

Agency Compliance Reporting of Algorithmic Tools

CALENDAR YEAR 2025

Overview

This is the fourth annual report on algorithmic tools as required by Local Law 35 of 2021 (“LL 35”)¹ and managed by the Office of Technology & Innovation (“OTI”) under Executive Order 3 of 2022 (“EO 3”)². Each year, OTI provides guidance to agencies on their reporting, ensures that agency materials are properly prepared for the public, and strives to make this work more comprehensive and accessible. This year’s report – which contains a record number of agencies disclosing a record number of algorithmic tools – demonstrates our commitment to increasing transparency and public awareness about how and why New York City government deploys these advanced technologies.

While the definition of an “algorithmic tool” is established in LL 35, OTI breaks this down into three criteria:

1. **Data analysis:** The tool must use machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis;
2. **Decision-making:** The tool must make or assist in making decisions about and implementing policies; and
3. **Public impact:** These decisions must materially impact the rights, liberties, benefits, safety, or interests of the public, including their access to available city services and resources for which they may be eligible.

A tool must meet **all three** criteria in order to be eligible for reporting.

This report is also available as an open data set at <https://data.cityofnewyork.us/d/jaw4-yuem>. Questions or feedback about this report can be directed to Renata Gerecke at rgerecke@oti.nyc.gov.

Key Changes for 2025

OTI has made changes to the report based on public feedback from engagement sessions in Queens Public Library branches over the summer of 2025.³

Agencies were asked to add detail to the data descriptions and vendor description for each tool. OTI also added questions about the autonomy of each tool, how often each

¹ For the full text of LL 35, see:

<https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA29B34-9266-4B52-B438-A772D81B1CB5>

