Deploying Machine Learning for a Sustainable Future

Cary Coglianese
University of Pennsylvania Carey Law School

Deploying Machine Learning for a Sustainable Future

Cary Coglianese
University of Pennsylvania Law School

Abstract

To meet the environmental challenges of a warming planet and an increasingly complex, high tech-economy, government must become smarter about how it makes policies and deploys its limited resources. It specifically needs to build a robust capacity to analyze large volumes of environmental and economic data by using machine-learning algorithms to improve regulatory oversight, monitoring, and decision-making. Three challenges can be expected to drive the need for algorithmic environmental governance: more problems, less funding, and growing public demands. This paper explains why algorithmic governance will prove pivotal in meeting these challenges, but it also presents four likely obstacles that environmental agencies will need to surmount if they are to take full advantage of big data and predictive analytics. First, agencies must invest in upgrading their information technology infrastructure to take advantage of computational advances. Relatively modest technology investments, if made wisely, could support the use of algorithmic tools that could yield substantial savings in other administrative costs. Second, agencies will need to confront emerging concerns about privacy, fairness, and transparency associated with its reliance on Big Data and algorithmic analyses. Third, government agencies will need to strengthen their human capital so that they have the personnel who understand how to use machine learning responsibly. Finally, to work well, algorithms will need clearly defined objectives. Environmental officials will need to continue to engage with elected officials, members of the public, environmental groups, and industry representatives to forge clarity and consistency over how various risk and regulatory objectives should be specified in machine learning tools. Overall, with thoughtful planning, adequate resources, and responsible management, governments should be able to overcome the obstacles that stand in the way of the use of artificial intelligence to improve environmental sustainability. If policy makers and the public will recognize the need for smarter governance, they can then start to tackle obstacles that stand in its way and better position society for a more sustainable future.

Deploying Machine Learning for a Sustainable Future

Cary Coglianese[*]

In the face of extraordinary environmental challenges created by a warming planet and an increasingly complex, high-tech global economy, government needs to become smarter about how it makes and implements environmental policy. Specifically, government needs to build a robust capacity to analyze large volumes of environmental and economic data using machine-learning algorithms. It needs, in other words, to move toward algorithmic environmental governance.

Businesses have already demonstrated how algorithms can lead to more accurate and better optimized decisions across a wide range of functions, including medical treatments, fraud identification, and self-driving cars.[1] To meet the demands of a sustainable future, government will need to use these same kinds of algorithmic tools for improving environmental management. In the hands of responsible environmental officials, machine-learning algorithms can promote more efficient use of scarce resources and the design of more cost-effective solutions to persistent and new environmental challenges.

What Is Algorithmic Environmental Governance?

An algorithm is simply a series of computational steps. In this most basic sense, algorithms have long helped environmental decision makers. But machine-learning algorithms—sometimes referred to as artificial intelligence or predictive analytics—are different. They take advantage of modern digital computing power to analyze vast quantities of data—Big Data—to produce highly accurate predictions. In contrast to conventional statistical analysis, they work by a process of

---

[*] Edward B. Shils Professor of Law and Political Science, and Director, Penn Program on Regulation, University of Pennsylvania Law School. This paper was prepared as a chapter for Daniel Esty, ed., *A Better Planet: Forty Big Ideas for a Sustainable Future* (Yale University Press).

“learning” on their own. With enough computing power, machine-learning algorithms can do their work at lightning speed.

To appreciate more fully how machine-learning algorithms work, it helps to contrast them with standard statistical techniques for making predictions, such as regression analysis. With conventional techniques, a human analyst selects both the variables to include in a mathematical model and the model’s functional form. Machine-learning algorithms, by contrast, do the selecting of variables and functional forms on their own. Humans establish an objective that a learning algorithm is supposed to meet—namely, what it should predict—and the algorithm essentially takes things from there.[2]

Although machine-learning algorithms can be structured in different ways, the most intuitive way to understand how they work is by visualizing a computational process that rapidly tries out all possible combinations of variables from a large dataset using a host of different functional forms until it finds the best match—that is, the function and variables that yield the most accurate predictions.[3] Machine-learning algorithms “train” on existing data but then are tested and applied with new data.[4] Through this basic process, machine-learning robots help navigate self-driving cars, identify spam in email inboxes, and play difficult games, such as chess and Go.

