastically reshaped society around the world. Because of the growing perception of machine learning as a technology capable of addressing large problems at scale, machine learning applications have been seen as desirable interventions in mitigating the risks of the pandemic disease. However, machine learning, like many tools of technocratic governance, is deeply implicated in the social production and distribution of risk and the role of machine learning in the production of risk must be considered as engineers and other technologists develop tools for the current crisis. This paper describes the coupling of machine learning and the social production of risk, generally, and in pandemic responses specifically. It goes on to describe the role of risk management in the effort to institutionalize ethics in the technology industry and how such efforts can benefit from a deeper understanding of the social production of risk through machine learning.

## MACHINE RISK SOCIETY

Ulrich Beck begins his 1992 book *Risk Society* by saying that “the social production of wealth is accompanied by the social production of risks.”<sup>1</sup> He describes how economic and social inequality overlaps with vulnerability to the risks produced by the scientific and technological practices that underlie the wealth of the modern world. The COVID-19 pandemic illustrates Beck’s thesis in excruciating detail, as do the applications of machine learning that attempt to address the many concerns of the pandemic. Indeed, machine learning has become a key technological practice that produces, and redistributes, risk across society.

The risk of pandemic disease is socially produced—the SARS-CoV-2 virus traveled on jet planes and cruise ships through

global networks of trade and tourism,<sup>2</sup> but the shareholders of transportation companies have been insulated from the risk to their business through enormous government cash bailouts. Populations that may be most vulnerable to disease have the fewest medical resources,<sup>3</sup> and countries with capacity to manufacture any coronavirus vaccine are less likely to be able to afford it for their own citizens.<sup>4</sup> In the US, many of those most likely to come into contact with others during lockdown—the essential grocery clerks, home health workers, warehouse workers, and delivery workers—are the least likely to have sick leave and work-from-home policies.<sup>5</sup> Meanwhile, those who can afford to use delivery services while they work from home are able to enjoy greatly reduced risks of contracting COVID-19, and are more likely to receive timely and adequate care should they fall sick. Wealth and risk are distributed unevenly and unfairly by social



and technological structures that seem to operate autonomously. These structures are pervasive and deeply intertwined with technologies of bureaucratic control in vivid (as automated decision systems for welfare services<sup>6</sup> and predictive policing<sup>7</sup>) and banal forms (as accounting norms, quarterly shareholder dividends, and unscalable unemployment insurance portals running on COBOL).<sup>8</sup>

The risks that define contemporary life are not just socially produced, but are actively managed by technocrats and technologies that calculate risk thresholds for pollutants and automotive safety, by a financial system that creates a market for risk so that it can be hedged against, and by business managers who offer customers the ability to outsource infection risk by arbitraging weak and outdated labor laws. Caitlin Zaloom, an anthropologist who studies risk, identifies two modes of understanding risk from a sociocultural perspective—how groups understand risk as a threat or vulnerability, and how groups attempt to exert control over an uncertain future by limiting exposure to risk.<sup>9</sup> We apply those understandings of risk to applications of machine learning in the COVID-19 pandemic to ask how those who build machine learning applications construct and act upon risk. This necessarily leads us to also ask how they maintain and build socioeconomic power by exerting control over the distribution of risk. This framing extends ongoing conversations about the social implications of machine learning beyond those of algorithmic bias<sup>10,11</sup> to point to a type of harm that occurs not when machine learning “gets it wrong” due to various forms of bias. Rather, we interrogate what happens when machine learning “gets it right” by acting as expected but doing so while reinforcing and exacerbating social and economic inequity.<sup>12</sup> Despite recent debates within the machine learning community showing how resistant some quarters are to acknowledging the social, cultural, and economic dimensions of the field (against the ample evidence from others),<sup>13</sup> the question of how to properly bound which aspects of sociotechnical systems machine learning practitioners have agency upon and responsibility for has become vitally important.

Pandemics are not strictly a microbial phenomenon, they follow the contours of society the contagion encounters,<sup>14</sup> some of which shift medical risk toward the impoverished or oppressed. As each of us are made responsible to minimize our own risk of spreading the coronavirus, it becomes all too clear that “responsibility” is, to some extent, coterminous with a social and economic status that allows us to externalize risk onto others. In recent years, this capacity for nearly autonomous, unquestioned risk externalization has been mediated through machine learning applications to a remarkable extent. Machine learning, because of its versatility in dealing with problems across domains, because of the wealth of financial and computing resources at its practitioners’ disposal, and because teams of machine learning engineers can collaborate remotely, has found no shortage of potential applications to the current crisis. Therefore, we suddenly find our personal and societal responses to the pandemic emerging through machine learning applications that act as a ubiquitous tool for distributing societal, health, and economic risks.

Machine learning applications already shape the distribution of risk across the labor, health, and surveillance landscapes that are central to distributing risk as part of our pandemic

response. Gig worker platforms, instant delivery, supply chain management, and automated scheduling applications determine who is exposed to how much risk from the disease, while the workers themselves attempt to navigate continually shifting affordances on those platforms.<sup>15</sup> The machine logics of these systems are determinative of risk, whether human technocrats could have orchestrated these processes in identical ways or not, and early evidence suggests some of the ways that such applications are proving brittle in their response to the COVID-19 pandemic.<sup>16</sup> These applications financialize and arbitrage risk from app users (whether individual consumers or enterprise users) to gig workers, waged staff, and warehouse laborers, with stockholders reaping the difference.