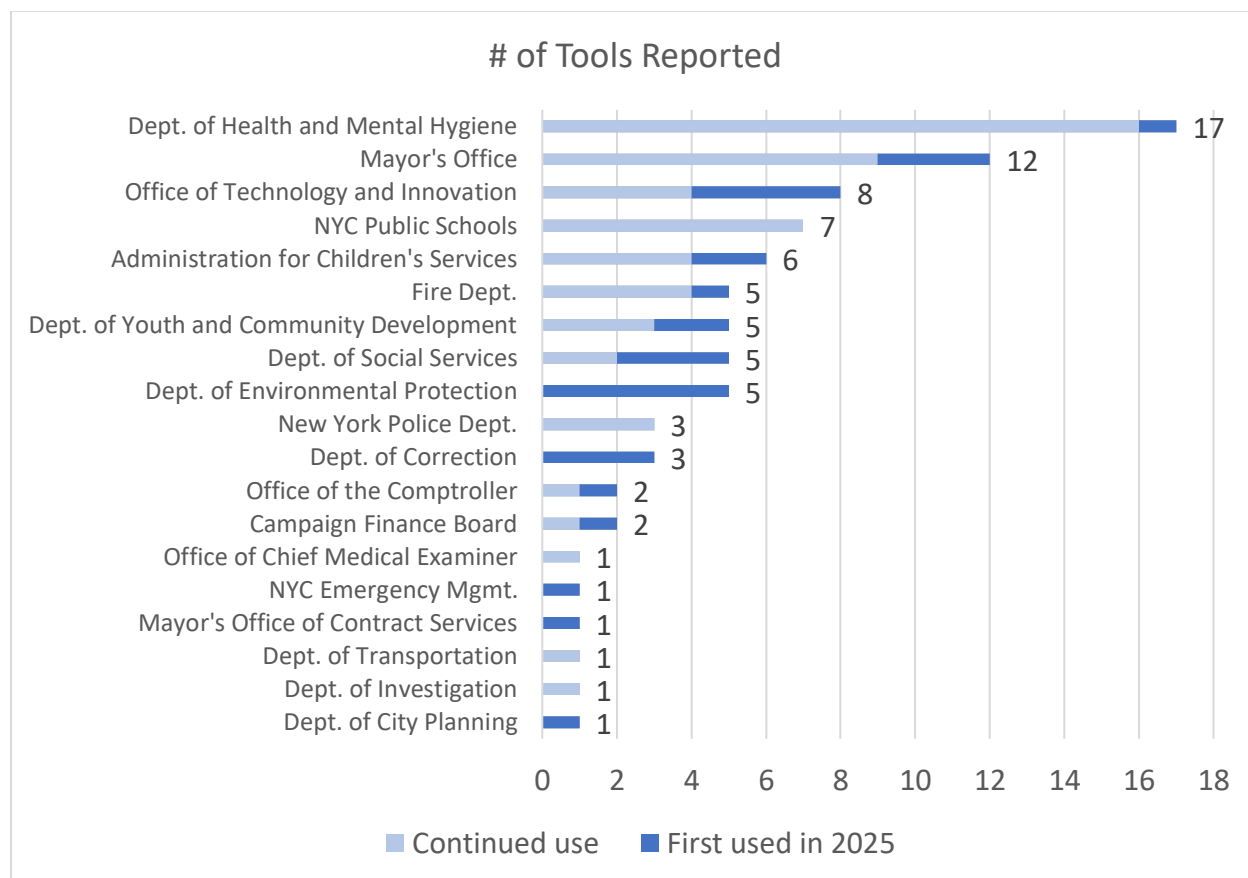
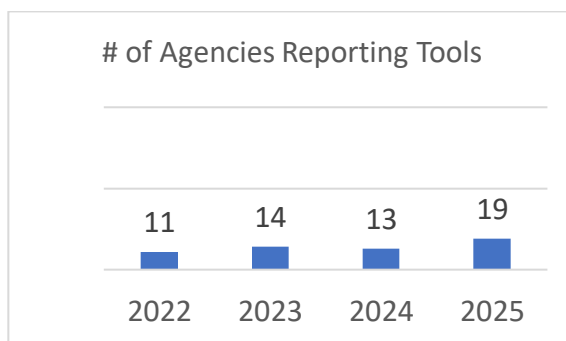
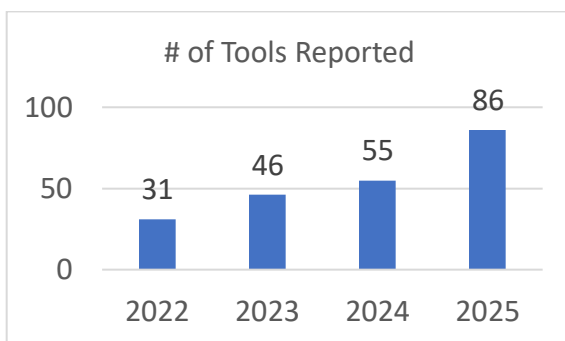
² For the full text of EO 3, see: <https://www.nyc.gov/office-of-the-mayor/news/003-002/executive-order3>

³ Four sessions were held at different Queens Public Library branches, with 10-30 attendees at each session. Future sessions are being planned for other boroughs.

tool is used, and whether there is a website where the public can learn more information about each tool.

Summary of Agency Reports

In 2025, a record 58 agencies participated in the reporting process. Of those agencies, 19 reported using a total of 86 algorithmic tools – both also record numbers. The 56% increase in reported tools used is the largest to date.



The following table (continued on the next page) summarizes reporting from City agencies for 2025.

Agency	Tools Reported
Administration for Children's Services	6
Board of Correction	0
Board of Standards and Appeals	0
Borough President of Manhattan	0
Business Integrity Commission	0
Campaign Finance Board	2
Civil Service Commission	0
Civilian Complaint Review Board	0
Commission on Human Rights	0
Commission on Racial Equity	0
Commission to Combat Police Corruption	0
Conflicts of Interest Board	0
Department for the Aging	0
Department of Buildings	0
Department of City Planning	1
Department of Citywide Administrative Services	0
Department of Consumer and Worker Protection	0
Department of Correction	3
Department of Cultural Affairs	0
Department of Design and Construction	0
Department of Environmental Protection	5
Department of Finance	0
Department of Health and Mental Hygiene	17
Department of Housing Preservation and Development	0
Department of Investigation	1
Department of Parks & Recreation	0
Department of Probation	0
Department of Records and Information Services	0
Department of Sanitation	0
Department of Small Business Services	0
Department of Social Services	5
Department of Transportation	1
Department of Veterans' Services	0
Department of Youth and Community Development	5
Economic Development Corporation	0
Equal Employment Practices Commission	0
Financial Information Service Agency	0

