OCTOBER 2021

AI Strategy

The New York City Artificial Intelligence Strategy

Mayor's Office of the Chief Technology Officer

nyc.gov/cto

First published

October 13th, 2021

Version

File version 1.9.1014-12

Type

Typeset in *Libre Baskerville* and *Public Sans*, with accents of *Space Mono*.

Online

https://on.nyc.gov/ai-strategy

Accessible HTML version:

https://www1.nyc.gov/assets/cto/#/publication/ai-strategy

Acknowledgments

The Mayor's Office of the Chief Technology Officer (NYC CTO) would like to thank the varied internal and external stakeholders who shared their expertise in the development of this document. A full list of organizations, institutions, offices, and agencies consulted is listed in *Supplement B*.

In addition, we would like to acknowledge those who contributed to or reviewed all or part of the text in detail: Dr. Meg Young, Dr. Eric Corbett, Dr. Hanna Wallach, Stefaan Verhulst, Dr. S. Craig Watkins, Dr. Brandie Nonnecke, and Dr. Chris Riley. This work was significantly improved by their insights.

The NYC AI Strategy builds upon prior work on AI in NYC by the New York City Algorithms Management and Policy Officer and the New York City Economic Development Corporation, as well as related work by the Mayor's Office of Data Analytics, the Mayor's Office of Economic Opportunity, and the Chief Privacy Officer and Mayor's Office of Information Privacy, and we would like to acknowledge and highlight their contributions.

Note

The NYC AI Strategy is subject to applicable laws, rules, and regulations, including City procurement rules and processes. The City reserves all rights, including rights to postpone, cancel, or amend the AI Strategy at any time. The City shall not be liable for any costs incurred in connection with the AI Strategy.

Message from the CTO

Let's talk about AI.

New technology often generates excitement, inspires hope, and even ignites fear. We imagine the myriad ways tech can change life for the better. At the same time, we understandably worry about how it might be abused.

These responses to AI, in particular, make sense. Artificial intelligence and machine learning are rapidly assuming integral roles in everyday life, but the general public and some key decision makers do not yet understand them well. That's why it is necessary to broadly deliver a base level of understanding about this revolutionary technology. The NYC AI Primer aims to do just that.

We must also implement approaches, tools, collaborations, and governance to ensure that the use of this technology is appropriate. In the age of AI, digital rights are human rights.

In the coming years, governments, private companies, academic institutions, and others will make choices about AI that reverberate far into the future. AI's impact in the 21st century promises to be akin to that of the internet in the 20th century and electricity in the 19th.

As a global epicenter of innovation and home to nearly nine million people, New York City has a key role to play in shaping this future. Through the NYC AI Strategy, we are laying out the next steps needed to make the most of artificial intelligence, to protect people from harm, and to build a better society for all.

We hope you'll join us.

John Paul Farmer

John Paul Farmer

Chief Technology Officer,
The City of New York

Table of Contents

Executive Summary

It is not hyperbole to say that AI will change the world. In fact, artificial intelligence (AI) is changing the human experience today, driving sweeping social, economic, and technological transformation that affects us all. An umbrella term encompassing a range of technologies from sophisticated to simple, AI is already being applied to make predictions, inferences, recommendations, and decisions with data. It is used in a wide array of domains, including health and medicine, transportation, finance, law enforcement, and social services; there are also various types of applications, from automating administrative tasks to enabling computers to process human language, understand images and video, or recommend products or content online. AI is also used in many key societal functions today, such as informing delivery or inspection routes, hunting for evidence of disease, and determining eligibility for loans or benefits. Like software itself, AI will touch virtually every area of life in the years ahead, including everything from basic scientific research to the operation of various products and services, and its impacts will be felt from a personal to a societal scale. For these reasons, the City of New York believes that an ecosystem approach grounded in digital rights is necessary to maximize benefits, minimize harms, and ensure the responsible application of AI.

New York City has distinctive characteristics that shape and should continue to guide the local AI ecosystem, each grounded in the sheer scale and diversity of the local economy and its participants. The Big Apple is a global capital for finance, fashion, media, and many other industries that increasingly use AI; a hub for technology startups, investors, and major technology companies; and a leader in AI education, research, and development, as well as the study of the intersection of technology and society. Underlying this institutional and intellectual diversity, New York City is home to one of the largest and most diverse populations in the world. The thoughtful use of AI can be a powerful tool to adapt services and infrastructure to better serve the public in terms of economic opportunity, social equity, environmental sustainability, and more. AI presents new opportunities for the city's businesses and new job prospects for residents, and it can offer the potential for a better quality of life for all.

In addition to its promise, the perils of AI have come into sharper focus in recent years. There are risks to using AI in inappropriate ways, as discussed in detail in *Supplement A: The NYC Artificial Intelligence Primer*. Using AI is often more than a technical endeavor; especially in societal and government domains, it frequently raises social, political, economic, ethical, and policy questions and can pose serious risk of harm to individuals or communities, sometimes involving explicit, outright discrimination. Further, AI can present risk to New Yorkers' work and job quality, and security, as a growing number of processes, tasks or job functions are being automated, changing the landscape of work and the skills needed for employment and career growth. Navigating these challenges will require careful thought and planning, including robust engagement with the community.

To inform this Strategy, over fifty City agencies and external organizations were interviewed or consulted. Selected findings from these conversations are organized into five thematic areas that warrant increased City attention: data infrastructure, applications of AI, policy and governance, cross-sector partnerships, and business, education, and the workforce. Opportunities identified in these areas include: establishing a Citywide data strategy; building on existing governance structures to encourage responsible and effective use; building capacity and expertise within government; identifying and pursuing ways in which AI can be used for social benefit; encouraging and facilitating external partnerships with academia, industrial research labs, businesses, non-profits, and communities; ensuring private sector use of AI technology respects the rights of New Yorkers; and supporting equitable access to opportunity.

New York City's approach to questions at the intersection of technology and society is grounded in the framework of digital rights.

This Strategy emphasizes the need to consider the city's AI ecosystem as a whole. This approach aims to account for the full range of stakeholders working in, training for, building, buying, using, governing, and impacted by AI, and to characterize a healthy ecosystem as one in which a broad mix of individuals and organizations across sectors and communities work in concert to promote well-being, fairness, and opportunity for all. It is important to consider how AI relates to government decision-makers and technical leaders; external organizations from technology startups and incumbents to hospitals, schools, universities, non-profit advocacy and community organizations; and the general public. It is the City's job to help cultivate this ecosystem to work towards shared goals of public good.

Like software itself, AI will touch virtually every area of life in the years ahead.

New York City's approach to questions at the intersection of technology and society is grounded in the framework of "digital rights," which include privacy, accountability, trust, transparency, and fairness and non-discrimination, among others. Although there is broad agreement on these abstract principles, there are two key challenges that society and governments must now face: first, determining what these principles mean in real-world situations and how they can be implemented in concrete projects, and second, the reality that these rights are often in tension with each other. Navigating these trade-offs presents important ethical questions that go beyond technology.

To maximize the net positive impact for New Yorkers, it is necessary to ensure that public- and private-sector decision-makers and the general public are literate in what AI actually is and what it is not, as well as in key considerations around its use. A foundational understanding of how AI works is critical to developing better policy, recognizing opportunities and risks, and evaluating claims made by others. This grounding is a precursor to productive discussions about AI, and this Strategy devotes significant effort towards this goal in *Supplement A: The NYC AI Primer*. This baseline understanding facilitates assessments of AI's realistic potential for impact, both positive and negative. The City will focus on meeting the need for such broad literacy, both directly and through cross-sector collaboration.

This Strategy aims to lay a foundation for a healthy AI ecosystem in New York City. AI is a complex, far-reaching, and evolving subject, and there is much work to do. This initial effort prioritizes establishing a baseline of information about AI to help local and regional decision-makers work from an accurate and shared understanding of the technology and the issues it presents; outlining key components and characteristics of New York City's AI ecosystem today; and framing a set of areas of opportunity for City action. This Strategy is a first step, not a comprehensive answer, and the ideas here are structured to be adapted and expanded upon going forward. Much is still unknown about AI, and both society and the field are still in the early stages of this story.

New York City is a gateway to America and the ideals it represents — a symbol of opportunity to people across the country and around the world. In this spirit, the City must proactively participate in and respond to technology-driven change in a balanced manner. The City must recognize the potential in these technologies and the momentum behind their ongoing development and use, and it must work to ensure AI use across society upholds and strengthens human rights, civil liberties, and democratic principles; offers equitable opportunity; and is governed appropriately.

In approaching the topic of AI, New York City must be measured, responsive, informed, and humble. Through such an approach, this Strategy aims to develop an AI ecosystem that is good for New York City, for society, and for humanity.

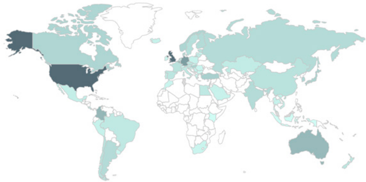
Introduction

Artificial Intelligence (AI) is driving transformation across all areas of society today. An umbrella term encompassing a range of technologies both sophisticated and simple that are used to, among other things, make predictions, inferences, recommendations, or decisions with data, AI is used in many products and services that people use, interact with, or are impacted by every day. It is used to help diagnose medical conditions and care for patients, to predict demand for transportation and run ride hailing services, to optimize energy consumption of devices or large systems and perform energy management, to detect fraudulent financial transactions and manage financial services, and to recommend products or content and operate consumer internet and e-commerce platforms. Like software itself, AI will touch virtually every area of life in the years ahead, in ways both visible and invisible, and its impacts will be felt from a personal to a societal scale.

AI offers myriad opportunities for New York City. The thoughtful use of AI can be a powerful tool to adapt existing City services and infrastructure to better serve the public in terms of economic opportunity, social equity, sustainability, and more. AI presents new opportunities for New York City's businesses and new job prospects for residents, and it can offer the possibility of a greater quality of life for all. But there are also significant risks associated with AI. In recent years, the promise and perils of AI have come to greater public awareness in the context of a broadening conversation about wider societal changes and inequities. There are risks to using AI in inappropriate ways, either in situations where it may be inherently ill-suited or where its implementation is not rigorous and careful. Using AI is often not solely a technical endeavor; it can raise social, political, economic, ethical, and policy questions and can pose serious risk of harm to individuals or communities, sometimes involving explicit, outright discrimination by race, gender, disability, age, language, or other personal attributes. Further, AI can present risk to New Yorkers' work and job security, as a growing number of tasks or job functions are being automated, changing the landscape of work and the skills needed for employment and growth. Navigating these challenges will require careful thought and planning, including robust engagement with impacted communities.

AI has become an active and major focus area at both the national and international levels. Dozens of countries around the world now have national AI strategies, including the UK, France, Germany, Finland, Scotland, China, Japan, and India. The United States also has a national strategy, including a National AI Initiative housed at the recently launched AI.gov. The National AI Initiative Act of 2020 became law on January 1, 2021, providing for a *coordinated program across the entire federal government* to "accelerate AI research and application for the Nation's economic prosperity and national security."[1] The efforts also include the establishment of a National AI Initiative Office in the White House Office of Science and Technology Policy; a National AI Advisory Committee and Research Resource Task Force; the creation of a network of National AI Research Institutes; expanding the mission of the National Institute of Standards and Technology to include advancing standards and guidelines for AI; and a directive for the National Science Foundation to produce a National Academies AI Impact Study on Workforce, among others.[2] The City must engage in a similarly comprehensive manner.

New York City is recognized the world over as an epicenter of innovation and progress. Given this leadership position, it is important for the City to proactively participate in and respond to the opportunities and challenges AI presents from a balanced stance. The City recognizes the realistic potential of AI and the momentum behind its ongoing development and use, and it must work to ensure that as these technologies are used increasingly across society they uphold and strengthen human rights, civil liberties, and democratic principles; offer equitable opportunity; and are governed by informed, interdisciplinary collaboration.



The OECD, an intergovernmental organization with 38 member countries, has devoted significant energy to AI, including, shown here, mapping hundreds of AI policy initiatives in over 60 countries and territories, as well as the EU.
Source: OECD

1 See https://www.ai.gov/.

2 Ibid. For a summary of these and other initiatives, see "*Summary of AI Provisions from the National Defense Authorization Act 2021*," Stanford Institute for Human-Centered AI, 2021, available at https://hai.stanford.edu/policy/policy-resources/summary-ai-provisions-national-defense-authorization-act-2021.

The broad aim of this Strategy is to lay a foundation for a healthy cross-sector AI ecosystem in New York City. This is a complex, far-reaching, and evolving subject, and there is much work to do. This initial effort prioritizes establishing a baseline of information about AI to help ensure local and regional decision-makers are working from an accurate and shared understanding of the technology and the issues it presents; outlining key components and characteristics of New York City's AI ecosystem today; and framing a set of areas of opportunity for City action. This Strategy is a first step, and the ideas here are structured to be expanded and adapted going forward.

Scope

Several important AI topics are out of scope for this Strategy. These include policy areas that are outside the purview of local government, such as defense and military applications, international policy issues, and policy related to large consumer internet platforms.[3] Also out of scope are physical robots, a complex subject in its own right that cannot be fully addressed here, and "artificial general intelligence" or "strong AI," a notion of AI that attempts to fully emulate a person, which is still theoretical, though widely depicted in science fiction.[4]

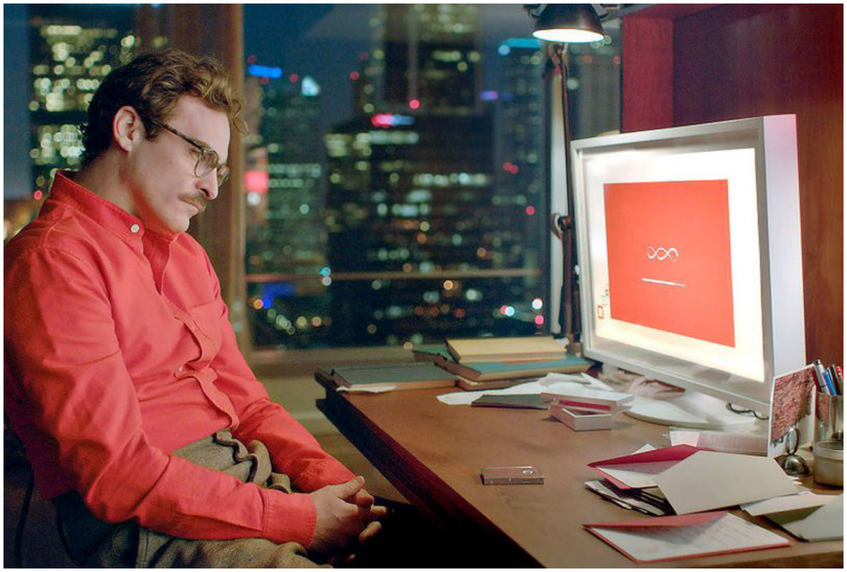
This Strategy does discuss data more generally, as the issue is deeply intertwined with AI. This includes "data science," a topic related to but distinct from AI, although the title "data scientist" can refer to a wide range of responsibilities in either area. Data science broadly refers to extracting and using insights from data, even including simple methods like calculating basic statistics (e.g., averages, percentiles) or using dashboards. It places more emphasis on one-off analyses and on data visualization (e.g., with charts) rather than building systems, though it can also include sophisticated statistical analysis and techniques that are also considered to be part of AI.[5]

3 There are locally important issues that involve online platforms, such as misinformation and disinformation about public health, elections, and other topics, but discussion is omitted here as the policy debate about these matters is currently more active at the federal or international level.

4 M. Jordan, "*Artificial intelligence — the revolution hasn't happened yet*," Harvard Data Science Review, 2019.

5 On data science, see, e.g., D. Donoho, "*50 Years of Data Science*," Journal of Computational and Graphical Statistics, 2017, available at https://doi.org/10.1080/10618600.2017.1384734; on data visualization, see E. Tufte, *The Visual Display of Quantitative Information*, second edition, Graphics Press, 2001.

Tension between workers and machines or technology has had a long history. Charlie Chaplin's silent comedy *Modern Times* (1936) focused on people's struggles and desperate conditions in the Great Depression, which in Chaplin's view were due in large part to industrial efficiency and automation. Here, an iconic image of Chaplin's character being swallowed by a machine he works on.
Source: Janus Films



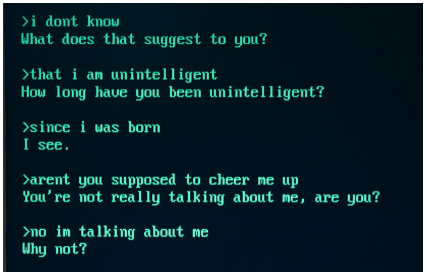
AI has long been a potent topic in literature and science fiction, with roots tracing to, among others, Shelley's *Frankenstein* (1818) and, later, Asimov's *Robot* series (1940s). Here, Joaquin Phoenix in *Her* (2013), a film using a human character's relationship with an AI virtual assistant (voiced by Scarlett Johansson) to examine themes about human relationships and technology.
Source: Warner Bros

What is Artificial Intelligence?

It is important to have a foundational understanding of how AI works to facilitate better policy, recognize both opportunities and risks, and evaluate claims made by others. One of the chief difficulties in this topic is that claims — both positive and negative — are often exaggerated to the point of being misleading, and "AI" is often used more as a marketing term than a precise description of the techniques used. This section provides a brief overview; *Supplement A: The NYC AI Primer* goes in depth and includes concrete, practical examples, along with a discussion of ethical and policy considerations that arise in AI, and serves as a foundation for the rest of the document and its approach.

The term "artificial intelligence" was first coined in the 1950s to describe efforts by computer scientists to produce general human intelligence and behavior in computers; these early efforts to create AI systems were largely "rule-based" to attempt to simulate human reasoning.[6] One famous early example was ELIZA, a system that used simple rules and pattern matching to simulate conversing with a psychotherapist.[7] "AI" now largely refers to the use of an approach called *machine learning* (ML), a way to write "software by example" by providing the computer with illustrative examples to "learn" from. It is used in cases where it is too difficult to specify exact rules for carrying out a task, but where it is possible to collect examples of the task being performed. These tasks are often narrow and domain-specific, as outlined above, and often unrelated to simulating human intelligence or behavior.

In traditional (non-AI) software, developers tell a computer exactly how to carry out a given task using precise, fixed instructions. This sequence of instructions is called an *algorithm*.[8] This approach works well for tasks like sorting a list of names or typesetting a book, but does not work well for problems like differentiating between photos of dogs and cats, reading the handwritten address on an envelope, or identifying fraudulent credit card transactions. Intuitively, there is far too much variation in these cases to handle



ELIZA, a 1966 system that used simple rules and pattern matching to simulate conversing with a psychotherapist.
Photo: Marcin Wichary, Creative Commons BY 2.0

6 For some standard references on AI, see: N. Nilsson, *The Quest for Artificial Intelligence*, Cambridge University Press, 2009; S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, fourth edition, Pearson, 2020. For some early historical references (the latter of which coined the term 'artificial intelligence'), see: A. Turing, "*Computing machinery and intelligence*," Mind, 1950; J. McCarthy, M. Minsky, N. Rochester, and C. Shannon, "*A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*," 1955, available at http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. References on machine learning are cited in *Supplement A*.

7 J. Weizenbaum, "*ELIZA — A Computer Program for the Study of Natural Language Communication Between Man and Machine*," Communications of the ACM, 1966.

8 See, for example: T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, third edition, MIT Press, 2009; T. Roughgarden, *Algorithms Illuminated*, Soundlikeyourself Publishing, 2020; D. E. Knuth, *The Art of Computer Programming*, 2011, series under ongoing development at https://www-cs-faculty.stanford.edu/~knuth/taocp.html.

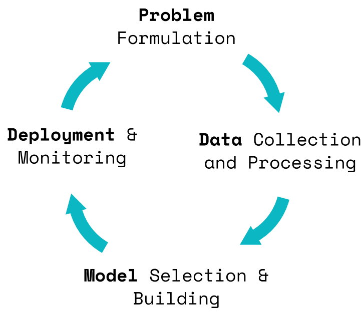
Training a computer to categorize images is a standard example, called image classification, that illustrates many general aspects of ML and computer vision. Here, it is formulated as binary classification (categorizing inputs, in this case images, into one of two categories, in this case cats and dogs), and the model's outputs (shown in the boxes) are probabilities of which category seems correct. A cutoff or confidence threshold must be decided on to turn probabilities into predicted classifications, with the choice of cutoff determining overall system performance. The cactus highlights one subtle danger — since the model does not actually "know" what it is "looking" at, and any image can be provided as input, the model is then forced to pick between the two specified categories even though neither one applies. Though seemingly obvious, such issues often show up in more serious situations, including those that need to assign people to a fixed set of demographic categories.
Source: Illustration by NYC CTO; photos sourced from Unsplash and used with permission.

with explicit rules, even though some of these tasks are easy for humans.

Machine learning, in contrast, uses data together with certain mathematical techniques to create computer programs.[9] These techniques largely come from the fields of statistics, probability, mathematical optimization, and computer science, though increasingly tools from economics, the social sciences, and other areas in applied mathematics are used as well.[10] The computer is given a description of the task to be performed; data in the form of examples of what the correct results look like; a mathematical way of formalizing or expressing assumptions about how the data relates to the task called a "model"; and a learning algorithm indicating how to improve at the task by trial-and-error. The result is a "trained model" which can take new inputs for which the correct output or result is not known and guess the correct output. This guessed output is called a "prediction."

9 This document focuses on a particular form of machine learning called "supervised learning"; there are also other areas in machine learning, including "unsupervised learning," "reinforcement learning," and more.

10 Because machine learning draws on so many other fields, there are often multiple pieces of jargon referring to the same concepts, depending on the academic training of the person speaking or even the venue in which an academic publication appears. In addition, many pieces of technical jargon from statistics and machine learning also have colloquial uses that can cause confusion, including "discrimination," "bias," "prediction," and others.



In particular, ML is about *generalizing* from particular examples provided as data, so this "training data" clearly needs to be chosen to facilitate this generalization. Similarly, the generalization (learning) process must strike a careful balance between picking up enough patterns in the data that are generally relevant while not paying too much attention to unrepresentative quirks or outliers that happen to be in the data. Put differently, a trade-off must be struck between effectively replicating the past and generalizing effectively to the unknown future. This trade-off is partly determined by the metric the AI system is told to optimize for. If this metric is chosen carelessly, the system can end up, for example, overlooking or underemphasizing certain people or groups because they are more difficult or expensive to deal with, which can lead to unfair outcomes.

As machine learning is increasingly widely used, there are now specialized and high-level tools that make trained models for particular tasks available to engineers without any machine learning expertise. For example, to detect faces in photos, there is no need to collect data and build a whole model from scratch; in addition to a significant variety of open source software, there are commercial cloud services and tools based on AI that can carry out this task for a fee.[11] In other words, applying AI to a task today may not be a particularly difficult process requiring significant ML modeling and implementation expertise.

Despite this, it is extremely important for decision-makers to have a working understanding of the details of how AI works, as laid out in *Supplement A*. In particular, the Supplement walks through the lifecycle of an AI project using the running example of (algorithmic) mortgage lending, and highlights examples of what is involved and what can go wrong. In addition, it provides a detailed discussion of key ethical considerations and describes how to think about issues like accountability, fairness and non-discrimination, community engagement, and ethical trade-offs. A rigorous engagement with these details is central to ensuring any use of AI is both effective and responsible.

The reader may wonder how AI is relevant, in practice, to their specific role. For better or worse, AI will increasingly arise in a multifaceted fashion for local stakeholders. For example, in the context of City government, an agency may need to consider AI from several different angles: in policymaking about how AI applications do or will impact New Yorkers in their domain; in informing or educating consumer or workers about those impacts; in building and deploying AI to streamline agency operations to be more responsive; in evaluating vendors for technology procurements, among others. This complex mix of policy-making, public engagement and education, practical deployment, procurement, and in-house development and application can arise in many agencies, from those that deal with transportation to those that deal with public health, businesses, social services, or infrastructure.

For example, at the Taxi and Limousine Commission (TLC), which may at first glance seem relatively traditional, AI potentially arises in policymaking about Taxi companies, and in building and deploying AI education and public engagement campaigns for TLC-Licensed Drivers as they interact with AI systems while

11 A broad suite of AI services are available from the large cloud platforms: Google Cloud Platform, Amazon Web Services, and Microsoft Azure. These services can also be at different levels of abstraction. For example, one can manage everything manually, one can use "fully managed" AI services that allow training and building custom models but handle other aspects, or one can use services that carry out specific high-level tasks like labeling objects in images or identifying celebrities and can be used by software engineers with no understanding of the services' internal workings. There are also many other such services from startups and other technology companies that offer specialized AI-powered products in areas like sales and marketing, advertising, anti-money laundering, computer vision, process automation, and many others.

performing their duties. As another example, NYC Cyber Command is engaged in City policymaking but also both builds and procures AI systems.

In addition, when considering practical use, even if there are commercial systems available or a third-party vendor can be hired, it is critical to have sufficient understanding to select the right product or vendor, to ask the right questions, to tell whether the specific use case is appropriate, and to think through the ethical, policy, and governance issues that arise.[12] The section on AI in City Government below provides an overview of City agency relationships to AI today.

12 For a guide to creating or evaluating machine learning systems in public agencies, see J. Kleinberg, J. Ludwig, and S. Mullainathan, "*A guide to solving social problems with machine learning*," Harvard Business Review, 2016.

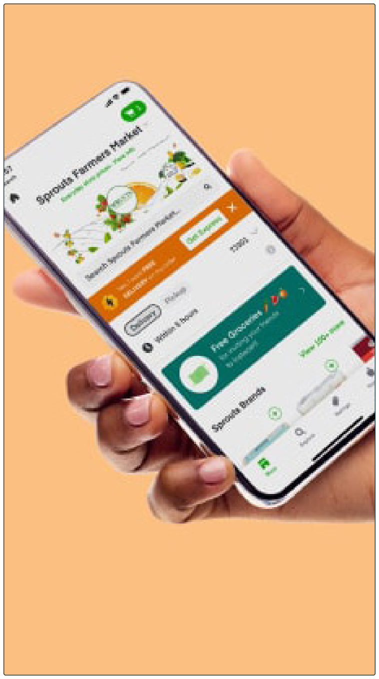
Where is AI used today?