The tech companies that are driving the technological response to the COVID-19 pandemic have arranged their business models and organizational practices around building products that distribute upsides and downsides according to socio-historical patterns, as interpreted by machines, yet lack the capacity to “understand just how pervasively... technology is being used to marginalize many groups of people,” as leading machine learning researcher Timnit Gebru and colleagues have observed.<sup>13,17</sup> Machine learning applications are ultimately exercises in distributing attention and resources, both in how we understand the risks we face and how those risks are distributed across society.

## RISK AND MACHINE LEARNING

Increasingly, the calculation, production, and management of risk has been accomplished through the application of machine learning techniques. Machine learning is used to understand threats and vulnerabilities, and also as a means of exerting control over such threats. Indeed, risk is foundational to machine learning. Loss functions, central to the applied power of machine learning, were developed to analyze and minimize risk. All objective functions can be thought of as minimizing the risk of a prediction being wrong, but machine learning has been applied to more “human readable” understandings of risk across a vast array of domains. These include predicting cardiovascular risk,<sup>18</sup> estimation of genetic risk factors,<sup>19</sup> consumer credit risk,<sup>20</sup> and risk of individuals attempting suicide.<sup>21</sup> Machine learning is used not only to understand the nature of threats to health, safety, and finance but also to intervene in these threats by allocating scarce resources toward interventions that minimize risk the most for those who control or own the algorithmic tools. In this way, a rideshare platform produces risk for drivers who take on the liability of car payments and vehicle maintenance, while minimizing risk for the platform itself through its freedom from maintaining a fleet of vehicles for the service it ostensibly offers.<sup>22</sup>

Machine learning, then, is deeply implicated in the social production of risk that Beck and Zaloom describe above. By seeing machine learning as productive of risk, it becomes possible to recognize the responsibilities machine learning practitioners hold for the ways they produce risk. In constructing representations from data, in crafting classifiers and evaluating their utility, and in optimizing for desired performance behaviors, machine learning shapes and distributes risk across society.<sup>23</sup> Machine learning can identify for bankers which individuals are at the

greatest risk of not repaying loans, for police which neighborhoods are at the greatest risk of certain kinds of criminal activity, and for doctors (and insurers) patients who are at the greatest risk of diabetes, heart disease, cancer, pneumonia, or COVID-19. And in all too many cases, machine learning actively *produces* risk while also distributing it unevenly across society, as predictive policing algorithms focus police attention on already over-policed (largely Black and Latinx) neighborhoods.<sup>24</sup>

Algorithmic methods for deciding where risk lies would be a purely academic exercise if not for the real-world impacts to people's lives that this entails. Machine learning applications operate upon theoretical constructs that are not directly observable—creditworthiness, health, and recidivism cannot be directly measured—but rather are inferred from proxies for such constructs.<sup>25</sup> Models built from such proxies, because they are operationalized as part of algorithmic decision-making systems, often seem like a concrete instantiation of the theoretical constructs they purport to represent but may instead become like self-fulfilling prophecies, actually producing the phenomena they purport to measure. And they are often used in ways that produce concrete negative impacts to people's lives.

Individuals classified as risky loan applicants have trouble buying homes and building wealth. Neighborhoods classified as high-risk crime areas get policed more heavily and see more arrests for petty crimes.<sup>26</sup> Over time, heavily policed neighborhoods see more re-arrests, and therefore steeper penalties for residents who become re-offenders.<sup>27,28</sup> Patients classified as being at high risk of certain diseases may receive life-saving early testing and have better long-term outcomes, but may also see higher healthcare costs and dangerous side effects from maintenance medications,<sup>29</sup> be seen as having pre-existing conditions by their insurers, or be triaged to a lower degree of medical urgency by a racially biased algorithm trained to seek price efficiency.<sup>30</sup>

To be clear, these impacts are not a first-order result of machine learning but neither are they entirely external to machine learning. Rather, they are the result of complex sociotechnical responses to risk in specific domains. The role of machine learning in producing “risk” as an actionable construct for some and not for others cannot be ignored, as machine learning practices rely on, recreate, and often amplify already-existing patterns of how risk is distributed across society, irrespective of whether that risk is fairly or justly distributed. This is particularly apparent in how racial disparities are algorithmically encoded in many machine learning applications, as when over-representation of non-white inmates in criminal justice records leads to the over-estimation of risk for non-white defendants in pretrial detention models.<sup>31</sup> However, it is also evident in how such applications themselves participate in the perpetuation of unjust institutions, such as the carceral system.<sup>7</sup> On top of this, the role that machine learning plays in producing crime, for some and not for others,<sup>32</sup> cannot be ignored.

Machine learning practitioners have invested significant effort in adjusting technical systems to blunt downstream risks, but the role machine learning might play in the reproduction of patterns of risk is also operative in less-apparent ways as well. Even when datasets are balanced,<sup>33</sup> optimization functions are constrained to minimize bias for disadvantaged groups,<sup>34</sup> and instances of

algorithmic bias are measured and mitigated,<sup>35</sup> a long tail<sup>36</sup> of social effects remain embossed on the data creation and collection processes that underlie the machine learning economy.

