

②

white paper series: machine learning

# SOCIAL BIAS

*In Machine Learning and  
Early Recovery*



## EXECUTIVE SUMMARY

Machine learning is a branch of artificial intelligence that uses various data sources to train an algorithm for a specific purpose. Machine learning simplifies and automates diagnostic or identification processes in various industries like healthcare, finance, and urban planning. Recently, data scientists have started using trained machine learning algorithms to speed up the early recovery process after natural disasters. As a result of climate change, scientists and disaster professionals must plan for considerable sea level rise and increasing severity and frequency of natural disaster events. The more frequently natural disasters occur, the greater the need for faster recovery and stronger resilience. Government entities like the Federal Emergency Management Agency (FEMA) have yet to start utilizing machine learning to standardize property damage assessments in the natural disaster recovery process.

Machine algorithms can assess property damage based on collecting and labeling perishable data such as photos of damaged buildings. The data used to train an algorithm is typically perishable, meaning that the data becomes less relevant over time as those affected by natural disasters work to fix their homes. Perishability creates temporal restraints for data collection. Although this technology can accelerate the relief and recovery process, the use of highly technological tools without assessing their impacts on people can further existing inequalities in the recovery process. Human-induced biases in machine learning appear in data input, collection, algorithmic frameworks, and the application of models. If machine learning is to be relied on by organizations as a method to understand damage and prioritize resources, there must be deliberate action to control bias.

## 2.1 Introduction

The field of disaster response and recovery has grown steadily over the last few decades as climate change has caused more frequent and severe natural disasters. As this field has grown, emergency management professionals have begun to explore the use of data science techniques that can predict the severity of a disaster before it happens and helps assess damages in communities post-disaster. In the context of federal disaster recovery, damage assessment is a foundational step in an aid-distribution process that guides resource distribution.

Damage assessment as a tool is deployed by both the private and public sectors based on community capacity or grant requirements. Assessments are currently conducted by on-the-ground responders. Their efforts help provide decision makers key insights and summarizations of post-disaster conditions. The limitations to understanding damage through property assessment are not necessarily visible, as social infrastructure such as community networks become fractured in line with physical damage post-disaster. Integrating data science technology into fields like emergency management could aid in making communities more resilient, as technology helps shorten the recovery timeline through faster processing speeds. Rather than human input assessment, technology such as machine learning can streamline and reduce bias.

Perishable data becomes less relevant over time as those affected by natural disasters work to fix their homes and thus eliminate pertinent data. It becomes imperative for emergency and first responders to collect data within hours to days after a natural disaster, a constraint acknowledged by residents and responders. Machine learning algorithms that assess property damages rely on collecting and annotating perishable data such as photos of damaged buildings.

This paper discusses how machine learning tools can store and analyze perishable data from natural disasters by integrating data collection, damage assessments, and aid recovery operations.

## 2.2 Early Recovery Demands

There are additional issues with the integration of data driven tools into early recovery. Under the constraints of weather, time, road access, local demand, and disaster damage, emergency managers are pressured by private and public organizations, such as FEMA or insurance companies, to increase recovery speeds and resource allocation distribution. Local emergency managers, or prominent community leaders, may be tasked with undertaking data collection to inform planning support for faster recovery aid and management. There are many dimensions that contribute to the demand for faster early recovery. For example, conversations with local residents and leaders from Southeastern Louisiana highlighted local concerns for how recovery from previous disasters would directly affect a vulnerable community's ability to recover from the next.

Another pressure to increase the speed of decision making stems from delayed and uncoordinated long term recovery. In the wake of Hurricane Ida, we see early recovery action influencing long term recovery effectiveness. As mentioned previously, local residents whose homes were most impacted struggled with the ability to receive relocation cash assistance or rental assistance until a damage assessment could be conducted. Faster early recovery speeds may eliminate the induced trauma of displacement and evacuation through faster access and distribution of life essential aid and resources.<sup>1</sup> While the process will likely continue to rely on human input for verification of damage, technology can better inform actors on the likelihood of damage for rapid deployment of people and resources.

