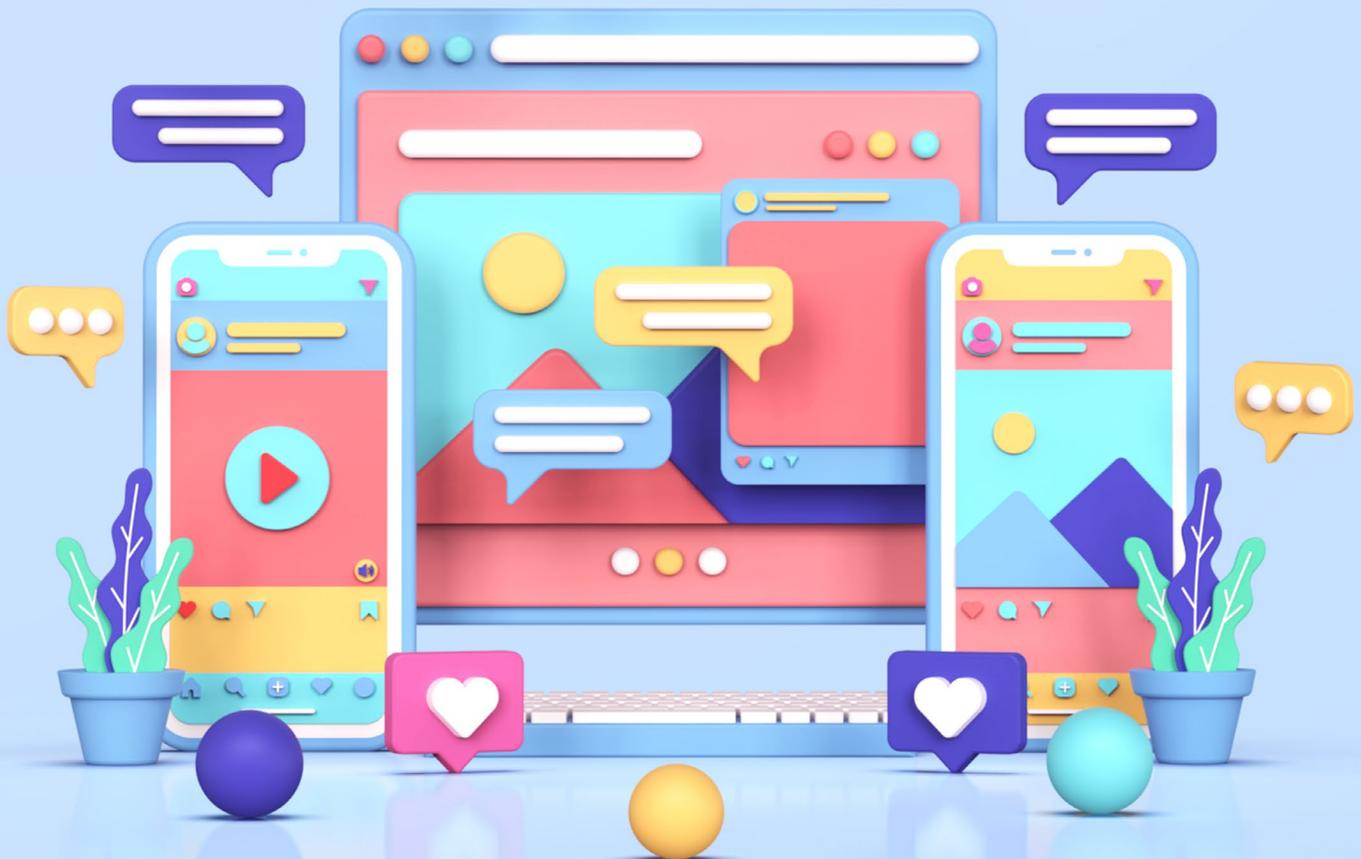


# Data, Policy and the Disaster of Misrepresentation and Mistrust

By Monica Sanders



# Abstract

Using big data and current technology is increasingly becoming part of the practice of emergency and disaster management. With respect to marginalized populations, there are flaws in the way technologies are designed and how data is collected that have negative, disparate impacts on these groups. The organizations that produce the tools and products sometimes have diversity problems that impact the resources they provide. As emergency and disaster management professionals and organizations increase their use of these tools, it is important to understand the biases associated with them, the causes and how it will impact their ability to serve groups in most need of their assistance.