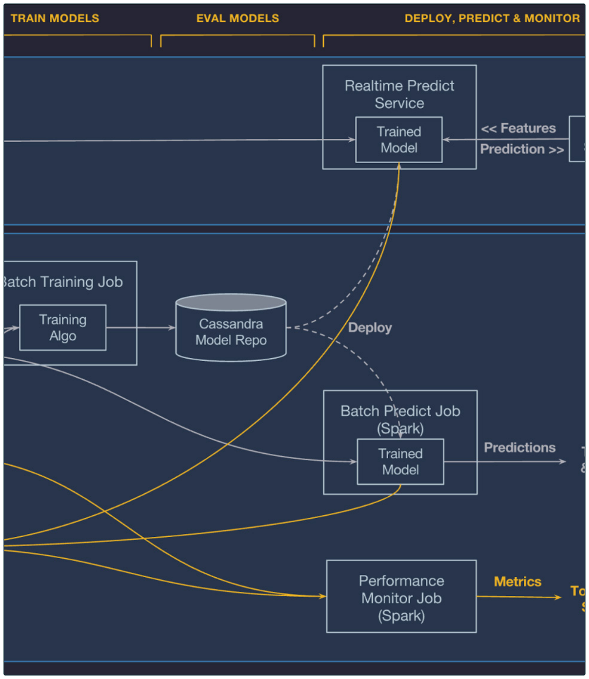
To understand the scope of AI and how interrelated its potential and risks can be, it is helpful to have in mind a diverse set of examples of AI applications in both the public and private sectors. Note that the descriptions here are intended to be illustrative and do not evaluate or discuss the impacts of these uses.

Across sectors, AI systems are used to automate or streamline internal processes, such as classifying customer support requests to route them to the correct department, filling out or parsing forms, detecting anomalous behavior like fraud or cyber intrusions, or screening, hiring, or evaluating employees.

In transportation, rideshare companies like Lyft and Uber use AI to predict demand, dynamically adjust pricing, and dispatch drivers; self-driving cars use a broad range of AI systems to operate.[13] In education, AI can be used to detect plagiarism, personalize lessons for students, or monitor engagement with online learning. AI is used in housing and home lending, from biometric or facial recognition systems used to secure buildings to systems that determine whether or with what terms mortgages should be approved. More generally, AI is used in a wide range of functions in financial services, for both businesses and consumers, such as detecting credit

13 See, e.g., "*Lyft Engineering Blog – Machine Learning*," available at https://eng.lyft.com/tagged/machine-learning; "*Uber AI*," available at https://www.uber.com/us/en/uberai/; J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "*Towards fully autonomous driving: Systems and algorithms*," IEEE Intelligent Vehicles Symposium, 2011.





Consumers increasingly use AI powered applications to request cars, order food, and more. *Photo: Instacart*

Here, a partial system diagram of an internal platform used to manage the full ML lifecycle: manage data; train, evaluate, and deploy models; make predictions; and monitor predictions (in this case, for ride hailing services). The lifecycle is discussed further in *Supplement A*. *Source: Uber Engineering*

card fraud or predicting market trends at hedge funds or other institutions.[14]

Health and medicine is one of the most significant areas for AI, as it is used to help design drugs, model disease, predict how proteins will fold, identify tumors in medical images, and help discover vaccines. It is also used to assess health insurance claims and to help screen patients into different types of care programs.

In government, AI is also used in social services, law enforcement, and criminal justice. For example, in New York City, the criminal courts system uses an AI-based tool to provide a release assessment determining which individuals can be released before trial; the police department uses facial recognition technology to identify suspects and other systems to detect patterns in crime; and other social service agencies use systems to prioritize cases for human review.[15] More generally, AI has also been applied to many problems in the public interest,[16] such as reducing the risk of lead poi-

14 For example, see M. Lopez de Prado, *Advances in Financial Machine Learning*, Wiley, 2018.

15 See Luminosity and the University of Chicago's Crime Lab New York, "*Updating the NYC Criminal Justice Agency Release Assessment Final Report*," 2020, available at https://www.nycja.org/publications/updating-the-new-york-city-criminal-justice-agency-release-assessment, and NYC Algorithms Management and Policy Officer, "*Summary of Agency Compliance Reporting*," 2020, available at https://www1.nyc.gov/assets/ampo/downloads/pdf/AMPO-CY-2020-Agency-Compliance-Reporting.pdf.

16 Some examples selected from R. Ghani, "*Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services*," Testimony to Artificial Intelligence Task Force, Committee on Financial Services, U.S. House of Representatives, 2020.

DeOldify, an open source computer vision model for automatically colorizing and restoring old images and film; here, a photo from 1911. For plausible colors, the model must distinguish between hair, clothes, plants, and so on, and some details like colors of clothes are necessarily just guesses.
Source: Jason Antic github.com/jantic/DeOldify

soning,[17] improving educational outcomes for students at risk of not graduating from school on time,[18] improving health and safety conditions in rental housing,[19] and identifying and diagnosing errors in human physicians' decisions about testing patients for heart conditions.[20]

Some of the most visible applications are in the areas of processing human language (for search engines, automated machine translation, or chatbots) or images (for image search or facial recognition) as well as to recommend items in online stores or content platforms. These general application areas of AI are referred to as natural language processing or NLP (AI applied to language), computer vision or CV (AI applied to images or video), and recommender systems, respectively.[21]

From these examples, it should be clear that AI's applications are hugely diverse and cannot be summarily classified as entirely positive or entirely negative. To think through the implications, nuance is needed; *Supplement A* provides a detailed discussion of how to think through such examples.

17 E. Potash, J. Brew, A. Loewi, S. Majumdar, A. Reece, J. Walsh, E. Rozier, E. Jorgenson, R. Mansour, and R. Ghani, "*Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning*," KDD, 2015. See also J. Abernethy, A. Chojnacki, A. Farahi, E. Schwartz, and J. Webb, "*ActiveRemediation: The search for lead pipes in Flint, Michigan*," KDD, 2018.

18 H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, and K. L. Addison, "*A machine learning framework to identify students at risk of adverse academic outcomes*," KDD, 2015.

19 K. Rodolfa, J. Zanzig, E. Salomon, K. Ackermann, L. Haynes, "*Preventing San Jose Housing Violations*," available at http://www.datasciencepublicpolicy.org/projects/public-safety/san-jose-housing/.

20 S. Mullainathan and Z. Obermeyer, "*Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*," National Bureau of Economic Research (NBER) Working Paper No. 26168, 2021.

21 D. Jurafsky and J. Martin, *Speech and Language Processing*, third edition (in progress), MIT Press, 2021, draft available at http://stanford.edu/~jurafsky/slp3/, and, e.g., R. Fergus, "*Computer Vision: Course Syllabus*," New York University, 2020.

Building a Healthy AI Ecosystem for NYC

The broad goal of this Strategy is to ensure New York City is equipped to meet the opportunities and challenges AI presents with robust and holistic steps that will support a healthy local ecosystem — one in which a broad mix of organizations across sectors work in concert to promote well-being, equity, and opportunity for all New Yorkers. In a healthy ecosystem:

→ Decision-makers across sectors are informed about what AI is and how it works, and are well-equipped to weigh the range of technical, social, and ethical considerations that come into play in its use.

→ Government use of AI is productive, fair, and accountable. Government policymaking is informed by diverse perspectives, and interdisciplinary and domain expertise. It incorporates public engagement where needed, and takes care to avoid unintended consequences of policies, which is a particular risk in this complex and rapidly evolving area.

→ Industry and other organizations, such as hospitals, schools, and universities use AI responsibly, with careful attention to ethical considerations like fairness, accountability, and privacy. Here, too, teams work to incorporate diverse views, both in terms of individual demographics and disciplinary and domain expertise. Small businesses have an equitable opportunity to participate and grow in the local ecosystem.

→ Residents are protected from harm (including disparate impact), empowered to participate in informed decision-making, and have appropriate ways of holding institutions accountable.[22] Residents are also prepared for workforce disruptions and have equitable opportunities to participate in a changing workforce.

→ Cross-sector collaboration is actively promoted to foster learning, innovation, and responsible use.

22 Accountability can take different forms in different contexts, such as being able to understand, examine, or even contest decisions made by AI systems in industry or government. See *Supplement A* for further details.

New York City's AI ecosystem today

An ecosystem approach accounts for the full range of stakeholders interacting with the topic of AI and aims to account for a range of contextual factors that shape the local ecosystem as a whole. These interactions can take a variety of forms, including working in, training for, building, buying, using, governing, and being impacted by AI. It is necessary for the City to consider this array of factors holistically to be able to effectively plan for the multifaceted ways in which AI is having impact across society.

While a detailed mapping of the local AI ecosystem is beyond the scope of this Strategy, the diagram below outlines key categories that should be taken into account.

INDUSTRY	EDUCATION	CIVIL SOCIETY	GOVERNMENT	RESIDENTS & COMMUNITIES
→ Cloud platforms → AI consumer and business products → AI developer products and services → AI consumer and business users → Industrial research labs → Investors → Startups → Small businesses	→ K-12 → Colleges and universities → Research institutes → Public training programs → Private training programs → Internship programs → Summer camps → After-school programs → Online courses	→ Non-profits → Advocacy groups → Think tanks → Foundations → Civic tech groups	→ Mayor's Office → NYC agencies → NYC City Council → Additional NYC government → State and federal governments → International governments → Intergovernmental groups → Legal, standards, and policy bodies	→ Community organizations → Mutual aid groups → Individual people

The remainder of this section highlights some key attributes of and examples from the local ecosystem.

New York has a large and diverse population that can drive inclusive growth and innovation.

The foundation of any ecosystem is its people; AI is no different. Home to more than eight million New Yorkers, New York City is the largest city in the United States and one of the largest cities in the world. It is also among the most diverse. The city is home to over three million immigrants from all over the world; hundreds of languages are spoken here.[23] New York is a "majority-minority" city, and over half of its residents are women.[24] Almost one million New Yorkers report living with a disability,[25] and the metropolitan area has the largest LGBT population in the United States.[26]

As AI is increasingly integrated into the city's economy and society, New Yorkers can offer uniquely diverse perspectives and skills. These can power distinct business ideas that serve different market segments in novel ways, but they can also help ensure that local AI efforts in general are more fair, equitable, and forward-thinking. These contributions can include being a part of teams building, using, or governing AI; creating or investing in new AI businesses; contributing to education and research in the field; and participating and advocating in the public realm. Across the ecosystem, New Yorkers' diverse voices can strengthen the way AI is developed, used, and considered, with positive outcomes for ecosystem health.

The city's economy is large and diverse, both of which offer unique opportunities for growth in AI.

New York City's gross metropolitan product is $1.66 trillion;[27] if the city were its own country, its economy would be roughly the twelfth largest in the world. These numbers increase further when looking at the broader metropolitan and tri-state areas with which the city is intertwined. Further, New York City is a global capital for a uniquely diverse set of industries, including the arts, media, and publishing, life science and health, finance, fashion, restaurants, and hospitality. Indeed, many industry leaders are headquartered

23 NYC Mayor's Office of Immigrant Affairs, "*State of Our Immigrant City: Mayor's Office of Immigrant Affairs Annual Report*," 2020, available at https://www1.nyc.gov/assets/immigrants/downloads/pdf/MOIA-Annual-Report-for-2020.pdf.

24 U.S. Census Bureau, "*American Community Survey 5-Year Estimates*," 2019, available at https://data.census.gov/cedsci/profile?g=1600000US3651000.

25 Ibid.

26 F. Newport and G. Gates, "*San Francisco Metro Area Ranks Highest in LGBT Percentage*," Gallup, 2015, available at https://news.gallup.com/poll/182051/san-francisco-metro-area-ranks-highest-lgbt-percentage.aspx.

27 Research and Statistics Group, New York Federal Reserve Bank, "*New York City Economic Indicators*," 2021, available at https://www.newyorkfed.org/medialibrary/media/research/regional_economy/charts/Regional_NYC.

here. These characteristics create significant opportunities for AI investment and growth in New York. The diversity of the local economy also means that, as AI is integrated across the ecosystem, it is more likely to be grounded in domain-specific needs, constraints, and broader social considerations. If actively harnessed, this contextualization can help create AI businesses that have unique, innovative elements and consider broader impacts.

EXAMPLE: Financial services and fintech

The financial services industry has long been a key sector in the city's economy. It has also been among the earliest adopters of AI and ML tools for a wide range of uses. These companies range from the largest banks, asset managers, and brokerages, such as Goldman Sachs, Morgan Stanley, BlackRock, and JPMorgan Chase, to more specialized firms that focus on technical investing, such as Two Sigma, DE Shaw, Jane Street, and others. These companies have long served as magnets for top technical talent, and employees from these companies have in turn gone on to projects in other domains and have invested in or founded new companies. The sector has continued to evolve, with the increasingly important "fintech" (innovative and emerging technology applied to financial services) sector being an important area of growth for the city as well as a key application domain for AI.

Manhattan's business districts are home to many of the world's largest financial services firms.
Photo: Florian Wehde

EXAMPLE: Mount Sinai and AI in healthcare

The Icahn School of Medicine at Mount Sinai and the Mount Sinai Health System have worked to integrate AI into healthcare — for both clinical and research purposes — for years.[28] These applications include predicting patient disease risk, foreseeing disease progression, improving inpatient safety, assessing population health across the health system, and targeting at-risk patients with precision medicine. For example, Mount Sinai's AI-powered system helps clinicians identify and prioritize patients at risk for conditions such as cardiopulmonary deterioration, malnutrition, and falls. Developed by Mount Sinai's Clinical Data Science team, the system supports clinical decision-making for thousands of patients each day.

New York City has a thriving tech sector and investment in AI is established and growing.

New York is the world's second most valuable tech ecosystem, at $147 billion, and the sector is rapidly expanding.[29] Prior to the COVID-19 pandemic, the city's tech sector employed more than 330,000 people across the region.[30] Despite losses across the economy during the crisis, tech led job growth in the city in 2020.[31] In 2020 alone, Apple, Amazon, and Facebook together added 1.6 million square feet of office space in the city, adding to the 1.7 million square feet Google acquired in recent years.[32]

The city is also home to a thriving community of startups and small businesses. More than 9,000 startups and over 100 incubators are located in New York,[33] and the city sees over $70 billion in startup valuation and exits per year.[34] In 2019, Brooklyn's startup growth rate was second only to that of San Francisco, at over 350% since 2008; the broader city has seen growth at over 300%.[35] Venture capital investment in the city is increasing, reaching over $15 billion in 2020, and on track to exceed that level in 2021.[36]

And there is significant AI investment in place in the city today. According to the Center for Security and Emerging Technology at Georgetown University, over $9 billion of funding in AI is already

28 See https://icahn.mssm.edu/research/artificial-intelligence.

29 TechNYC, "*NYC Tech Snapshot*," available at https://www.technyc.org/nyc-tech-snapshot.

30 B. Viney, J. Bowles, and L. Gallagher, "*Tech Disrupted: How Coronavirus is Challenging NYC's Tech Ecosystem*," Center for an Urban Future, 2020, available at https://nycfuture.org/research/tech-disrupted.

31 S. Amandolare, E. Dvorkin, and C. Shaviro, "*Preparing New Yorkers for the Tech Jobs Driving NYC's Pandemic Economy*," Center for an Urban Future, 2020, available at https://nycfuture.org/research/preparing-nyers-for-tech-pandemic.

32 TechNYC, "*NYC Tech Snapshot*," available at https://www.technyc.org/nyc-tech-snapshot.

33 Ibid.

34 NYC Economic Development Corporation, "*Emerging Tech*," available at https://edc.nyc/industry/emerging-tech, and Startup Genome, "*Global Startup Ecosystem Report*," 2018, available at https://startupgenome.com/reports/global-startup-ecosystem-report-gser-2018.

35 J. Bowles, E. Dvorkin, N. Sharp, and C. Shaviro, "*Brooklyn's Growing Innovation Economy*," Center for an Urban Future, 2019, available at https://nycfuture.org/research/brooklyns-growing-innovation-economy.

36 J. Glasner, "*New York Sees Startup Funding Spike in 2021*," Crunchbase News, 2021, available at https://news.crunchbase.com/news/new-york-sees-startup-funding-spike-in-2021/.

invested in New York City,[37] and AI is a key investment area for venture capital firms.[38] In 2020, 13% of the US AI workforce was located in New York, and that figure is growing rapidly.[39]

EXAMPLE: Flatiron Health

Flatiron Health, founded in New York City in 2012, is a technology company with the mission of improving lives by learning from the experience of every cancer patient. From a technical perspective, this learning is largely done through data infrastructure and statistical and machine learning techniques, though the company also has teams of oncologists and other domain experts. Today, Flatiron is one of the leaders in what is called "real world data" or "real world evidence" for oncology data in the US. It partners with hundreds of cancer centers and the world's top developers of oncology therapeutics, as well as collaborating with researchers and regulators to accelerate R&D and impact treatment for patients worldwide. In 2018, Flatiron was acquired for nearly $2 billion and became an independent affiliate of the Roche Group, a multinational healthcare and pharmaceuticals company.

EXAMPLE: Hugging Face

Hugging Face, founded in 2016, is a leading AI startup based in New York City and Paris. The company builds tools and a platform to help AI teams build, train, and deploy state-of-the-art models for natural language processing, building on standard open-source AI technologies like Facebook's PyTorch and Google's TensorFlow. Hugging Face has raised $60 million in financing, including $40 million in March 2021. Its tools are used by thousands of organizations, including Microsoft, Facebook, and Google, and it is an example of an AI company focused on building tools to help others use AI rather than using AI for specific applications themselves.

EXAMPLE: Local Minority and Women-Owned Businesses

It is important to note that local AI companies include small and medium-sized enterprises — many of which are certified Minority and Women-Owned businesses. These companies offer a wide range of AI-related products and services, from those that assist

37 J. Olander and M. Flagg, "*AI Hubs in the United States*," Georgetown CSET, 2020, available at https://cset.georgetown.edu/publication/ai-hubs-in-the-united-states/.

38 Ibid.

39 Ibid.

clients to streamline business processes, to those that work to organize and structure data, and more.

New York City is already a hub for both industrial and academic AI research and development as well as work at the broader intersection of tech and society.

Companies and academic institutions alike have invested heavily in AI research and development in New York City. The wide array of leading research universities located in New York are all actively working to advance the field in myriad ways, and several have large, dedicated centers devoted to AI, data, and related topics, such as Columbia's Data Science Institute and NYU's Center for Data Science. Additionally, the city is home to a range of leading industrial research labs working at the cutting edge of AI and related technology development, such as Microsoft Research, Google Brain, and Facebook AI Research. Unlike many academic disciplines, a large amount of original research in AI is conducted at industrial research labs — such as DeepMind and OpenAI, as well as the local labs just mentioned — in addition to universities. This is partly because large technology companies have the resources — data, computational power, finances, and staff expertise — to do certain types of work.[40] There is disagreement and debate in the AI community about the benefits and drawbacks to this situation, such as concerns over the reproducibility of industrial research or potential corporate conflicts of interest.[41]

EXAMPLE: Technology, society, and advocacy

One distinctive aspect of the ecosystem is that the city is a global center for academic, industry, and advocacy expertise on the relationship between technology and society. Several local academic institutions have created research labs focused on the intersection of technology and society, including NYU's GovLab and AI Now Institute and Cornell Tech's Urban Tech Hub and Digital Life Initiative. Local advocacy groups like the Data & Society Research Institute work at the intersection of advocacy and emerging technology. And a number of local civic tech organizations — such as BetaNYC or Silicon Harlem — work to support public engagement

40 See, e.g., R. Gelles, Z. Arnold, N. Luong, and J. Melot, "*PARAT–Tracking the Activity of AI Companies*," Georgetown CSET, 2021, available at https://cset.georgetown.edu/publication/parat-tracking-the-activity-of-ai-companies/; J. Olander and M. Flagg, "*AI Hubs in the United States*," Georgetown CSET, 2020, available at https://cset.georgetown.edu/publication/ai-hubs-in-the-united-states/.

41 B. Haibe-Kains et al, "*Transparency and reproducibility in artificial intelligence*," Nature, 2020; H. Stower, "*Transparency in medical AI*," Nature Medicine, 2020; A. Belz, S. Agarwal, A. Shimorina, and E. Reiter, "*A systematic review of reproducibility research in natural language processing*," arXiv preprint arXiv:2103.07929, 2021; W. Heaven, "*AI is wrestling with a replication crisis*," MIT Technology Review, 2020.

and equitable access to tech opportunity. In addition, large technology companies, especially Microsoft and Google, have prominent local teams focused on the ethics of AI. Together, these organizations make New York City a global center for work in what is sometimes called "responsible AI," and they offer a vibrant source of expertise and thought leadership on these subjects to the local ecosystem.[42]

A related aspect is the local presence of Black-led racial justice organizations focused on the ways in which the design, deployment, and governance of advanced technical systems can negatively impact Black communities. This includes national groups working locally in New York, such as Color of Change, the nation's largest online racial justice organization, and AI for the People, a newer organization focused on using art and culture to educate lawmakers about racial justice implications of ML use in public life. It also includes local organizations such as the SAFE Lab at Columbia School of Social Work, which conducts research on "the ways in which youth of color navigate violence on and offline,"[43] and the Ocean Hill-Brownsville Alliance, which works at the neighborhood level to advocate for residents' rights and agency in their interactions with AI and related technologies.

The city has robust educational infrastructure in place to support ongoing AI literacy and skill-building needs, and particular investment in place in public- and public-interest resources.

New York City is home to a wealth of world-class universities, with the City University of New York's (CUNY) twenty-five colleges and schools being a crown jewel of public education that lead the country as engines of economic mobility.[44] The City has further demonstrated investment in foundational tech education in its public K-12 system via the citywide CS4All program. New York is also home to a broad network of public and private technology training providers, offering programs across a wide spectrum of topics and levels, from library-based training programs to private bootcamps.[45] These resources, taken together, offer the city tremendous

42 For a summary of many of the topics concerning this community, see, e.g., AI Now Institute reports available at https://ainowinstitute.org/reports.html.

43 See https://safelab.socialwork.columbia.edu/content/about.

44 S. Reber and C. Sinclair, "*Opportunity engines: Middle-class mobility in higher education*," The Brookings Institution, 2020, available at https://vtechworks.lib.vt.edu/bitstream/handle/10919/98982/OpportunityEnginesFinal.pdf.

45 For details on available programs, see https://nycfuture.org/research/plugging-in and https://www1.nyc.gov/assets/cto/#/project/digital-inclusion-initiatives .

opportunity to grow literacy and skill in order to responsibly leverage AI and to do so in a way that prioritizes equity and inclusion.

Across these institutions, a number of initiatives and programs have taken shape locally that are invested in preparing New Yorkers to contribute to "public interest technology" in particular. CUNY, Columbia University, Cornell Tech, and New York University are all part of the Public Interest Technology University Network (PIT-UN) founded by New America, a think tank based in Washington, DC and New York City. PIT-UN is a collaboration among dozens of higher education institutions committed to building the field of public interest technology, and "growing a new generation of civic-minded technologists and digitally-fluent policy leaders" through coursework, experiential learning, and research opportunities.[46]

EXAMPLE: City University of New York (CUNY)

The City University of New York serves students across its twenty-five colleges and schools, the majority of which offer Computer Science degree programs. In 2019, as the largest urban public university system in the United States, CUNY served 241,080 degree-seeking undergraduates and 30,162 graduate students, and awarded over 57,000 degrees.[47] CUNY's student body is highly diverse. It is the only university system in the country in which Asian, Black, Hispanic, and white students each constitute more than 20% of the student body.[48] Sixty-one percent of students receive Pell grants and 42% are from households earning less than $20,000 annually.[49] Most CUNY undergraduates (57%) and graduate students (65%) are women.[50]

CUNY is engaged in a range of public interest technology efforts, including varied experiential learning programs and efforts to support interdisciplinary research and curricula. These include an internship program, coordinated by the Department of Information Technology and Telecommunication (DOITT), through which students can work in agencies across City government as application and web developers, data analysts, technical support, and other roles, and a program with BetaNYC's Civic Innovation Lab,

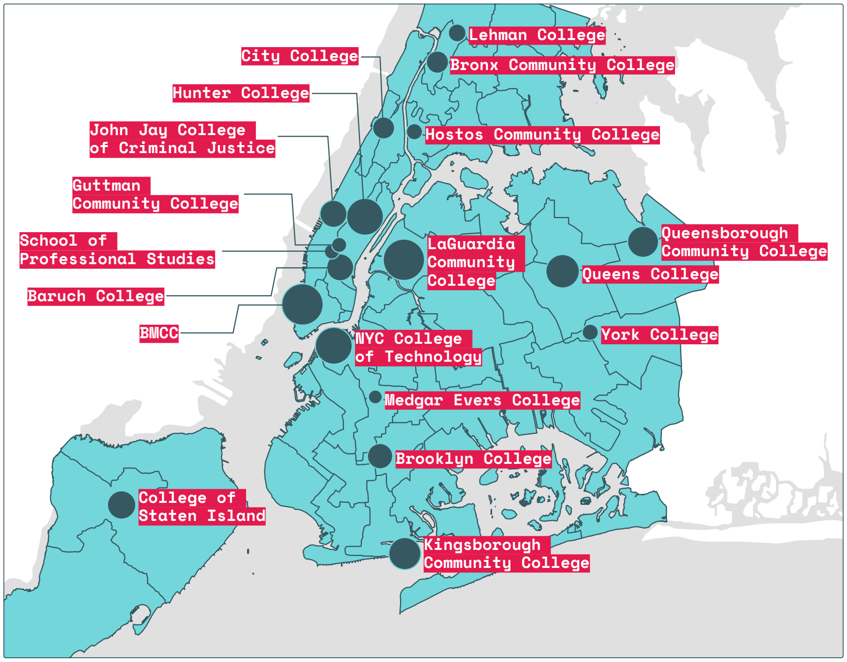
46 New America, "*PIT-UN*," available at https://www.newamerica.org/pit-un/about/, and https://www.newamerica.org/pit/press-releases/31-million-awarded-grow-field-public-interest-technology-leading-colleges-and-universities/.

47 Office of Institutional Research, City University of New York, "*Current Student Data Book by Subject*," available at http://www.cuny.edu/about/administration/offices/oira/institutional/data/current-student-data-book-by-subject/#Enrollment.

48 Ibid.

49 Ibid.

50 Ibid.



The City University of New York (CUNY) is the largest public urban university system in the United States, serving over 260,000 undergraduate and graduate students — as well as nearly 200,000 registrations in adult and continuing education programs — via colleges spread across all the five boroughs. Here, individual colleges are sized to indicate student enrollment.
Source: Map by NYC CTO; data from CUNY https://www.cuny.edu/about/administration/offices/evaluation/data-methods/data-hub/maps/

which offers students training and experience supporting local community boards' digital and data analysis needs.[51] CUNY and SUNY are also partners with various industry stakeholders in the financial sector, led by Microsoft's National Council of Artificial Intelligence for Responsible AI in Finance, and are working to develop student training opportunities, such as hosting a hackathon addressing ethical issues in fintech.

Additionally, the City's "CUNY 2x Tech" initiative — led by the Department of Small Business Services (SBS) and its Tech Talent Pipeline (TTP) industry partnership — works to grow the higher education pipeline of local tech talent, and has successfully doubled the number of tech bachelor's degrees awarded by CUNY since 2017.[52] The program supports a tech-specific internship program, industry sponsored led project-based courses, and a program known as the Tech-in-Residence Corps that sponsors industry professionals teaching in-demand advanced topics in the classroom. There are more details on TTP's efforts in Section 5.