## 2.3 Perishable Data

A limitation related to damage assessments and early recovery is the lack of data that would enable a comprehensive analysis of damage. In general, rebuilding occurs almost immediately after the period of sheltering or relocation is over. The measures used to address damage may also inhibit the ability to understand the extent of damage. To study disaster recovery at the community level, researchers have an inherent need for the rapid collection of damage

## DEFINITIONS

### MACHINE LEARNING

A branch of Artificial Intelligence and Computer Science in which computer systems are able to learn and adapt without following human instructions. Data patterns are inferred by model algorithms and statistical models.

### IMAGE CLASSIFICATION

Labeling and classification of digital photos. For example, an image of a damaged home is classified based on the extent of aesthetic and structural damage.

### OBJECT DETECTION

Object detection can locate specific features within an image. For example, an image including yard debris and a home will only detect damage to the building. Bounding boxes are utilized to identify the specific object feature as distinct from others.

### COGNITIVE BIAS

The tendency for people to perceive information in different and distorted ways based on their own experiences and preferences. How information is framed, as well as the context in which it is given (or lack thereof) affects ones perception of reality.

### ACCURACY

An overall metric that determines how often the machine learning algorithm correctly identifies images or object in the model.

### PRECISION

A categorical metric that determines the number of correct observations the model predicts over the total number of correct and incorrect observations of a class.

### RECALL

A metric that determines how often the machine learning algorithm correctly identifies images or object in the overall model.

data. The data, known as perishable data, are measures that are vital to capture, analyze, and understand immediately following a disaster event when contextual information is clear in the memory of residents. However, the prioritization of needs by local actors with limited capacities in the aftermath of disasters means collecting perishable data for research purposes is not of the utmost importance.

A temporal gap exists around the timeline for perishable data capture while a capacity-gap furthers organizational issues. Therein lies the outlet and potential for utilizing data science methods to increase prioritization and local support through rapid and efficient collection methods. While data science tools can capture that perishable data, they can also be leveraged to store, understand, and communicate damage data in a timely manner.

## 2.4 Machine Learning

Based on early recovery demands and perishable data constraints, deep learning is one emerging technology that has the potential to be adapted for equitable, efficient operations. Deep learning is a type of machine learning that learns through neural networks. The name neural network refers to its algorithms as modeled from human cognitive function. Convolutional neural networks are a special focus of deep learning aimed to identify and classify images. All deep learning models have one thing in common—the data used to train the algorithm is understood through weights of certain characteristics. These weighted characteristics, also referred to as weights, instruct the model to make decisions. As a process for identification or classification, machine learning works best when provided clear and measurable data.<sup>2</sup> Altogether, the complexity and novelty of deep learning triggers barriers to entry outside of typical data science fields. Civicly minded analytics, apart from data technologists, may help bridge the informational knowledge gap.<sup>3</sup> Civic analytics can include pre-existing roles in government who learn to incorporate data science into decision making processes.

As the uses of machine learning continue to expand, the urban planning field turns to deep learning to improve the planning practice. Urban planners

will need education training in data science and awareness of the biases in machine learning to navigate complex environments. Multilayered urban problems already garner immense interagency coordination and require profound organizational knowledge. As a field classically trained with skills in research, mapping, and community engagement, urban planning both pursues yet lacks comprehensive data science knowledge related to machine learning. If urban planners continue to turn towards data science as an approach to tackling complex, interconnected issues, they must recognize the inherent limitations and biases of data-driven decisions using neural networks.<sup>4</sup>

## 2.5 Damage Assessment with Machine Learning

Integrating various sources of information from communities and infrastructure affected by natural disasters into rapid damage assessment increases the accuracy of learning-based models. Training models that combine data from multiple hurricane events can accurately predict the estimated damages of a test-case hurricane event.<sup>5</sup> Collecting images from a combination of sources increases machine learning algorithms' accuracy, precision, and recall. Incorporating pictures posted to social media platforms of damages due to earthquakes from multiple events<sup>6</sup> with similar images from Google optimizes the performance of deep learning models.<sup>7</sup>