## Suggested Citation

Sanders, Monica. "Data, Policy, and the Disaster of Misrepresentation and Mistrust." *Homeland Security Affairs: Pracademic Affairs* 1, Article 6. (May 2021). [www.hsaj.org/articles17234](http://www.hsaj.org/articles17234).

## Introduction

When Katrina struck the Gulf Coast, Facebook was still a social networking site with pictures, posts, and the occasional "check-in" or "poke." We understood populations and income disparities via census reports and traditional academic studies. In 2020, the year of COVID-19, extreme wildfires and more hurricanes than the system had names, responders had at their disposal crowdsourcing, mobility and browser data, mapping, drone footage, and 'citizen science' from which to draw data. Social media and technology became a common part of emergency and disaster management.

Also common to all of these events was the disparate impact on poor, Black, and Brown communities. African American, Latinx, and Indigenous people are dying from COVID-19 at two to three times the White rate at this writing time. The same groups' children would experience pandemic-related educational setbacks from schooling at home more than other groups. In the criminal justice, intelligence, and law enforcement worlds, these groups are some of the most over surveilled and over-profiled. This commonality has led to mistrust of information about the pandemic, data collection and big government interventions in all their forms. There is good reason for mistrust. Despite being watched more than other groups, these communities are often overlooked and underserved. There are historical and socio-cultural reasons for these disparities, much of it rooted in this country's history and some of it in discriminatory impacts related to emergency and disaster management.<sup>1</sup> This will be more of a challenge as the field increasingly uses big data and technology.

Another part of the problem is the nature of some of the data collected and how it was collected. For example, a great deal of mobility data is taken from cell phone call records and GPS, scooters, bike rentals, ride-hailing services, and some social media geo-tagging. Some of

these sources tend to be urban and used by the more affluent. Others are taken from major mobile phone systems and vehicle GPS subscriptions. Rural people, the poor, and those who use short-term contracts or “pay as you go” phone services, as is the case with many Latinx and African American people, would not be adequately included in any analysis or policy stemming from this data. The result is undercounting, misrepresentation, and mistakes associated with these technologies which have the dual outcome of underserving these communities during emergencies and further engendering distrust in institutions.

In this essay, I will look at the evolution of social media and other data sources in the emergency and disaster management field from Hurricane Katrina until recent times. I will also review the larger context of oversurveillance and sources of mistrust of technology in BIPOC communities. Together, this evolution and increased use of big data, social media, and other technologies, particularly within agencies with responsibilities for disaster management, has the potential to further harm marginalized communities. Addressing some of the core issues with big data, AI and machine learning could help curb some of these problems before they happen.

## Early Challenges and Ongoing Research

In our work to understand data, social media and disasters, I used desk research to review trends in the respective uses of technology vis a vis several points in time in disaster history. Those were social media use and availability from Hurricanes Katrina versus Sandy as illustrated in the work from Sarah Estes Cohen. I also looked at research, Internet trade resources and journalism about public institutions and data collection generally to understand mistrust of these institutions, particularly in BIPOC communities. Finally, to make some early inferences about all of these forms of data production and collection, I reviewed the Federal Emergency Management Agency (FEMA) and Small Business Administration (SBA) programs which are or will be data driven.<sup>2</sup>

One of the challenges and areas for further research beyond the scope of this article, is that some, but not all datasets are open for public review.<sup>3</sup> Those that are open for review are from selected programs or collected on an episodic basis showing, for example, data sets for the individual assistance program but not for mitigation, procurement or other elements. Upon reviewing the available information, while it is clear that Open API tools for standards are being used<sup>4</sup>, it is not clear which implementation tools are being used. As I will highlight later in this essay, it is in those tools (i.e. collection and analysis for implementation) where the bias and counting issues often occur. The research team has submitted requests to assist in our ongoing research to both agencies.

Concerning social media and other aspects of algorithmic bias, the next step in our ongoing work is to interview emergency managers and humanitarian responders for their perspectives and use cases about the overarching theme of this essay. Now that we understand the history and socio-cultural underpinnings, how do we address the ongoing issues of misrepresentation and mistrust alongside the progress and beneficial aspects of using big data, social media and other technological tools in disaster management?