Agency	Tools Reported
Fire Department	5
Landmarks Preservation Commission	0
Law Department	0
Mayor's Office	12
Mayor's Office of Contract Services	1
Mayor's Office of Criminal Justice	0
New York Police Department	3
NYC Emergency Management	1
NYC Housing Authority	0
NYC Public Schools	7
Office of Administrative Trials and Hearings	0
Office of Chief Medical Examiner	1
Office of Collective Bargaining	0
Office of Payroll Administration	0
Office of Technology and Innovation	8
Office of the Actuary *	0
Office of the Comptroller *	2
Public Advocate	0
Public Design Commission	0
School Construction Authority	0
Taxi & Limousine Commission	0
Total	86

* New to reporting this year.

Table of Contents

Administration for Children’s Services (ACS)	2
Campaign Finance Board (CFB).....	10
Department of City Planning (DCP)	13
Department of Correction (DOC).....	15
Department of Environmental Protection (DEP).....	18
Department of Health and Mental Hygiene (DOHMH)	24
Department of Investigation (DOI)	43
Department of Social Services (DSS)	45
Department of Transportation (DOT).....	51
Department of Youth and Community Development (DYCD)	52
Fire Department (FDNY).....	58
Mayor’s Office (MO)	64
Mayor’s Office of Contract Services (MOCS)	79
New York Police Department (NYPD)	80
NYC Emergency Management (NYCEM).....	84
NYC Public Schools (NYCPS).....	85
Office of Chief Medical Examiner (OCME).....	94
Office of Technology and Innovation (OTI)	95
Office of the Comptroller (Comptroller)	105
Appendix.....	108

Administration for Children's Services (ACS)

Accelerated Safety Analysis Protocol (ASAP) Tool

Department: Division of Policy, Planning, and Analysis

First Used: May 2018

Updated in 2025: Yes

Purpose Type: Performance evaluation

Computation Type: Ranking

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Individuals

Vendor(s): None

Tool Description

Predictions of Severe Harm (identifying likelihood of substantiated allegations of physical or sex abuse within the next 24 months) are based on a machine learning methodology and are calculated for all children involved in active investigations early in the investigation (day 10). An investigation is assigned a numeric likelihood of this outcome based on the child in the case with the highest likelihood. The ACS Quality Assurance unit in the Division of Child Protection reviews about 3,000 active investigations annually, selecting those with the highest likelihood of severe harm.

Tool Purpose

The Quality Assurance (QA) unit in the Division of Child Protection at ACS has the capacity to review about 3,000 investigation cases out of about 50,000 investigations annually. ACS developed this predictive model to support the selection of cases for QA review. Open investigations involving children with the greatest likelihood to experience future severe harm – substantiated allegations of physical or sex abuse in the following 24 months – are selected for review. The tool does not support decisions about services or interventions for individuals or families involved with ACS, beyond the selection of the case for this additional QA review.

If the QA review team identifies gaps in routine, required documentation or practice, the team speaks with the field office conducting the investigation and follows up to make certain these gaps have been addressed. Scores are not shared with staff in the QA unit or the investigative unit. The model only supports the decision about which investigations are prioritized for review by the QA unit.

Update Description

Historically, the model made predictions only on investigation cases; however, over time, the proportion of child protection cases that are tracked to Collaborative Assessment, Response, Engagement & Support (CARES) has increased, and it has become necessary

to identify the risk of severe harm for children in these cases as well. Updated model predictions include CARES cases.

The model was also retrained on new cases to tackle data drift resulting in part from policy changes and practices.

The model no longer collects identifying information; the only identifier is the case number.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model on ACS historic administrative data about closed investigations from 2012 to 2022. The training set included about 815,284 observations (80 percent of the historic data). The model was tested on closed investigations from January 2012 to November 2022 with 205,241 observations (20 percent).

Input Data

Predictions are based on administrative data about prior and current child welfare involvement including investigations triggered by a New York State Central Register call and time spent in foster care. Only ACS administrative data are used in the model.