Two vital needs have emerged during this project's research into artificial intelligence algorithms for



Image 2.1: A manufactured home in Dulac, Louisiana damaged from Hurricane Ida

damage assessment. First, there is a need for an abundance of training data and imagery specifically curated to assess infrastructural damage in post-natural disaster communities.<sup>8</sup> Open-sourced datasets, such as Crisis NLP and the Qatar Computing Research Institute, contain images collected from social media after natural disasters; however, the pre-annotated images do not align with FEMA Preliminary Damage Assessment guidelines.<sup>9</sup> Second, the post-natural disaster images accessible through media sources, open-sourced datasets, and stock photography outlets results in severely imbalanced training datasets for machine learning models<sup>10</sup>. In general, open-sourced datasets contain an overrepresentation of images classified as severely damaged. To correct for the underrepresentation of other damage types, CrisisNLP datasets perform Google image searches to assist with training machine learning models. While correcting for underrepresentation is considered good practice for increasing overall model accuracy, sourcing images from Google does not equitably represent the communities impacted by natural disasters.

Similarly, datasets sourced from social media also

contain a large portion of imagery unrelated to assessing damage to a home. The poor signal-to-noise ratio of relevant images sourced from social media can be addressed by direct human-input selection or creating a separate machine learning program to filter for useful images.<sup>11</sup> Regardless of the methods one uses to select relevant images for input into a damage assessment model, biases can be introduced into the created dataset.

## 2.6 Biases of Machine Learning

All machine learning models require training data that is generated and collected from human experiences.<sup>12</sup> Human experiences and memories of past events affect thinking, behaviors, and the decision-making of current events. To err is human; errors in mental processing and interpretation of information define cognitive bias. Everyone exhibits cognitive bias. It occurs when we self-select news sources that reflect our political viewpoints or assume another person's beliefs and opinions.<sup>13</sup> When biases refer to

an individual's or group of people's social identities, such as race, gender, or religion, we identify these as social biases. Since all data is generated or affected by human decision-making, all data is inherently biased.

There is a need to design machine learning methods with intention that can control for biases. Many types of biases exist within machine learning processes; each type has its own potential method for control. Social bias refers to an individual being in favor or against others based on their race, gender, or other social identities. Machine learning can control for social bias during the data collection process. During the data collection process, the inclusion of historically disadvantaged communities, such as the Houma Nation in Louisiana, into training datasets ensures equitable representation within the model. In damage assessment, an assessor or local official with prior knowledge of a neighborhood might classify damage differently based on location; this type of bias is known as confirmation bias. Confirmation bias occurs when individuals go into a decision-making process with subjective thoughts about their tasks. Confirmation bias is controlled for during the annotation process by standardizing labeling protocols with clear guidelines.

To further support the adoption of machine learning processes, there are a host of limitations that can produce bias and inequity and must be addressed. Machine learning assemblage, or the process of creating a machine learning cycle from collection to training, holds different biases than its application in the field. Researchers and planners must protect local communities from undue harm or negligence through



Image 2.2: The last remnants of a stilted home in Chauvin, Louisiana after Hurricane Ida swept through the Houma Nation

assembling and applying machine learning with transparent, robust methodology.

## 2.7 Reducing Bias of Machine Learning Algorithms for Damage Assessment

The use of machine learning algorithms in disaster recovery is all but inevitable. Machine learning models can better predict future events and accelerate recovery.<sup>14</sup> However, no two natural disasters impact communities in the same manner. Each subsequent natural disaster generates more data than past disasters as new prediction techniques become operational. Before, during, and after disasters, digital information is collected from sensors, satellite and surveillance imagery, drones, smartphones, and many other Internet-connected devices. Survivors of natural disasters also use social media platforms to communicate with relief and recovery professionals, often posting images with location data<sup>15</sup>, which allows the targeting of recovery efforts. All relevant data points must be collected and analyzed quickly to ensure their usefulness in recovery operations. Perishable data must also be collected and made available to data science researchers as inputs in machine learning models, such as those used for preliminary damage assessments.