# Social Media and Disasters

When Hurricane Katrina struck in 2005, Facebook was still a social networking tool for college students and those just out of college. There was no widespread social media use and the data being collected from usage were not used in disaster management work. Juxtapose this history with October of 2012, when Hurricane Sandy made landfall. Many in the emergency management community acknowledge that the management of that storm constituted a shift in technology use in disaster management.<sup>5</sup> At that point, people of all ages were using Facebook and Twitter, Tumblr, and YouTube. The Federal Emergency Management Agency (FEMA) and the City of New York had online presences to monitor the public's reaction to the storm and provide public safety information. Before Sandy, the city reached more than three million people with its social media presence. FEMA also used social media heavily, combining it with traditional media for various public outreach initiatives. According to GOVtech, the agency reached more than six million people with a tweet and more than 300,000 people on Facebook during the Sandy response. Researchers monitoring the evidence and literature after the event were able to distinguish at least fifteen different kinds of social media use. That would inform practitioners and policy makers about how the public received disaster information to how people reunited after disasters.<sup>6</sup> The experience led to a first responders social media working group in the Department of Homeland Security.

In addition to marking new territory in social media use, Sandy also brought in a new era of using predictive modeling to track storms and forecast flooding.<sup>7</sup> It was well reported that the United States' models were behind Europe's in predicting the storm's landfall,<sup>8</sup> but it was models from the New Jersey-based Stevens Institute of Technology and NASA DISASTERS that would be used to help officials issue more accurate flash flood warnings. This same work would be used in mitigation, offering an analysis of the long-term impact on critical and public infrastructure.

Today, we see an increased blending of technology, data, and disaster management. In the 2020 Wildfires, drones were used to drop fire retardants and set backfires in areas too dangerous for human firefighters to work. They were also used to map the wildfires along with satellite information from NASA and other federal agencies.<sup>9</sup> The combined information from the mapping, speed, location, and other data from the drones helped with predictions that would ultimately get the fires under control.

These uses indicate that predictive mapping and drone usage have matured in recent years. With large amounts of "good" datasets, scientists can better predict natural disasters.<sup>10</sup> For example, seismic data can help predict earthquakes<sup>11</sup>, while recorded rainfall and past path analysis can do the same with floods and hurricanes, respectively. A multinational, interdisciplinary review of academic papers about machine learning or AI driven analyses of each of these kinds of natural hazards, as well as tsunamis, signaled a growing interest in applied and practical applications of the technology.<sup>12</sup>

Thus, it would seem fitting when Google decided to build a flood prediction system in India and use Google Maps and Search to warn people. This issue is that the design and implementation focus was on warning people about issues with structures. There was no

accounting for how people would react to the warnings or the means by which those warnings would come. The same can be said for crowdsourcing information taken from Twitter and mobile apps by University of Dundee researchers. While the increased use of technology has numerous public benefits, it has heightened fear and mistrust in specific public segments. As we ‘modernize,’ we could inadvertently further ‘marginalize’ if we do not understand the population’s apprehensions. Much of that is rooted in how the advent of “big data” generally has underserved or harmed some of the communities we in the emergency and disaster management practice need to help most.

## The Marginalized and Mistrust in Institutions

Whether it be algorithmic bias, facial recognition errors and how law enforcement agencies deal with them, or the bad press surrounding the technology industry’s treatment of Black employees, there is not much cause to engender trust in tech institutions among Black, Indigenous and people of color (BIPOC). One of the more predominant issues surrounding trust is the over surveillance of BIPOC and other marginalized communities by law enforcement agencies. Facial recognition technology and its flaws concerning accurate identification of people with darker skin tones are better-known problems. In New Jersey, a Black man recently won a lawsuit after spending ten days in jail after the technology, combined with predictive analysis, misidentified him.<sup>13</sup> Although New York and New Jersey did use technology and social media to help the general public during Sandy, there are continuing problems with using some of the same technologies to police, often inappropriately, marginalized communities.