51 For more on these programs, see: https://www1.cuny.edu/sites/matters/2019/04/05/promoting-technology-in-the-public-interest/, https://www.cuny.edu/employment/student-jobs/internships/cuny-internship-programs/, and https://beta.nyc/programs/civic-innovation-lab/.

52 For more on the CUNY 2x Tech program, see https://www.techtalentpipeline.nyc/cs-doubling.

EXAMPLE: Community-based training programs

In addition to being home to a rich array of non-profit training providers,[53] New York City has the largest network of public computer centers in the country — which collectively offer more than 2,500 hours of technology programming per week.[54] These organizations deliver programs to residents in a wide range of topics, from basic digital literacy to coding, data literacy, media literacy, and digital rights, among many others. They can also play an important role in public engagement efforts by serving as trusted community institutions with deep local knowledge, and they represent a critical resource for educating and engaging New Yorkers on AI. For example, Queens Public Library partnered with NYU's Center for Responsible AI and Peer-to-Peer University to launch "We Are AI," a five-week course to "introduce the basics of AI, discuss some of the social and ethical dimensions of the use of AI in modern life, and empower individuals to engage with how AI is used and governed."[55]

City government is poised for leadership and impact.

Finally, City government itself offers both a strong foundation for leadership in AI policy and governance, and significant opportunity to innovate in the responsible use of AI for the public good. With over 330,000 employees, a municipal budget of nearly $100 billion,[56] and agencies working in a wide range of domain areas, the City can launch initiatives with significant impact, and offer important lessons to municipalities across the globe. The next section outlines a partial map of City agencies and how they relate to AI; efforts to date and specific opportunities available are described in more detail in the Findings and Opportunities section that follows.

53 E. Dvorkin with S. Amandolare, J. Chambers, and C. Shaviro, "*Plugging In: Building NYC's Tech Education and Training Ecosystem*," Center for an Urban Future, 2020, available at https://nycfuture.org/research/plugging-in.

54 NYC Mayor's Office of the Chief Technology Officer, "*The New York City Internet Master Plan*," 2020, available at https://www1.nyc.gov/assets/cto/downloads/internet-master-plan/NYC_IMP_1.7.20_FINAL-2.pdf, and NYC Mayor's Office of the Chief Technology Officer, "*Citywide Public Computer Centers*," 2019, available at https://data.cityofnewyork.us/Social-Services/Citywide-Public-Computer-Centers/cuzb-dmcd.

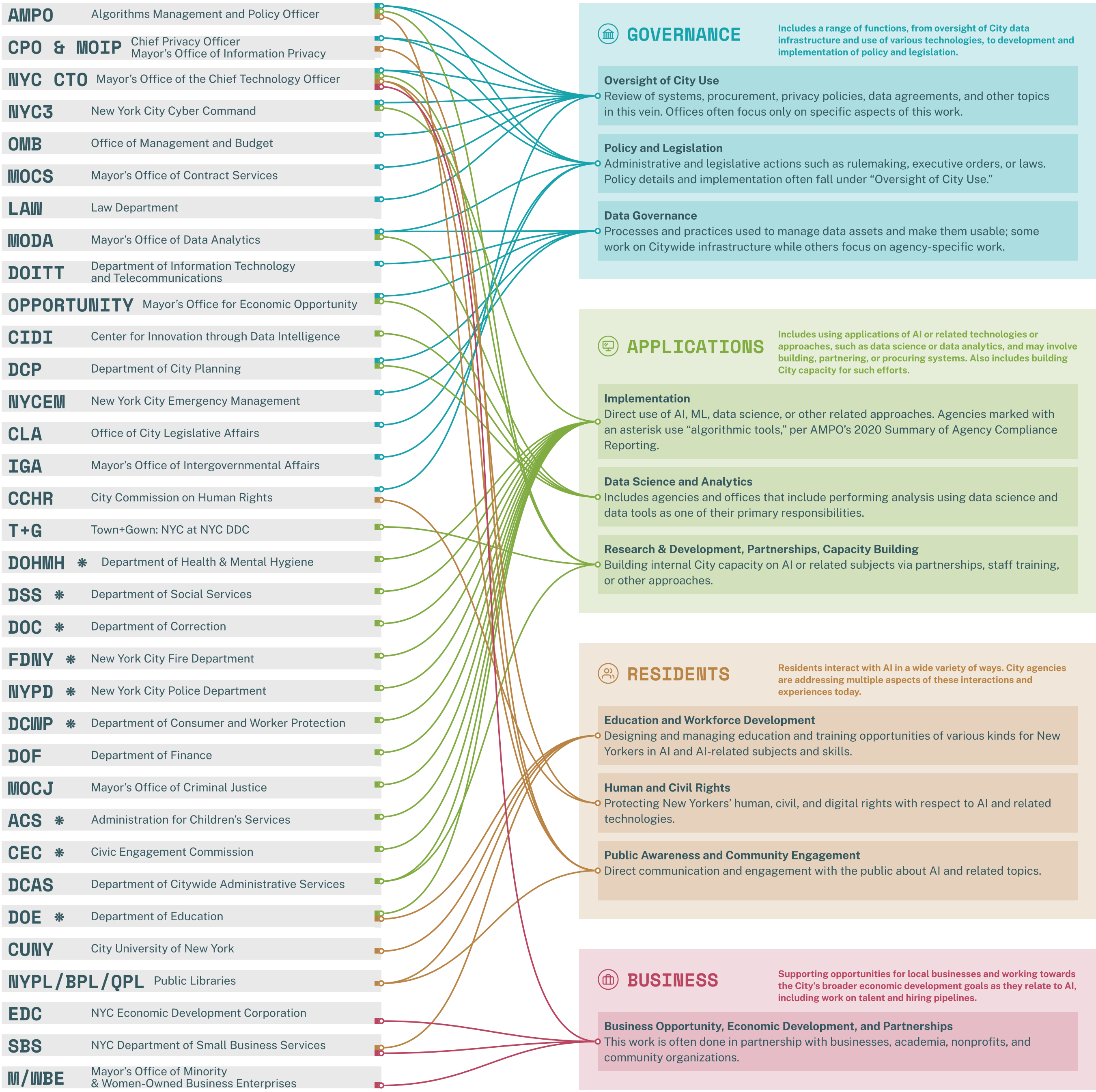
55 NYU Center for Responsible AI, "*We Are AI*," available at https://dataresponsibly.github.io/we-are-ai/.

56 NYC Office of Management and Budget, "*June 2021 Adopted Budget, Fiscal Year 2022*," 2021, available at https://www1.nyc.gov/site/omb/publications/finplan06-21.page.

AI in NYC government

New York City government is made up of almost 150 agencies and offices that have or will have a wide range of roles and perspectives on AI and the local ecosystem. These roles can include governing data or AI, using AI, addressing AI's use and impacts across society, interacting with or educating residents, and focusing on jobs, economic development, and business opportunities. Each of these agencies will have a particular vantage point on how AI relates to their work moving forward. It is also important to keep in mind their widely variable scales: some consist of only a handful of people, while others have tens of thousands of employees organized into large, discrete divisions that may themselves operate with a high degree of autonomy and with their own internal technology teams and policy goals.

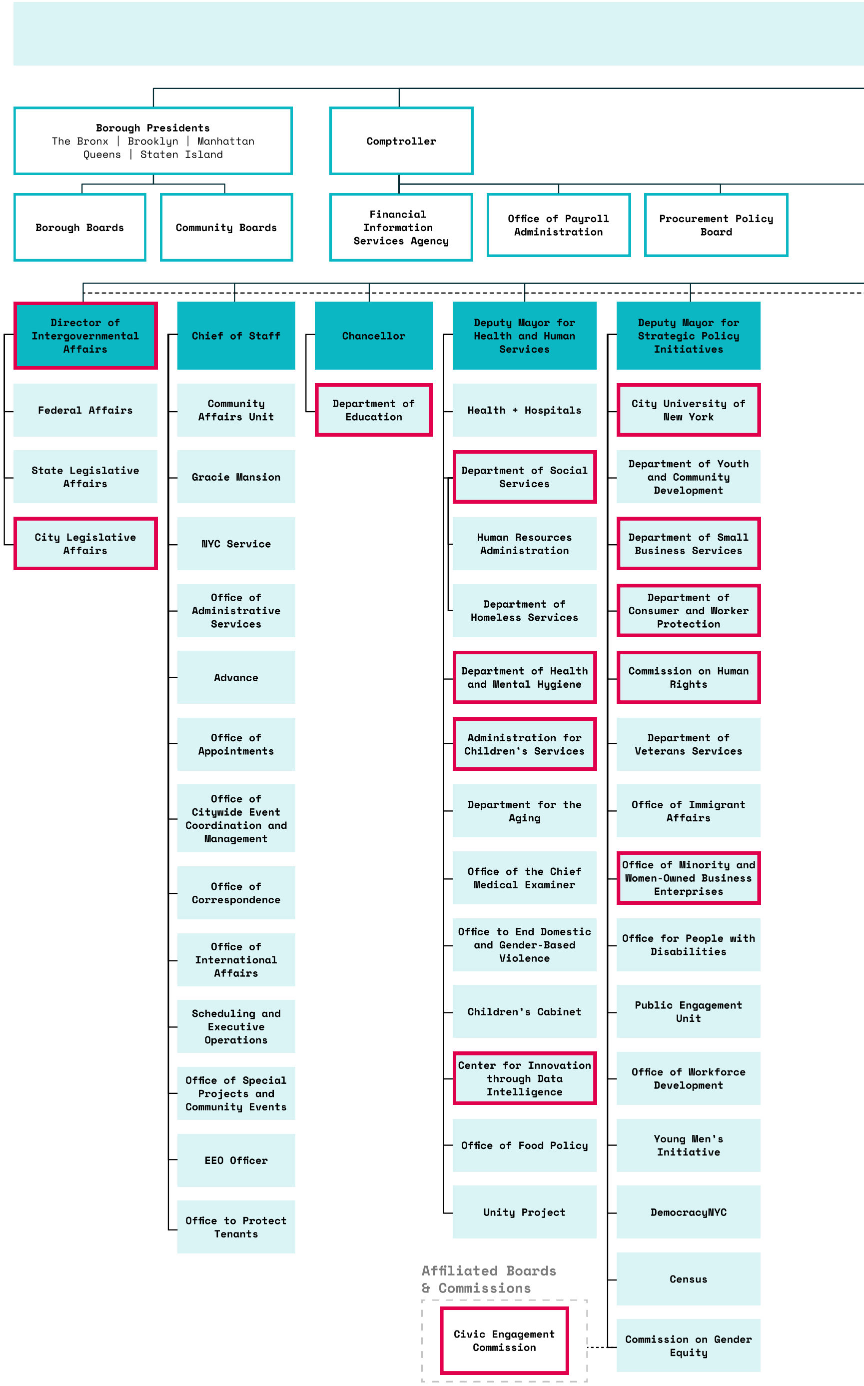
The diagram below details how City agencies relate to AI today. This includes agencies that may not consider their work "AI" but use related tools from data science and statistics.

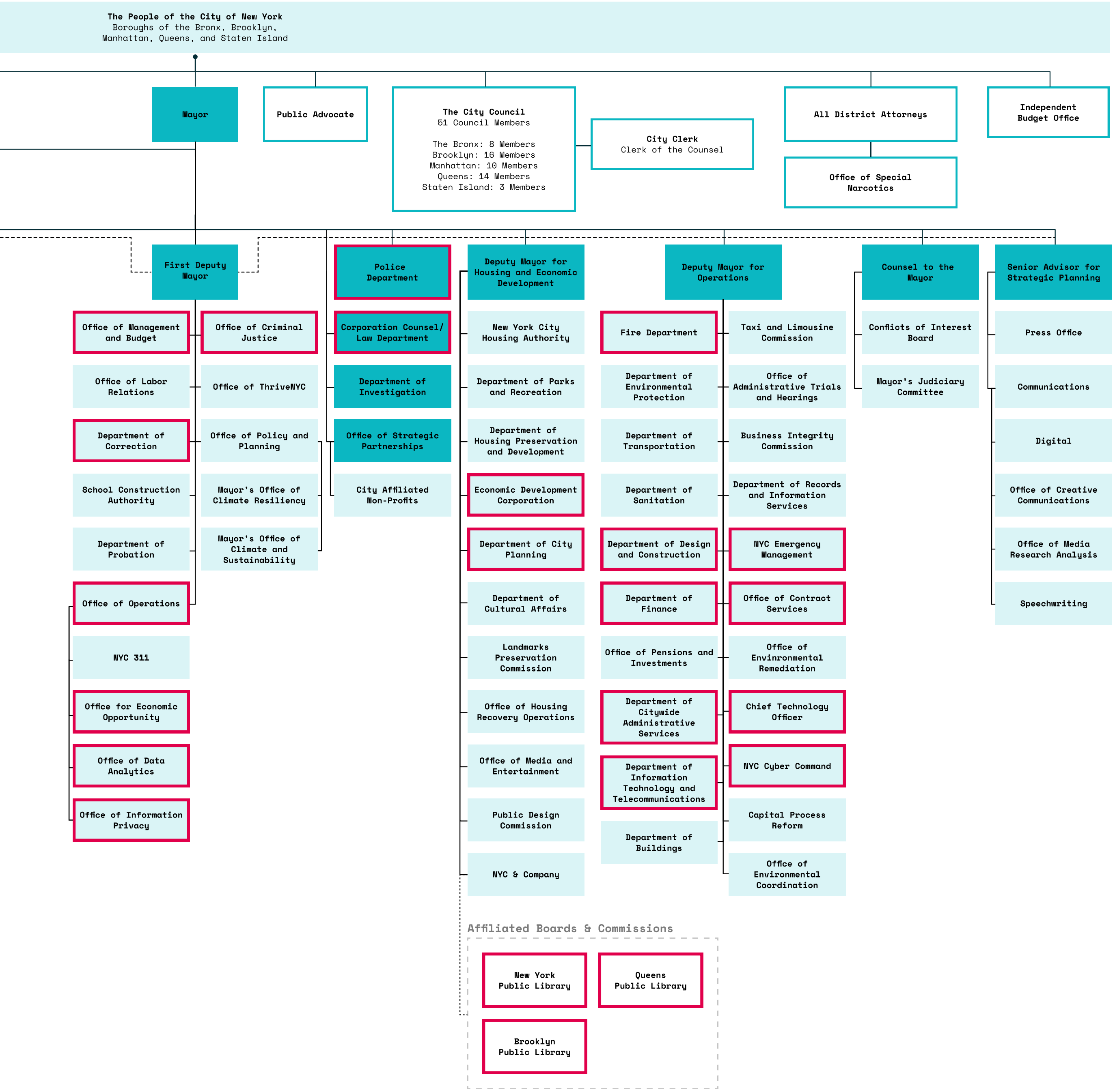


* agency use described in 2020 AMPO report

RIGHT: Simplified organizational chart of the NYC government. Offices in darker blue report directly to the Mayor. Agencies highlighted in red have some relationship to the topic of AI, as described in the figure on the previous page.

Note that this is not a comprehensive organizational chart of NYC government. For more details, see https://www1.nyc.gov/office-of-the-mayor/org-chart.page.





The People of the City of New York
Boroughs of the Bronx, Brooklyn, Manhattan, Queens, and Staten Island
Mayor
Public Advocate
The City Council
51 Council Members
The Bronx: 8 Members
Brooklyn: 16 Members
Manhattan: 10 Members
Queens: 14 Members
Staten Island: 3 Members
City Clerk
Clerk of the Counsel
All District Attorneys
Office of Special Narcotics
Independent Budget Office
First Deputy Mayor
Office of Management and Budget
Office of Labor Relations
Department of Correction
School Construction Authority
Department of Probation
Office of Operations
NYC 311
Office for Economic Opportunity
Office of Data Analytics
Office of Information Privacy
Office of Criminal Justice
Office of ThriveNYC
Office of Policy and Planning
Mayor's Office of Climate Resiliency
Mayor's Office of Climate and Sustainability
Police Department
Corporation Counsel/ Law Department
Department of Investigation
Office of Strategic Partnerships
City Affiliated Non-Profits
Deputy Mayor for Housing and Economic Development
New York City Housing Authority
Department of Parks and Recreation
Department of Housing Preservation and Development
Economic Development Corporation
Department of City Planning
Department of Cultural Affairs
Landmarks Preservation Commission
Office of Housing Recovery Operations
Office of Media and Entertainment
Public Design Commission
NYC & Company
Deputy Mayor for Operations
Fire Department
Department of Environmental Protection
Department of Transportation
Department of Sanitation
Department of Design and Construction
Department of Finance
Office of Pensions and Investments
Department of Citywide Administrative Services
Department of Information Technology and Telecommunications
Department of Buildings
Taxi and Limousine Commission
Office of Administrative Trials and Hearings
Business Integrity Commission
Department of Records and Information Services
NYC Emergency Management
Office of Contract Services
Office of Environmental Remediation
Chief Technology Officer
NYC Cyber Command
Capital Process Reform
Office of Environmental Coordination
Counsel to the Mayor
Conflicts of Interest Board
Mayor's Judiciary Committee
Senior Advisor for Strategic Planning
Press Office
Communications
Digital
Office of Creative Communications
Office of Media Research Analysis
Speechwriting
Affiliated Boards & Commissions
New York Public Library
Queens Public Library
Brooklyn Public Library

Findings and Opportunities

To better understand the opportunities and challenges in the local ecosystem today, the NYC Mayor's Office of the Chief Technology Officer (NYC CTO) conducted interviews with more than fifty stakeholders, including representatives from over twenty City agencies and over thirty organizations across the private, academic, and non-profit sectors. A full list is included in *Supplement B: The Voices that Shaped this Strategy*.

These conversations were necessarily diverse in a number of respects. Within City government, NYC CTO met with agencies that are using AI today and those considering using it in the future; those responsible for policymaking and governance in AI and related areas; and those involved in consumer and worker protection, education, business opportunity, economic development, and the workforce. The individuals spanned a variety of roles and functions, including agency heads and their deputies, CIOs and CTOs, staff responsible for managing and analyzing internal and interagency data, and a range of other relevant program and policy staff members. They cut across the domains of transportation, public health, education, small business, climate and sustainability, city planning, emergency management, and more.

Externally, NYC CTO met with an array of universities, research and advocacy organizations, and civic technology groups, with a particular focus on groups working to engage New Yorkers on issues related to AI or to understand the ways in which these technologies impact and are embedded in communities and the broader social world. In addition, NYC CTO spoke with a variety of companies involved in creating and using AI in New York City, including leading industrial research labs, the local startup community, and local AI employers, large and small, including leaders of a number of minority and women-owned businesses, some of which deploy AI in the public sector.

These conversations touched on a wide range of topics related to AI, and the information that emerged from them guided the organization of these findings into five thematic areas:

1. City data infrastructure;
2. AI applications within the City;
3. City governance and policy around AI;
4. Partnerships with external organizations;
5. Business, education, and the workforce.

Public engagement and City capacity building were key topics that arose repeatedly in conversations with stakeholders; discussion of these can be found in *Sections 3* and *5*, respectively.

In each of these categories, findings are followed by a set of emergent opportunities for the City. The detailed items included within each of these should, at this stage, be viewed as informed but illustrative — rather than a comprehensive list of everything required or desirable. As the City refines its approach to AI, and builds capacity in the space, further development of the particular actions the City takes will be needed. The closing *Next Steps* section details a set of initial commitments the City will undertake in the near term.

It is important to emphasize that all of the areas addressed here are intertwined. Accordingly, it is crucial that decision-makers across different roles and functions engage with them holistically, rather than in a piecemeal fashion.

Further, as noted, AI and its uses are rapidly evolving. There is uncertainty around how it will develop and the opportunities and challenges that will emerge. While it is critical that the City begin to take holistic action to address the AI-driven transformations already underway, it is also important to maintain humility and be agile in the face of this uncertainty, and it will be important for the City to regularly review and update this Strategy.

1. City data infrastructure

Findings

New York City's existing approach to data is wide-ranging, if fragmented, and can be built upon to support both AI applications and other important improvements.

Because of the fundamental way in which AI and related topics relate to data, a strategy around potentially using AI in any sector must begin with a strong strategic approach to data in general. Within City government, this approach should include both a City-wide component as well as plans at the level of individual agencies or key domain areas like health, education, and transportation. Although outlining a full approach is out of scope here, a few important aspects are highlighted here, followed by examples of existing work that should be built on.

Some of the key strategic areas one would want to consider include data collection, acquisition, procurement, standards, cleaning, sharing, ownership, access, usage agreements, warehousing, and analytics; digital rights concerns, such as privacy and security, relate to many of these.[57] While a full discussion of each of these topics is out of scope, a few illustrative points are discussed here. Several agencies, including the Mayor's Office for Economic Opportunity (NYC Opportunity), NYC Emergency Management (NYCEM), DOITT, and the Department of City Planning (DCP), have also previously prepared reports on City data infrastructure with recommendations that will remain relevant for some time.[58]

Because of how the composition of the data affects what happens downstream (see *Supplement A*) — from very simple analyses to machine learning — collecting and curating data thoughtfully is a foundational first step to any responsible AI initiative. This must include paying careful attention to how data is shaped by its collection method, including who is and is not reflected in the dataset and in what proportions. For example, agencies must account for

57 For some references that touch on some of these topics, see, e.g., M. Kleppman, *Designing Data-Intensive Applications*, O'Reilly Media, 2017.

58 For example: Geographical Information Systems (GIS) Committee, "*Promoting GIS Data Exchange and Shared Services through a Common Infrastructure: Recommendations*," City of New York, 2012; NYC Mayor's Office for Economic Opportunity, "*Information Sharing and System Modernization in New York City: A report to NYC City Council for Local Law 75 (2018)*," 2018, available at https://www1.nyc.gov/assets/opportunity/pdf/specialinitiatives/local-law/LL75-info-sharing-final.pdf.

the fact that only making surveys available digitally may exclude the nearly 40% of New Yorkers who have insufficient internet access,[59] and that only making surveys available in English and Spanish, or even in all Local Law 30 languages, will exclude information from certain populations that may be important to measure.

Similarly, it is important to ensure that data, once collected, is stored in a "machine readable" form that is conducive to its use (as opposed to paper or other unstructured forms), and that it follows certain data standards so it can be related to other datasets or shared with others to facilitate a wide range of uses. For example, the General Transit Feed Specification is a common format for public transportation agencies (including in New York) to share transit information. Prior to its introduction, there was no uniformity among transit agencies' data feeds, and it is the broad adoption of this standard that has enabled widely used applications like Google Maps. In the city, this can manifest in very basic ways. For instance, if all agencies that dealt with data on physical buildings used a common identifier, such as the designated building identification number ("BIN"), to refer to buildings consistently, a range of different analyses would suddenly become possible or much easier to do. Although the BIN is an identifier defined and created by the City and is used by some agencies, it is not used universally. In other cases, even lacking standardization like storing a single "name" field as opposed to "first name" and "last name" separately can be a significant impediment to agency — and especially interagency — work that requires combining many different datasets, which are often owned by different agencies.[60] Though these steps are not necessarily sophisticated, they are prerequisites for all of what follows.

Finally, it can be important to attend to steps needed to share data across agencies, particularly for those agencies that have frequent needs in this regard. Where steps such as data sharing agreements or appropriate consent are not in place, interagency data sharing can take far longer than necessary and prevent benefits from reaching New Yorkers in a timely manner, or at all.

59 Per the New York City Internet Master Plan, 40% of New Yorkers lack the combination of home and mobile broadband service — the comprehensive connectivity the City considers required today. Twenty-nine percent lack a mobile data plan. Eighteen percent of residents lack access to both home and mobile service. For more on these figures, and the City's approach to universal broadband, see the "*New York City Internet Master Plan*," Mayor's Office of the Chief Technology Officer, available at https://www1.nyc.gov/assets/cto/downloads/internet-master-plan/NYC_IMP_1.7.20_FINAL-2.pdf.

60 NYC Mayor's Office of Operations, "*BINs Working Group: Observations and Recommendations*," 2013.

EXAMPLE: City Planning, Economic Opportunity, and data infrastructure agencies

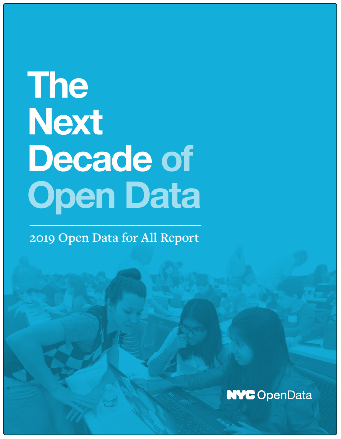
An array of City agencies possess a wealth of practical knowledge about opportunities and pain points. Although some of the larger agencies, such as DOE or DOHMH, have significant internal data operations, there are also several agencies that focus on interagency or Citywide infrastructure and datasets, including MODA, NYC Opportunity, DOITT, CIDI, DCP, and NYCEM. Agencies like DCP or DOB that own key datasets that are used by many other offices throughout the City also possess significant insight and experience.[61]

61 For descriptions of these agencies and acronyms, see diagram on page 33, which also describes where in the City government these agencies reside.

EXAMPLE: NYC Open Data

The NYC Open Data program, maintained by MODA and DOITT, makes thousands of agency datasets available for use by the general public, researchers, and companies, as well as other agencies. City employees often use Open Data for interdepartmental and interagency data sharing. The Open Data Law, passed in 2012, requires that all data suitable to be made publicly available be shared on a citywide portal and establishes an annual compliance process to that end. In 2019, NYC Open Data published a ten-year strategy for strengthening the compliance process, improving the usability of the portal, and building the community of data users.[62]

62 NYC Open Data, "*The Next Decade of Open Data*," 2019, available at https://opendata.cityofnewyork.us/wp-content/uploads/2019/09/2019_OpenDataForAllReport.pdf. MODA published an update to this plan in 2020, available at https://opendata.cityofnewyork.us/wp-content/uploads/2020/09/2020_OpenDataForAllReport_Full.pdf.