Government officials are beginning to see the value of machine-learning algorithms.[5] When addressing environmental problems, government leaders must rely on accurate predictions to inform their decisions. They could benefit from the superior predictive power and speed of machine learning. To see how algorithms could improve environmental governance, consider a few examples:

- *Identifying toxic chemicals*. The U.S. Environmental Protection Agency (EPA) faces the daunting challenge of determining which chemicals out of tens of thousands could cause cancer and should be banned. Conducting animal tests or even in vitro analysis on every chemical is simply not feasible. To select which chemicals to study further, EPA and other government agencies have built a massive dataset on toxic chemicals. EPA analysts have shown that they can use machine-learning tools to analyze those data and make predictions about whether any particular new chemical is likely to have toxic effects, saving the agency substantial time and resources while also protecting the public.
- *Targeting facilities for environmental inspections*. In any given year, EPA has the resources to inspect no more than about 10 percent of all facilities in the United States that operate with a

water discharge permit.[6] Machine-learning tools can dramatically increase the efficiency of inspection targeting, enabling regulatory agencies to direct their limited number of inspectors toward facilities more likely to have compliance and environmental problems. After all, sending inspectors to facilities that are faithfully complying with the law is not a smart use of limited inspection resources. Researchers at Stanford have shown that EPA could improve the efficiency of its Clean Water Act inspection targeting by as much as 600 percent with machine-learning algorithms.[7]

- *Predicting areas with climate-related flood risks*. As climate change unfolds, coastal areas face heightened flood risks. Deciding where to undertake climate resilience actions, such as constructing levees or reforming building codes in coastal cities, will be greatly aided by accurate predictions of the areas facing the greatest risk. Cities can make more accurate infrastructure plans and better resource allocation decisions with machine learning.[8]

In these and other ways, algorithmic tools can become an essential component in a policy strategy for a sustainable future. Algorithmic tools not only can help better inform traditional regulatory functions, but they can also go further to support fully automated environmental compliance monitoring systems that integrate remote-sensing technology or infrared cameras to provide real-time information about emissions of pollutants. Over the longer term, real-time monitoring combined with machine-learning analysis could potentially support a type of automated performance-based regulatory system that would afford polluting facilities greater flexibility in the management of their environmental operations.[9]

Why Society Needs Algorithmic Environmental Governance

Algorithmic tools are needed because environmental agencies face increasing demands due to changing technologies and a changing climate. Government will most likely need to meet its additional demands with the same or even fewer resources. By investing in computing technology and the right kind of human analytic capacity to support machine learning, government agencies should be able to save money and improve performance by better allocating scarce human resources and facilitating more flexible and refined environmental policies—even perhaps to the point of regulating by robot.[10]

Three factors can be expected to drive the need for algorithmic environmental governance: more problems, less funding, and growing demands.

More problems. The number and volume of potentially hazardous chemicals and technologies will only continue to grow, at rates beyond environmental regulators' capacities for testing and monitoring all possible risks. The sheer number of pollution sources will also likely expand. For example, although the United States' growing reliance on natural gas for energy will help reduce planet-warming carbon dioxide emissions, it will also bring with it the challenge of preventing fugitive methane emissions—small leaks of an even more potent greenhouse gas from any point in the vast production and distribution chain for natural gas.[11] Similarly, the advent of 3-D printing will usher in an era of distributed manufacturing that will increase the number of smaller polluting sources throughout the country.

These and other technological and economic changes will occur at the same time as climate change continues to wreak havoc on the planet. Society's future will depend on smart climate change mitigation policies—and it will also need smart climate change adaptation decisions. Machine learning can help improve decision-making about infrastructure planning, flood and storm response, public health monitoring, and natural resource and agricultural management.