Some examples are outlined in the Brennan Center for Justice’s tracking of the New York Police Department’s (NYPD) use of technology.<sup>14</sup> It uses social media monitoring to track hashtags and generally monitor certain groups or individuals. There have been many instances of racial bias accusations and false positives for identifying criminals caused by misinterpreting data. The breadth of the NYPD’s technology use is wide and shocking if you happen to come from a marginalized group. According to the Brennan Center, it has at its disposal: facial recognition technology, video and social media monitoring, predictive policing, cell-site simulators, automated license plate readers, domain awareness systems, drones, x-ray vans, gunshot detection systems, body cameras, a DNA database, and surveillance towers. Every one of these has some element of race, ethnicity, or gender bias attached to its use. While the NYPD is being used as an exemplar, these issues could happen in any jurisdiction with the funding to leverage this technology amount. In addition to the case in New Jersey, there have been similar problems in Florida.<sup>15</sup> Given that police and fire departments are part of or work closely with emergency management, it is not unimaginable that the tools would be used in multiple scenarios. The problem is that the bias associated with them can also appear in multiple scenarios.

In each of these technologies, there is an algorithmic core that lends itself to a discussion about algorithm bias. We know that algorithmic bias is when artificial intelligence (AI), whether it be machine learning or deep learning, develops biases based on race, gender, or religion.<sup>16</sup> An example is a photo app incorrectly tagging people with darker skin. Many times, it is caused by the data from which it learns. A sorting program that only receives photos of white or lighter-skinned people might mistakenly eliminate darker people because it cannot recognize those features.

Other times, the bias comes from the data scientist or algorithm writer.<sup>17</sup> It is common knowledge that the tech industry is a male-dominated field. Because of this factor, applications, programs, and AI may be disproportionately “masculine” in their learning. A natural language program used in job applications, for example, may not select a female engineering candidate because it was exposed to mainly male pronouns in association with that kind of work. Essentially, algorithmic bias is yet another extension of human bias. The tools we develop inherit our societal, racial, and gender perspectives.<sup>18</sup>

Where AI and big data could alleviate some of the social ills in our world as their designers undoubtedly intended them to, they can also amplify them. In a health care system that already disproportionately underserves people of color, it has been reported that AI bias means chronically sick Black people are less likely to be referred for complex care than equally sick white people.<sup>19</sup> Given that AI is involved in more than 200 million Americans’ managed care, this is a severe under-representation issue. Like the one mentioned earlier in this essay and used in various law enforcement, security, and other industries, facial recognition systems misidentify Asian and African descent people roughly 100 times more than white people.<sup>20</sup> Native Americans had the highest misidentification rate of any group, according to the Washington Post article. There are stories of such problems in algorithm usage in systems ranging from beauty contests to job application portals. Take the pronounced, society-wide impact of algorithmic bias and combine it with historical mistrust of government and institutions; it is clear why some groups will generally mistrust the technology, particularly in times of crisis.

## Problems with Data Collection and Analysis

In order to understand the issues with “big data” and marginalized communities, it is important to understand how it is defined. This section begins with a brief overview and then continues with examples of how the nature of the data and methods by which it is collected leads to undercounting.

For those who are sticklers for the purest definitions, it refers to software for data sets that exceed traditional databases’ capabilities. For others, the term big data is shorthand for predictive analytics. A good example of either type is the New York Stock Exchange. It represents a massive data set that is continuously in flux and beyond a traditional database. Though in flux, it moves in somewhat predictable historical patterns and responds to certain factors in predictable ways. That means it can be used as a predictor or be the subject of predictive analytics.

The more extensive definition of big data refers to the typologies and characteristics of the big data’s information. There is structured data, which is any data that can be stored, accessed, and processed in the form of a fixed format.<sup>21</sup> This could be a database or survey results. Any data with an unknown form of structure is classified as unstructured data. In addition to the size being huge, unstructured data poses multiple challenges in its processing for deriving value out of it. Think of “Google Search” results as an example of unstructured data.

Semi-structured data has aspects of both. When thinking of big data, it is vital to acknowledge the four characteristics “V’s” or as many technologists and data scientists call them: Volume, or the sheer amount of information; Variety, meaning the different types of information as collected from diverse sources; Velocity, or the speed at which it is being generated; and Variability or Veracity, which refers to inconsistencies in the data.<sup>22</sup> Some include a fifth, Value, because of its meaning to companies in predicting consumer behavior or making more efficient hiring decisions. Some of the more widely used methods involve semi-structured data.