EXAMPLE: City data sharing agreements

The use of data sharing agreements is an important part of the City's overall data management infrastructure. There has been significant work on this topic by a number of agencies. One key example is the leadership of the Chief Privacy Officer and the Mayor's Office of Operations in creating the Citywide Data Integration Agreement (CDIA), which is a master framework agreement that all City agencies have signed, which specifies many of the required privacy, data security, and other terms appropriate for multi-agency data sharing agreements. Agencies can append the CDIA (and incorporate its terms, by reference) to any new bilateral or multi-agency data sharing agreement. Another example is NYC Opportunity's work on the Interagency Data Share Agreement, established to facilitate data sharing for the Worker Connect initiative.[63]

Some agencies have interagency agreements specifically for sharing sensitive data with other agencies, as permitted by law. The Center for Innovation through Data Intelligence (CIDI), under the auspices of the Deputy Mayor for Health and Human Services, holds agreements with a set of health and human services agencies within the City to facilitate data sharing for a range of research efforts using very sensitive data like health records. There have additionally been a variety of efforts to bring in data from external sources to support City goals. For example, the City's Recovery Data Partnership was launched in 2020 as a first-of-its-kind effort for community, non-profit, and private organizations to share data with the City in order to aid in COVID-19 response and recovery efforts.[64] Similarly, at the agency level, the TLC has made the rideshare permits it issues to private sector companies contingent upon their sharing trip data in order to inform the City's transportation policymaking.

63 For more on Worker Connect, see https://www1.nyc.gov/site/opportunity/portfolio/worker-connect.page.

64 For more on the RDP, see https://www1.nyc.gov/site/analytics/initiatives/recovery-data-partnership.page.

EXAMPLE: Department of Finance, geospatial data intelligence, and 3D mapping data

There is also significant potential in a range of unusual datasets that the City may not yet be actively using. In one notable example, DOF's Geospatial Data Intelligence team used high-resolution 3D mapping data (LiDAR point data[65]) acquired from openly available sources to conduct a pilot project that uniquely identified horizontal and vertical changes (e.g., new construction) in buildings not yet identified by DOF Assessors or permitting data. This approach was able to identify millions of dollars in property assessment changes that could otherwise have gone unrecognized. As a result of the pilot, additional LiDAR data will be acquired from 2021-2023, and it is estimated each capture may yield about $10 million in assessment changes. This data will also be available for and facilitates modeling and planning for environmental problems like heat island effects, urban runoff, and storm surge impacts.

65 For background on LiDAR, see, e.g., https://oceanservice.noaa.gov/facts/lidar.html.

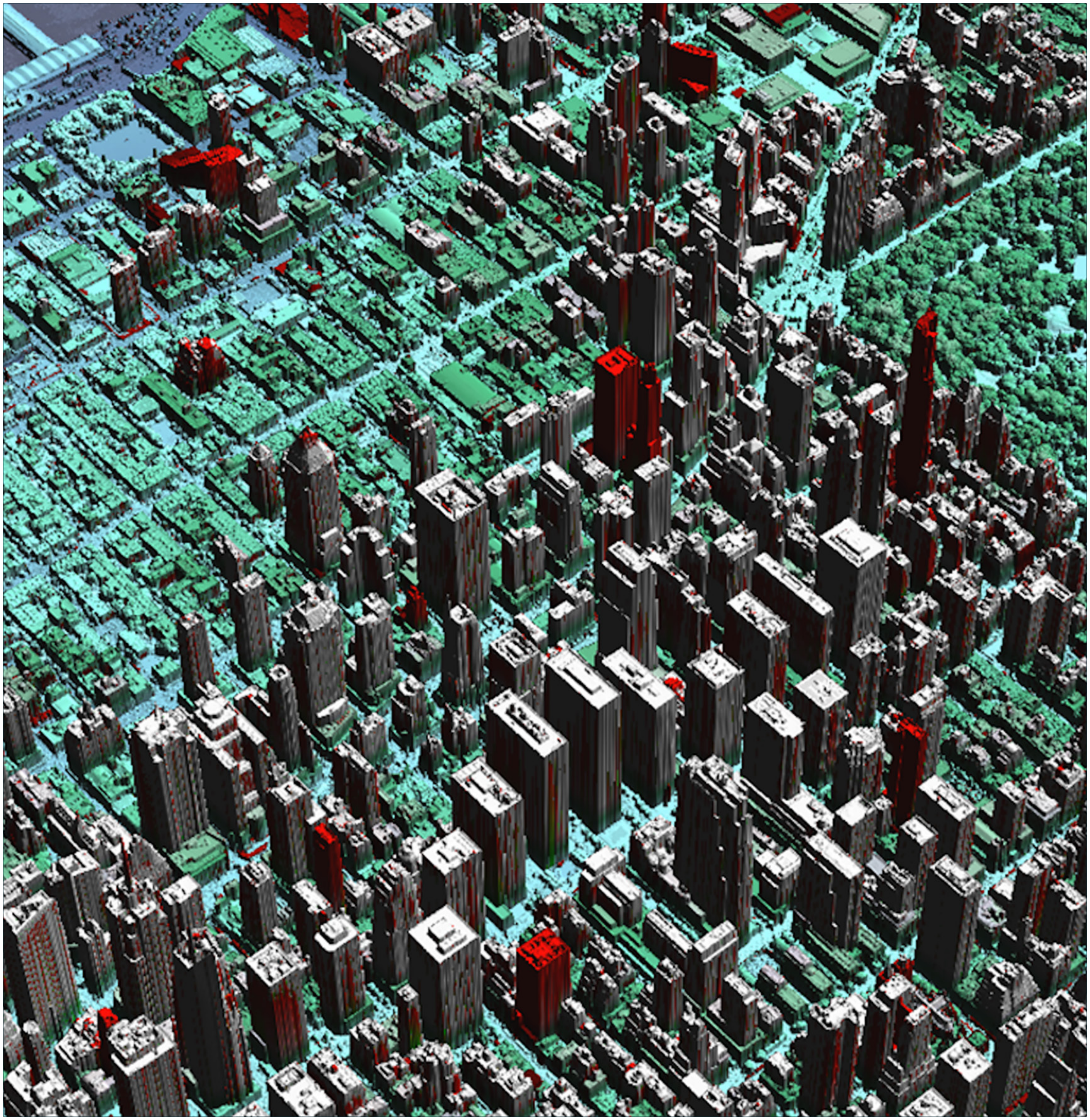
Opportunities

Establish a Citywide data strategy and foster consistency among agency-specific strategies.

Better data infrastructure, collection, cleanliness, interoperability, and standards will yield a wide range of benefits above and beyond their potential use in actual AI applications. The City would benefit from, for example, establishing a central framework, directing agencies to develop data strategies, and coordinating appropriate working groups of agency technical leadership to ensure that those strategies are consistent where needed. Some of the specific topics and issues that a data strategy should address were touched on in the Findings for this section.

Data strategy is not specific to AI. Indeed, a robust data strategy and ecosystem is also important for numerous other policy frameworks and City goals, so when considering potential execution and ownership, the relationship of topics addressed by a data strategy to other high-level strategic areas (e.g., emergency management)

FACING PAGE: 3D LiDAR-derived image of midtown Manhattan looking northwest; Central Park is shown in the top right. This image shows change detection of new buildings highlighted in red, using 2010 and 2014 nDSM LiDAR data.
Source: NYC Department of Finance



should be taken into account. In recent years, this challenging area has seen increased focus and exploration from governments around the world, as well as from longstanding institutions — from banks to hospitals — in the private sector.

Identify common agency pain points and create centralized guidance and template documents that agencies can rely on in their efforts.

There are a number of issues around data that are common across agencies and would benefit from a more centralized approach, rather than requiring agencies to build this expertise in-house. For example, many agencies are rightly concerned about sensitive data they may need to share with cloud providers when using cloud services; creating common guidance can increase agency confidence and streamline their operations, while also helping agencies avoid mistakes and steer clear of risks. The City can study potentially useful approaches, such as centralizing expertise in various aspects of understanding data procurement and templatizing data efforts, integrating digital rights considerations, for the purpose of supporting agencies.

Cultivate more data engineering focus and expertise across City agencies.

Data engineering, roughly speaking, refers to the work involved in making data usable. This involves a set of skills distinct from data science, data analysis, and machine learning. A very common mistake is to hire data scientists or AI experts without having first invested in requisite data engineering and infrastructure, because the former rely on the latter. Importantly, data engineering work also produces broad benefits not limited to AI; even very basic types of analysis that have significant benefits to the City become prohibitively difficult to do when the data is not in a usable form.

Continue to improve standards for and links between the data sent into NYC Open Data to make it more reliable and usable.

The Open Data Law requires that each agency have an Open Data Coordinator (ODC), who is responsible for identifying agency datasets that are appropriate to share publicly and ensure that data is accurate. Each year, NYC Open Data works to improve standards for metadata and documentation and training for ODCs as part of the compliance process. Improved documentation helps to ensure that datasets are presented in context, are less likely to be misinterpreted by users, and can help facilitate ethical use.[66] Continued efforts to improve compliance and documentation standards, as well as creating more links between agency datasets, will save a significant amount of time and effort and help users unlock more value from the City's public data assets.

Boost usability of NYC Open Data for particular tasks and problem domains.

NYC Open Data currently hosts thousands of datasets, but many users, both inside and outside City government, have found it difficult to navigate all of these and to know which combination of datasets is relevant for projects on particular topics like small businesses, air pollution, or transportation. The City, using expertise in these areas within offices like MODA, can indicate which public datasets are relevant to particular topics of interest, how the data fits together, and what appropriate and inappropriate uses of that data might be.

2. City applications

Findings

Many opportunities exist to make good use of AI in City government, but most agencies will require support to identify, assess, and realize them.

Although AI projects must be approached with a great deal of care, as discussed above in the *What is Artificial Intelligence?* section and in *Supplement A*, there are a wide range of potential uses that would benefit New York City and its residents in several different ways. This section describes several qualitatively different uses of AI or

66 T. Gebru, J. Morgenstern, B. Vecchione, J.W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, "*Datasheets for datasets*," arXiv:1803.09010v7, 2020.

related techniques in the City and outlines how they can each serve as useful templates for other agencies and applications.[67]

EXAMPLE: Cyber Command, cyberdefense, and infrastructure applications

NYC Cyber Command is a centralized organization created by Mayoral Executive Order to lead the City's cyber defense efforts, working with roughly 150 agencies and offices to prevent, detect, respond to, and recover from cyber threats. To manage this work, Cyber Command uses custom machine learning systems built in-house and hosted on Google Cloud Platform[68] to collect very large-scale datasets on network activity that are used to flag anomalous behavior via binary classification models. This system was built using over 2.3 petabytes (2.3 million gigabytes) of historical training data and is used to process over 11 billion events each day (hundreds of thousands of model predictions per second), with the system running 24/7 with 99.99% or better uptime.

This sort of infrastructural application does not involve people or communities at all, and the key questions around fairness discussed in *Supplement A* do not arise. There is also simply no question of considering human teams as an alternative; it would clearly be impossible to process data at this scale. However, it is important that the models be interpretable to human operators at the agency itself in order to ensure that things are working correctly and potential incidents can be effectively resolved. Both the technical and organizational approach will be a strong model for other agencies dealing with infrastructure and topics like climate change, especially in situations where it is unrealistic to process the data at the scale needed in any other way.

In addition to internally developed models, Cyber Command is also actively working with third party vendors to access, evaluate, and use other cybersecurity tools, some of which actively use ML for data analysis and detecting malicious activity.

67 The Algorithms Management and Policy Officer (AMPO) in the Mayor's Office of Operations has recently published the first Citywide directory of "algorithmic tools" (defined in AMPO policies), many of which use AI. The examples provided in this strategy may not appear in the report based on the specific criteria established by AMPO policies for public reporting. See "*Summary of Agency Compliance Reporting*," 2020, available at https://www1.nyc.gov/assets/ampo/downloads/pdf/AMPO-CY-2020-Agency-Compliance-Reporting.pdf, and "*AMPO Policies*," 2020, available at https://www1.nyc.gov/site/ampo/resources/policies.page.

68 See https://cloud.google.com/customers/nyc-cyber-command/.

EXAMPLE: Citywide Administrative Services, energy billing, and robotic process automation

The Department of Citywide Administrative Services (DCAS) has an Energy Supply team in the Division of Energy Management (DEM) that is responsible for paying all the City's electricity, gas, and steam bills. DEM receives over 15,000 bills to process each month. The unit also handles budgeting for these expenses, so it is important to monitor what is being spent. However, bills can sometimes have irregular charges that can be difficult to detect, and the billing structure is complex, in addition to the volume being too large to manage solely with human review.

Previously, the bill review team utilized existing reports in an in-house utility billing and reporting system, together with human review, for anomaly detection. Because of the complexity of the logic and number of special cases needed, tackling this problem with traditional software engineering was prohibitively difficult. Instead, DCAS IT built an ML model (using Google's TensorFlow) trained on millions of historic billing records, weather data, and facility information — including year built, square footage, and usage types. The variety in the training data allows the model to factor in the time of year and ambient temperature and to generalize from similar buildings.

In many cases, the model can predict the expected bill very accurately (within cents); where there are discrepancies between the actual and predicted bill amounts that exceed a 10% threshold, the system flags the bill for human review. This system has saved millions of dollars, with DEM receiving $4.5 million in refunds in 2020 and 2021 from over 100 detected billing anomalies from multiple utility companies.

While this is a highly specific use case, this sort of "robotic process automation" (RPA) application — possibly with humans in the loop to review model predictions — is likely applicable in numerous agencies. This could include, for instance, agencies that need to process or prioritize a large number of forms in relatively predictable ways, especially in contexts that are not highly sensitive.

These implementations need not be complicated: It will often be possible to create such systems with high-level tools that do not require deep expertise or manually building custom models. However, even where the purely technical aspects are more straightforward, it is nevertheless important to ensure that the use case does not raise digital rights concerns.

Finally, it is worth noting that this use case falls under the umbrella of "anomaly detection" or "outlier detection," the general term for tasks that involve identifying observations that are rare or unusual in some respect. While the Cyber Command example above is very different in its specifics, it is also an anomaly detection application; this is a good illustration of how the same technical concepts can be applied in a wide variety of cases.

EXAMPLE: Mayor's Office of Criminal Justice, human bias, and social impacts

"Pretrial risk assessment" refers to the process of determining a defendant's risk of failure to appear or to commit other violations that occur prior to a trial. Though this has historically been done by human judges, predictive models are increasingly used to inform these assessments. In this highly sensitive situation, where there is great potential for unfairness and harm, one must approach the area with significant concern over the ethics and other behavior of the system, as discussed in detail in *Supplement A*.

In 2019, the Mayor's Office of Criminal Justice (MOCJ) and the NYC Criminal Justice Agency (CJA) partnered with the University of Chicago's Crime Lab New York (CLNY), social science research firm Luminosity, and behavioral science design firm Ideas42 to evaluate and update the assessments in use in the city, with a special focus on addressing racial or other disparities present.[69] In addition to interviewing a wide range of stakeholders in the criminal justice system, the researchers built a predictive model to assess a variety of components of the existing process and recommended updates to improve both its performance and fairness. They found that the updated process "recommends a far greater number of people for [release on recognizance] while maintaining the current high court

69 See Luminosity and the University of Chicago's Crime Lab New York, "*Updating the NYC Criminal Justice Agency Release Assessment Final Report*," 2020, available at https://www.nycja.org/publications/updating-the-new-york-city-criminal-justice-agency-release-assessment.

appearance rates, substantially reduces the disparity in recommendation rates when considering race/ethnicity and sex, and dramatically reduces the magnitude of false positive rates and the differences in false positive rates based on race/ethnicity and sex."[70] In other words, judges have previously released defendants without any pretrial conditions 65% of the time, while the updated CJA assessment would recommend 89% of defendants for release, and because of the higher accuracy of the tool, this can be done without any projected increase in missed court appearances and while improving performance on racial equity criteria.

70 Ibid.

Recognizing the sensitive nature of any decision-making aids used in a criminal justice context, the project team took a number of special precautions in developing and implementing this tool. From the beginning, the team had the explicit goal of reducing the use of jail and reducing the existence of racial disparities in the assessment process. Two separate research teams were retained to independently analyze the data, produce candidate models, and review all analyses, and a nationally-sourced Research Advisory Council that included leading experts in algorithmic tools and racial bias in criminal justice reviewed each step of the process.

In addition, the team solicited feedback on the process from a variety of stakeholders and worked to ensure transparency in the process, analysis, and field application of the tool. For example, the calculation is public and static so that it can be recalculated by hand in court, and the paperwork given to court parties about each individual includes the source data used to achieve a score to allow for human corrections or departures from the tool's recommendations. This paperwork also included explicit language to emphasize how a person's score translates to their likelihood to return to court, using phrasing that clearly communicates the high rates of overall success. The team further published a detailed research report that includes analysis output and decisions made during the tool construction process, and explicitly reports on trade-offs in play, and the tool's impact on racial disparities. Finally, the team published an ongoing validation plan, monitors outcomes regularly, and offers extra support for individuals with less favorable scores, to help increase return to court when released.

There are several important takeaways from this example. First, systems can be successfully used to measure, identify, and mitigate human biases in existing processes, as long as the project is approached thoughtfully. Such sensitive projects can benefit from agencies collaborating with external experts (discussed further in *Section 4*, below), and there is value in experts creating institutions with different institutional goals and organizational structures than traditional academia, which typically focuses on publications and proofs-of-concept rather than partnerships and direct impact. Finally, such uses of machine learning can help surface trade-offs that were previously left implicit. Deciding to use this system in practice required identifying a clear baseline and making a determination that, even if the new system is imperfect, the impacts on actual people are a significant improvement.

EXAMPLE: Environmental Protection, Health and Mental Hygiene, and noise

Noise pollution is among the leading quality of life challenges for urban residents in the United States. Approximately 90 percent of adults in New York City are exposed to excessive noise levels, meaning beyond the limit of what the EPA considers to be harmful, and noise is the most common complaint for callers to 311.[71] Excessive noise exposure can cause effects on health, including sleep disruption, hearing loss, hypertension, and heart disease, and studies have revealed negative impacts of noise on the learning of children in the form of decreased memory, reading skills and lower test scores.[72]

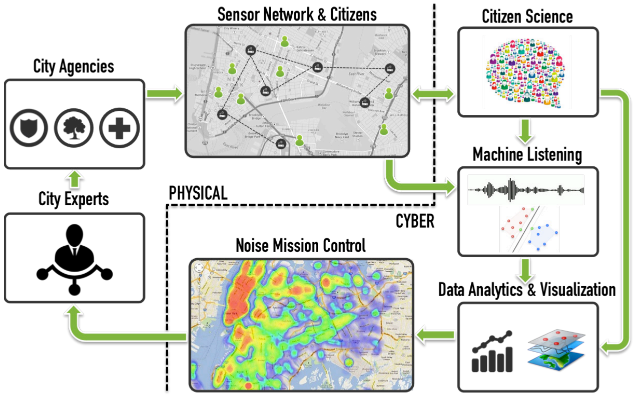
In 2016, a team led by NYU's Center for Urban Science and Progress (CUSP) launched a first-of-its-kind comprehensive research initiative focused on the monitoring, analysis, and mitigation of noise pollution. This project – which involves large-scale noise monitoring – leverages the latest in ML, data analysis, the Internet of Things, and citizen science reporting to more effectively monitor, analyze, and mitigate urban noise pollution. Known as Sounds

71 NYU Center for Urban Science and Progress, et al., "*About SONYC*," available at https://wp.nyu.edu/sonyc/; NYU News, "*NYU Launches Research Initiative to Combat NYC Noise Pollution*," 2016, available at https://www.nyu.edu/about/news-publications/news/2016/november/SONYC_Launch.html; C. Mydlarz, M. Sharma, Y. Lockerman, B. Steers, C. Silva, and J. Bello, "*The life of a New York City noise sensor network*," Sensors, 2019.

72 M. Hammer, T. Swinburn, and R. Neitzel, "*Environmental noise pollution in the United States: Developing an effective public health response*," Environmental Health Perspectives, 2014; M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld, "*Auditory and non-auditory effects of noise on health*," The Lancet, 2014; A. Bronzaft and G. Van Ryzin, "*Neighborhood Noise and Its Consequences: Implications for Tracking Effectiveness of NYC Revised Noise Code*," School of Public Affairs, Baruch College, CUNY, 2007.

of New York City (SONYC), this multi-year project is conducted in partnership with the Department of Environmental Protection (DEP), DOHMH, Department of Transportation (DOT), DDC, Community Affairs Unit (CAU), and NYC CTO.

The SONYC project involves a very different use of ML than the examples in this section. In particular, it uses a distributed network of both sensors and individuals for large-scale noise monitoring. These low-power sensors record street sounds, and ML is used to process the audio and train the sensors to recognize and differentiate between different types of noise (e.g., jackhammers, sirens, music, yelling, or barking). In the later phases of the project, rather than recording audio, the sensors will use ML to recognize individual sound sources and produce statistical reports on sound levels and types, combining it with data from 311 complaints and other citizen noise reports to form an "acoustic model" of the city that City enforcement officials can use to more strategically identify and mitigate noise. Notably, this work uses humans-in-the-loop and special ML techniques to create custom models that are tailored to specific locations and work robustly with noise sources not seen during training.



A high-level diagram of SONYC (Sounds of New York City), a novel "cyber-physical" system for monitoring noise pollution, designed by NYU in collaboration with a range of NYC agencies.
Source: New York University

In this kind of application, the privacy of the audio recordings is an obvious concern to those considering New Yorkers' digital rights, and the project includes a framework to address this. In addition, the team worked with the City to identify communities with high levels of exposure and partnered with community organizations to deploy sensors in residential areas. SONYC shares data with the City via APIs and the Recovery Data Partnership (see *Section 1*, above) and supports local communities in improving noise conditions.

EXAMPLE: High school admissions, matching, and other algorithmic approaches

The examples presented in this section all use ML in different ways. There are in fact a host of other sophisticated technical approaches to or existing techniques for solving practical problems that use the same mathematical building blocks but may not be ML or AI (as described in *Supplement A*). These can include mathematical optimization, resource allocation, matching, planning, routing, simulation, mechanism design,[73] and more, and all of these approaches should also be actively considered as part of an AI strategy whether they are "AI" or not. For example, the matching algorithm used by DOE for NYC high school admissions is not AI or ML, despite being a technical approach to a problem of resource allocation, but many points raised in *Supplement A* still apply.[74]

EXAMPLE: DOF and computer vision applications

DOF's Geospatial Data Intelligence Group uses street and Right of Way imagery (images acquired from a vehicle) of building facades as the basis for automatic extraction of building components using machine learning. These images are acquired from DOF Assessors as well as a geospatial dataset from Cyclomedia. Two recent pilot projects — both of which use ML models called convolutional neural networks, which are common in computer vision applications — aim to identify the absence or presence of staircases in front of buildings, which is helpful in verifying whether or not buildings have an occupied basement, and to identify whether a store is vacant or not. The automatic monitoring of store vacancies may give the City a faster understanding of the changing commercial landscape of NYC and allow DOF Assessors to focus on the most important changes in neighborhoods.

Today, the City's capacity to make use of AI in these ways is uneven.

While some agencies have the staff, knowledge, computer hardware, existing contracts, and other resources needed to leverage AI, others do not. Because needs, agency sizes, and missions vary,

73 T. Börgers, *An Introduction to the Theory of Mechanism Design*, Oxford University Press, 2015; see also "*Mechanism Design for Social Good*," at https://www.md4sg.com, for work on current applications in the public interest.

74 This is a method from the field of market design known as "deferred acceptance"; see A. Abdulkadiroğlu, P. Pathak, and A. Roth, "*The New York City School Match*," 2005. The National Medical Resident Matching Program is run similarly, and these techniques were recognized with the 2012 Nobel Prize in Economics, see A. Roth, "*The Theory and Practice of Market Design*," at nobelprize.org. The FCC uses related methods, based on auctions, to allocate US broadband spectrum, and these methods were recognized with the 2020 Nobel Prize in Economics, see P. Milgrom and I. Segal, "*Designing the US Incentive Auction*," and P. Milgrom, "*Auction Theory Evolving: Theorems and Applications*," at nobelprize.org.

capacity needs will also be specific to each organization. While agencies do have the option to supplement internal staff with consultant engagements, these can potentially lead to missed opportunities for internal capacity growth or even weakened governance if learnings or resources are not fully transferred to City staff during or after an engagement. The broader issue of workforce development and City capacity is discussed further in *Section 5*, below.

Opportunities

Select a set of agencies to conduct an internal review for potential AI applications based on template examples.

The examples included in this section's *Findings* outline a range of different types of applications where the City could consider productively using AI: infrastructure applications, streamlining agency operations, making human decisions more transparent and rigorous, directly addressing social problems or delivering social services, and more.

The City can select a diverse mix of agencies — focused on anything from climate change to social services — to conduct internal reviews that identify potential applications of AI or ML in service of their agency's mission, along with any questions or concerns they may have and a high-level sense of resources that may be required. A central interdisciplinary team, including existing practitioners like those mentioned in the examples just described, can help review these potential projects and make recommendations for how best to proceed in the various cases, as well as identify common pain points that may be better addressed outside individual agencies. This should include recommending partnerships with external teams or experts that could help execute the projects, as well as helping to navigate policy considerations and ethical risks.

Bring together agency technical leadership with experience using AI methods and create practical guidance materials for Citywide use.

In 2019, Executive Order 50 created the position of Algorithms Management and Policy Officer (AMPO) in the Mayor's Office of Operations, and outlined a range of functions and tasks for that office. Among these was to "research new developments and best practices…" and to "remain current…" with respect to algorithmic tools. In support of this mandate, and to build City agency capacity to make use of AI tools, key personnel involved in the projects described above, as well as other projects and systems detailed in the AMPO report, could be convened to discuss their experience with these projects. This might include a discussion of organizational and bureaucratic components (such as getting buy-in or support from non-technical agency leadership and OMB, and determining resources needed), as well as discussion of technical approaches used. Such a group might produce value through a report targeted to leadership of City agencies that describes their experiences and makes recommendations about how to approach such projects in other agencies. This group can also serve as an informal community of practice that agencies can approach with candid questions on an ongoing basis.

Support agencies' ability to work with standard ML software and hardware.

Partly because of its mathematical nature, ML systems are typically built with different technologies and software packages than some agencies may be familiar with. In particular, it is standard to use open source libraries — such as scikit-learn, tidyverse, PyTorch, and TensorFlow — and usually in the programming languages Python or R. While Python is a general purpose language suitable for many tasks, R is a specialized statistical language commonly used by statisticians, data scientists, and social scientists. Agencies who are used to using technologies from vendors like Microsoft may not realize both that Microsoft's own offerings are compatible with these (in Azure, for instance) and that it is important to use standard tools for ML work (for a range of reasons, including maintainability, feature support, overall performance, and others).

In consultation with NYC CTO and other practitioners at agencies like those described in this section, agency CIOs and CTOs can proactively educate themselves and their teams on these technologies and what would be involved in using them or contracting with vendors that use them. Finally, for certain newer techniques like "deep learning," specialized hardware (specifically, GPUs) may be needed; many people access such hardware through standard cloud services offered by Amazon, Microsoft, Google, and others.

3. City governance

Findings

AI and its applications are evolving rapidly, and society is grappling with these changes on an ongoing basis. There is a pressing need for the City to remain responsive to the changing technological and social landscape in its policy and governance efforts.