Less funding. Budgetary resources devoted to environmental protection appear unlikely to increase significantly in the foreseeable future. If at least some governmental enforcement and monitoring functions can be entirely automated by combining algorithmic tools with advances in remote sensing, the cost savings for government could be substantial.[12] According to one estimate, greater reliance on machine-learning forecasting to screen chemicals for toxicity could save close to $1 million per toxic chemical identified.[13]

Growing demands. As the private sector continues to innovate with optimizing algorithms and other technologies, it will likely increase public demands for more precise but flexible environmental policies. Individuals are already growing accustomed in their private lives to the precision that machine-learning algorithms make possible, such as the customized recommendations from companies such as Amazon, Netflix, Google, and Apple. Why not make regulatory obligations customized too? Many business leaders would undoubtedly prefer that government shift away from a reliance on crude, one-size-fits-all rules to more cost-effective regulatory systems that micro-target industrial facilities and impose customized performance targets on each.[14]

Building Capacity for Algorithmic Governance

Making the move to algorithmic environmental governance will not be easy. EPA and other governmental agencies will face four main obstacles if they are to take full advantage of the predictive potential of machine-learning algorithms. These obstacles can be overcome, making algorithmic governance fully realizable, but it will require making deliberate investments and responsible management choices.[15]

First, government must invest in its information infrastructure. Unfortunately, too many government agencies at present are woefully behind the curve when it comes to computing power. According to an analysis by the U.S. Government Accountability Office, three-quarters of current spending by the federal government on information technology goes to supporting "legacy systems"—that is, to "increasingly obsolete" systems that are dependent on "outdated software languages and hardware."[16] Although upgrading information technology obviously will require capital investments, relatively modest technology investments, if made wisely, could support the use of algorithmic tools that could yield substantial savings in other administrative costs.

Of course, the kind of infrastructure needed to support algorithmic environmental governance goes well beyond computing power: it also entails large quantities of data. Fortunately, EPA and various state agencies have undertaken substantial efforts in recent years to transfer many paper-based reporting systems to electronic filing systems, which means that facility-level data can increasingly be archived in digital form.[17] As agencies come to rely more on remote sensing instruments for monitoring pollution, those data could also be fed into digital archives. The government will, of course, need to manage all the information it amasses so that environmental data can be linked with other datasets and analyzed by machine-learning algorithms.[18] In a study conducted at the Penn Program on Regulation, we found that machine learning markedly improved the accuracy of inspection targeting when facilities' records in both EPA and Occupational Health and Safety Administration datasets could be combined with publicly available financial data.

Second, government will need to address growing concerns about privacy, fairness, and transparency associated with its reliance on Big Data and algorithmic analyses.[19] Individuals worry, for example, that seemingly innocuous and totally uninformative bits of data can, with the aid of machine-learning tools, yield remarkably accurate predictions about private aspects of their lives, such as their sexual orientations. Concern also exists that biases already contained in human-generated data—say, racial biases in police arrest records—will become baked into the outputs of algorithmic analyses that rely on those data. Others worry that machine-learning algorithms are insufficiently transparent due to the inherent difficulty in explaining exactly how they achieve their forecasts.

These varied concerns have arisen to date about machine learning in a variety of contexts outside of environmental governance—for example, use by social media companies or criminal courts. Yet government officials can expect similar questions to arise with respect to algorithmic environmental governance, and so they should design and deploy algorithms responsibly to avoid these concerns. Data access and security protocols can help address privacy concerns. Biases can be identified and addressed through an emerging array of statistical techniques.[20] The "black box" nature of machine-learning algorithms should also not prevent governments from providing sufficient transparency.[21] With thoughtful planning and responsible management, governments should be able to address any concerns that arise over the use of machine-learning tools to improve environmental sustainability.