When looking at the discussion around disproportionate impacts on marginalized communities, it is the first and last of those “V’s” where the problems come into play. This means there is not enough variety in the collection methods or there are problems with the standards and implementation, or veracity of the data. Here are some of the more prominent data collection methods and corresponding data sets: browser data from cookies and trackers, social media harvesting, employer databases, satellite imagery, games, and cell phone location data.<sup>23</sup> For each kind of collection, there is a potential problem when dealing with marginalized groups. Social media harvesting and browser data are sometimes collected with the same biased tools or individuals with the same implicit bias as noted in the algorithms themselves, leading to similar misrepresentation or underrepresentation of marginalized people. In another example, a study from Stanford and Carnegie Mellon noted that older and nonwhite Americans are less likely to be captured in the mobility data used in COVID-19 tracing.<sup>24</sup> One of the problems is the vendors would not disclose the collection sources. However, previous studies have shown that call records collection and GPS traces, typical collection methods, undercount people in rural areas and low-income areas where smartphone usage is not high.<sup>25</sup> The same is valid for low-income Black and Latinx users; they may not have a monthly cellular plan but opt for “pay as you go” options where information is not stored or aggregated in the same way.

So how do we prevent some of this bias? Making sure the data going into an algorithm is as representative as possible is probably *the* most crucial factor in preventing machine learning bias. Placing all the different types of groups of data into a dataset can be challenging. Ensuring that the data is segmented to be sure that it is correctly grouped and managed is a great deal of work. But when balancing that against the consequences to society, and more selfishly, our tool’s accuracy, it is worth it. When we have insufficient data about one group, we can weigh groups against each other to compensate. But be aware this can lead to new, unexpected biases. Working through all of these issues requires monitoring, reviewing, and a great deal of real-world testing.

Another serious issue contributing to bias and the ensuing mistrust in institutions is one of representation in organizations that produce big data. In 2020, Trust Radius, the organization which has previously produced a Women in Tech report, issued its inaugural People of Color in tech report.<sup>26</sup> According to the report, Black, Latinx, and Native American people make up less than 5% of the tech workforce. The report took an honest view of the problem with the broader term “people of color” because it causes the representation of Asian and South Asian tech professionals to skew diversity numbers favorably. Google recently made headlines for hiring a well-respected AI expert of African descent but then firing her when she raised ethical issues concerning algorithmic bias, facial recognition programs as well as workplace issues.<sup>27</sup> If the faces of tech and faces behind tech as we know it are not representative, it should not be surprising that the tools they build will be more susceptible to the kinds of bias explored in this essay. Acknowledging this factor in hiring and team-building decisions is another bias mitigation technique.

# Why It Matters

Agencies and organizations using “big” or “bespoke” data and technologies from these same sources and vendors to support programs or create policy will not have equitable or even consistent outcomes. Among the federal agencies engaged in embracing big data as part of their modernization efforts are those with missions that impact disaster and emergency management: FEMA, HUD, and the Small Business Administration (SBA). FEMA has launched several initiatives as part of its data-driven decision-making work. That began with the Foundations for Evidence-Based Policymaking Act, signed into law in January 2019.<sup>28</sup> Those initiatives include data transparency work and information sharing with agencies with which it frequently coordinates, like the Army Corps of Engineers. The Enterprise Data Analytics Modernization Initiative focuses on sharing information like generators, gasoline, and other critical supplies and then predicting areas of most need. It will also be implemented in financial programs like preparedness grants and public assistance programs, even as it attempts to address equity in its operational work.<sup>29</sup> FEMA also deployed big data analytics to improve the return on investment for building code standards, making a stronger case for their implementation in the process.<sup>30</sup>

There are many benefits to come from this work. First, FEMA’s systems need modernization and better security practices, especially after the inspector general report about the data leak of disaster survivor information in 2019.<sup>31</sup> Streamlining information about resources and improving placement and distribution of those resources is also a valuable effort. Decades of hazard mitigation advocacy efforts have been centered around strengthening and enforcing building codes to improve resiliency and lessen the cost of disasters.

The potential problems come with the biases, particularly the undercounting biases associated with cell phone data not being recognized. The impact on communities can range from resources not being prepositioned in or directed to the most marginalized areas. Grant program design and implementation may not include race and poverty indicators or gender and poverty because of the algorithmic bias related to misidentifying or misrepresenting individuals with these characteristics. There is a risk that relying on big data too much, too quickly could deepen inequity and lessen efficiency.