There is a need for data scientists to have comprehensive imagery datasets of damaged structures following natural disasters to train, validate, and test machine learning algorithms.<sup>16</sup> To have an inclusive nationwide recovery process, community partners need to communicate with the technical knowledge and capacity of data science teams to build equitable machine learning models for future disaster events. In fact, recovery can have different meanings depending on who uses the term.<sup>17</sup> Developing an algorithm to accelerate the recovery process and rapidly assess property damage without incorporating the social context of local communities has the potential to exacerbate the economic inequities currently exhibited in the United States. Neither an emergency manager nor an algorithm can determine a household's ability to rebuild and recover



Image 2.3: Image classified as "none" for damage assessment by open-sourced database

simply from analyzing images of damaged homes without first understanding the local context and social networks available to a specific community. Therefore, machine learning algorithms that detect damage should not be the sole determinant of prioritization given highly connected, wealthier, or well designed households may recover more quickly and effectively compared to less connected, lower income, or less structurally sound households.

#### Annotation Protocols

Standardizing annotation protocols for categorizing damage assessment images controls human-induced biases in datasets. Annotation standardization based on FEMA's Preliminary Damage Assessment Guide can help to reduce prejudices and increase model precision through clear definitions of categorical damage. Minor damage can be defined or classified with clear differences than other levels of damage such as severe. Rather than use personal bias of damage, standardization eliminates human error. Collection of data, annotation of images, and training of machine learning models for damage assessment is best completed in preparation for a natural disaster, rather than post-disaster. Exposure to traumatic events can lead to cognitive biases through changes to an individual's locus of control or how one perceives the control one has over external events.<sup>18</sup> Therefore, once disaster strikes, cognitive biases can be amplified, particularly for local emergency personnel.

Properly categorizing the distribution of damage assessment through combining data points from multiple disaster events also controls for biases in datasets when training a machine learning algorithm. The severity of a single natural disaster can lead to an overrepresentation of houses assessed as having major or destroyed levels of damage. Overrepresentation of damage categories can have real-world consequences when distributing aid such as household assistance funding. A household with income levels below or near the federal poverty level may not have the ability to recover from minor levels of damage from a natural disaster.

Elements of social recovery – providing shelter and long-term housing, food and financial assistance, resilience, and psychological support – should

be prioritized at the same level as economic and infrastructure systems recovery. One's perception of recovery is determined by how well they return to normal or begin a "new normal" after recovering from the mental, financial, and physical impacts of natural disasters. Compressing the timeline for damage assessments does not exclusively accelerate recovery for communities. Our team sees the need for data science, artificial intelligence, and urban planning professionals to work together to improve the future of equitable natural disaster recovery.

## 2.8 Implementation Considerations

In joining the fields of data science and urban planning, those applying and training machine learning algorithms must consider the implications for planning decision support. For damage assessment processes, machine learning may eliminate potential social bias held by assessors or local emergency responders. On the other hand, it does have the potential to direct resources based on inaccurate or misleading machine learning results. While algorithmic



Image 2.4: This home in the Garden District celebrates Mardi Gras in Louisiana after the neighborhood suffered damage from flooding

bias may occur, there are direct steps to take to prevent imprecise damage assessments.