## MACHINE LEARNING AND COVID-19

Given the profound threat of COVID-19, it is crucial to consider how the application of machine learning to the social challenges of a global pandemic can produce and distribute risk across society. These risks are socially constructed, as are specific harms produced by machine learning systems, and so we must ask what is being done to make sure that one is not amplifying the other.

Since the pandemic was declared in early March of 2020,<sup>37</sup> hundreds of articles have been published to pre-print archives, such as aRxiv, bioRxiv, and medRxiv reporting potential advances in machine learning applications for combatting the pandemic. These applications include natural language processing for combing through the existing literature on COVID-19,<sup>38</sup> machine learning models that attempt to infer the asymptomatic spread of the virus,<sup>39</sup> models of the effect of quarantine policies on viral transmission,<sup>40</sup> facial recognition applications for use in emergency room triage,<sup>41</sup> and deep learning for COVID-19 diagnoses through medical imaging.<sup>42</sup> These studies point to the ways machine learning participates in how risk is both understood and managed, through attempts to understand the nature of the threat and to respond to it.

More recently, a host of machine learning applications have been developed to track, treat, and limit the spread of the virus.<sup>43,44</sup> These applications include automated contact tracing used to notify those who have been exposed that they ought to self-quarantine (for which machine learning algorithms assist in estimating the strength of contacts between people based on Bluetooth signal strength),<sup>45,46</sup> but also natural language processing-based early warning systems (available on a subscription basis) for outbreaks,<sup>47</sup> and computer vision systems that detect mask-wearing and crowding on public transportation systems to inform potential riders that they may want to choose alternate modes of transportation during busy times.<sup>48</sup>

While these efforts are doubtless well intentioned, each suggests a machine learning intervention into the already-existing distribution of risks and potential harms for society.<sup>49</sup> In the US we are seeing the most severe cases of COVID-19 striking Black and Latinx communities in drastically disproportionate ways,<sup>50</sup> and it is clear that this is in large part because of how risk and inequality map onto each other along racial dimensions.<sup>51</sup> This is in part because of long-standing disparities in health outcomes for Black and Latinx communities, environmental racism that places environmental determinants of health near Black and Latinx communities,<sup>52</sup> as well as the disproportionate number of Black and Latinx workers in job roles that cannot be filled at home and do not have adequate sick leave or healthcare.<sup>53</sup> For contact tracing, the ability to make use of a notification to minimize one's own risk by self-quarantining is far too dependent on one's personal wealth and capacity to afford to stay home (either because of a generous workplace sick leave policy, the ability to work from home, or one's own savings). Any contract tracing is a *sociotechnical* system that depends on how different parts of social life fit together—telling people they *should* stay

home does not mean that they *will be able to* stay home. For contact tracing to work at all, its designers must be attuned to the context of social life in which such systems can produce harmful, difficult-to-foresee effects<sup>54</sup> that replicate or amplify inequalities already present in society. Rather than individualizing the risk through contact tracing systems currently being proposed, attending to the contextual use of such a system could collectivize risk by identifying and emphasizing the necessary forms of social support for self-quarantine and medical care: adequate sick leave and quarantine leave policies, robust testing, and economic relief that targets individual workers over large companies.

Automated contact tracing notifications are useful to those who already possess the means to work from home and are less useful to those who cannot self-quarantine without losing their job. An automated update about a crowded subway car will not help someone who has no other means of transportation and cannot be late to return home to care for a child. Others, meanwhile, will maintain their freedom of movement—the delta between who can move about freely and who cannot will have profound consequences for those who have and those who do not have compatible mobile devices, or who lack the resources to maintain their income or care for their dependents while unable to work or cohabitate with others.<sup>55</sup> In this way, automated technologies that construct and manage the shared risks of the pandemic can perniciously codify and reinforce unjust social and economic dynamics that are the context in which the infectious disease spreads. Machine learning applications may construct risk such that individuals can act upon it to their own advantage without addressing the social conditions that make such risks unevenly distributed, thereby presenting a false sense of risk reduction. Those who build and deploy automated tools to track and treat the pandemic should not treat risk as if it were flat across the population, but as something that they are actively engaged in constructing and distributing and are responsible for doing so justly.

## MANAGING RISK IN THE TECH INDUSTRY

Understandings of how the production and distribution of risk through algorithmic technologies leads to sociotechnical impacts of machine learning are still in their infancy, and such technologies are only just beginning to be thought of as capable of being brought under any sort of governance regime.<sup>56–59</sup> Professional organizations have attempted to mitigate harmful impacts from applying data-driven and machine learning solutions to the problems of the COVID-19 pandemic. Some of these are quite robust, if non-binding, sets of recommendations,<sup>60</sup> while others attend to a limited definition of privacy rights without attending to the range of sociotechnical impacts discussed above.<sup>61</sup>

Over the past 2 years, we have been studying how those inside of Silicon Valley tech companies, which build the machine learning models that are most likely to directly affect people, go about understanding the impacts of machine learning and developing organizational practices to manage the effect they have on how risk is distributed across society.<sup>62,63</sup> Under the mantle of “ethics,” Silicon Valley companies manage the emerging risks their products and services pose for individuals, society, and for their own firms. In response to the unprece-

dent public health challenges represented by the COVID-19 pandemic, many of these companies are rushing to play a role in producing technological solutions. In the rush to produce solutions, however, it is even more important to think through the lessons the tech industry has learned from managing organizational risk in the years leading up to the current crisis, and not to jettison those lessons out expediency. Reading the potential applications of technology to the current crisis through the recent history of tech ethics, several lessons stand out: context is key, the upside benefits of technology are in tension with downside risks, and leadership and organizational culture matters. These lessons are made explicit through the applications of machine learning to the COVID-19 pandemic, but can also be extended to examining how machine learning, and digital technologies more broadly, produce and distribute risk across society.