HUD and the SBA also have embarked on similar work. HUD has opened its datasets to the public, researchers, and other interest parties for free online. It includes everything from its assessment of American Indian Housing Needs<sup>32</sup> to the Neighborhood Mapping Tool it uses to identify “areas of greatest need.”<sup>33</sup> The agency is also considering FEMA’s approach, using big data analytics to improve service delivery or risk and budget management analysis.<sup>34</sup> The SBA has a similar digital strategy emerging as part of the same government-wide initiatives that are now law. It has opened its program datasets to the public.<sup>35</sup> In both cases, the information is clear, though the collection methods are not entirely clear. Unless there is a policy to mitigate or eliminate algorithmic and collection bias, policy based on this data will miss population segments or entire groups. The same could be true concerning outside organizations using the information.

## Conclusion

Just as the world of big data, artificial intelligence, and technology, in general, is broad, inspiring, and overwhelming, so is the world of disaster and emergency management. A movement to leverage more technology to make our work safer and efficient should not come at the cost of our collective ability to value humanity. When in the role of consuming data, be sure to question vendors and the tools they provide. Ask questions about how datasets are collected. Leverage that information against the “ground truths” volunteers, staff, and local organizations will already know about any given area’s needs and vulnerabilities. Use the human intelligence to fill any potential gaps in the artificial intelligence to generate more accurate outcomes. Socialize the limitations of these tools in training by including them as factors in case studies and scenario gaming.

When in the role of being a creator or producer of data and technology tools, be intentional in your operations. Review some of the existing research about how to mitigate against bias and for fairness in machine learning and AI tools. The AI Fairness 360 Toolkit<sup>36</sup> can be used to test the quality and integrity of algorithms as well as set metrics for datasets. Another way to deal with fairness and overcome collection bias is to ensure that a wide variety of population segments and vulnerability indicators are included in datasets.

Consistent with other calls for more diversity in the field, as more disaster and emergency management organizations and agencies begin to hire data scientists, engineers, and other tech professionals, make sure to hire a diverse workforce to avoid inherent algorithmic bias.

## About the Author

Monica Sanders is an Associate Professor of Sociology and Criminal Justice at the University of Delaware. Her research interests include law, technology and vulnerable groups. Previously, she created and taught a course on disaster law and policy at the Washington and Lee University School of Law. She also has faculty affiliations with the Georgetown University Law Center and Emergency and Disaster Management Program. Professor Sanders also served as Senior Legal Advisor for International Response and Programs at the American Red Cross, focusing on international disaster response and humanitarian assistance principles. Previously, she was a Senior Committee Counsel for both the House of Representatives and Senate Committees on Homeland Security. She serves on a number of advisory boards, including the Institute for Building Technology and Safety. Professor Sanders received her degrees from the University of Miami (B.S.), the Catholic University of America (JD), Harvard Law School Project on Negotiation (Cert.) and University College London (LL.M). She may be reached at [monica.sanders@georgetown.edu](mailto:monica.sanders@georgetown.edu).