Output Data

Rank ordered list of open investigation cases involving children with the highest likelihood to experience future severe harm, defined as substantiated allegations of physical or sex abuse in the following 24 months to be reviewed by a special QA Review Team.

Caseloads Projection

Department: Division of Policy, Planning, and Analysis

First Used: July 2024

Updated in 2025: Yes

Purpose Type: Resource allocation

Computation Type: Forecasting

Autonomy: Informative

Frequency of Use: Quarterly

Population Impacted: Geographic space; Individuals (ACS staff)

Vendor(s): None

Tool Description

The model estimates weekly caseloads by area for the next 12 months from the run date (day 14 of the prevention case).

The model is a time series model that forecasts caseloads citywide. The forecast analysis is performed at a more granular geography for the borough level. The model accounts for seasonality, trend, and variation in data to accurately project future caseloads.

Staffing decisions are made based on projected caseloads, historical attrition rate, and current team size.

Tool Purpose

Caseloads projections help identify how busy certain neighborhoods are likely to be in the next 12 months to optimize workforce distribution across different boroughs in the city. As social workers complete their training program, they are assigned to field offices. The right number of Child Protective Service Social Workers needs to be assigned to each borough based on the workload.

Update Description

ACS retrained the model on more recent ACS historic caseloads data from July 2021 to December 2024. The model was tested on caseloads from January 2025 to November 2025.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model was trained on ACS historic caseloads from July 2021 to December 2024 and tested on caseloads from January 2025 to November 2025 to forecast caseloads over the next 12-month period.

Input Data

Predictions are based on administrative data on investigations and Family Service Units' involvement is extracted from CONNECTIONS (a child welfare computer system that provides for the documentation of information about families and children in New York State). A snapshot of all open investigations was taken at the start of the week (every Monday). This value is treated as the average caseload for the week. Only ACS administrative data are used in the model.

Output Data

Forecast of number of open investigations involving children by borough.

Family Team Conference Model (FTC)

Department: Division of Policy, Planning, and Analysis

First Used: February 2025

Purpose Type: Resource allocation

Computation Type: Ranking

Autonomy: Supervised

Frequency of Use: Monthly

Population Impacted: Individuals (Families ACS serves)

Vendor(s): None

Tool Description

The Family Team Conference (FTC) is a structured planning conference facilitated by ACS with families. The Prevention Reassessment Planning FTC is held with families that have been engaged in prevention services for approximately 100 days. The purpose of this conference is to assess whether the current services are effectively meeting the family's needs, determine whether additional support or intervention is necessary, and strengthen collaboration between ACS, the families they serve, and provider agencies to promote better outcomes.

The FTC model was developed for the Division of Prevention Services to support timely, data-informed decision-making for families engaged in child welfare prevention services. The model produces a risk score for each child approximately 100 days after the start of a new prevention service - well in advance of the FTC-Reassessment Planning Conference. This prediction estimates the likelihood that the child will be the subject of an indicated New York State Central Register investigation within the next 24 months from the observation date.

Tool Purpose

The Division of Prevention Services (DPS) uses these predictions to help identify and prioritize active prevention cases that may benefit from enhanced attention during the FTC Reassessment Planning Conference. By surfacing families with elevated predicted risk, the model enables senior staff in DPS to assign ACS Child and Family Specialists (CFS) to these conferences, where they may tailor discussions about targeted service adjustments or strengthened supports.

The FTC model supports the core purposes of the FTC Reassessment Planning Conference, which include: 1. Evaluating the Family's Progress: Highlight families whose risk trajectory suggests the need for closer examination of progress over the previous six months. 2. Reassessing Risk and Safety: Provide data-driven insight into factors

associated with elevated likelihood of future indicated allegations. 3. Adjusting Service Plans: Inform decisions on modifying service plans to address emerging needs, service gaps, or barriers to engagement. 4. Supporting Family Voice in Decision-Making: Ensure families have the opportunity to participate meaningfully in discussions informed by clear, evidence-based information. 5. Strengthening Collaboration Across Supports: Equip CFS staff and provider agencies with consistent, actionable information to better coordinate planning and interventions.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model on ACS historic administrative data about closed investigations from 2014 to 2021. The training set contains 138,712 cases (80 percent of the historic data) ending between Jan 2014 and December 2021. The test set consisted of 34,862 cases (20 percent) ending between Jan 2014 and December 2021. The data was split such that no family appears in both sets.