### Context Is Key

Any framework to identify and manage risk within an organization must deliberately and methodically consider the context in which it operates. Context includes not just the social milieu at which a product is targeted, but also the intentions, worldview, and the necessarily partial knowledge of those who build the product.<sup>64</sup> Including producers of technology in any consideration of context is important because the way they are positioned in the world can have an outsized influence on what are chosen as problems to be solved,<sup>65</sup> what data are selected to serve as proxies for unobservable phenomena,<sup>25,66</sup> and what forms of risk are visible. This is particularly true for the use of technology to manage the risks COVID-19 presents to society.

Technological applications for the pandemic offer a compelling set of technical challenges for engineers and designers, but not all technical challenges present opportunities to improve outcomes. A deep learning tool for hypoxia detection intended for use in emergency room triage sounds useful, but given the speed at which health professionals can visually detect hypoxia (by noting pallor of complexion or blue lips),<sup>41</sup> such a technological intervention is an *additional* step on top of what needs to happen in triage already, not necessarily a *time-saving* tool for busy hospital workers. Recent research demonstrates the importance of recognizing the additional work integrating new tools into the existing practices of workplaces they are intended for requires from all those who interact with them.<sup>67</sup>

Given the enthusiasm for using COVID to accelerate the adoption of artificial intelligence (AI) in healthcare settings,<sup>47</sup> it is reasonable to ask whether solving relatively simple diagnostic (but scientifically interesting) tasks is really as useful as predicting and managing resources that vulnerable human caretakers truly need, such as adequate personal protective equipment.<sup>68</sup> Perhaps diagnostic applications of AI receive so much attention because diagnosing is a type of activity that powerful and economically valued physicians do, whereas supply chain management that keeps the nurses, janitors, and technicians—who have far more contact with patients than physicians—safe is mundane and less economically valued. There is less money to be had in keeping nurses alive than in displacing physician labor, for no reason other than how risk has been financialized and distributed. Technologists should ask themselves: if an automated tool for tracking and treating a pandemic is not useful



for the most vulnerable, then in what sense is it useful enough to merit investment and justify the effort of integrating it into existing practices?<sup>67</sup>

### Upsides and Downsides

Inside product and legal teams at tech companies, it has often been easier to argue for limiting the risk of a product that might harm users than it is to argue for changes to a product that benefit users. This is particularly true if those positive outcomes for users or society cannot be straightforwardly accounted for in the company's bottom line. This is at least in part because limiting the riskier aspects of a product aligns with mechanisms companies already have to limit their liabilities. Conversely, additional investment in a product "merely" to generate a social good is seen as reducing the return on investment because it raises the cost of investment without raising the financial return in a manner that can be booked. Amid the current pandemic, however, this tendency is inverted, and within the closed loop of technology vendors and enterprise clients, it may appear there is only upside to adapting the ad-serving and data brokerage surveillance apparatus to the purpose of contract tracing or epidemiological modeling. However, this upside comes with significant downside risks that COVID-19 tracking systems might pose to individuals and groups outside that closed loop in the near and distant future.<sup>69</sup>

These risks are often framed as threatening privacy, and there are various technical methods for limiting such risks for individuals, including differential privacy, encrypted computation, and decentralized computation (particularly for contact tracing and other diagnostic applications). Yet such technical methods have trade-offs in terms of accuracy and time-to-market—a not insignificant issue given the urgent demand and short contracting windows for such systems. But "privacy" is an inadequate frame for these risks, because it individualizes responsibility to manage one's own data without attending to how these risks are produced through the design of technical systems and their integration into society. Furthermore, privacy unburdens the legal system from adequately protecting the rights of those who might be harmed by the systematic misuse of personal data, and forestalls any possibility of "collective determination over the infrastructures and institutions that process data and that determine how it will be used."<sup>70</sup> Having a framework in place to work through the internal and external risks to the firm, and to society, is crucial for maximizing the upside benefits of any machine learning applications.

To accomplish these aims, such a framework would need to be supported by the necessary resources to achieve the substantive outcomes that are desired. For a company building an intervention as sweeping as contact tracing, success depends on having financial, social, emotional, and medical resources already in place to enable people to deal with the risk a digital contact tracing app assigns to them. A smartphone alert is useless if people do not have the ability to isolate themselves without suffering or failing those who depend on them, which means that the distribution of those capacities also determines the distribution of the upsides and downsides of a contract tracing intervention.<sup>71</sup> Machine learning applications are (thus far) not useful for rectifying this type of risk distribution despite that there are plenty of data proxies for economic inequity. Because so many

machine learning applications are built around arbitrating risk from an advantaged to a disadvantaged party we should expect the same from pandemic solutions that are not subject to critical interrogation.