# Notes

1. Sue Sturgis is the editorial director of Facing South and the Institute for Southern Studies. Sue Sturgis, “Recent Disasters Reveal Racial Discrimination in FEMA Aid Process.” Facing South. September 24, 2018. <https://www.facingsouth.org/2018/09/recent-disasters-reveal-racial-discrimination-fema-aid-process>.
2. Monica Sanders, “TDAI Speaker Series, AI, Big (or Bespoke) Data and Disasters” TDAI Speaker Series, 28 Jan. 2021, [www.youtube.com/watch?v=wJ6Hb3yyzOQ](http://www.youtube.com/watch?v=wJ6Hb3yyzOQ), (See Q&A for description of datasets and review issues).
3. OpenFEMA, API Documentation, 13 Oct. 2020, [www.fema.gov/about/openfema/api](http://www.fema.gov/about/openfema/api).
4. “About,” OpenAPI Initiative, 17 Dec. 2020, [www.openapis.org/about](http://www.openapis.org/about). (Explanation of standard and origins used in AI and data tools).
5. Sarah E. Cohen, “Sandy Marked A Shift For Social Media Use In Disasters,” 2013, <https://www.govtech.com/em/disaster/Sandy-Social-Media-Use-in-Disasters.html>. FEMA and City of New York both make unprecedented use of social media throughout disaster cycle.
6. Brian J. Houston, et al., “Social Media and Disasters: a Functional Framework for Social Media Use in Disaster Planning, Response, and Research,” *Wiley Online Library*, John Wiley & Sons, Ltd, 22 Sept. 2014, <https://www.onlinelibrary.wiley.com/doi/abs/10.1111/disa.12092>.
7. “Hurricane Sandy Newsletter,” Stevens Institute Of Technology, 2013, <https://www.stevens.edu/research-entrepreneurship/research-centers-labs/davidson-laboratory/hurricane-sandy-newsletter>.
8. Scott Johnson, “Why European Forecasters Saw Sandy’s Path First,” *Ars Technica*, 2021, <https://arstechnica.com/science/2012/12/why-european-forecasters-saw-sandys-path-first/>.
9. David Helvarg, “Fireball-Dropping Drones and The New Technology Helping Fight Fires, 2020, <https://www.nationalgeographic.com/science/2020/10/fireball-dropping-drones-new-technology-helping-fight-fires/#close>. A variety of technologies were used to help fight and track 2020 Wildfires in California, Oregon and Washington State.
10. Joshi Naveen, “How AI Can And Will Predict Disasters,” 2019, <https://www.forbes.com/sites/cognitiveworld/2019/03/15/how-ai-can-and-will-predict-disasters/?sh=2e877a445be2>.
11. James Vincent, “Google And Harvard Team Up To Use Deep Learning To Predict Earthquake Aftershocks,” The Verge, 2018, <https://www.theverge.com/2018/8/30/17799356/ai-predict-earthquake-aftershocks-google-harvard>.
12. F. Martínez–Álvarez and A. Morales–Esteban, “Big Data and Natural Disasters: New Approaches for Spatial and Temporal Massive Data Analysis,” *Computers & Geosciences*, Pergamon, 6 May 2019, [www.sciencedirect.com/science/article/pii/S009830041930411X](http://www.sciencedirect.com/science/article/pii/S009830041930411X).
13. Joe Jurado, “Black Man Files Lawsuit After Being Jailed Due To Error In Facial Recognition Technology,” *The Root*, 2020, <https://www.theroot.com/black-man-files-lawsuit-after-being-jailed-due-to-error-1845967661>. The article goes in-depth to explain that there were no overlapping factors between the victim and the profile generated by the technology.
14. Angel Diaz, “New York City Police Department Surveillance Technology,” 2019, <https://www.brennancenter.org/our-work/research-reports/new-york-city-police-department-surveillance-technology>. Chart showing breadth of NYPD surveillance apparatus and problems associated with every technology.
15. Kashmir Hill, “Another Arrest, and Jail Time, Because of a Bad Facial Recognition Match,” January 3, 2021, <https://www.sun-sentinel.com/consumer-reviews/sns-nyt-jail-time-bad-facial-recognition-match-20210103-vk3ypyxp3zczxbkdhpwqv6hir4-story.html>.
16. Ben Dickson, “What Is Algorithmic Bias?” 2018, <https://bdtechtalks.com/2018/03/26/racist-sexist-ai-deep-learning-algorithms/>.