The City has developed an initial foundation for AI governance and policymaking that can be strengthened and leveraged for the next phases of work, but this next phase must at its core acknowledge how much is still unknown.

There is opportunity to build on existing structures and policy developments as the City moves toward greater maturity in its governance of AI. The City must adopt a holistic, iterative, interdisciplinary, evidence-driven, and technically grounded approach to policymaking in AI to reduce the possibility of inconsistency, gaps, or conflict across individual efforts and avoid, to the extent possible, unintended negative consequences. As one simple example, the City of San Francisco had to amend an ordinance on facial recognition when it was found that it inadvertently banned the use of standard smartphones that include a version of the technology.[75]

75 T. Simonite and G. Barber, "*It's Hard to Ban Facial Recognition Tech in the iPhone Era*," Wired, 2019, available at https://www.wired.com/story/hard-ban-facial-recognition-tech-iphone/.

Procurement and contracting are central topics in AI governance.

The decision to build systems in-house or to rely on vendor products or services can have a significant impact on the visibility and control the City has in the systems it employs. There are unique considerations when contracting AI versus other technologies, as discussed in *Supplement A*.

Today, some agencies have the capacity in place to build systems in-house, but others must rely on vendor contracts to make use of AI. The City has robust procurement rules and highly structured citywide and agency-level processes for soliciting proposals and negotiating contracts. There is strong expertise at key agencies to evaluate bids and negotiate contracts that result in fair and responsible systems. The City can build its internal expertise more widely as a broader set of agencies begin to integrate AI.

Several agency representatives interviewed for this Strategy raised the potential value of new, central shared resources to help agencies navigate some of the unique issues that arise in procuring datasets or AI technologies, for several reasons. These could include building and maintaining institutional expertise, offering tools to support rigorous evaluation of (often unfounded or exaggerated) vendor claims about their use of "AI," and helping streamline agency efforts via, for example, offering an "AI rider" that could be added to contracts to help tackle concerns about malfunctioning or even unethical AI systems (see *Supplement A*).

For example, NYC Opportunity has requested that vendors provide "plain language documentation" about the underlying logic of their coding that can be understood by a broad audience. This effort aims to obtain information that is both more comprehensive than a white paper or a high-level description of functionality, and more comprehensible than code or technical specifications.

EXAMPLE: Algorithms Management and Policy Officer

As noted, the Mayor's 2019 Executive Order 50 created the position of Algorithms Management and Policy Officer in the Mayor's

Office of Operations. This new position was charged with developing a policy framework and set of management practices centered around fair and responsible use of "algorithmic tools" by City agencies. Key among those management practices are protocols around assessing tools, providing members of the public with channels for inquiry and complaint, and critically, establishing public reporting and transparency processes, all of which relate closely to the opportunities identified later in this section.

For the latter, in 2021, the AMPO released the results of its initial agency compliance reporting effort, summarizing sixteen algorithmic tools being used by nine City agencies and offices.[76] (The AMPO's focus is on "algorithmic tools" rather than AI per se; there are algorithmic tools that do not use AI and there are AI systems that are not considered algorithmic tools.) This is a first-in-the-nation report on government use of algorithmic tools, with participation from the nearly 100 agencies and offices covered by Executive Order 50, and will serve both as a critical foundation for ongoing work within New York City as well as a model and inspiration for other cities around the world.

The AMPO builds on a broader ethos of responsible data governance and use within the City, also reflected in the mission of offices like MODA, NYC Opportunity, the Mayor's Office of Information Privacy (MOIP), NYC CTO, and by the goals of the NYC Open Data program and EquityNYC.[77] The AMPO is also advised by a Steering Committee consisting of senior representatives or leaders of a number of different offices and agencies.

Importantly, several of these offices have established networks of official liaisons within every City agency: Each agency has an AMPO Liaison, an Agency Privacy Officer, and an Open Data Coordinator. These networks are a key resource that could be used in new ways in the future — for example, to disseminate education, training, and best practices across the entire City government (City staff training is discussed further in *Section 5*, below).

76 NYC Algorithms Management and Policy Officer, "*Summary of Agency Compliance Reporting*," 2020, available at https://www1.nyc.gov/assets/ampo/downloads/pdf/AMPO-CY-2020-Agency-Compliance-Reporting.pdf.

77 For more on NYC Open Data and EquityNYC, see https://opendata.cityofnewyork.us/ and https://equity.nyc.gov, respectively.

EXAMPLE: NYC Internet of Things Strategy

In March 2021, NYC CTO published an Internet of Things (IoT) Strategy for New York City.[78] The NYC IoT Strategy describes the landscape of IoT usage across society, explores treatments of the technology in educational and policy settings, outlines the state of New York City's IoT ecosystem, and establishes a set of critical near-term actions toward creating a healthy, cross-sector IoT ecosystem in New York City – one that is productive, responsible, and fair. The NYC IoT Strategy also follows an approach grounded in digital rights and is built around six key principles: governance and coordination, privacy and transparency, security and safety, fairness and equity, efficiency and sustainability, and openness and public engagement.

78 NYC Mayor's Office of the Chief Technology Officer, "*The New York City IoT Strategy*," 2021, available at https://www1.nyc.gov/assets/cto/downloads/iot-strategy/nyc_iot_strategy.pdf.

This strategy builds on years of recent work, including technology pilots like parts of the SONYC project described in the previous section. Following this publication, the City, led by NYC CTO, is actively engaging with the public about the strategy, has established an internal Smart City Collaborative to foster information sharing and collaboration among agencies, and is developing an impact assessment on the responsible use of IoT technologies targeted at a general audience.

The city's ecosystem has unique characteristics that make it ripe for innovation and leadership in participatory, human-centered approaches to AI.

New York City's rich civic technology community has fostered a variety of existing initiatives for engaging the public in making government more transparent and responsive. These include efforts to engage city residents in making and improving City data, applications, and services — such as the SONYC project described in *Section 2* — as well as varied efforts to educate New Yorkers and build awareness on key issues and initiatives. This work happens both within City government through offices like the Office of the Public Advocate (OPA), MODA, NYC Opportunity's Service Design Studio, AMPO, NYC CTO, and through public libraries and community centers, among other entities. It is also a focus beyond the City, in non-profits, research institutes, university labs, and community organizations dedicated to design, technology, data, and public participation. This unique nexus of organizations and talent across sectors positions the City to adopt and adapt participatory approaches in AI.

Crucially, best practices and standards for robust public engagement in AI are not yet agreed upon and are themselves the active subject of current research. The complexity of these systems will require new, innovative methods to enable robust and meaningful participation. The fact that public engagement in AI is at a very early stage presents an excellent opportunity for New York City to creatively explore and help define this space, but these initiatives should be approached in an experimental spirit.

Opportunities

Promote experimental, empirical policymaking.

The City can establish an internal working group to support a holistic and coordinated approach to AI policymaking and procurement, consisting of both agencies procuring AI and those focused on technology policy. This working group can convene a community of practice to iterate policies and protocols, elevate best practices, and seek input on emerging matters of shared concern. It should be informed by structured engagements across sectors and with communities to maintain visibility into emerging concerns.

Support robust visibility into internal systems.

Whether built in-house or purchased via contract, it is important that the City has robust and meaningful visibility into each system it employs by understanding how it works, how accurate it is, under what conditions, and how its performance was initially evaluated. Agencies should have a clear understanding of the uses and limitations of system outputs, train staff members that interact with the systems to recognize situations when errors are likely, and monitor for performance changes over time. In some cases, this work could benefit from appropriately designed partnerships with local experts; the topic of partnerships is discussed further in the next section. This will require a collaborative effort between policy and oversight offices, such as AMPO, CPO and MOIP, and NYC CTO, as well as offices focused on procurement.

Work to incorporate ongoing, regular review of deployed systems.

Some important concerns with AI systems are that they can be built around an inherently faulty premise, they can turn out not to work at all in practice, and their performance can change and possibly degrade over time, for reasons described in *Supplement A*. As a result, there is a unique and acute need to incorporate ongoing monitoring of deployed systems — whether procured or built in-house — rather than on one-time upfront reviews, even if those are more exhaustive. This is different from other technology governance regimes that reasonably rely on upfront review rather than ongoing monitoring.

Adapt models for community engagement and participatory approaches.

The City can build on existing programs and partnerships to engage New York City's rich local community of civic technologists and community organizations to innovate in public participation in the design, use, governance, and policymaking related to AI systems. This can include broad-based public education efforts, fostering participatory approaches to use case identification, engaging the public throughout system engineering, or other activities. A mix of agencies, including both technical and policy offices, as well as existing agencies focused on community and public engagement, can share learnings on how to most effectively engage the public throughout the AI lifecycle. A practical case study on using participatory design — and the benefits of this approach — for an AI system used to manage allocation of food donations is described in *Supplement A*.

Community members in a NYC[x] Co-Labs participatory workshop on access to tech education in Inwood and Washington Heights.
Source: NYC CTO

To understand the impacts of AI systems that directly deal with people, it is typically necessary to have both quantitative and qualitative information about how those systems interact with different populations, and to account for how those interactions may vary across populations. Although sometimes the issues in question are specific to AI, they often relate to broader social, historical, or other factors, such as income, education, health, internet access, or any number of other areas.

4. Partnerships

Findings

New York City is home to world-class AI research and talent that can complement City capacity, but both agencies and external stakeholders desire stronger conduits to collaboration.

The City's capacity to make use of AI can also be productively enhanced through partnerships with a range of external groups. As noted in the above ecosystem description, the city is home to a tremendous amount of AI research activity and talent across academia, non-profits, and the private sector. Moreover, because many of the key policy and ethical questions around AI are still open research questions (see *Supplement A*), applied research collaborations are essential.

Today, both the City and these entities face difficulty in identifying and navigating partnerships. The City's needs and organizational structure are often opaque to external groups, and avenues to engage are unclear. Likewise, the City is often unaware of projects and resources available externally. There can be a range of administrative and logistical challenges to navigating these kinds of partnerships. Administrative and legal approval processes can be lengthy and involve multiple entities and steps. Timelines for the respective parties often do not align, due to differing timeline needs for outputs or because academic year calendars dictate student involvement in City work. External partners also often have different needs or approaches to data sharing or publishing findings. There are structures and programs in place in the City today, as well as in other cities, that address some of these challenges in other domains, and that can be viewed as building blocks for a comprehensive approach.

EXAMPLE: Town+Gown:NYC and academic research services procurement

Town+Gown:NYC is a citywide research program, established by the Department of Design and Construction (DDC) in 2009, that has created a variety of structures and events to coordinate City research needs with university capacities related to the built environment. This includes a citywide Master Academic Consortium Contract that all City agencies can access to procure faculty-led academic research services from fifteen institutions, including CUNY, Columbia, NYU, Fordham, and many others.

A Town + Gown event. *Source: NYC Department of Design and Construction*

Town+Gown:NYC also has an experiential learning component that facilitates "in-kind exchanges" between agencies and students from the academic consortium and other schools, and it conducts knowledge-sharing events based on both faculty- and student-led work.[79]

79 For more on the Town+Gown program, see https://www1.nyc.gov/site/ddc/about/town-gown.page.

Although Town+Gown describes itself as being focused on the built environment, the program can serve as a broader model for academic collaborations that can also be targeted to areas such as data science and AI. In fact, many agency research needs can already be accommodated under the Master Academic Consortium Contract itself, as DDC considers the "built environment" to be a wide-ranging "inter-discipline" that includes technology as one key component.

Results from two recent projects under Town+Gown:NYC. The first, "Pathways to Carbon-Neutral NYC," provided the most comprehensive analysis to-date of scenarios for NYC's energy supply and demand through midcentury. The second, "NYC Stormwater Resiliency Plan," is the first-ever citywide analysis of rainfall-based flooding and enabled the City to develop new plans. *Source: NYC Department of Design and Construction*

EXAMPLE: Austin, Texas, Good Systems, and the Master Interlocal Agreement

In a related vein, the City of Austin and the University of Texas, Austin (UT Austin) had a range of ongoing collaborations, but found that it was difficult to keep track of the numerous ways in which these collaborations were happening. In addition, they found that research agreements being processed as business or procurement contracts significantly slowed down contract negotiation, that there were no agreed-upon best practices for research engagements, and that there were many missed opportunities to connect government offices with research expertise.

To address these challenges, the city and university created a "Master Interlocal Agreement." This agreement formalizes these research collaborations and provides a consistent, citywide framework for engaging in them. Active for five years, the agreement authorizes funds for research activities, and allocates responsibility appropriately between the city and the university. Importantly, basic terms and conditions regarding data sharing and intellectual property have already been negotiated in the ILA to streamline processing, while leaving room for project-specific term negotiation. This framework also makes it possible to conduct collaborative projects on sensitive topics and with sensitive data. Some of the projects enabled by the ILA include determining optimal locations and times for lane closures on arterial streets to minimize traffic delays, and testing the use of automated object recognition/tracking in video data streams to support the assessment of pedestrian safety.

In addition to projects that directly build on the ILA, UT Austin has created a program called Good Systems,[80] a seven-year, $10 million collaborative applied research effort to design AI technologies that benefit society.[81] Many of these projects include collaborations with a number of different agencies in the City of Austin.

In addition to existing programs like Town+Gown, initiatives like this, and similar partnerships in other cities and jurisdictions, can provide helpful templates that New York City can build on and learn from.

80 For more on Good Systems, see https://bridgingbarriers.utexas.edu/good-systems/.

81 Corridor News, "*City of Austin, University of Texas Formalize Research Partnership*," 2020, available at https://smcorridornews.com/city-of-austin-university-of-texas-formalize-research-partnership/. For more on the origin and structure of the Master Interlocal Agreement, see also "*City of Austin & UT Austin Teaming Workshop*," at https://utexas.app.box.com/s/yblyz9jr3qagkaku4h9ixzr5kzy1cjck.

EXAMPLE: NYC[x] Innovation Fellows and volunteer talent

The City has also implemented a variety of successful programs for bringing practitioners from the private sector into its work on a short-term basis to meet internal needs, while offering valuable public interest experience to external volunteers. For example, NYC CTO, in partnership with the non-profit US Digital Response, established the NYC[x] Innovation Fellows program in 2020 to embed three- to five-person teams to work on concrete projects for a wide range of agencies to rapidly solve specific challenges through the use of lean, agile, and user-centered methodologies.[82] These projects have addressed topics including hate crimes, equitable broadband adoption, language translation, services for the aging, and CityPay, which is a DOF payment portal that handles $21 billion in transactions annually.

82 For more on this program, see https://www1.nyc.gov/assets/cto/#/project/nyc-x-innovation-fellows.

EXAMPLE: Public interest technology and other research labs

As noted, there are a number of local research labs at universities (or groups of universities) aimed at working on problems in the public interest, many of which intersect with AI. These include NYU's GovLab, Center for Urban Science and Progress, and AI Now Institute; Cornell Tech's Urban Tech Hub and Public Interest Tech Studio; and a range of projects at Columbia University's School of International and Public Affairs and Data Science Institute. There are also other leading organizations, such as the University of Chicago's Crime Lab New York and Ideas42, both of which have worked with the City on criminal justice projects, described in *Section 2*.

In addition to universities, a great deal of AI research occurs in industrial research labs. As previously described, New York City is home to a number of internationally renowned labs, including Facebook AI Research (FAIR), Microsoft Research (which includes a prominent team focused on "Fairness, Accountability, Transparency, and Ethics in AI"), and Google Brain. In addition to publications, these labs also publish critical open source software — particularly TensorFlow and PyTorch — that power the broader AI ecosystem. These labs, in addition to the universities they collaborate with, also attract top AI talent from around the world to live and work in New York City.

As noted, CUNY, as the largest urban public university system in the country, plays a crucial and unique role in the local ecosystem, due to its sheer scale, wide range of both educational and research programs, and enormously diverse student body. In addition, as described, the Public Interest Technology University Network (PIT-UN) is a collaboration among dozens of higher education institutions — including CUNY, Columbia, Cornell and NYU, as well as Harvard, MIT, Stanford, and other leading academic institutions across the country — committed to building the field of public interest technology by "growing a new generation of civic-minded technologists and digitally-fluent policy leaders."[83]

83 See https://www.newamerica.org/pit/press-releases/31-million-awarded-grow-field-public-interest-technology-leading-colleges-and-universities/.

EXAMPLE: NYCEDC RFEI for NYC Center for Responsible Artificial Intelligence

In May 2019, NYCEDC released a request for expressions of interest (RFEI) to secure proposals for the development and operation of the NYC Center for Responsible Artificial Intelligence (AI), an innovation, collaboration, and applied research space designed to support the creation of responsible data science and AI in New York City.[84] The Center was to convene leaders from NYC startups, large companies, government, community, and academia to participate in a set of programs that would establish ethical practices and build trust in AI. It aimed to serve as a nexus and driver of responsible innovation for the city's growing AI ecosystem, building on New York City's reputation as a leader of inclusive and ethical implementation of cutting-edge technologies.

84 NYCEDC's announcement of this effort can be found at: https://edc.nyc/press-release/nycedc-seeks-proposals-develop-and-operate-nycs-center-responsible-artificial.

These objectives were to be achieved through a physical space housing four key activities that integrate ethics as a core component of AI and data applications: 1) An Applied Research Lab to create best practices and tools that integrate ethics when developing and testing real-world AI applications; 2) Data Collaboratives to develop and test data sharing models to protect privacy and enable responsible innovation; 3) A Talent and Education Program to develop rigorous ethics curricula for data and computer science students and industry professionals; and 4) An AI for Good Program to provide space and resources to support companies that create and use AI for social good.

The City planned to partner with local and global leaders to invest up to $7 million to support these programs that would ensure innovations in data and AI technologies benefit New Yorkers. NYCEDC was in discussions with RFEI respondents prior to the COVID-19 pandemic, but procurement for the Center was paused until further notice due to COVID-related budget realities.

Opportunities

Designate a central City team to match agency needs to academic institution capacity and support ongoing partnerships.

In universities, there is significant value in centralized research offices, often under the university's Office of the Vice Provost for Research. These offices can serve as a central point of contact for navigating large universities. A similar central point of contact within the City would significantly ease administrative burden both for external stakeholders and City agencies, and could be modeled on the Town+Gown:NYC program or other such initiatives.

Such efforts should prioritize diversity and inclusion along demographic lines, but also take into account different types of institutions, domains, and disciplines; public universities and community colleges should be included, as well as non-technical disciplines and New York City students and faculty across a wide range of backgrounds.

Foster partnership opportunities with external experts.

The City can take advantage of New York City's rich local research and practitioner community and expand its capacity to use and plan for AI by fostering partnership opportunities with external experts. Forging and managing these partnerships can be eased by central coordination, as noted in the Findings for this section.

Moreover, partnerships can be designed for mutual benefit to support ongoing value and use: offering local researchers and practitioners valuable applied research or civic tech opportunities while delivering needed expertise to City efforts. Such partnerships offer the added benefit of fostering greater engagement across sectors, which can support greater awareness and alignment around shared goals.

The City could explore expanding on successful models like the MOCJ partnership with Luminosity and CLNY, and the NYC[x] Innovation Fellows program, to broaden the organizations and areas of expertise it can bring in.

Create ongoing structures for engagement across different groups in the city's AI ecosystem as well as with City agencies.

To strengthen and deepen relationships and mission-alignment across departments, and sectors, the City can act as a convener within the local ecosystem — gathering researchers, civic tech and advocacy groups, industry practitioners, and community groups to share work in a bi-directional way, and discuss emerging issues and opportunities. There are a variety of models the City might use in this regard, all of which should emphasize interdisciplinarity, equity, and inclusion. For example, major international AI conferences are increasingly moving to a hybrid model where there will be multiple "watch hubs" around the world for participants to gather with other researchers locally, while participating in the conference itself virtually. This trend accelerated significantly during the pandemic. The City could consider providing physical space and convening the local ML community to participate in these conferences together, as these could evolve into major gatherings that organically benefit the overall ecosystem in a variety of ways. Because of the high level of interest in AI across a broad range of stakeholder groups, these conferences are attended by many thousands of people, as well as a huge number of corporations, startups, and other organizations, and serve as major recruiting and social events, in addition to showcasing academic research.

Participants at Neural Information Processing Systems (NeurIPS), a leading machine learning and AI conference. NeurIPS and other major conferences attract many thousands of attendees globally. *Source: Mila – Quebec Artificial Intelligence Institute*

5. Business, education, and the workforce

"It is in keeping with the American tradition — one which has made the United States great — that new frontiers shall be made accessible for development by all American citizens."

Vannevar Bush
"Science: The Endless Frontier"
Former Director, US Office of Scientific Research and Development

Findings

There are multiple local efforts in place to develop the local AI workforce, and the city is home to strong institutions and programs that can be leveraged for this effort; strategic and operational coordination of these efforts can boost their impact.

From the growing use of AI in hiring and management, to the increasing automation of tasks and roles, to the expansion of jobs involved in creating, integrating, and managing AI and related systems, AI is impacting New Yorkers' working lives in a wide variety of ways, and these impacts only stand to increase in the years to come. Accordingly, the City's approach to AI education and workforce development should take into account the full range of needs throughout a person's life, preparing New Yorkers to thrive in workplaces and roles impacted by AI growth in all of these various ways.[85]

AI is relevant in the City's approach to K-12 education, digital literacy programming, reskilling initiatives, internship programs, post-secondary training, and many other efforts. The divergent and shifting landscape also makes the need for a coordinated, strategic approach particularly pressing. The COVID-19 pandemic has recently accelerated movement on some of these fronts. For example, the past year has incentivized many companies to do more automation, which can impact both job security and job quality — across both white-collar and blue-collar roles.[86] Further, there is research to suggest that the populations hardest hit by the pandemic are also at most risk of job loss via automation — for example, those working in food preparation and other front-line occupations.[87] Across the City's efforts, it will be critical to address inequities that exist in terms of both the impact AI is having in the workplace, and in terms of access to information, skills, and opportunity.

85 For some discussion of the use of AI on hiring and the workforce (sometimes referred to as "algorithmic hiring" and "algorithmic workforce management"), see, e.g., M. Raghavan, S. Borocas, J. Kleinberg, and K. Levy, "*Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*," ACM FAccT, 2020; A. Mateescu and A. Nguyen, "*Algorithmic Management in the Workplace*," Data & Society, 2019, available at https://datasociety.net/wp-content/uploads/2019/02/DS_Algorithmic_Management_Explainer.pdf; "*AI Now 2019 Report*," section 1, AI Now Institute, available at https://ainowinstitute.org/AI_Now_2019_Report.pdf; D. Lynch, "*Hiring troubles prompt some employers to eye automation and machines*," The Washington Post, 2021, available at https://www.washingtonpost.com/business/2021/05/19/automation-labor-economy/.

86 Ibid. Also see, e.g., K. Roose, "*The Robots are Coming for Phil in Accounting*," The New York Times, March 6, 2021, available at https://www.nytimes.com/2021/03/06/business/the-robots-are-coming-for-phil-in-accounting.html.

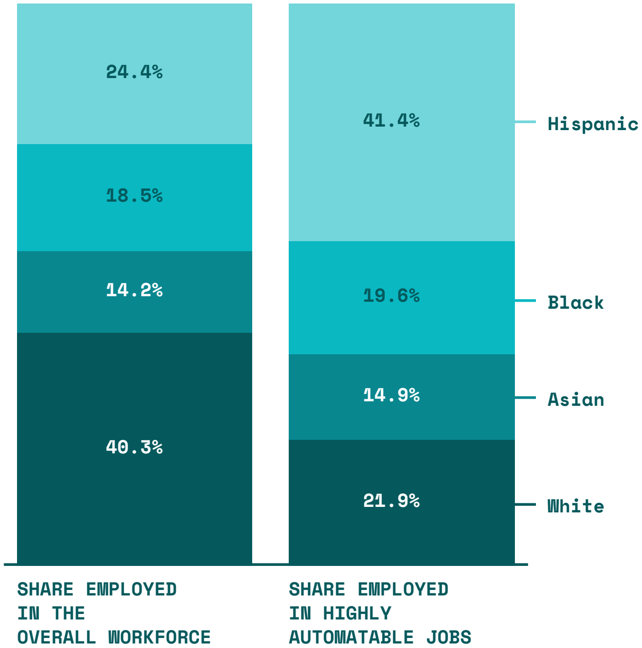
87 E. Dvorkin, C. Shaviro, and L. Gallagher, "*Upskilling for an Equitable Recovery: Hardest Hit New Yorkers Most Vulnerable to Automation*," Center for an Urban Future, 2021, available at https://nycfuture.org/research/automation-and-equitable-recovery.

It is important that the City's efforts recognize that "AI jobs" are diverse, and can include not just those focused on building AI models, but also data cleaning, design research, product and project management, and other roles. These roles are often given a wide range of nonstandard job titles, and exist in companies in a variety of sectors, not just technology companies.[88] This can make this labor market challenging to navigate both for employers and candidates. Further, reskilling or upskilling existing and mid-career workers is a key need as well as an important tool for fostering "bilingualism" in the local ecosystem. Supporting the development of individuals and teams that combine technical skills with domain expertise can also be an impactful way to both effectively and responsibly leverage AI.

88 See, for example, "*Lecture 13: ML Teams and Startups*," Full Stack Deep Learning, 2021, available at https://fullstackdeeplearning.com/spring2021/lecture-13/.

The City's own workforce is not exempt from this broader transformation, and will have an ongoing need to develop and maintain the skills to use, govern, and plan for AI effectively and responsibly. Indeed, many of the opportunities outlined in this Strategy will rely on the ongoing development of skills and capacity among City workers. So it will be vital to include the civil service in the City's thinking.

New York City has a robust infrastructure in place for education and training that can be leveraged for these efforts — from its public schools, to its leading universities, to its community-based training programs. And there are numerous programs in place today to address AI skills needs. There is opportunity to better connect the myriad programs in place to ensure that residents can progressively develop skills and access resources as they move toward and through employment opportunities and that employers have ready access to diverse talent.



A 2021 analysis from Center for an Urban Future found that highly automatable jobs are more often held by people of color.
Source: Adapted from table in "Upskilling for an Equitable Recovery," Center for an Urban Future (CUF), 2021

EXAMPLE: CS4All

In 2015, the City's Department of Education launched the Computer Science for All (CS4All) initiative, toward bringing Computer Science education to every elementary, middle, and high school by 2025, with an emphasis on female, Black, and Latino students. The program offers training for teachers and administrators, as well as curriculum materials that can be integrated according to individual schools' goals — with a broad focus on foundational skills and concepts, including data and algorithms, among others.[89]

EXAMPLE: Tech Talent Pipeline

The City's Tech Talent Pipeline (TTP) engages with local tech industry stakeholders to understand their workforce needs and establish an inclusive pipeline of New Yorkers equipped to fill them. This includes developing training programs, working to increase the number of CUNY Computer Science graduates, and connecting New Yorkers to tech internships and jobs.[90] Many CUNY schools have also offered training through the Tech-in-Residence Corps, a collaboration with TTP to bring industry practitioners into schools to develop courses and teach in-demand skills needed to enter the workforce.[91] TTP's focus in addressing AI-related needs to date has been on the delivery of industry-aligned advanced courses such as Foundational AI, Data Science, Machine Learning, Deep Learning, Ethical AI, Natural Language Processing, research capstone projects, and more.