Admittedly, machine learning is not the complete solution for damage assessment. For example, tools such as the Rapid Integrated Damage Assessment model, developed by the National Disaster Preparedness Training Center (NDPTC), must rely on other output steps to accurately identify damage results for improved accuracy. The tool does not rely on just machine learning outputs to assess damage, rather it is the last step to capture damage likelihoods. Therefore, machine learning should supplement damage determinations in the initial parts of early recovery with the understanding that machine learning outputs are not the final determination for damage. One researcher presses that "because AI can cause considerable harm to individuals and groups, it is not sufficient to leave their development and regulation to those without expertise in this area."<sup>19</sup> Outputs of severe or moderate damage should be validated in two ways. The first is auditing the outputs through monitoring results. When testing an image dataset on a model, there are no checks and balances on the system unless built in through human review.<sup>20</sup> Additionally, damage can be validated through a two way feedback. Residents in some areas, such as the St. Charles Parish in Southeastern Louisiana, allows residents to review their households' damage score post-disaster. In one case, a family home that was destroyed was initially determined to be minorly damaged. The family was provided the opportunity to challenge the score, and was justly awarded aid and recovery support. Empowering local voices through a feedback system enhances results through verification or contestation of damage. It also allows emergency managers and planners more outlets to capture data that is overlooked through street-level machine learning, notably damage inside of homes.

Algorithmic bias makes machine learning applications in the real world questionable. Many cases of improper machine learning detection and misclassification have been uncovered through practice.

While feedback systems may be effective for damage validation, there are limitations with this method. First, it assumes people know about the neural network



Image 2.5: Despite the appearance of damage, this local flower shop is back in operation just five months after Hurricane Ida destroyed neighborhood buildings

process, can access its information, and have the time, internet access, and other resources to contest damage scores. Therefore, there should be control measures outside of local feedback. Monitoring matters because in some instances, researchers have observed egregious racial and gender biased outcomes from machine learning.<sup>21</sup> Researchers and local responders must be able to not only conduct machine learning damage assessments, but also to properly manage its outcomes. There are many tools that help understand machine learning bias such as Audit-AI or AI Fairness 360.<sup>22</sup> Altogether, an auditing process is an essential step towards implementation of neural network tools.

It is notable that all on-the-ground assessments lack the ability to understand damage located inside of a structure or home. Natural disasters such as fires or hurricanes present varying damage inside of houses and buildings that may be overlooked by aerial, street-level, or in-person drive-by assessment. Fires can cause soot buildup and unsafe air qualities in and around a home not completely visible in imagery. In hurricane events, flooding infiltrates lower level floors which tend to hold infrastructure systems like HVAC systems, fire protection, electrical networks, and even plumbing systems. Water damage can even impact walls and facade elements of a building through mold or staining. Some damage is life threatening, such as



Image 2.6: A boarded up community recreation center with no apparent damage stands out with colorful murals of the rising seas levels in New Orleans

electrical shut offs or improper heating and cooling. In alleviating the burden to capture, process, and diagnose visible damage, local officials should also develop tools for faster in-unit or in-house damage detection.

There are positive impacts of machine learning if organized and applied correctly. By controlling human bias and relying on data-driven tools, the speed in which data is processed and managed can reduce the local burden placed on emergency management. As the process to diagnose and determine damage rapidly increases, it allows emergency managers the ability to focus their attention on other pertinent needs. Rather than use ground methods to holistically assess damage, technology can reduce personnel needed to capture and process the data. The reduced local burden could benefit a community by heightening the ability to reach more community members, especially the most vulnerable. Rather than wait for residents to access convoluted communication channels to express damage, local officials can begin to assess damage to prioritize areas regardless of perceived demand to recover. Data-driven tools in

this application can reduce mismanagement of aid being directed to the most vocal or most connected networks, steering large organizations and government entities toward the residents that need more assistance and support.

## 2.9 Conclusion

The current damage assessment process is human-reliant and burdensome. Incorporating machine learning into rapid damage assessment can eliminate the potential for human error and prejudice, controlling bias more than ever before. Taking a note from the planning and the data science fields respectively, machine learning can enhance damage assessment processes and applications in early recovery.

In doing so, the combination of the two fields' expertise and knowledge can increase capacity building of local disaster recovery networks. On the ground organizations and local emergency response professionals collect and process data in scattered ways to organize aid, distribute resources, and promote speedy recovery. Due to the nature of highly perishable

data, technology based tools offer a unique solution to these complex problems in both encouraging the sharing and storage of perishable data and reducing problematic assessment practices, all while reducing burdens on local recovery networks.