17. Megan Garcia, "Racist in the Machine: The Disturbing Implications of Algorithmic Bias," *World Policy Journal*, vol. 33 no. 4, (2016): 111-117; *Project MUSE* [muse.jhu.edu/article/645268](https://muse.jhu.edu/article/645268).
18. Karan Praharaj, "How Are Algorithms Biased?" Medium, 2020, <https://towardsdatascience.com/how-are-algorithms-biased-8449406aaa83>.
19. Heidi Ledford, "Millions Of Black People Affected By Racial Bias In Health-Care Algorithms," *Nature.Com*, 2019, <https://www.nature.com/articles/d41586-019-03228-6>.
20. Drew Harwell, "Federal Study Confirms Racial Bias Of Many Facial-Recognition Systems, Casts Doubt On Their Expanding Use," 2019, <https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>.
21. Adam Jacobs, 2009, <https://dl.acm.org/doi/fullHtml/10.1145/1536616.1536632>.
22. "What Is Big Data? | Oracle", Oracle.Com, 2022, <https://www.oracle.com/big-data/what-is-big-data.html>.
23. Villanovau.Com, <https://www.villanovau.com/resources/bi/6-ways-companies-can-collect-your-data/>; See also Louise Matsakis, "The WIRED Guide to Your Personal Data (and Who Is Using It," *Wired. Conde Nast*, 2019, <https://www.wired.com/story/wired-guide-personal-data-collection/>.
24. Amanda Coston et al., *Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy*, (November 2020), <https://doi.org/https://arxiv.org/pdf/2011.07194.pdf>.
25. Nishant Kishore et al., "Measuring Mobility to Monitor Travel and Physical Distancing Interventions: a Common Framework for Mobile Phone Data Analysis," *The Lancet Digital Health* 2 (11), (2020) [https://doi.org/10.1016/s2589-7500\(20\)30193-x](https://doi.org/10.1016/s2589-7500(20)30193-x).
26. 2020 People of Color in Tech Report, Trust Radius, September 21, 2020, <https://www.trustradius.com/vendor-blog/people-of-color-in-tech-report>.
27. Olivia Solon, and April Glaser, "Google Workers Mobilize against Firing of Top Black Female Executive," NBCNews.com. NBCUniversal News Group. December 4, 2020. <https://www.nbcnews.com/tech/internet/google-workers-mobilize-against-firing-top-black-female-executive-n1250038>. Black women also only represent about .07 percent of Google's technical workers. Timnit is also the founder of Black in AI.
28. Paul Ryan, "H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018," Congress.gov. January 14, 2019. <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
29. "FEMA Looks to Build Trust in Data Sharing after IG Found It 'Overshared' Records," Federal News Network, May 13, 2020. <https://federalnewsnetwork.com/big-data/2020/05/fema-modernizes-its-data-analytics-with-financial-transparency-in-mind/>; National Advisory Council Report to the FEMA Administrator, Federal Emergency Management Agency, Nov. 2020, [www.fema.gov/sites/default/files/documents/fema\\_nac-report\\_11-2020.pdf](http://www.fema.gov/sites/default/files/documents/fema_nac-report_11-2020.pdf).
30. News, Concrete. 2020, "Home," Concrete Products, December 7, 2020, <http://concreteproducts.com/index.php/2020/12/07/fema-big-data-proves-high-building-code-standards-roi/>. FEMA conducted a nationwide study of hazard resistant codes standards.
31. "Survivor Privacy Incident," n.d. Survivor Privacy Incident | FEMA.gov. Accessed January 3, 2021, <https://www.fema.gov/survivor-privacy-incident>. Sensitive and personally identifiable information of people who received assistance in the Transitional Sheltering Assistance Program was "overshared" according to the agency.
32. This is the descriptor for Native Americans or Indigenous people used on HUD's website.
33. "DATA SETS: HUD USER," n.d. DATA SETS | HUD USER, Accessed January 3, 2021. [https://www.huduser.gov/portal/pdrdatas\\_landing.html](https://www.huduser.gov/portal/pdrdatas_landing.html). HUD's open data set website.

34. "HUD, USDA Throwing a Little Spaghetti against the Wall to See What Big Data Projects Stick," 2019, Federal News Network, December 16, 2019, <https://federalnewsnetwork.com/ask-the-cio/2019/12/hud-usda-throwing-a-little-spaghetti-against-the-wall-to-see-what-big-data-projects-stick/>.
35. "Open Government Data Sources: The U.S. Small Business Administration," n.d. Small Business Administration, Accessed January 3, 2021, <https://www.sba.gov/about-sba/sba-performance/open-government/digital-sba/open-data/open-data-sources>.
36. "AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias," *IEEE Xplore*, BM Journal of Research and Development, 18 Sept. 2019, [ieeexplore.ieee.org/abstract/document/8843908](http://ieeexplore.ieee.org/abstract/document/8843908).

---

## Copyright

Copyright © 2021 by the author(s). Homeland Security Affairs is an academic journal available free of charge to individuals and institutions. Because the purpose of this publication is the widest possible dissemination of knowledge, copies of this journal and the articles contained herein may be printed or downloaded and redistributed for personal, research or educational purposes free of charge and without permission. Any commercial use of Homeland Security Affairs or the articles published herein is expressly prohibited without the written consent of the copyright holder. The copyright of all articles published in Homeland Security Affairs rests with the author(s) of the article. Homeland Security Affairs is the online journal of the Naval Postgraduate School Center for Homeland Defense and Security (CHDS).

Cover photo created by syifa5610, [www.freepik.com](http://www.freepik.com).