Therefore, product managers, engineers, policy teams, legal consults, executive boards, marketing teams, and user experience researchers need to ask themselves if the necessary social support is in place where their product is intended to be used. They also would need to ask themselves if their product will conceive of and distribute risks in a manner that is just, and what measures, metrics, and other signals they would need access to know that such questions have been answered adequately. While some companies have begun to invest in the capacity to ask such questions,<sup>62,63</sup> the industry as a whole does not yet have such frameworks in place.

### Leadership Matters

Approaches that can identify and redistribute risk so that it does not fall disproportionately on those least able to bear its burden exist,<sup>72</sup> but are often not required to be brought into practice. Inside tech firms, choices about investing in upside benefit over limiting downside risk get made at the top, as are decisions about how much time, energy, and cost to invest into context alignment. As our research indicates, without clear signals from the CEO that efforts to enact good governance, build products with responsible safeguards, and prioritize users' rights would not be scrapped for a bottom-line calculation, those efforts cannot amount to much.<sup>62</sup> In recent years, activist employees and a vigilant public have provided a check<sup>73</sup> on some companies when their CEOs pursue contracts with repressive regimes or military applications. However, not all companies have experienced this kind of pressure internally (the possible reasons for this run the gamut from a lack of diversity inside to an environment in which such dissent is actively discouraged), and others have not proven responsive when faced with pressure. Some of the most recalcitrant of these companies are in the running for contract tracing application contracts, and will require a different form of pressure to conform to expectations for the responsible deployment of such a system (if this is even possible).

Leadership matters when it comes to the government response to pandemic disease, and when it comes to provisioning contact tracing algorithmic systems from tech companies, too. Many companies, big<sup>74</sup> and small,<sup>75</sup> are racing to build contact tracing applications and related infection tracing tools for governments to use. Some will be safer for users, in terms of digital rights, than others. Having options is great, but there is little obligation for federal, state, or local governments to choose the option that is best able to preserve digital rights. As it stands in the US, there are very few legal protections for civil rights and liberties that might be harmed by the misuse of the kind of data contract tracing systems depend on. Draft legislation, like the NY State Geolocation Tracking Ban<sup>76</sup> would provide a modicum of protection from unreasonable police use of such data, but government contracts for contract tracing systems will be written under the legal frameworks we have now, not the ones we wish we had.

Trust in those at the very top—of companies, of public health institutions, and (perhaps most importantly) national

governments—to do what only leadership can do is paramount. Only those who hold final authority can prioritize alternatives that distribute risk equitably rather than toward those who are most vulnerable. In practice this can mean only building in contexts where the necessary social infrastructure for product goals to be met can be ensured. It can also mean refraining from turning the tools of emergency response to COVID-19 toward other, more nefarious purposes. In the absence of regulation, or other forms of social pressure, the incentives for pursuing opportunities to squeeze additional profit by selling data or licensing a machine learning product to unsavory actors can be difficult to resist. Similarly, the resolve needed to dismantle emergency tools when the crisis passes is currently the limiting factor for leadership in determining how the risks we all face of getting sick, losing loved ones, and losing our livelihoods unfolds.

## CONCLUSION

Historically, risk distribution is a key conceptual and economic feature of machine learning applications. This type of risk is not a natural phenomenon, such as where lightning might strike, but a social construction of technocratic systems through which people must pass to have access to the economy, justice system, and health care. Machine learning has already shaped the landscape on which our society is responding to the COVID-19 pandemic by distributing risk, and the pandemic has accelerated the role that data-driven technology has in directly determining the conditions of our lives. However, despite the efficiency and utility promised by machine learning applications, there lurks a fundamental challenge: are machines good for distributing risk in the ways we actually should distribute it?

Consider for a moment, whether it is possible to build machine learning applications that distribute risk *up* the socioeconomic ladder rather than *down*. Not just “is it conceivable,” but would those of us involved in the research and development of these technologies know what that looks like and how to get there? This would, in part, look like subverting existing power hierarchies, as demonstration projects tracking white-collar crime zones have done.<sup>32</sup> It would also look like reconfiguring *who* builds machine learning applications (machine learning teams in industry, and corporate ethics teams, have had notorious difficulties retaining Black and Latinx team members)<sup>77</sup> and how they are governed. If not, then is it just to build high-tech pandemic solutions that distribute risk only downward?

Approaches that can identify and redistribute risk so that it does not fall disproportionately on those least able to bear its burden exist,<sup>72</sup> but are not required to be brought into practice. Inside tech firms, choices about investing in upside benefit over limiting downside risk get made at the top, as do decisions about how much time, energy, and cost to invest into context alignment. As our research indicates, without clear signals from the CEO that efforts to enact good governance, build products with responsible safeguards, and prioritize users’ rights would not be scrapped for a bottom-line calculation, those efforts cannot amount to much.<sup>62</sup> In recent years, activist employees and a vigilant public have provided a check<sup>73</sup> on some companies when their CEOs pursue contracts with repressive regimes or military applications. But not all companies have experienced this kind of pressure internally, and others have not

proven response when faced with it. Some of the most recalcitrant of these companies are in the running for contracts applying machine learning to the COVID-19 pandemic, and will require a different form of pressure to conform to expectations for the responsible deployment of such a system—if this is even possible.