EXAMPLE: Supporting diversity in the local tech workforce

Acknowledging the "historic and systematic barriers for Black, Latina, Indigenous, low-income, first-generation, and other marginalized people" in AI and machine learning, Cornell Tech's Break Through AI program encourages women and nonbinary college students in the New York area to get training and build a portfolio of industry-relevant experience to help them get entry level jobs.[92] Free to participants, the program also offers mentoring and support through the job application process.[93] This program is part of a broader, national Break Through Tech (BTT) program, which

89 For more on the CS4All initiative, see http://cs4all.nyc/.

90 For more on the range of TTP initiatives, see https://www.techtalentpipeline.nyc/.

91 For more on the Tech-in-Residence Corps program, see https://www.techtalentpipeline.nyc/tech-in-residence-corps.

92 For more on this program, see https://tech.cornell.edu/impact/break-through-tech/break-through-ai/.

93 Ibid.

"work[s] at the intersection of academia and industry to propel more women and underrepresented communities into technology degrees and careers.[94] Break Through Tech New York, the flagship BTT partnership between Cornell Tech and CUNY, offers a broader array of internship and training programs to support women entering technology roles in the city.[95]

In a similar vein, with support from the PIT-UN and New America, CUNY's College of Staten Island (CSI) St. George extension has established intentional learning communities with a focus on public interest technology, toward building tangible pipelines from high school into higher education. The goal of this work is to build diversity into the tech talent pipeline and to prepare students with technology skills marketable in the current and future workforce. Students take the majority of their general education courses that lead to an Associate's Degree (and beyond) with a focus on public interest technology. The program provides students with the opportunity to interact with social justice-focused tech entrepreneurs through the CSI Tech Incubator.[96]

EXAMPLE: LifeSci NYC

A $1 billion investment in life sciences research and development (R&D) and related innovation led by NYC EDC, LifeSci NYC is expected to create nearly 16,000 new jobs by 2026 — many accessible to New Yorkers without advanced degrees.[97] This builds on the existing city ecosystem of nine major research centers, over 50 hospitals, more than 100 research centers that have collectively been awarded more than $2 billion in annual NIH federal research funding, a highly talented and diverse workforce, and industry-leading companies that have attracted more than $1 billion in annual venture capital investment.[98] This effort places a specific focus on connecting existing researchers and institutions to the resources needed to advance programs to commercialization, unlocking space for life sciences companies to grow within the city, and building a pipeline for the talent and workforce needed to support these companies across the five boroughs.

94 Break Through Tech, "*Mission & Vision*," at https://www.breakthroughtech.org/mission-vision/.

95 For more on these programs, see https://www.breakthroughtech.org/where-we-work/new-york/.

96 A. Adams, "*Building a Pipeline into PIT at College of Staten Island, CUNY*," 2021, available at https://www.newamerica.org/pit-un/blog/building-a-pipeline-into-pit-at-college-of-staten-island-cuny.

97 For more on LifeSci NYC, see https://edc.nyc/program/lifesci-nyc.

98 Ibid.

EXAMPLE: Analytics Exchange and other City professional development programs

MODA's Analytics Exchange (AnEx) is an internal community of practice that convenes quarterly to provide City analytics professionals with opportunities to showcase their work, learn new skills, and discuss shared interests and concerns. There are additionally a range of external sources that might be tapped to build capacity within City staff, such as online courses or publicly available training materials offered by technology companies,[99] and DCAS offers a range of citywide professional development resources.[100]

Governments, experts, and the public at large currently have limited visibility into how or to what extent private sector companies are using AI, what their workforces look like, what roles are changing due to automation, and where companies may have emerging needs and interests.

Further, some of the ways in which private sector use of AI is impacting New Yorkers' lives can be particularly hard to discern and track. For example, identifying where and how AI is impacting the quality of New Yorkers' work, how it may affect populations differently, and how community values can shift with respect to AI systems from one context to the next can all be crucial for both policy and design decisions, but are often opaque, absent dedicated research.[101] There is opportunity to learn more about the state of private sector AI use and its impacts, and about the AI workforce, to foster better planning and policymaking, support effective workforce development initiatives, and to create more transparency in the field, broadly.

EXAMPLE: Data on AI and the workforce

There has been some important initial work to address some of the blind spots around what is happening with AI in the private sector.[102] For the last several years, Stanford's Institute for Human-Centered AI (HAI) has produced an annual "AI Index" report that tracks and collates a wide range of data on AI activity around the world.[103] Similarly, Georgetown's Center for Security and

99 See, e.g., Google AI's collection of professional educational materials at https://ai.google/education/ and Microsoft's at https://www.microsoft.com/en-us/ai/ai-school.

100 For more on DCAS training programs, see https://www1.nyc.gov/site/dcas/agencies/citywide-training-and-development.page. DCAS offerings include its Citywide IT Training Program, which "*...provides access to over 200 high-quality technical training courses and 60 certifications from industry leaders [and also operates] Organizational and Executive Development programs*" (Ibid). The City's current list of subsidized degree programs is available at: https://www1.nyc.gov/site/dcas/agencies/mayors-graduate-scholarship-participating-schools.page.

101 Expectations of transparency, for instance, can be quite different in different contexts. For example, researchers have found that people expect different sorts of transparency from systems used for medical diagnosis as opposed to hiring decisions. See also, The Alan Turing Institute, "*Project ExplAIn*," available at https://www.turing.ac.uk/news/project-explain.

102 M.R. Frank, D. Autor, J.E. Bessen, E. Brynjolfsson, M. Cebrian, D.J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, and D. Wang, "*Toward understanding the impact of artificial intelligence on labor*," Proceedings of the National Academy of Sciences, 2019, available at https://www.pnas.org/content/pnas/116/14/6531.full.pdf. See also, S. Athey, "*Testimony, Committee on the Budget, U.S. House of Representatives Hearing on Machines, Artificial Intelligence, and the Workforce: Recovering and Readying Our Economy for the Future*," 2020, available at https://hai.stanford.edu/news/susan-athey-how-ai-can-aid-us-economic-recovery.

103 See https://aiindex.stanford.edu/report/.

Emerging Technology (CSET) has produced a series of reports about AI and the workforce.[104] In a recent report, they find that "the key centers employing AI professionals are San Francisco (27 percent), New York (13 percent), Seattle (nine percent), and Los Angeles, Boston, and Washington-Baltimore (roughly five percent each). While San Francisco hosts the largest fraction of the AI workforce, the region has the lowest AI employee growth rate at 18 percent, while East Coast hubs grow at between 30 and 57 percent. Additionally, the West Coast attracts a significant portion of its talent from universities based in the eastern United States. In contrast, New York, Boston and the Washington-Baltimore region attract only about five percent of their talent from the West Coast."[105]

Despite being able to obtain these valuable statistics, CSET also observes that "a lack of good data on the U.S. artificial intelligence workforce limits the potential effectiveness of policies targeted at growing and cultivating this cadre of talent."[106] In a similar vein, a Brookings Institution report noted that "To date, there have been no large scale, systematic studies in the U.S. on how robots and AI affect productivity and labor in individual firms or establishments … This is because the data are scarce. […] While [some newly available] data are a promising step, there is still a need for a large-scale survey of technology use across multiple sectors of the economy. Congress should fund the U.S. Census Bureau to collect this data."[107] Echoing this, a recent report in *Proceedings of the National Academy of Sciences* also discusses the "barriers that inhibit scientists from measuring the effects of AI and automation on the future of work, [which] include the lack of high-quality data about the nature of work."[108]

As noted, the National AI Initiative Act of 2020 includes a commitment to study AI's current and future impacts on the U.S. workforce, across sectors.[109] This will include an investigation into "research gaps and data needed" to better understand and plan for these impacts — making it a promising new effort to watch.[110]

104 See, for example, "*U.S. AI Workforce: Labor Market Dynamics*," "*AI Hubs in the United States*," "*PARAT–Tracking the Activity of AI companies*," "*U.S. Demand for AI Certifications: Promise or Hype?*," and "*AI and Industry: Postings and Media Portrayals*," all available at https://cset.georgetown.edu/publications/.

105 J. Olander and M. Flagg, "*AI Hubs in the United States*," Georgetown CSET, 2020, available at https://cset.georgetown.edu/publication/ai-hubs-in-the-united-states/.

106 Ibid.

107 R. Seamans, "*Robot census: Gathering data to improve policymaking on new technologies*," The Brookings Institution, 2021, available at https://www.brookings.edu/research/robot-census-gathering-data-to-improve-policymaking-on-new-technologies/.

108 M. R. Frank, D. Autor, J. Bessen, E. Brynjolfsson, M. Cebriana, D. J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, D. Wang, H. Youn, and I. Rahwan, "*Toward understanding the impact of artificial intelligence on labor*," Proceedings of the National Academy of Sciences, 2019.

109 See https://www.ai.gov/strategic-pillars/education-and-training/ and https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1217.

110 Ibid.

Regulatory and enforcement approaches to AI are nascent and need to be approached in a measured way, taking into account the real potential for unintended negative side effects.

The City of New York can play an important policy role in promoting positive approaches to AI in the local private sector and in protecting New Yorkers in their interactions with it. Additionally, the City has existing engagements with other municipalities across the country, and world, aimed at sharing best practices, and identifying opportunities for alignment and collaboration.

The City also has long-established structures and programs to protect New Yorkers as consumers and workers, and promote small business interests, including those of minority and women-owned business enterprises (M/WBEs).[111] Likewise, the City is developing robust technology policy capacity across a range of offices, as noted. There is opportunity to bring these capacities together, as AI and related technologies increasingly impact New Yorkers in these areas, to ensure City actions stay current with the technology's development and industry practices on an ongoing basis.

111 For example, the Department of Consumer and Worker Protection, the Department of Small Business Services, and the Mayor's Office of Minority and Women Owned Business Enterprises, among others.

Opportunities

Monitor workforce trends and needs on an ongoing basis, and strategically support existing or build new programs.

As a first step, there is clear opportunity for New York City to establish a robust, ongoing source of data about the state of AI use and its impacts on the city's workforce. Paired with continued industry engagement through the Tech Talent Pipeline and other efforts, such data can support the City's work to design programs that are responsive to evolving industry needs and practices, anticipate and plan for job and skill displacement, and effectively address equity and underrepresentation in the local AI workforce. Here, the City can explore a variety of options — including piloting its own survey, or looking for opportunities to leverage existing or emerging efforts, such as the U.S. Census, or follow-on actions from the National AI Initiative study. The City can, further, engage in targeted research efforts to better understand AI impacts and residents' and workers' experiences and expectations in key areas of concern.

The City can also work to ensure the programs it supports are strategically aligned, and that there is continuity among individual offerings, so that learners have clear pathways to progressively build their skills, and ultimately obtain and advance in jobs. And it should explore the need to broaden its approach to AI skill development to include the full range of emerging roles and skills. Here, particular attention should be given to identifying and addressing upskilling or re-skilling needs, and to retaining existing workers, and their domain expertise. Finally, the City can continue to leverage its robust investment in K-12 Computer Science education to build the foundational skills and competencies needed for the next generation of AI use, while seeking insight on emerging needs via internal and external efforts.[112]

Increase involvement of minorities, women, and historically underrepresented groups in AI.

The City should keep a keen eye toward equity and inclusion, and toward supporting pathways that build and uphold New Yorkers' social and economic well-being. Equitable and inclusive participation across the wide range of roles and functions in AI has numerous benefits for ecosystem health. It can support innovative approaches and serve new market segments, make a broad range of local AI efforts more fair and forward-thinking, and help ensure social and economic gains are fairly shared.

The City can continue to build on existing efforts through CUNY, the Tech Talent Pipeline, the Break Through Tech program, and others, to support diversity and inclusion in New York City's AI workforce. City government can also look for opportunities to engage local M/WBE vendors in its own AI work. In 2019, the City's Charter was updated to increase the discretionary award limit for M/WBE contracts from $150,000 to $500,000 — easing procurement of M/WBE vendors for City agencies.[113]

112 K-12 AI education has been a major recent focus in China, in particular — see, e.g., D. Liu, "*China ramps up tech education in bid to become artificial intelligence leader*," NBC News, January 4, 2020, available at https://www.nbcnews.com/news/world/china-ramps-tech-education-bid-become-artificial-intelligence-leader-n1107806.

In 2018, the Association for the Advancement of Artificial Intelligence and the Computer Science Teachers Association launched a working group to develop "*national guidelines for teaching AI to K-12 students*" for the United States, which may be a useful resource to the City, when released. For more on that effort, see https://ojs.aaai.org//index.php/AAAI/article/view/5053 and http://stelar.edc.org/projects/22569/profile/developing-k-12-education-guidelines-artificial-intelligence.

113 State of New York, "*Assembly Bill A8407*," 2019, available at https://www.nysenate.gov/legislation/bills/2019/a8407.

Build internal staff capacity to assess, use, and work with AI topics.

The City can expand on existing programs to increase both agency leadership and staff capacity to responsibly leverage AI. An appropriate set of City experts with experience in AI, including, for example, NYC CTO, AMPO, and Cyber Command, should work with agencies to clarify distinctions between AI, ML, statistics, and other terms, and establish a common framework of understanding, to avoid ad-hoc processes and approaches arising in different agencies. Such an approach might build on existing networks like MODA's Analytics Exchange and should proactively identify opportunities to expand Citywide professional development resources.

Protect New Yorkers' digital rights.

The City can explore opportunities to develop or leverage data on AI's impact on the local workforce and engage stakeholders toward addressing AI-related concerns in broader planning for the future of work — both inside and outside of City government. The City could also further develop how it protects the digital rights of consumers, workers, small businesses, and more with respect to AI.

The City can pair technical and technology policy agencies with agencies with responsibilities in consumer, worker, and small business protection, both to assess private sector use of AI and related technologies and to identify and address emerging issues. The City should regularly engage with local research and advocacy organizations to build knowledge and capacity in this area. In many arenas like, for example, employment discrimination, there are already legal frameworks for what is and is not permissible, and the City has existing functions to address these questions; the City can explore ways to supplement its existing structures with the appropriate expertise and resources.

FACING PAGE
Photo: Ed Reed, New York City Mayoral Photography Office

Next Steps

Eighty years ago, the filmmaker Charlie Chaplin wrote: "In this world there is room for everyone, but we have lost the way. […] Machinery that gives abundance has left us in want. Our knowledge has made us cynical; our cleverness, hard and unkind. We think too much and feel too little. More than machinery, we need humanity. More than cleverness, we need kindness and gentleness. Without these qualities, all will be lost."[114]

114 C. Chaplin, *The Great Dictator*, United Artists, 1940.

Although artificial intelligence is a technical topic, the most pressing questions about AI, at least as they relate to government and society at large, are at heart not technical questions at all. They are fundamentally questions about people, their lives and their relationships, and what human civilization will evolve to look like. That is why they are so difficult and, at times, contentious. In embarking on the process of developing an ecosystem-wide framework for approaching AI that centers digital rights, New York City is taking important steps to address these questions, and has a special opportunity to lead, both at home and around the world.

The foundation for these efforts is a grounding in the details of what AI is and is not, and a basic understanding of how it works and how it can go awry. *Supplement A* provides a detailed discussion of these issues. The City urges leaders and practitioners across the ecosystem to review this summary, with a view toward building a shared knowledge base and a productive alignment for the work ahead.

The City welcomes input on this AI Strategy from any interested individuals and organizations at:
https://our.cityofnewyork.us/a/ai-strategy-feedback.

Modernize the City's data infrastructure.

The City will continue to work to identify common agency pain points, including needs for centralized guidance and template documents that agencies can rely on in their efforts. In addition, the City will work to improve data standards and practical usability of data, as well as address digital rights issues, where relevant.

Identify and pursue beneficial areas to use AI.

The City will convene agencies that are already using AI to distill practical lessons learned for the benefit of one another and agencies at more nascent stages. In addition, several agencies will be selected as part of an internal review to determine how best to operationalize at the agency level the approach described in this Strategy, including identifying key agency questions and needs.

Strengthen City capacity to ensure effective and responsible use of AI, including robust public engagement.

The City will conduct a public engagement process on this Strategy. In addition, the City will identify key policy areas that could benefit from a more iterative, experimental approach that includes external collaborations, will develop strategies to provide relevant AI training and education within City government, and will explore opportunities to pilot participatory and other approaches.

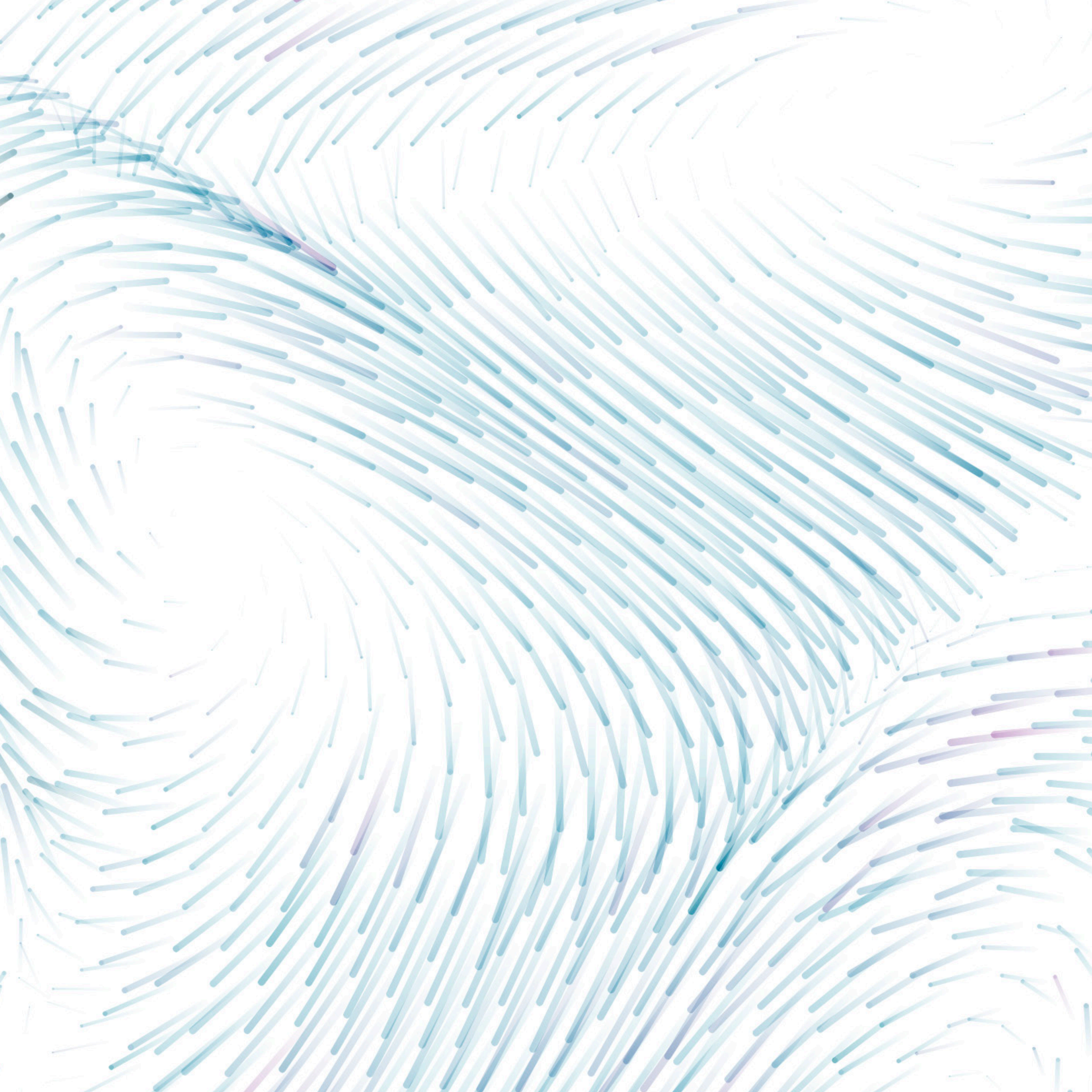
Grow productive external partnerships.

The City will convene key agency stakeholders with relevant experience in establishing productive academic and applied research partnerships. Together, the group will work to formulate a Citywide approach to streamlining such partnerships and consider establishing new structures for ongoing external engagement.

Protect New Yorkers' digital rights and foster equitable opportunity across the ecosystem.

The City will explore opportunities to develop or leverage data on AI's impact on the local workforce and engage stakeholders toward addressing AI-related concerns in broader planning for the future of work — both inside and outside of City government. The City will further develop how it protects the digital rights of consumers, workers, small businesses, and more with respect to AI.

Supplement A:
NYC AI Primer



NYC AI Primer

Artificial intelligence (AI) is changing the human experience today, driving sweeping social, economic, and technological transformation that affects us all. The City of New York believes that an approach grounded in digital rights is necessary to maximize its benefits, minimize its harms, and ensure its responsible application. Moreover, establishing a clear understanding of what AI is, how it works, and what some of the key practical and ethical considerations are around its use is foundational to building a healthy AI ecosystem for New York City.

One of the chief difficulties in the discourse on AI broadly is that claims — both positive and negative — are often exaggerated to the point of being misleading, and "AI" is also often used more as a marketing term than a precise description of the techniques used. Even among those working in the field, there can be inconsistency or disagreement with regard to scope, definitions, and priorities. To facilitate better policy, recognize both opportunities and risks, and evaluate claims made by others, New York City decision-makers require greater clarity on what "AI" means, what components can make up a system, the wide range of ways considerations like performance and accuracy, fairness, accountability, privacy, and security can come into play, and the complexity of weighing these factors against each other in any given situation.

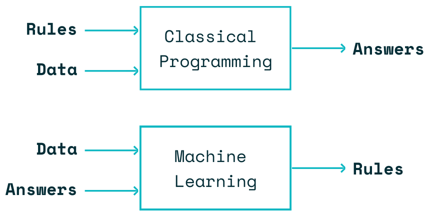
This document aims to help provide this foundation, primarily for an audience of technical, policy, or other decision-makers who are in or interact with New York City government. Importantly, this is a rapidly evolving field, and this should not be taken to be a comprehensive or final account. Ongoing engagement will be required to ensure local stakeholders are keeping pace with the technology, its use, and its consideration across society as each of these aspects continues to develop.

"We are drowning in information and starving for knowledge."

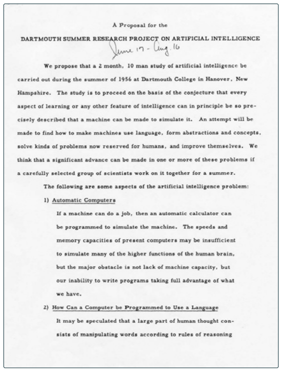
Rutherford D. Rogers
Former Chief of Research Libraries
The New York Public Library

Computer chess was long one of the most visible testing grounds for AI technology. Shown here, World Chess Champion Garry Kasparov plays IBM's Deep Blue in 1996. *Photo: Laurence Kesterson*

AI is an umbrella term encompassing a range of technologies both sophisticated and simple that are used to, among other things, make predictions, inferences, recommendations, or decisions with data. The term “artificial intelligence” was first coined in the 1950s to describe efforts by computer scientists to produce general human intelligence and behavior in computers; these early efforts to create AI systems were largely “rule-based” to attempt to simulate human reasoning.



A simplified diagram comparing traditional software to machine learning, showing ML as a kind of new software programming paradigm. From F. Chollet, *Deep Learning with Python*, O’Reilly, 2017.



A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1) Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2) How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning

Initial proposal for the “Dartmouth Summer Research Project on Artificial Intelligence,” generally considered to be the founding of AI as a research discipline; a plaque commemorating the 50th anniversary of the event.

In traditional (non-AI) software, developers tell a computer exactly how to carry out a given task using precise, fixed instructions. This sequence of instructions is called an algorithm.[1] This approach works well for tasks like sorting a list of names or typesetting a book, but does not work well for problems like differentiating between photos of dogs and cats, reading the handwritten address on an envelope, or identifying fraudulent credit card transactions. Intuitively, there is far too much variation in these cases to handle with explicit rules, even though some of these tasks are easy for humans.

1 See, for example: T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms,* third edition, MIT Press, 2009; T. Roughgarden, *Algorithms Illuminated*, Soundlikeyourself Publishing, 2020; D. E. Knuth, *The Art of Computer Programming*, 2011, details at https://www-cs-faculty.stanford.edu/~knuth/taocp.html.

Among practitioners and specialists, "AI" now largely refers to the use of an approach called machine learning (ML), a way to write "software by example" by providing the computer with illustrative examples to "learn" from.[2] Machine learning uses data together with certain mathematical techniques to create computer programs. These techniques largely come from the fields of statistics, probability, mathematical optimization, and computer science, though increasingly tools from economics, the social sciences, and other areas in applied mathematics are used as well.[3] The computer is given a description of the task to be performed; data in the form of examples of what the correct results look like; a mathematical way of formalizing or expressing assumptions about how the data relates to the task called a "model"; and a learning algorithm indicating how to improve at the task by trial-and-error. The result is a "trained model" which can take new inputs for which the correct output or result is not known and guess the correct output. This guessed output is called a "prediction," which in this context is a technical term and often does not refer to predicting the future. The question of what kinds of outputs can in practice be effectively "predicted" with ML is itself a subtle topic and the subject of ongoing research.[4]

Because references to statistics or statistical language are so pervasive in AI, it is worth briefly addressing a potential point of confusion. In the context of government, the word "statistics" often refers to what are sometimes called "administrative statistics," or government's definition of relevant measures (such as poverty, employment, or race) and subsequent measurement to facilitate governance.[5] These include anything from measurements (or "statistics") related to people to those related to agricultural production.[6] While not unrelated, this is different from the kind of "inferential statistics" or "statistical inference" used in AI.

The process of building an AI system is described in more detail with concrete, practical examples below, along with associated ethical and policy considerations that arise.

2 This document focuses on a particular form of machine learning called "supervised learning"; there are also other areas in machine learning, including "unsupervised learning" and "reinforcement learning." For additional general reading, see the *Further References* section at the end of this document.

3 Because machine learning draws on so many other fields, there are often multiple pieces of jargon referring to the same concepts, depending on the academic training of the person speaking or even the venue in which an academic publication appears. In addition, many pieces of technical jargon from statistics and machine learning also have distinct colloquial uses that can cause confusion, including "discrimination," "bias," "prediction," and more.