## ENDNOTES

---

1. Hu, Z., Sheu, J. Y., Wei, C. (n.d.). Post-disaster relief operations considering psychological costs of waiting for evacuation and relief resources. *Transportmetrica* (Abingdon, Oxfordshire, UK). Taylor & Francis.
2. Kleinberg, J., Ludwig, J., & Mullainathan, S. (2016, December 8). A Guide to Solving Social Problems with Machine Learning. *Harvard Business Review*. <https://hbr.org/2016/12/a-guide-to-solving-social-problems-with-machine-learning>
3. Kontokosta, C. E. (2021). Urban informatics in the science and practice of planning. *Journal of Planning Education and Research*, 41(4), 382–395.
4. Ibid.
5. Chen, Y., & Ji, W. (2021). Rapid Damage Assessment Following Natural Disasters through Information Integration. *Natural Hazards Review*, 22(4), 04021043. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000504](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000504)
6. (Nepal, 2015 and Ecuador, 2016)
7. Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage Assessment from Social Media Imagery Data During Disasters. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 569–576. <https://doi.org/10.1145/3110025.3110109>
8. Ibid
9. Ibid
10. Amini, A., Soleimany, A. P., Schwarting, W., Bhatia, S. N., & Rus, D. (2019). Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 289–295. <https://doi.org/10.1145/3306618.3314243>
11. Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage Assessment from Social Media Imagery Data During Disasters. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 569–576. <https://doi.org/10.1145/3110025.3110109>
12. Alelyani, S. (2021). Detection and Evaluation of Machine Learning Bias. *Applied Sciences*, 11(14), 6271. <https://doi.org/10.3390/app11146271>
13. Cherry, K. (2020, July 19). What Is Cognitive Bias? [Psychology]. Very Well Mind. <https://www.verywellmind.com/what-is-a-cognitive-bias-2794963>
14. Elichai, Am. (2018, December 13). How Big Data Can Help in Disaster Response. *Scientific American*. <https://blogs.scientificamerican.com/observations/how-big-data-can-help-in-disaster-response/>
15. Ibid
16. Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage Assessment from Social Media Imagery Data During Disasters. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 569–576. <https://doi.org/10.1145/3110025.3110109>
17. Quarantelli, E. L. (1999). The Disaster Recovery Process: What we Know and Do not Know From Research. 19.
18. Croft, J., Martin, D., Madley-Dowd, P., Strelchuk, D., Davies, J., Heron, J., Teufel, C., & Zammit, S. (2021). Childhood trauma and cognitive biases associated with psychosis: A systematic review and meta-analysis. *PLOS ONE*, 16(2), e0246948. <https://doi.org/10.1371/journal.pone.0246948>
19. White, James M ; Lidskog, Rolf. (2021). Ignorance and the regulation of artificial intelligence. *Journal of Risk Research*. Abingdon: Routledge.
20. Ibid
21. Cerrato, P., Halamka, J., & Pencina, M. (2022). A proposal for developing a platform that evaluates algorithmic equity and accuracy. *BMJ health & care informatics*, 29(1).
22. Ibid



---

## Social Bias: Machine Learning and Early Recovery

Authors: Kiley Fitzgerald, Jeffery Pritchard, and Danielle Stewart

---

### about this project

This project is a joint effort by students and faculty within the Master of Urban and Regional Planning program at the University of Michigan and the National Disaster Preparedness Training Center (NDPTC) as a Capstone project for the Winter 2022 semester.

A key focus of the University of Michigan team is to work in a manner that promotes the values of equity, uplifting local voices, transparency and honesty. As a result, the outcomes of this capstone aim to speak to both our collaborators at the NDPTC and the local communities impacted by disasters across the United States. Our responsibilities as researchers will also include the implementation and/or recommendation of innovative solutions to issues surrounding machine learning, damage assessments, prioritization determinations, and social infrastructure networks.