## REFERENCES

1. Beck, U. (1992). *Risk Society: Towards a New Modernity* (Sage Publications), p. 260, Theory, culture & society.
2. Liu, A. (2020). Blaming China for Coronavirus Isn’t Just Dangerous. It Misses the Point | Andrew Liu (The Guardian). <https://www.theguardian.com/commentisfree/2020/apr/10/blaming-china-coronavirus-pandemic-capitalist-globalisation-scapegoat>.
3. Savali, K.W. (2020). Where’s the Data? We Need to Know How Many Black People Are Dying from COVID-19 (Essence). <https://www.essence.com/health-and-wellness/covid-19-data-black-people-coronavirus/>.
4. Plotkin, S., Robinson, J.M., Cunningham, G., Iqbal, R., and Larsen, S. (2017). The complexity and cost of vaccine manufacturing – an overview. *Vaccine* 35, 4064–4071.
5. Nguyen, A. (2020). Paid Sick Leave Flattens the Curve - Data & Society: Points [Internet]. <https://points.datasociety.net/paid-sick-leave-flattens-the-curve-3298b618810e>.
6. Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (St. Martin’s Press). <https://www.amazon.com/Automating-Inequality-High-Tech-Profile-Police/dp/1250074312>.
7. R. Benjamin, ed. (2019). *Captivating Technology: Race, Carceral Technology, and Liberatory Imagination in Everyday Life* (Duke University Press), p. 1.
8. Kelly, M. (2020). Unemployment Checks Are Being Held up by a Coding Language Almost Nobody Knows (The Verge). <https://www.theverge.com/2020/4/14/21219561/coronavirus-pandemic-unemployment-systems-cobol-legacy-software-infrastructure>.
9. Zaloom, C. (2004). The productive life of risk. *Cult. Anthropol.* 19, 365–391.
10. Hajian, S., Bonchi, F., and Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-Aware Data Mining (ACM Press), pp. 2125–2126. <http://dl.acm.org/citation.cfm?doi=2939672.2945386>.
11. Buolamwini J., Gebru T.. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Proceedings of Machine Learning Research. 2018.
12. Birhane, A., and Cummins, F. (2019). Algorithmic injustices: towards a relational ethics. arXiv <http://arxiv.org/abs/1912.07376>.
13. Johnson, K. (2020). AI Weekly: A Deep Learning Pioneer’s Teachable Moment on AI Bias (VentureBeat). <https://venturebeat.com/2020/06/26/ai-weekly-a-deep-learning-pioneers-teachable-moment-on-ai-bias/>.
14. Shah, S. (2016). *Pandemic: Tracking Contagions, from Cholera to Ebola and beyond* (Farrar, Straus and Giroux).
15. Rosenblat, A. (2020). Gig Work Used to Be a Recession-Proof Safety Net. Not Anymore (Fast Company).
16. Heaven, W.D. (2020). Our Weird Behavior during the Pandemic Is Messing with AI Models (MIT Technology Review). <https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/>.
17. Gebru, T., and Denton, E. (2020). FATE/CV Tutorial. <https://sites.google.com/view/fatecv-tutorial/home>.
18. Weng, S.F., Reys, J., Kai, J., Garibaldi, J.M., and Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? Liu B. *PLoS One* 12, e0174944.
19. Kruppa, J., Ziegler, A., and König, I.R. (2012). Risk estimation and risk prediction using machine-learning methods. *Hum. Genet.* 131, 1639–1654.