Input Data

Predictions are based on administrative data about prior and current child welfare involvement at the start of a case. This includes New York State Central Register (SCR) investigations and time spent in foster care. Only ACS administrative data are used in the model.

Output Data

A monthly rank ordered list of open prevention cases involving children whose families have the highest likelihood of being in and indicated SCR investigation within 24 months from date of observation.

Housing Prioritization

Department: Division of Policy, Planning, and Analysis

First Used: April 2023

Purpose Type: Resource allocation

Autonomy: Informative

Population Impacted: Individuals

Updated in 2025: No

Computation Type: Ranking

Frequency of Use: Monthly

Vendor(s): None

Tool Description

The model estimates a risk score for a child receiving prevention services whose family will apply and be eligible for a homeless shelter within 12 months from the start of service (day 14 of the prevention case).

The model helps predict the risk of application for homeless shelters among families receiving prevention services. With a limited number of vouchers available, the risk model helps ACS prioritize housing assistance for those families at greatest risk of becoming homeless.

The service provider meets with the family to conduct a qualitative assessment of the family's housing needs and vouchers are offered based on their findings. This is one of multiple ways in which ACS can identify families potentially eligible for housing assistance.

Tool Purpose

The city has allocated 100 housing vouchers to families receiving ACS prevention services. The shelter application model identifies the likelihood of a family in prevention services applying for homeless shelter within 12 months beyond the current prevention case. The model uses a machine learning methodology and is calculated for all children in a prevention case. ACS prevention services uses the results as one of several ways of identifying possible families that service providers can assist in applying for shelter.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model on ACS historic administrative data regarding preventive services started between 2014 and 2020. The training set contains 140,242 observations (80 percent of the historical data) between January 2014 and December 2020. The test set consisted of 34,508 observations (20 percent) between January 2014 and December 2020. The data is split such that no family appears in both sets.

Input Data

Predictions are based on administrative data about prior and current child welfare involvement at the start of a case. This includes New York State Central Registry investigations and time spent in foster care. Only ACS administrative data are used in the model.

Output Data

Rank ordered list of open prevention cases involving children whose families have the highest likelihood of applying for a homeless shelter within 12 months of starting a prevention service.

Prevention Score Card

Department: Division of Policy, Planning, and Analysis

First Used: September 2021

Updated in 2025: Yes

Purpose Type: Performance evaluation

Computation Type: Ranking

Autonomy: Monitored

Frequency of Use: Yearly

Population Impacted: Individuals

Vendor(s): None

Tool Description

Predictions of Repeat Maltreatment (identifying the likelihood of being involved in a future indicated investigation within the next 24 months at the start of service) are based on a machine learning methodology and are calculated for all children receiving prevention services from ACS prevention service providers.

Tool Purpose

The Repeat Maltreatment model is used to make predictions on day 10 from the start of the prevention case to assess the risk of the family at the beginning of the service. A prevention case is assigned a numeric likelihood of an indicated investigation based on a New York State Central Register (SCR) within 24 months from the start of a prevention service.

The prevention providers are assessed for their performance based on the service needs/risk levels of the families they've served during the previous fiscal year.

The programs were sorted and ranked based on their average risk, and then divided into four quartiles by rank order: the top 25 percent of programs are classified as the Very High-Risk Cohort, the next 25 percent of programs as the High-Risk Cohort, the next 25 percent as the Medium-Risk Cohort, and the lowest 25 percent as the Low-Risk Cohort. Assignment to a cohort is not a way of performance assessment of the program but to group prevention service providers for fair comparisons based on the risk level of families they served.

Update Description

ACS retrained the model on recent ACS historic administrative data about closed investigations from July 2010 to June 2019. The training set included about 338,467

observations (80 percent of historic data). The model was tested on about 84,494 observations (20 percent). The tool was also updated such that identifying information is not included in the training, input, or output data, only a case number.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model on recent ACS historic administrative data about closed investigations from July 2010 to June 2019. The training set included about 338,467 observations (80 percent of historic data). The model was tested on about 84,494 observations (20 percent).