4 J. Kleinberg, J. Ludwig, S. Mullainathan, and Z. Obermeyer, *"Prediction policy problems,"* American Economic Review, 2015; S. Athey, *"Beyond prediction: Using big data for policy problems,"* Science, 2017; A. Narayanan and M. Salganik, *"Limits to Prediction,"* 2020, available at https://msalganik.github.io/cos597E-soc555_f2020/ .

5 A. Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, translated by Camille Naish, Harvard University Press, 1993.

6 The US government has a decentralized "Federal Statistical System," spanning 125 agencies engaged, to some degree, in collecting data and producing such descriptive statistics, with 13 agencies whose primary mission is statistical work; the best-known include the Bureau of Economic Analysis, Bureau of Labor Statistics, and the Census Bureau.

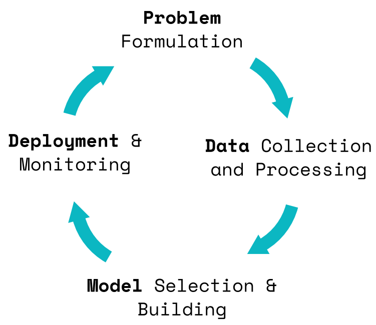
The AI lifecycle

The terms AI lifecycle or ML lifecycle describe the steps used to create a new machine learning system.

This generally involves the steps of identifying and precisely formulating a problem to be addressed; collecting and processing data; building a model; and deploying and monitoring the system.[7] In some cases, parts of this process are repeated iteratively: the team may collect new data and retrain or restructure the model every few months, or the system may automatically incorporate new data (learn) on an ongoing basis.[8] Because of the myriad challenges throughout this process, there are now entire companies that focus on offering products or services to aid other organizations even in single components of this lifecycle, from data labeling to system monitoring.

7 See, e.g., *Full Stack Deep Learning*, available at https://fullstackdeeplearning.com; *Stanford ML Systems Seminar Series*, available at https://mlsys.stanford.edu.

8 Systems that keep adapting (retraining) automatically over time are called "online" systems. There are pros and cons to this approach that need to be evaluated in context.



A simplified depiction of the lifecycle of an AI application.

This section outlines the AI lifecycle and discusses considerations that can arise at each stage. It will use mortgage lending[9] as a running example. Today, mortgage lenders often use ML algorithms to help make decisions about whether or not to approve a given loan application. This example was selected partly because it is both real and impactful, but also because it is simple while still exhibiting all of the different complexities discussed below.

The person or team going through the process below will be referred to as the "developer."

9 See, e.g., S. Trilling, *"Fair Algorithmic Housing Loans,"* Aspen Tech Policy Hub, 2020, available at https://www.aspentechpolicyhub.org/project/fair-algorithmic-housing-loans/. There are additional references on algorithmic lending and mortgage lending in citations through this document.

Problem formulation

Developers must first precisely formulate the problem.

This includes defining what the inputs and outputs are intended to be. In the case of mortgage loan decisions, the input would include a list of characteristics[10] of the loan application, potentially including information about the borrower, the financial terms of the loan, and details about the property, while the output could be taken to be "yes" (approved) or "no" (declined). This very common formulation is called "binary classification" because the input data is being classified into one of two classes or categories.[11]

There are other ways to formulate the problem. For example, the output could be a numeric[12] "risk score" from 0-100 that gives an estimate of the probability of loan repayment, and the developer would need to decide what role humans are intended to play in the process. It could be that the yes/no outputs are merely suggestions to help advise a human decision-maker; that the outputs are fully automated decisions; or that the system is designed to have three possible outputs, including a "maybe" option that prompts human review. Decisions like these are both necessary and subjective, and can have both practical and ethical implications.

It is essential to define what a model performing "well" means for the organization.

Often, this should be measured relative to some specified baseline (possibly the performance of a human team performing the same task), as systems can be flawed but still improve on the status quo enough that they are worth using. In mortgage lending, the lender may measure success based on corporate financial metrics (e.g., more loans get made with fewer borrowers defaulting on their loans); the system may also seek to behave similarly to humans but be faster or more transparent, or to improve on measures of equity and fairness. For example, researchers have found that although ML loan models do discriminate, they also can be up to 40% less discriminatory than face-to-face lending.[13]

"Some problems are better evaded than solved."
C. A. R. Hoare

10 This input list of characteristics can be visualized as a row in a spreadsheet, with one row per loan application and the columns corresponding to the different characteristics, such as borrower age and borrower income.

11 Binary classification is the formulation used for a very wide range of real systems, such as those that classify credit card transactions as valid or fraudulent or those that classify emails into valid or spam.

12 In addition to the output being one of a fixed collection of options (called classification) or a number (called regression), the outputs can also be much more complex structures. For example, in language translation systems, the input is a sentence in one language and the output is the correct translation in a different language; in a face detection system, the input is an image or video and the output may be the size and location of boxes that contain the faces in the image.

13 R. Bartlett, A. Morse, R. Stanton, and N. Wallace, *"Consumer-lending discrimination in the FinTech era,"* Journal of Financial Economics, 2021.

Beyond the technical aspects above, developers must consider the broader context in which the system will be deployed and if the concept motivating the system even makes sense.

It is important to understand that when systems perform poorly or have negative impacts in the real world, that can be caused by failures of problem formulation and conception rather than necessarily being a primarily technical issue.[14]

There are many points to consider at this stage, including the goals of the project, who it is intended to benefit, the stakeholders that should be consulted, auxiliary systems or processes that must be built around the system itself, and proactive consideration of possible malfunctions and their impacts on people. These are complex, interlinked, and context-specific, and there is no recipe for navigating them, so it is helpful to include a range of perspectives.

Depending on the application, engaging the public or other stakeholders may be a beneficial or necessary way to ensure this step is done effectively.

Community and public engagement, as well as participatory approaches, are covered in a subsequent section, but if it is appropriate to use engagement or participatory methods, this must be done in a thorough and careful way. For example, the developer must be willing to potentially reframe or even cancel the project altogether as a result of the input received, and it may be helpful to use quantitative or other specialized methods in order to effectively elicit and incorporate informed input. In short, there are a range of tools that can be brought to bear on this process that can complement or supplement familiar approaches like feedback forms and town halls.

14 See, e.g., V. Eubanks, *Automating Inequality: How high-tech tools profile, police, and punish the poor*, St. Martin's Press, 2018; D. Kolkman, *"'F**k the algorithm'?: What the world can learn from the UK's A-level grading fiasco,"* LSE Impact Blog, London School of Economics, 2020, available at https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/.

Data

Although discussions of AI and ML often emphasize models or algorithms, it is well-known among practitioners that issues around data are often much more determinative of the success or failure of a project and can take up the vast majority of the time, effort, and cost.

"The world is not the sum of all the things that are in it. It is the infinitely complex network of connections among them. As in the meanings of words, things take on meaning only in relationship to each other."
Paul Auster

One step is "data collection": For mortgage lending, a historical set of loan applications, each associated with a "ground truth" output value (such as approval status), must be obtained. This ground truth output may be available, may need to be inferred, or may require a new project in human annotation to produce — a process called "data labeling." Newer lenders would not have access to enough historical loan data to use ML at all, while older lenders may have paper records that need to be processed into a "machine readable" database, possibly requiring significant manual human effort. An important consideration is that for some tasks, the notion of "ground truth" may itself be somewhat or entirely subjective; this can both complicate the process of producing training data as well as result in downstream effects on the overall system built around it that may or may not be intended.[15]

Another step is "data cleaning," which generally refers to detecting and removing errors or inconsistencies in data. For example, loans may have been recorded with an inconsistent mix of 5-digit and 9-digit ZIP codes, some loan applications may include the borrower's gender while others don't, and even valid phone numbers can be written in many different formats. The properties may be represented by different brokers who provide data in inconsistent formats or based on different policies. Some data may be the result of forms completed by hand and later entered into a system, a common process which often introduces at least some errors. In addition, combining multiple datasets can be time consuming, error-prone, or even prohibitively difficult without standardized identifiers, such as social security numbers, license plate numbers, or Universal Product Codes.[16]

15 For example, see A. Jeffries and L. Yin, *"To Gmail, Most Black Lives Matter Emails Are 'Promotions',"* The Markup, 2020, available at https://themarkup.org/google-the-giant/2020/07/02/to-gmail-black-lives-matter-emails-are-promotions.

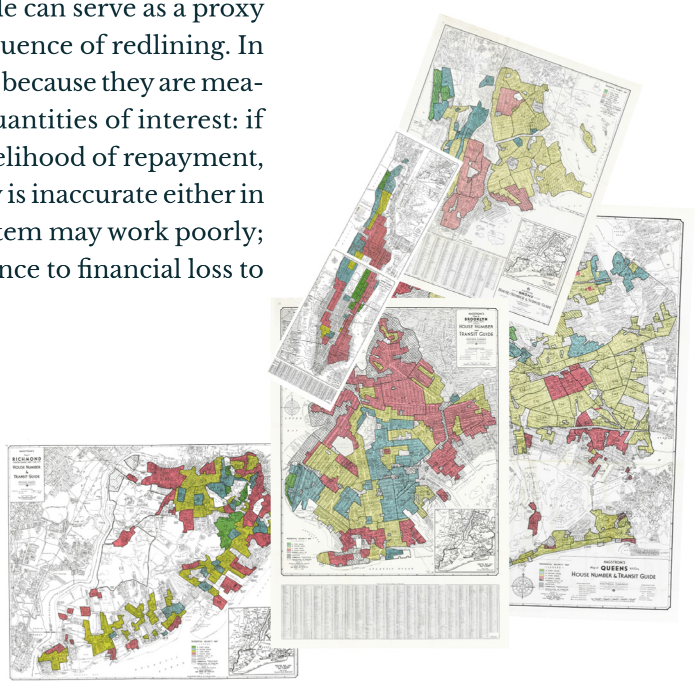
16 See https://www.gs1us.org/upcs-barcodes-prefixes/get-a-barcode/why-gs1-us.

These seemingly mundane tasks can be enormously time consuming and expensive and often require data engineering or domain expertise different from that needed to build models. For instance, some data may have been captured digitally at the source, possibly using an online form, but the specifics of how that form was designed — such as how a question was worded or how the input was validated — often shapes the actual data collected in ways that are opaque to a person looking at the data without that context. Lack of appropriate data for a task can lead to many problems, such as poor system performance or accuracy, data security breaches, or unfairness to particular groups, often for subtle reasons. For example, if there are very few people with certain attributes in the data, the outputs may be much less accurate for those groups because there is not enough data to go on.[17]

Certain values in the inputs or the output often serve as "proxies," which must be carefully considered to avoid undesired behavior.

This can be implicit or unintended: ZIP code can serve as a proxy for race, partly because of the historical influence of redlining. In other cases, developers explicitly use proxies because they are measurable and seem close enough to actual quantities of interest: if lenders are interested in borrowers' true likelihood of repayment, they may use credit scores instead. If a proxy is inaccurate either in general or in the context of the task, the system may work poorly; this can result in anything from inconvenience to financial loss to more serious harm.

17 For example, although facial recognition systems have been found to perform poorly for dark-skinned people in general, and dark-skinned women in particular, there have been subsequent studies showing that simply training the systems with datasets that are more representative appear to give significantly better performance on these affected groups. See, e.g., C. Romine (Director of Information Technology Laboratory at NIST), *"Testimony in Hearing on Facial Recognition Technology (Part III): Ensuring Commercial Transparency & Accuracy,"* Committee on Homeland Security, US House of Representatives, 2020; P. Grother, M. Ngan, and K. Hanaoka, *"NISTIR 8280: Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects,"* NIST, 2019; R. Puri, *"Mitigating bias in AI models,"* IBM Research Blog, 2018. For initial work on the discrimination itself, see J. Buolamwini and T. Gebru, *"Gender shades: Intersectional accuracy disparities in commercial gender classification,"* ACM FAccT, 2018.



Redlining typically referred to racial discrimination against particular neighborhoods when providing services or benefits; mortgage lending is one such example with a long history. This was often done explicitly, with "bad" neighborhoods outlined in red. See, e.g., R. Rothstein, *The Color of Law: A Forgotten History of How Our Government Segregated America*, Liveright, 2017.

Shown here, redlining maps of NYC from 1938.

Models

Models codify the developer's assumptions about the data and task at hand.

The developer must next choose a model to use. Essentially, a model codifies a set of assumptions about the potential relationship between the inputs and the output, and can be thought of as a mathematical formula that combines the inputs with a set of adjustable numbers to produce a predicted output. These adjustable numbers are called "parameters" or "weights," to reflect that they indicate how much emphasis to put on different inputs for the given task. The types of models that are appropriate in a given situation are partly dictated by the problem formulation and the nature of the inputs and outputs; for example, one would use one type of model if predicting a yes/no binary output for mortgage loans and a different type if predicting a risk score or other value. Though some models can be very complex and have billions of parameters, extremely simple models (like linear regression, which dates back to the early 1800s) from traditional statistics can be very effective and are often used in practice as well.

"All models are wrong, but some are useful."

George Box
Former President
American Statistical Association

Training a model involves using the data and a learning algorithm to set the parameters associated with the model. At the beginning of the training process, these are set to random values, which will of course result in the model initially producing poor predictions that do not match the labeled outputs; these discrepancies are incorporated in a learning algorithm to iteratively tune the parameters, resulting in a trained model. This model can take any input data in the same form as the training data, such as a new loan application with all the same attributes encoded as in the training data, and produce a predicted output (loan decision).

The key goal in machine learning is generally to make good predictions on new data that have not been manually labeled or previously seen.

Unlike what is often the case in social science projects, the goal is typically not to better understand existing historical data for its own sake using statistical methods. The main goal is to be able to do “well” at deciding what to do with new loans that have not yet been approved or denied, and one must define what “well” means.

There are many ways to measure and define performance.

For instance, in binary classification, there are four possible outcomes of a prediction; in mortgage lending, these correspond to correctly approving a loan (“true positive”), correctly denying a loan (“true negative”), incorrectly approving a loan (“false positive”), and incorrectly denying a loan (“false negative”). In the simplest case, one might care about predictive accuracy, which would just count the percentage of correct predictions. However, false positives and negatives are different types of errors with potentially different risks or impacts, so they may need to be accounted for differently. False negatives could result in both revenue loss for the lender and greater difficulty buying property for the borrower; false positives could lead to borrowers being given loans they cannot repay. When there are multiple criteria, there are often trade-offs between them, and the developer needs to determine how to measure overall performance in line with organizational goals and potential broader impacts. To take the extreme case, one can easily drive either false positives or negatives to zero simply by denying or approving all loans, respectively, but this would obviously lead to the system performing very poorly on the whole.

In practice, developers will try out several different models on the same data and then choose one after seeing how well the models actually do on their specific problem with the data available. This decision can be based on several factors. In some cases, one may simply select the model with the best performance on the validation set (defined below). In other cases, developers may choose a different model that performs well enough but also has other benefits, such as being easier to understand, inspect, audit, implement, or maintain. Depending on the setting, even the most sophisticated, best-performing model may perform too poorly to use in practice; conversely, as noted above, even a model that performs relatively poorly may still significantly improve on the status quo (possibly traditional software systems or human teams previously used to perform the task) either in performance or other ways, and still be well worth using. In short, model performance should be evaluated relative to an appropriately chosen baseline or status quo, not in a vacuum.

Performance is measured with respect to some particular choice of "validation" data.

To actually determine which model does best, there must be a way to evaluate how well it performs on data it was not trained on. Roughly, this is done by setting some of the labeled training data aside as "validation data"; validation data are not used in training and are only used to measure the performance of the model.

The following is critical: When a performance metric is reported for a machine learning model, that number should be understood as being evaluated on a particular set of validation data. If this validation data is not chosen appropriately, those metrics can give a misleading picture of how the model will perform in the real world. Because of this, and the varying ways in which performance can be measured, claims from vendors or in media reports that a system is "99% accurate" are often incomplete or outright meaningless without further details and explanation.

Deployment and monitoring

Unlike traditional software like a web browser, the performance and behavior of machine learning models often changes over time, so it is critical to carefully monitor systems once they are deployed.

"Do not be too positive about things. You may be in error."
C. F. Lawlor

These changes over time can happen for many reasons. Among the major reasons are that the data on which the system is used is, or becomes, different from the training data used to build the model, potentially because the underlying phenomenon being modeled itself changes.[18] For example, there may be inflows or outflows of certain types of people in an area; changes to housing supply or

18 These are sometimes referred to as "drift" or "shift," depending on the details.

housing laws; or interest rates may rise to the degree that a different set of people are seeking mortgages. If a New York company buys a mortgage loan evaluation product from an outside vendor, it may be that that model is based on data from Florida, where loan data may look very different.[19] It is a virtual certainty that there will be many such shifts in the aftermath of COVID-19 in a range of domains, not only housing.

In other cases, actions taken based on the system's recommendations may be fed back into the system as new training data. This "online" re-training process is a means to keep a system up-to-date over time, but this process can also create the risk of feedback loops that could cause a model to, for example, increasingly only approve loans by existing homeowners, even if many first-time buyers are also likely to repay their loans.[20]

Finally, because of the complexity of engineering ML systems in general, there are a broad range of other practical considerations that arise in testing, monitoring, and maintaining these systems that must be considered.[21] As one example, there should not be any differences between the way data is processed to train the model and the way new data is processed before it is run through the model in production; any mismatches here can cause a range of potentially serious performance problems that can be difficult to detect and debug. Though this may seem obvious, this type of bug or mistake is very common in practice. As another example, it is important to be able to safely roll back to a previous version of a model that is known to work correctly.

19 This may sound far-fetched but is not hypothetical. In the context of medicine, see S. Lynch, *"The Geographical Bias in Medical AI Tools,"* Stanford Institute for Human-Centered AI, 2020, available at https://hai.stanford.edu/news/geographic-bias-medical-ai-tools, which summarizes research showing that *"most [ML] algorithms [for clinical diagnosis tasks] are trained on datasets from patients in only three geographic areas, and that the majority of states have no represented patients whatsoever."* This is partly driven by the difficulty and expense associated with producing good datasets for training; developers often gravitate to using the data that is most readily available.

20 See, e.g., D. Ensign, S. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian, *"Runaway feedback loops in predictive policing,"* ACM FAccT, 2018.

21 See, e.g., D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, and M. Young, *"Machine Learning: The high interest credit card of technical debt,"* Google Research, 2014; E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, *"The ML test score: A rubric for ML production readiness and technical debt reduction,"* IEEE International Conference on Big Data, 2017; M. Zinkevich, *"Rules of Machine Learning: Best Practices for ML Engineering,"* Google Research, available at http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf.

Ethics, governance, and policy

This section discusses a range of additional concerns about AI and ML, with a particular focus on ethics. It highlights several aspects of concern, including fairness and non-discrimination, and emphasizes public engagement as a key tool.

"They took each other's advice, opened one book, went over to another, then did not know what to decide when opinions diverged so widely."

Gustave Flaubert
Bouvard et Pécuchet

Ethics has become an increasingly prominent topic in AI in recent years. Work on this topic is often referred to as "AI Ethics," "responsible AI," or other similar terms. Not all these issues apply to every kind of AI application, and the discussion below is intended to give a feel for the topic, not to be fully comprehensive. It is worth emphasizing that all of the topics discussed here are the subject of active and recent research, so there is much that is not yet fully understood or settled.

It is above all important to make sure that the system in question actually works and accomplishes its goals. This is not as simple or as much of a given as it may sound.

Too often, systems are built around a flawed premise or simply do not work for their intended purpose; just because one can collect some training data and mechanically go through the motions of training a model does not mean it always makes sense to do so, or that the result should be taken seriously. Like many of the other potential failure modes described both above and below, this problem can manifest as anything from mere inconvenience or suboptimal performance to severe unethical behavior and impacts on real people and communities, sometimes including racism, sexism, or even matters of life and death.

To give a real example, a major electronic health record company sells an AI system for predicting whether patients will develop sepsis (a life-threatening condition that can arise in response to infection). This system is used by hundreds of hospitals around the country. In a recent study, researchers found that the model both performs substantially worse than the vendor reported, and poorly for clinical use in general: The tool misses two thirds of sepsis cases (high false negative rate) while also overwhelming doctors with false alerts (high false positive rate).[22] This is also not one of the instances of discrimination that are discussed further below; the system simply does not work as it should across the board. Such examples underscore the need to rigorously evaluate system design and performance from a range of different perspectives.

While it is helpful to have an underlying ethical framework to underpin one's approach, it is still typically unclear how to actually operationalize such principles in practice.

When discussing ethics, it is important to think about what high-level ethical principles or framework are being used. In recent years, at least 100 institutions, from corporations to academics to nonprofits to national governments, have published various sets of "AI principles." In a comparative study, Harvard researchers found that there is broad overlap in these, which include privacy, accountability, fairness and non-discrimination, human control of technology, and others.[23] Though there is significant consensus around such principles at a high level, society and the field are still at the early stages of determining how to operationalize them. In the context of local governments, such principles have been referred to as "digital rights," by analogy with human rights, and these rights and principles are discussed in more detail below. For this reason, this document focuses more on practical challenges than on motivating or explaining the principles themselves.

AI forces developers, and society at large, to make societal and policy values and goals explicit.

To build an AI system, as described in the preceding sections, developers must specify things like what the system should be optimizing for and how different types of errors should be weighed. When building a system that impacts people, this often involves making ethical and other policy values quantitative and explicit. These values are not specific to AI: they are implied in any policy or decision-making process (such as hiring decisions, college admissions, or patient treatment) previously conducted solely by humans and governed by organizational policies. When used for

22 A. Wong, E. Otles, J. Donnelley, A. Krumm, J. McCullough, O. DeTroyer-Cooley, J. Pestrue, M. Phillips, J. Konye, C. Penoza, M. Ghous, and K. Singh, *"External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients,"* Journal of the American Medical Association–Internal Medicine, 2021.

23 J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, *"Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI,"* Berkman Klein Center for Internet & Society, Harvard University, 2020, available at https://dash.harvard.edu/handle/1/42160420.

human decision-making, these typically were and continue to be implicit rather than explicit, and that implicitness in part can allow human decisions to remain unfair and inequitable.

As above, a mortgage loan AI system would need to specify the relative costs of wrongly denying a loan versus wrongly approving one, and possibly even explicitly break these down for different groups (by gender, race, or other attributes). This explicitness can cause discomfort, but it must be understood that only the explicitness, rather than the trade-offs themselves, are new, and making them explicit provides a significant collective opportunity to revisit and redesign how policies have been designed and implemented more broadly. Indeed, one of the potential uses of AI is to help inspect and evaluate how human decisions have been made.[24]

Accountability

Broadly, accountability in the context of AI refers to being responsible or answerable for the outputs, decisions, or impacts resulting from the use of an AI system or model.

This can take several different forms; the coverage here is not comprehensive but aims to give a feel for different ways in which this can be approached.[25]

One of the simplest forms of accountability is being transparent about the fact that an AI system is in fact being used to perform important functions or make impactful decisions. Although this may seem straightforward, it has been controversial in some contexts, such as the management of patient health.[26] Beyond this, one can consider providing transparency into specific aspects of the system, such as the data or models used.[27]

Another potential goal is to allow for some human intuition or understanding of what the model is doing, as opposed to just knowing how it performs on some validation data. There are different approaches to this, including just using much simpler models that are inherently easier to interpret, using additional technical methods

24 S. Mullainathan and Z. Obermeyer, *"Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care,"* National Bureau of Economic Research (NBER) Working Paper No. 26168, 2021; S. Mullainathan, *"Biased Algorithms are Easier to Fix than Biased People,"* The New York Times–Opinion, 2019, available at https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

25 In particular, see a recent GAO report on the federal government's approach to accountability, which *"identifies key accountability practices — centered around the principles of governance, data, performance, and monitoring"*: US Government Accountability Office, *"Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities,"* 2021, available at https://www.gao.gov/products/gao-21-519sp.

26 R. Robbins and E. Brodwin, *"An invisible hand: Patients aren't being told about the AI systems advising their care,"* STAT News, 2020, available at https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/.

27 Some of these ideas have informally been referred to as "nutrition labels" for ML; see, e.g., T. Gebru, J. Morgenstern, B. Vecchione, J. Vaughn, H. Wallach, H. Daumé, and K. Crawford, *"Datasheets for datasets,"* arXiv preprint arXiv:1803.09010, 2018; M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, D. Raji, and T. Gebru, *"Model cards for model reporting,"* ACM FAccT, 2019.

to "inspect" the inner workings of more complex models, or designing the system to allow users to see how the model's outputs vary as certain input attributes are changed.[28] These sorts of decisions may allow one to see that, for example, borrower age is or is not very relevant in predicting loan approval, or indicating to a borrower or loan officer that a loan would have been approved if the applicant's income were over some number. Which approach makes sense, and is amenable to different stakeholders, will depend on context; the expectations and resulting approach will be different across consumer finance, medical diagnosis, and criminal justice, and are still subject to ongoing research and debate.[29]

When humans are involved, they need to be considered part of the system itself.

A different way to approach accountability is to integrate human oversight via maintaining a role for humans in the ultimate decisions. For example, the system may only make suggestions or help focus the human's attention on the more ambiguous or difficult cases. These are sometimes referred to as "partially automated" or "human-in-the-loop" systems. In these cases, the way in which the human operators are trained to use, override, or ignore the system, as well as how the interfaces of the system are designed, play a critical role in overall system behavior. In addition, the way the system's suggestions are framed, described, or presented can have an outsized impact on how the human in question reacts to them. For example, a user may interpret a "green light/red light" display very differently than a risk assessment score displayed with five decimal points, and such seemingly minor details can in turn influence, sometimes dramatically, the ultimate behavior of the "whole system." For this reason, it is critical to engage experienced designers when building systems that include human interfaces.[30]

For example, if human operators are allowed to override or deviate from the recommendations or decisions of a mortgage loan model, what matters and must be evaluated is whether the overall mix of computer and human decisions satisfy the desired goals, rather than the model by itself in a vacuum.[31]

28 Some of these different topics are referred to as interpretability, explainability, or transparency.

29 See, e.g., C. Rudin, *"Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,"* Nature Machine Intelligence, 2019; K. Miller, *"Should AI Models Be Explainable? That depends,"* Stanford Institute for Human-Centered AI, 2021; B. Haibe-Kains et al, *"Transparency and reproducibility in artificial intelligence,"* Nature, 2020; H. Stower, *"Transparency in medical AI,"* Nature Medicine, 2020; J. Vaughn, *"Transparency and Intelligibility Throughout the Machine Learning Life Cycle,"* available at https://www.youtube.com/watch?v=I-TSjiXGfSI.

30 See, e.g., M. K. Lee, D. Kusbit, E. Metsky, and L. Dabbish, *"Working with Machines: The Impact of Algorithmic, Data-Driven Management on Human Workers,"* ACM/SIGCHI Conference on Human Factors in Computing Systems, 2015; E. Tufte, *Visual Explanations,* Graphics Press, 1997; D. Huff, *How to Lie with Statistics,* Norton, 1954.