20. Khandani, A.E., Kim, A.J., and Lo, A.W. (2010). Consumer credit-risk models via machine-learning algorithms. *J. Bank Finance* 34, 2767–2787.
21. Walsh, C.G., Ribeiro, J.D., and Franklin, J.C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clin. Psychol. Sci.* 5, 457–469.
22. Rosenblat, A. (2019). *Uberland: How Algorithms Are Rewriting the Rules of Work* (Univ of California Press).
23. Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM* 55, 78–87.
24. Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code* (Polity Press).
25. Jacobs, A.Z., and Wallach, H. (2019). Measurement and fairness. arXiv <http://arxiv.org/abs/1912.05511>.
26. Shapiro, T.M. (2006). Race, homeownership and wealth. *Wash U. J. Pol.* 53, 23.
27. Ensign, D., Friedler, S.A., Neville, S., Scheidegger, C., and Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. In *Proceedings of Machine Learning Research [Internet]* (ACM) <http://arxiv.org/abs/1706.09847>.
28. Richardson, R., Schultz, J.M., and Crawford, K. (2019). Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *NYU L. R* 94, 42.
29. Dumit, J. (2012). *Drugs for Life: How Pharmaceutical Companies Define Our Health* (Duke University Press).
30. Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 447–453.
31. Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine Bias (ProPublica). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
32. Clifton, B., Lavigne, S., and Tseng, F. (2017). White Collar Crime Risk Zones (The New Inquiry Magazine). <https://whitecollar.thenewinquiry.com>.
33. Ryu, H.J., Adam, H., and Mitchell, M. (2018). InclusiveFaceNet: improving face attribute detection with race and gender diversity. In *ICML '18 FATML Workshop (ACM)*, pp. 1–6.
34. Friedler, S.A., Scheidegger, C., and Venkatasubramanian, S. (2016). On the (im)possibility of fairness. arXiv <http://arxiv.org/abs/1609.07236>.
35. Žliobaitė, I. (2017). Measuring discrimination in algorithmic decision making. *Data Min. Knowl. Discov.* 31, 1060–1089.
36. Amrute, S. (2020). Bored techies being casually racist: race as algorithm. *Sci. Technol. Hum. Values* 45, <https://doi.org/10.1177/0162243920912824>.
37. World Health Organization. (2020). Virtual Press Conference on COVID-19 [Internet] (World Health Organization). [https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf?sfvrsn=cb432bb3\\_2](https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf?sfvrsn=cb432bb3_2).
38. Zhao, Y., and Zhou, C.C. (2020). Applying lexical link analysis to discover insights from public information on COVID-19. *Bioinformatics* <http://biorxiv.org/lookup/doi/10.1101/2020.05.06.079798>.
39. Magdon-Ismail, M. (2020). Machine learning the phenomenology of COVID-19 from early infection dynamics. arXiv <http://arxiv.org/abs/2003.07602>.
40. Dandekar, R., and Barbastathis, G. (2020). Neural Network aided quarantine control model estimation of global Covid-19 spread. arXiv <http://arxiv.org/abs/2004.02752>.
41. Hegde, C., Jiang, Z., Suresha, P.B., Zelko, J., Seyedi, S., Smith, M.A., Wright, D., Kamaleswaran, R., Reyna, M.A., Clifford, G.D., et al. (2020). AutoTriage - an Open Source Edge Computing Raspberry Pi-Based Clinical Screening System (Emergency Medicine). <http://medrxiv.org/lookup/doi/10.1101/2020.04.09.20059840>.
42. Rehman, A., Naz, S., Khan, A., Zaib, A., and Razzak, I. (2020). Improving coronavirus (COVID-19) diagnosis using deep transfer learning [internet]. *Infectious diseases (except HIV/AIDS)*. <http://medrxiv.org/lookup/doi/10.1101/2020.04.11.20054643>.
43. Cho, A. (2020). Artificial Intelligence Systems Aim to Sniff Out Signs of COVID-19 Outbreaks (Science | AAAS). <https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks>.
44. Sivasubramanian, S. (2020). How AI and Machine Learning Are Helping to Tackle COVID-19 (World Economic Forum). <https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/>.
45. Park, S., Choi, G.J., and Ko, H. (2020). Information technology-based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *JAMA* <https://jamanetwork.com/journals/jama/fullarticle/2765252>.
46. Haskins, C., Lee, S.M., Rajogapalan, M., and Dixit, P. (2020). We Need an “Army” of Contact Tracers to Safely Reopen the Country. We Might Get Apps Instead (BuzzFeed News). <https://www.buzzfeednews.com/article/carolinehaskins1/coronavirus-contact-tracing-google-apple?origin=shp>.
47. Lee, K.-F. (2020). Covid-19 Will Accelerate the AI Health Care Revolution (Wired). <https://www.wired.com/story/covid-19-will-accelerate-ai-health-care-revolution/>.
48. McDonough, A. (2020). MTA Wants Startups to Help Solve its Coronavirus Problem (CSNY). <https://www.cityandstateny.com/articles/policy/transportation/mta-wants-startups-solve-its-coronavirus-problem.html>.
49. Lieberman, E. (2020). Risk for “Us,” or for “Them”? The Comparative Politics of Diversity and Responses to AIDS and Covid-19 (Items: Insights from the Social Sciences). <https://items.ssrc.org/covid-19-and-the-social-sciences/democracy-and-pandemics/risk-for-us-or-for-them-the-comparative-politics-of-diversity-and-responses-to-aids-and-covid-19/>.
50. Johnson, A., and Buford, T. (2020). Early Data Shows African Americans Have Contracted and Died of Coronavirus at an Alarming Rate (ProPublica). <https://www.propublica.org/article/early-data-shows-african-americans-have-contracted-and-died-of-coronavirus-at-an-alarming-rate?token=6ztk8omKweU8rlcm8bBIBXRv8xDBAzfX>.
51. Krishnan, L., Ogunwole, S.M., and Cooper, L.A. (2020). Historical insights on coronavirus disease 2019 (COVID-19), the 1918 influenza pandemic, and racial disparities: illuminating a path forward. *Ann. Intern. Med.* M20–M2223.
52. Bullard, R.D. (1999). Anatomy of environmental racism and the environmental justice movement. In *Confronting Environmental Racism: Voices from the Grassroots*, R.D. Bullard, ed. (South End Press), pp. 15–39.
53. Blanchard, J., Haile-Mariam, T., Powell, N.N., Terry, A., Fair, M., Wilder, M., Nakitende, D., Lucas, J., Davis, G.L., Haywood, Y., et al. (2020). For us, COVID-19 is personal, Kline J.A., editor. *Acad. Emerg. Med.*
54. Aouragh, M., Pritchard, H., and Snelting, F. (2020). “The Long Tail of Contact Tracing” (Societal Impact of CT) · Issue #118 (GitHub). <https://github.com/DP-3T/documents/issues/118>.
55. McDonald, S. (2020). Contact-tracing Apps Are Political (Brookings). <https://www.brookings.edu/techstream/contact-tracing-apps-are-political/>.
56. Selbst, A.D. (2017). Disparate Impact in Big Data Policing (Social Science Research Network). <https://papers.ssrn.com/abstract=2819182>.
57. Reisman, D., Schultz, J., Crawford, K., and Whittaker, M. (2018). *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, 22.
58. Kaminski, M.E., and Malgieri, G. (2019). Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *SSRN Electron. J.* <https://www.ssrn.com/abstract=3456224>.
59. Moss, E., Watkins, E.A., Metcalf, J., and Elish, M.C. (2020). Governing with algorithmic impact assessments: six observations. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3584818>.
60. The Executive Committee of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2020). Statement Regarding the Ethical Implementation of Artificial Intelligence Systems (AIS) for Addressing the