Third, government will need to strengthen its human capital to ensure it has personnel who understand how to use machine learning responsibly. One problem is that the federal government is already facing a significant shortfall of talent, with more than a third of federal employees eligible to retire by 2020.[22] At EPA, a quarter of the workforce is currently eligible for retirement.[23] These demographic trends are creating major challenges for government agencies, providing yet another reason why these agencies should take advantage of algorithmic tools and Big Data in the future. It will allow them to do more with less.

The demographic shift occurring in the government's workforce provides an excellent opportunity to rebuild the government in an even more analytically sophisticated way. In the coming years, environmental agencies can bring on board new professionals with the skills or aptitudes to use machine-learning tools. Training government staff in quantitative analytic tools will need to become a priority too. Although algorithms can make possible considerable efficiencies in governmental policymaking and oversight, the responsible design and use of algorithms will depend on more than just technology. People and their effective management will still matter.[24]

Finally, to work well, algorithms will need clearly defined objectives. In environmental policymaking, certain questions about risk management—such as how safe is "safe enough"—remain only loosely defined. But if environmental officials seek to use machine-learning algorithms to optimize certain kinds of risks, they will need to define those risks with clarity and precision. They will also likely need to define how algorithms should make trade-offs between forecasting accuracy and other values, such as fairness.[25] Toward this end, environmental officials will need to continue to engage with elected officials, members of the public, environmental groups, and industry representatives to forge clarity and consistency over how various risk and regulatory objectives should be specified. At the same time that government officials will need to strengthen their analytic and technological skills, they will continue to need to strive for excellence in social engagement.[26]

The Algorithmic Imperative

Although the obstacles in the way of algorithmic governance are not trivial, they can be overcome with sufficient planning and action. The time to take this action is now. Algorithmic environmental governance offers no panacea, but it does promise to support a strikingly more accurate and efficient environmental stewardship. The need for smarter governance, driven by more complex problems, increased public demands, and perennially scarce resources, will make it imperative that environmental agencies rely more on machine-learning tools in the coming years. If policy makers and the public recognize the need for smarter governance now, they can then start to tackle obstacles that stand in their way and better position society for a more sustainable future.

Acknowledgements
The author gratefully acknowledges helpful comments on an earlier draft from Richard Berk, Dan Esty, Chris Lewis, and Roshie Xing, as well as from participants at a Vanderbilt University Law School workshop.

Notes

[1] Darrell M. West and John R. Allen, "How Artificial Intelligence Is Transforming the World," Brookings Institution (April 24, 2018), https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/

[2] Machine-learning algorithms are designed to yield predictions but not support causal explanations. For an excellent overview of how these algorithms work, see David Lehr and Paul Ohm, "Playing with the Data: What Legal Scholars Should Learn about Machine Learning," *University of California Davis Law Review* 51, no. 2 (2017): 653–717.

[3] For a more technical treatment of a common type of machine-learning algorithm, see Leo Breiman, "Random Forests," *Machine Learning* 45, no. 1 (2001): 5–32.

[4] Lehr and Ohm, "Playing with the Data."

[5] Susan Athey, "Beyond Prediction: Using Big Data for Policy Problems," *Science* 355, no. 6324 (February 3, 2017): 483–85.

[6] Miyuki Hino, Elinor Benami, and Nina Brooks, "Machine Learning for Environmental Monitoring," *Nature Sustainability* 1 (October 2018): 583–88.

[7] Hino, Benami, and Brooks, "Machine Learning." Additional research has shown that machine learning can improve the detection of noncompliance with natural resource regulations. Cleridy E. Lennert-Cody and Richard A. Berk, "Statistical Learning Procedures for Monitoring Regulatory Compliance: An Application to Fisheries Data," *Journal of the Royal Statistical Society* Series A 170, no. 3 (2007): 671–689.

[8] Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau, "Flood Prediction Using Machine Learning Models: Literature Review," *Water* 10, no. 11 (October 2018): 1536–76.