31 This is sometimes referred to as a "sociotechnical systems" approach.

Fairness

The term "fairness" in AI refers to the notion that systems should not discriminate with respect to certain personal attributes or protected characteristics, such as race, gender, age, or disability.

This can take different forms depending on the situation. For example, it may be that loan applications that are otherwise similar are declined at much higher rates for women than men,[32] or that a system is much less accurate for certain groups of people in a way that results in some kind of disparate impact or harm.[33] This topic is an area of significant concern and has received a great deal of attention in recent years.[34]

Though this behavior is sometimes referred to as "bias" or "algorithmic bias," this document uses "fairness" both to avoid confusion with other unrelated technical definitions of "bias" in AI, and to emphasize that the concern is ultimately about impacts on people rather than a narrower consideration of model behavior alone.[35]

For example, a model that by itself is "unbiased" in some technical sense can turn out to have unfair outcomes for people when actually deployed (sometimes because of decisions humans-in-the-loop make outside the model itself). On the flip side, it may be possible to use a model that is technically "biased" to promote equity and advance other goals.[36] For example, in some preliminary work, researchers partnering with the Los Angeles City Attorney's office found that they could have a possibly biased system result in equitable criminal justice outcomes across racial groups by, among other things, coupling technical considerations with a tailored social service intervention strategy. [37] Alternatively, in the mortgage context, a model may be designed to support lending decisions in a way that actively corrects historical racial disparities in homeownership rates.

In sum, narrowly focusing on developing AI models and algorithms that better account for fairness will generally not be sufficient to actually achieve more equitable decisions or outcomes,

32 See, for instance, L. Goodman, J. Zhu, and B. Bai, *"Women Are Better than Men at Paying Their Mortgages,"* Urban Institute –Housing Finance Policy Center, 2016.

33 S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities,* 2021, accessible at http://www.fairmlbook.org.

34 For overviews, see, e.g., J. Vaughn and H. Wallach, *"Machine Learning and Fairness,"* 2020, available at https://www.youtube.com/watch?v=7CH0xLWQLRw; H. Wallach and M. Dudik, *"Fairness-related harms in AI systems: Examples, assessment, and mitigation,"* 2021, available at https://www.youtube.com/watch?v=1RptHwfkx_k; K. Rodolfa, P. Saleiro, and R. Ghani, *"Bias and fairness,"* chapter of *Big Data and Social Science: Data Science Methods and Tools for Research and Practice*, available at https://textbook.coleridgeinitiative.org/chap-bias.html.

35 See, e.g., J. Buolamwini and T. Gebru, *"Gender shades: Intersectional accuracy disparities in commercial gender classification,"* ACM FAccT, 2018.

36 R. Ghani, *"Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services,"* Testimony to Artificial Intelligence Task Force, Committee on Financial Services, U.S. House of Representatives, 2020.

37 K. T. Rodolfa, E. Salomon, L. Haynes, I. Mendieta, J. Larson, and R. Ghani, *"Case study: Predictive fairness to reduce misdemeanor recidivism through social service interventions,"* ACM FAccT, 2020.

which is the real goal. Instead, efforts should work towards making entire systems — including the humans and organizations involved — and their ultimate outcomes and effects fair.[38]

There is no single definition of fairness and the notion and goal appropriate for the situation at hand must be determined through the development process; however, it is often not appropriate for the developers to make these decisions unilaterally, so broader stakeholder engagement is often necessary.

In some instances, there may be laws requiring that certain types of decisions are made fairly, including definitions of what is meant by "fair" in that domain; examples include the Fair Housing Act, the Equal Credit Opportunity Act, or the Uniform Guidelines on Employment Selection Procedures. In other cases, it may be necessary for the developer to choose what "fair" should mean, and there are a large set of formal criteria with different implications.[39]

In the context of lending,[40] one possible criterion is "race and gender blindness"; in other words, a requirement that the system not include race or gender as input features. This is straightforward to implement, but is not likely to actually avoid disparate outcomes, partly because other input features can serve as proxies for race or gender.[41] In particular, because of the historical effects of redlining, a seemingly simple piece of information like ZIP code will often serve as a proxy for race and lead to the system being racially unfair. In addition, not letting the model see this data may prevent the developer from using certain types of technical corrections to ensure fairness across those attributes.

A different definition could be "parity," meaning that the exact same number (or percentage) of loans should be approved or denied in each demographic group. On the one hand, this may ensure that protected groups would get loans; on the other, it may mean they are more often getting loans they cannot repay. Yet another definition might be that the decisions serve to reduce disparities in homeownership rates across racial groups. In short, there are dozens of ways one might define fairness, and these conflict with

38 For a discussion of practical implementation issues in fairness, see, e.g., C. Bakalar, R. Barreto, S. Bergman, M. Bogen, B. Chern, S. Corbett-Davies, M. Hall, I. Kloumann, M. Lam, J. Candela, and M. Raghavan, *"Fairness On The Ground: Applying Algorithmic Fairness Approaches to Production Systems,"* arXiv preprint arXiv:2103.06172, 2021.

39 A. Narayanan, *"21 fairness definitions and their politics,"* Conference on Fairness, Accountability, and Transparency, 2018; S. Verma and J. Rubin, *"Fairness definitions explained,"* IEEE/ACM International Workshop on Software Fairness, 2018.

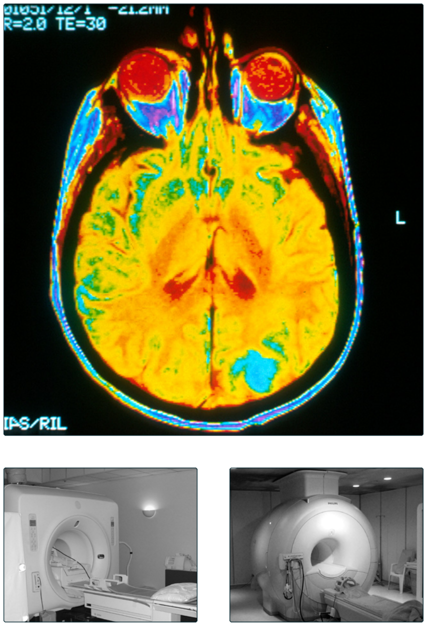
40 For a summary of several different fairness criteria for mortgage lending, see, for example, https://www.aspentechpolicyhub.org/wp-content/uploads/2020/07/FAHL-Cheatsheet.pdf.

41 See, e.g., Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, *"Dissecting racial bias in an algorithm used to manage the health of populations,"* Science, 2019.

each other in that a system will be fair under one definition but unfair by another. Ultimately, deciding on an appropriate standard is a context-specific policy decision that cannot be made on purely technical grounds, and needs to incorporate interdisciplinary expertise and values.

The root causes of model malfunctions that lead to unfair outcomes (even impacts that are disparate by race) often have nothing directly to do with race or other demographics.

Importantly, the types of effects above, though undesirable, are in no way specific to protected characteristics like race. For example, an AI system for medical diagnosis (say, determining whether an MRI scan contains a tumor or not) may be more accurate using images from Manufacturer A's hardware than it is using images from that of Manufacturer B. In itself, this has nothing to do with demographics. However, if the hospitals using Manufacturer B's hardware were resource-constrained and served a poorer patient population, this difference could serve as a proxy for class or income, and then the overall system may end up producing disparate impacts when actually deployed - and, for example, over- or under-diagnosing tumors at much higher rates for certain groups. In addition, lower-income populations tend to disproportionately include people of color, women, and other protected groups, so this differing performance by manufacturer is likely to produce other disparate impacts as well, even though this is far from the root "problem" in the system. But even if none of this were the case, this kind of system would still pose a serious concern, as all patients, including those not in protected classes, could be harmed by a model that is not performing well on their local equipment.



An MRI scan of the brain, and two MRI machines from different manufacturers.
Photo: MRI image by Dr. Leon Kaufman. University Of California, San Francisco, found on the National Cancer Institute, public domain image; Photo of GE Signa MRI machine by Wikimedia user "BrokenSphere," Creative Commons BY-SA 3.0; Photo of Philips 3T Achieva MRI machine by Wikimedia user "KasugaHuang," Creative Commons BY-SA 3.0

For this reason, the overall potential impacts of a system must be considered in evaluating whether a system is functioning appropriately or not.

The effects described above can arise at any stage of the AI lifecycle, including problem formulation, data collection and processing, and modeling. In some cases, the system's malfunction may be for straightforward reasons; for example, the medical diagnosis system may be inaccurate for images from Manufacturer B's hardware simply because there were not enough such images included in the training set. In this case, the solution may be as simple as getting more such images and retraining the model.[42] In other cases, the problem may be more subtle and not simply related to data.

Ultimately, this is a complex and evolving topic with no simple answers, much like most policymaking. Having said this, developers should become aware of some of the sources of unfairness and proactively consider them. For example, race, gender, class, and other factors are inextricably tied to data in many social and economic domains, including geographic and housing data, medical treatment, consumer and business finance, employment, and criminal justice. One would need to be much more vigilant in these areas and with such data than, for example, ML models used to help manage battery usage in a phone.

One example of a concrete way to identify potential issues is to carry out so-called "disaggregated evaluation," or an evaluation of model performance broken out by different subgroups in the data, demographic or otherwise.

This can reveal, for instance, if a model is performing well for the population overall but very poorly for some group that is a small percentage of the data. Although there are valid legal privacy concerns about the collection of sensitive data like race, gender, and disability, it is important that corporations and governments are able to carry out these sorts of analyses in order to ensure fairness or equity.[43]

Privacy and security

Privacy and cybersecurity are two of the most important digital rights topics, and they also apply to AI and ML. In some cases, these apply to the collection and use of data in general; for example, some data is considered more sensitive, either because it is "identifying" or because it concerns a sensitive topic, such as medical information about an individual.[44] In these cases, there may be tight-

42 See, e.g., A. Kaushal, R. Altman, and C. Langlotz, *"Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms,"* Journal of the American Medical Association, 2020.

43 See, e.g., S. Barocas, A. Guo, E. Kamar, J. Krones, M. Morris, J. Vaughan, D. Wadsworth, and H. Wallach, *"Designing Disaggregated Evaluations of AI Systems: Choices, Considerations, and Tradeoffs,"* arXiv preprint arXiv:2103.06076, 2021.

44 A detailed discussion of terms like "identifying" is out of scope here, but a core concept in privacy law is "Personally Identifiable Information" or PII. There is no standard definition of PII, and particular policy frameworks and jurisdictions have their own definitions. For some general discussion, see P. Schwartz and D. Solove, *"The PII problem: Privacy and a new concept of personally identifiable information,"* NYU Law Review, 2011.

er restrictions around the use of this data and more rigorous expectations of how the data should be protected. These are illustrated in traditional privacy frameworks like the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which includes a Privacy Rule (which defines notions such as Protected Health Information) and a Security Rule (which defines security expectations around electronic records). These topics are not discussed further here, as they are not specific to AI.[45] This section instead briefly highlights some other aspects of these issues.

Although more general issues about data privacy and cybersecurity also apply in the context of AI, there are also novel privacy and cybersecurity concerns that arise.

Some of these topics are technical and not discussed in detail here, but briefly, there are specialized privacy attacks that can apply to ML models. Two examples related to privacy are "membership inference" and "model inversion," which broadly involve learning things about the underlying data that was used to train a model given access only to the trained model itself.[46] When a model is used in a sensitive domain, this may be a concern. Similarly, "proxies" (discussed above) can implicitly introduce a form of privacy loss. In security, there are concerns about "software supply chains": because ML relies heavily on a shared set of resources in the form of datasets, models, and software libraries (which themselves depend on other, lower-level libraries), many of which are open source, there are questions about their vulnerability to digital supply chain attacks.[47] These examples are merely illustrative.

Privacy, security, fairness, accuracy, and other desirable goals or characteristics of systems are often in tension with each other. The trade-offs between these principles should be explicitly acknowledged, and developers must proactively and explicitly determine how best to navigate these trade-offs in any specific situation. This may require input from a range of stakeholders.

The issue of trade-offs between different aspects of AI systems, especially various digital rights, is one of the most fraught in the field.[48] For example, there can be trade-offs between privacy and

45 See P. Ohm, *"Broken promises of privacy: Responding to the surprising failure of anonymization,"* UCLA Law Review, 2009, for an overview of some recent issues in modern information privacy.

46 R. Binns, *"Privacy attacks on AI models,"* UK Information Commissioner's Office, 2019, available at https://ico.org.uk/about-the-ico/news-and-events/ai-blog-privacy-attacks-on-ai-models/.

47 A. Lohn, *"Poison in the Well: Securing the Shared Resources of Machine Learning,"* Georgetown Center for Security and Emerging Technology, 2021, available at https://cset.georgetown.edu/publication/poison-in-the-well/.

48 R. Binns, *"AI Auditing Framework: Trade-offs,"* UK Information Commissioner's Office, 2019, available at https://ico.org.uk/about-the-ico/news-and-events/ai-blog-trade-offs/.

accuracy (or other measures of model performance). Usually, the more data is used, the better the model will perform. This can include collecting or using data on a larger number of people or augmenting data about each person with additional demographic or other data, both of which could be determined to reduce how "privacy-respecting" the system is. On the other hand, avoiding collecting demographic data, including data that is sensitive, may degrade accuracy, and could result in faulty output or even harmful consequences (in, say, an inaccurate medical diagnosis).

Similarly, privacy and fairness can conflict in different ways. In one case, if developers find that its system is unfair due to insufficient training data on a particular demographic population, they may want to collect more data from such groups to increase model accuracy. Separately from this, in order to test whether an AI system is unfair or discriminatory in the first place, it is generally necessary to collect data on populations with protected characteristics (e.g., to carry out disaggregated evaluation, as described above). The developers would then face a trade-off between privacy (not collecting the data on characteristics) and fairness (collecting and using the data to test the system and make it fairer). This trade-off is not theoretical; indeed, lacking access to this data is cited by practitioners as one of the chief impediments to building fairer AI systems.[49]

49 K. Holstein, J. Vaughn, H. Daumé, M. Dudik, and H. Wallach, *"Improving fairness in machine learning systems: What do industry practitioners need?,"* CHI, 2019.

This is just one example; there are a range of other ways to balance privacy protections while enabling productive use or sharing of data, such as through synthetic data, de-identification, privacy-preserving data analysis algorithms (e.g., using secure multiparty computation), or governance structures like data sharing or confidentiality agreements where law permits the data to be shared.

There can also be trade-offs between accuracy and fairness, privacy and data security, explainability and accuracy, and so on. All in all, none of these rights can be taken to be a universal good.

Community engagement and participation

Public engagement is the work done by officials to meet and invite constituents into the processes of governance. This work takes many forms, from town halls and community boards to new technology-enabled modes of engagement such as crowd-sourcing apps, participatory democracy platforms, or social media. No matter the form, public engagement is guided by the democratic principle that decisions should be made with the public, not just for the public. This is especially important in contentious areas of public life with high stakes, such as public safety, public health, education, or child welfare.

Such engagement is also important in AI and can be an essential aspect of the responsible development, use, and governance of certain systems. In the public sector, it is particularly important for the public to be engaged because the government is responsible for ensuring that technology reflects the concerns, needs, and values of constituents, accurately accounts for impacts, and is deployed in an accountable manner, ideally in a way that supports a sense of trust, respect, and empowerment among constituents.

Especially because this is an emerging topic in the context of AI, determining when and how to do engagement, and what form it should take in each given situation, can be complex and challenging; best practices and standards for robust public engagement in AI are not yet agreed upon and are themselves the subject of active current research, and the complexity and novelty of these systems will likely require new, innovative methods to enable robust and meaningful participation.

Engagement plays an essential role for several practical reasons in addition to the high-level principles above. For example, community concerns that arise in engagement efforts can sometimes overshadow any benefits the system may have for the community involved, and when AI systems are deployed that do not reflect community needs, either in actuality or in perception, they may receive pushback from the community and ultimately not be adopted regardless of any other merits of the project or how careful the developers may have been with considering other ethical issues. There are many examples of this around the world, from residents organizing against the installation of biometric locks in their housing complexes to public protests over the way models were used to guess at what scores students were expected to get on admission exams.

There are also many applications that do not involve direct or meaningful human impacts — such as the algorithms used to optimize battery usage in a smartphone or many internal administrative applications in organizations — and in those cases, engagement may be unnecessary or can even be counterproductive.[50]

Public engagement should be considered with any system or process that uses computation to aid or replace decisions or policies that impact opportunities, access, liberties, rights, or safety.

Engagement can be used for different purposes and at different points in the AI lifecycle, depending on the context of the project. In the following example, the engagement was done before system deployment and to help design the system itself; in other cases, engagement has been done post-deployment.[51]

412 Food Rescue and participatory system design

412 Food Rescue is a non-profit located in Pittsburgh that matches donor organizations with expiring food to non-profit recipient organizations, and the organization decided to build an AI system to allocate donations because its existing manual approach was both time-consuming and inequitable. However, a difficult tradeoff quickly arose: because the donors tended to live in different and wealthier areas than recipients, increasing equity (allocating to recipients with greatest need) meant decreasing efficiency (longer distances to travel for volunteers).

412 partnered with researchers at Carnegie Mellon University to develop the system in a participatory way.[52] Researchers had stakeholders participate in each stage of the AI development lifecycle,

50 For administrative applications, attention to topics like impact on work, organizational structures and processes, and job security can also be important.

51 See, e.g., A. Brown, A. Chouldechova, E. Putnam-Hornstein, A. Tobin, and R. Vaithianathan, *"Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services,"* Proceedings of the CHI Conference on Human Factors in Computing Systems, 2019.

52 M. K. Lee, D. Kusbit, A. Kahng, J. Kim, X. Yuan, A. Chan, D. See, R. Noothigattu, S. Lee, A. Psomas, and A. Procaccia, *"WeBuildAI: Participatory framework for algorithmic governance,"* ACM: Human-Computer Interaction, 2019.

including determining which input features they felt were important (such as travel time, income level, or food access), voting on which model predictions best reflected what they would consider an appropriate equity-efficiency trade-off, and more. Importantly, the process of "engagement" or "participation" is not simply a matter of hearing people's opinions; here, it involved implementing a structured way to elicit data on people's preferences as well as a rigorous method (in this case, based on the theory of social choice from economics) to aggregate individual opinions into an overall policy. When stakeholders were interviewed after system implementation, researchers found that they felt the system was fair, in part due to the participatory approach used and in part because of the actual outcomes achieved. In this case, stakeholders needed to be engaged before the system was deployed; although many social domains exhibit complex trade-offs of this type in which there is no obvious "right" answer, it can be possible to directly incorporate varied views in the design to both achieve better outcomes and earn the trust of the people involved.

Conclusion

To build a healthy AI ecosystem for New York City, local decision-makers must work with a clear understanding of the technology and key practical and ethical considerations around its design, use, and governance. This AI Primer aims to equip decision-makers with a helpful foundation as they begin to engage with AI in an increasingly broad range of ways — to build teams, evaluate products, outline governance measures, formulate policy, or simply work to further develop their own knowledge and capacity.

Using and assessing AI can be a highly complex endeavor, and each case requires a detailed evaluation of a range of goals and factors. These can both be in tension with each other and be highly contingent on the context of each case. In that sense, this Primer can serve as an aid, but there can be no explicit prescription for how to make sound decisions, just as there can be no universal formula that guides policymakers when navigating difficult trade-offs. Indeed, as we have emphasized, many key issues that arise will often not be about AI itself or technical details at all. Moreover, this field, and many of the particular aspects described here, are very rapidly evolving, and it is not uncommon to see research or reporting that upends the status quo in a given area. The breadth and range of teams and use cases is further fueled by the fact that access to this technology, even at relatively large scale, is increasingly widely available and not limited solely to large corporations or governments, or even to individuals with significant technical expertise or formal training.

In AI, more so than many other areas, there are still far more questions than answers. For all these reasons and others, it will be critical to continue to learn and to evolve this framework as these broader efforts progress.

Further references

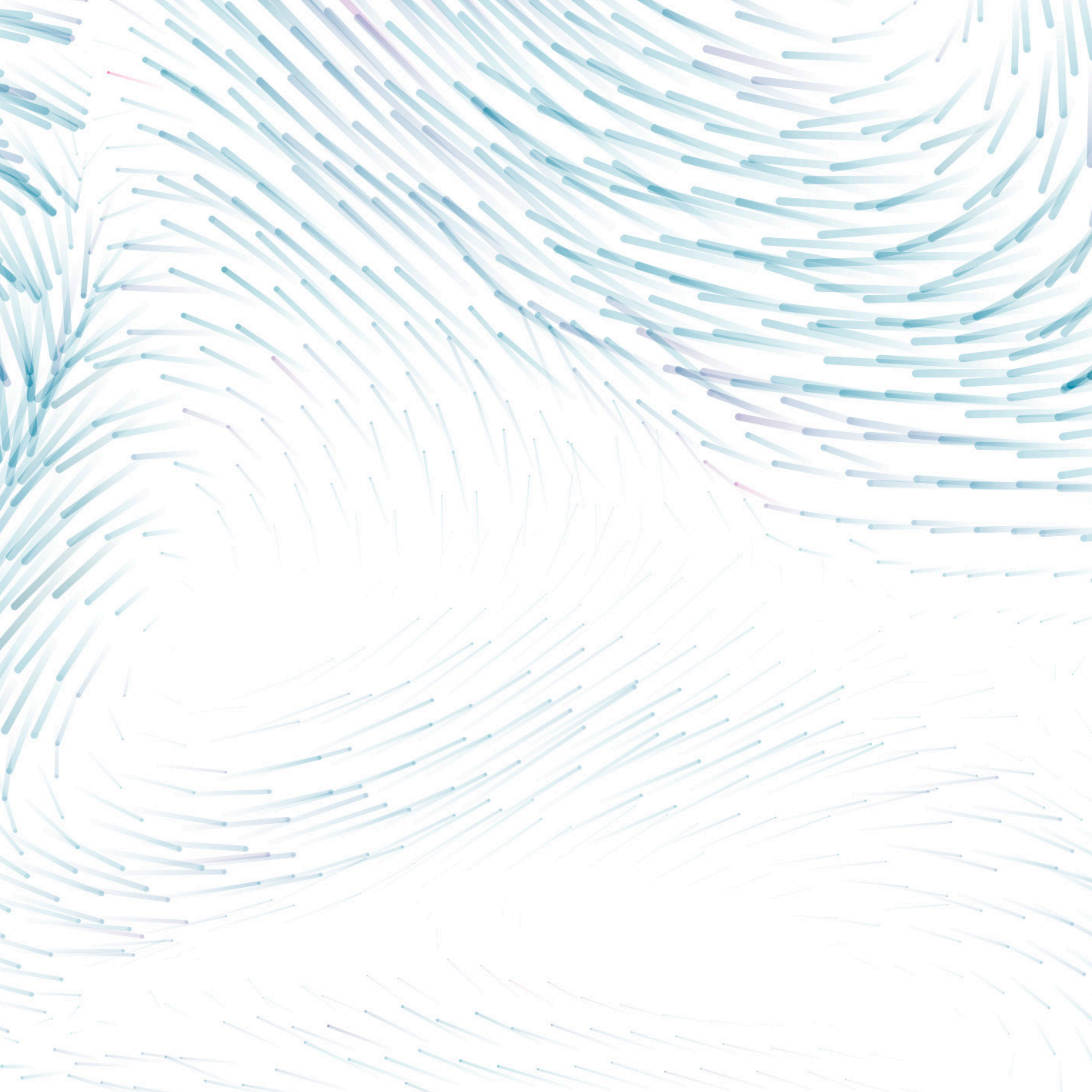
For some common texts on machine learning, see C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006; T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning,* second edition, Springer, 2016; I. Goodfellow, A. Courville, and Y. Bengio, *Deep Learning*, MIT Press, 2016; M. Hardt and B. Recht, *Patterns, Predictions, and Actions: A story about machine learning*, 2021, available at https://mlstory.org.

For a more general textbook on AI, see S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, fourth edition, Pearson, 2020.

There are also books focused on implementation, such as F. Chollet, *Deep Learning with Python*, O'Reilly, 2021; A. Géron, *Hands-On Machine Learning with scikit-learn, Keras, and TensorFlow*, O'Reilly, 2019; J. Howard and S. Gugger, *Deep Learning for Coders with Fastai and PyTorch: AI Applications Without a PhD*, O'Reilly, 2020.

For a general perspective on these areas, see M. Jordan, *"Artificial intelligence—the revolution hasn't happened yet,"* Harvard Data Science Review, 2019, and M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*, Farrar, Straus and Giroux, 2019.

Supplement B:
The Voices that Shaped this Strategy



The Voices that Shaped this Strategy

The inclusion of external organizations here should be taken to be purely for purposes of transparency and not to indicate their endorsement of the document's contents.

- → Algorithms Management and Policy Officer
- → Center for Innovation through Data Intelligence
- → City Commission on Human Rights
- → Department of City Planning
- → Department of Citywide Administrative Services
- → Department of Consumer and Worker Protection
- → Department of Education
- → Department of Design and Construction - Town+Gown Program
- → Department of Finance
- → Department of Health and Mental Hygiene
- → Department of Information Technology and Telecommunications
- → Department of Small Business Services
- → Department of Transportation
- → Mayor's Office of Climate Resiliency
- → Mayor's Office of Climate and Sustainability
- → Mayor's Office of Criminal Justice
- → Mayor's Office of Data Analytics
- → Mayor's Office for Economic Opportunity (NYC Opportunity)
- → Mayor's Office of Information Privacy
- → Mayor's Office of Minority and Women Owned Business Enterprises
- → Mayor's Office of Operations
- → New York City Cyber Command
- → New York City Economic Development Corporation
- → New York City Emergency Management
- → New York City Police Department
- → New York City Taxi and Limousine Commission

- AI Now Institute
- AI for the People
- BetaNYC
- Center for Democracy and Technology - GRAIL Network
- Center for an Urban Future
- City University of New York (CUNY)
- Columbia University - Data Science Institute
- Cornell Tech
- Data & Society
- Data for Black Lives
- Facebook AI Research
- Glitch
- Google
- M/WBE GovTek Connect
  - Capstone Strategy Group
  - Compulink Technologies
  - Data Conversion Lab
  - QED National
  - Sygma Technology
  - UAO Consultants
- Microsoft
- NYU - Center for Responsible AI
- NYU - Center for Urban Science and Progress
- NYU - GovLab
- Princeton University - Center for Information Technology Policy
- R Street Institute
- Silicon Harlem
- Tech:NYC
- Twitter - Machine Learning, Ethics, Transparency and Accountability (META) group
- Union Square Ventures
- University of California, Berkeley - CITRIS Policy Lab
- University of Chicago - Crime Lab
- University of Texas, Austin - Good Systems

fin.

NYC CTO