- COVID-19 Pandemic - IEEE Spectrum (IEEE Spectrum: Technology, Engineering, and Science News). <https://spectrum.ieee.org/the-institute/ieee-news/statement-regarding-the-ethical-implementation-of-artificial-intelligence-systems-ais-for-addressing-the-covid19-pandemic>.
61. Ienca, M., and Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat. Med.* 26, 458.
  62. Metcalf, J., Moss, E., and Boyd, D. (2019). Owning ethics: corporate logics, Silicon Valley, and the institutionalization of ethics. *Soc. Res.* 86, 449–476.
  63. Moss, E., and Metcalf, J. (2020). Ethics Owners: A New Model of Organizational Responsibility in Data-Driven Technology Companies (Data and Society Research Institute).
  64. Kaeser-Chen C., Dubois E., Schürer F., Moss E.. Translation Tutorial: Positionality-Aware Machine Learning. :1.
  65. Harding, S. (1986). *The Science Question in Feminism* (Cornell University Press).
  66. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., and Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT\* '19 (ACM Press), pp. 59–68. <http://dl.acm.org/citation.cfm?doid=3287560.3287598>.
  67. Sendak, M., Elish, M.C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., O'Brien, C., et al. (2020). "The human body is a black box": supporting clinical decision-making with deep learning. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (ACM), pp. 99–109. <http://dl.acm.org/doi/10.1145/3351095.3372827>.
  68. CDC (2020). Strategies to Optimize the Supply of PPE and Equipment (Centers for Disease Control and Prevention). <https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/index.html>.
  69. Sloane, M., and Cahn, A.F. (2020). Today's COVID-19 Data Will Be Tomorrow's Tools of Oppression (The Daily Beast). <https://www.thedailybeast.com/todays-covid-19-data-will-be-tomorrows-tools-of-oppression>.
  70. Goldenfein, J., Green, B., and Vijjoen, S. (2020). Privacy versus health is a false trade-off. *Jacobin Magazine* <https://jacobinmag.com/2020/04/privacy-health-surveillance-coronavirus-pandemic-technology>.
  71. Alkhatib, A. (2020). We need to talk about digital contact tracing. <https://ali-alkhatib.com/blog/digital-contact-tracing>.
  72. Peppin, A. (2020). Data-driven responses to coronavirus are only as good as the trust we place in them. <https://www.adalovelaceinstitute.org/data-driven-responses-to-coronavirus-are-only-as-good-as-the-trust-we-place-in-them/>.
  73. Holpuch, A. (2019). Google the Latest Company to Face Worker Revolt over Immigration Policy. *The Guardian* <https://www.theguardian.com/us-news/2019/aug/23/google-us-workers-speak-out-against-employers-controversial-issues>.
  74. Brandom, R., and Robertson, A. (2020). Apple and Google Are Building a Coronavirus Tracking System into iOS and Android (The Verge). <https://www.theverge.com/2020/4/10/21216484/google-apple-coronavirus-contract-tracing-bluetooth-location-tracking-data-app>.
  75. Ward, J., and Sottile, C. (2020). A Facial Recognition Company Wants to Help with Contact Tracing. A Senator Has Questions. *NBC News* <https://www.nbcnews.com/tech/security/facial-recognition-company-wants-help-contact-tracing-senator-has-questions-n1197291>.
  76. S.T.O.P. (2020). Welcomes Introduction of NY Geolocation Tracking Ban (S.T.O.P. - Surveillance Technology Oversight Project). <https://www.stopspying.org/latest-news/2020/4/10/stop-welcomes-introduction-of-ny-geolocation-tracking-ban>.
  77. Harrison, S. (2019). Five Years of Tech Diversity Reports—And Little Progress (Wired). <https://www.wired.com/story/five-years-tech-diversity-reports-little-progress/>.

#### About the Authors

**Emanuel Moss** is a doctoral candidate in Anthropology at the CUNY Graduate Center and researcher for the AI on the Ground Initiative at Data & Society. He is also a research assistant for PERVADE, an NSF-funded multisite project. **Jacob Metcalf** is a researcher for the AI on the Ground Initiative at Data & Society. He is a PI for PERVADE, an NSF-funded multisite project and also serves as the Ethics Subgroup Chair for the IEEE P7000 Standard.