[9] Of course, performance-based regulation needs to be well designed and adequately monitored to prevent gaming by regulated entities. Cary Coglianese and Jennifer Nash, "The Law of the Test: Performance-Based Regulation and Diesel Emissions Control," *Yale Journal on Regulation* 34, no. 1 (2017): 33–90.

[10] Cary Coglianese and David Lehr, "Regulating by Robot: Administrative Decision-Making in the Machine Learning Era," *Georgetown Law Journal* 105, no. 5 (2017): 1147–223.

[11] James Bradbury et al., "Clearing the Air: Reducing Upstream Greenhouse Gas Emissions from U.S. Natural Gas Systems," working paper, World Resources Institute (2013), https://wriorg.s3.amazonaws.com/s3fs-public/clearing_the_air_full_version.pdf. On the role of remote sensing and machine learning for methane leak detection, see Levente Klein et al., "Distributed Wireless Sensing for Methane Leak Detection Technology" (2017), https://www.osti.gov/servlets/purl/1435713.

[12] Cynthia Giles, "Next Generation Compliance," *Environmental Forum* (September–October 2013): 22–26.

[13] Matthew T. Martin et al., "Economic Benefits of Using Adaptive Predictive Models of Reproductive Toxicity in the Context of a Tiered Testing Program," *Systems Biology in Reproductive Medicine* 58, no. 3 (2012): 4–6.

[14] For a related discussion, see Anthony Casey and Anthony Niblett, "Self-Driving Laws," *University of Toronto Law Journal* 66, no. 4 (2016): 429–42.

[15] Cary Coglianese, "Optimizing Regulation for an Optimizing Economy," *University of Pennsylvania Journal of Law and Public Affairs* 4, no. 1 (2018): 1–13.

[16] U.S. Government Accountability Office, Federal Agencies Need to Address Aging Legacy Systems: Hearing Before the H. Comm. on Oversight and Gov't Reform, 114th Cong. (2016) (statement of David A. Powner, Director, Information Technology Management Issues), https://www.gao.gov/assets/680/677454.pdf.

[17] Giles, "Next Generation Compliance."

[18] Robert L. Glicksman, David L. Markell, and Claire Monteleoni, "Technological Innovation, Data Analytics, and Environmental Enforcement," *Ecology Law Quarterly* 44, no. 1 (2017): 41–88.

[19] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

[20] The statistical remedies will vary depending on whether bias exists in the data or in the algorithm.

[21] Cary Coglianese and David Lehr, "Transparency and Algorithmic Governance," *Administrative Law Review* 71, no. 1 (2019): 1–56.

[22] U.S. Government Accountability Office, Federal Workforce: Sustained Attention to Human Capital Leading Practices Can Help Improve Agency Performance: Hearing Before the H. Comm.

on Oversight and Gov't Reform, 115th Cong. 3-4 (2017) (statement of Robert Goldenkoff, Director of Strategic Issues), https://www.gao.gov/assets/690/684709.pdf.

[23] Brady Dennis, Juliet Eilperin, and Andrew Ba Tran, "With a Shrinking EPA, Trump Delivers on His Promise to Cut Government," *Washington Post* (September 8, 2018), https://www.washingtonpost.com/national/health-science/with-a-shrinking-epa-trump-delivers-on-his-promise-to-cut-government/2018/09/08/6b058f9e-b143-11e8-a20b-5f4f84429666_story.html.

[24] Daniel C. Esty, "Regulatory Excellence: Lessons from Theory and Practice," in Cary Coglianese, ed., *Achieving Regulatory Excellence* (Washington, DC: Brookings, 2017): 133–47.

[25] Richard Berk, "Accuracy and Fairness for Juvenile Justice Risk Assessments," *Journal of Empirical Legal Studies* 16, no. 1 (2019): 175–94.

[26] Cary Coglianese, "Regulatory Excellence as 'People Excellence,'" *The Regulatory Review* (October 23, 2015), https://www.theregreview.org/2015/10/23/coglianese-people-excellence/.