Input Data

Predictions are based on administrative data about prior and current child welfare involvement at the start of a case. This includes SCR investigations and time spent in foster care. Only ACS administrative data are used in the model.

Output Data

The model is used for generating a scorecard of prevention service providers by categorizing prevention programs based on the average risk profile of the cases they served during the assessment year. These groupings of program cohorts provide context for understanding the scorecard, as it allows for performance comparison of programs that accepted and served families with similar risk profiles.

Service Utilization Model

Department: Division of Policy, Planning, and Analysis

First Used: January 2025

Purpose Type: Resource allocation

Computation Type: Forecasting

Autonomy: Supervised

Frequency of Use: Biweekly

Population Impacted: Individuals (Families ACS serves)

Vendor(s): None

Tool Description

Predictions of Repeat Maltreatment (identifying the likelihood of being involved in a future indicated investigation within the following 24 months towards the end of service) are based on a machine learning model and are calculated for all children

receiving prevention services from ACS providers. A prevention case is assigned a numeric likelihood of an indicated investigation by the New York State Central Register (SCR) within 24 months of the end of a prevention service.

Tool Purpose

The Repeat Maltreatment model is a low-risk Service Utilization Model and is used by Department of Preventive Services (DPS) managers to identify the low-risk cases in prevention and to understand the length of service of these cases. DPS utilizes these reports to understand which programs have cases with a high length of service and are low risk. This report and the analysis are utilized internally for assessment purposes to understand program utilization and capacity.

The information from this report is not shared with providers.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

ACS trained the model on ACS historic administrative data about closed investigations from January 2014 to December 2021. The training set included about 138,712 observations (80 percent of the historic data). The model was tested on the remaining 34,862 observations (20 percent).

Input Data

Predictions are based on administrative data about prior and current child welfare involvement at the end of a prevention service case. It includes SCR investigations and time spent in foster care. Only ACS administrative data are used in the model.

Output Data

Rank ordered list of open prevention cases involving children with the lowest likelihood to experience repeat maltreatment, defined as substantiated allegations of abuse or neglect in the following 24 months to be reviewed by Preventive Services managers to identify the low-risk cases and to understand the length of service of these cases.

Campaign Finance Board (CFB)

Adobe Creative Suite

Department: Public Affairs

Version effective March 2026

First Used: February 2024

Purpose Type: Information presentation

Autonomy: Informative

Population Impacted: Individuals (People engaging with NYC Votes creative ad campaigns and voter engagement and education materials across digital and print)

Updated in 2025: Yes

Computation Type: Data transformation

Frequency of Use: Project-based throughout the year

Vendor(s): Adobe

Tool Description

The CFB uses Adobe Creative Suite to edit images and uses some features that use generative AI to allow designers to edit images with reduced manual work.

Tool Purpose

The CFB uses Adobe Creative Suite to facilitate efficient graphic design that appear in some content, including NYC Votes creative ad campaigns and for voter engagement and education materials across digital and print. For example, Adobe Photoshop is used to remove figures in background photos who did not provide consent for usage. It is used to modify background elements that would be distracting (e.g., external telephone wires, shadows). As another example, the “Generative Recolor” feature in Adobe Illustrator allows designers to generate color palettes and apply those new colorways to patterns and artwork. Previously, designers needed to do this manually by individually clicking each element and applying a new color one at a time; “Generative Recolor” does this more efficiently.

Update Description

Improvements to photo generation/modification, vector graphic creation, design efficiency.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

According to Adobe’s website, “generative AI Image models were trained on licensed content, such as Adobe Stock, and public domain content where copyright has expired.”

Input Data

Images.

Output Data

Visual content; primarily voter engagement and education materials across digital and print.

Vendor Information

Vendor Name(s): Adobe

Vendor Type(s): Procurement/Paid pilot

This is a platform service.

Zoom

Department: Public Affairs

First Used: July 2020

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Regularly/monthly

Population Impacted: Individuals (People who attend CFB public meetings)

Vendor(s): Zoom

Tool Description

Zoom is a virtual meeting platform; Zoom has an auto closed-caption function that uses AI.

Tool Purpose

The CFB uses Zoom for public meetings, including Board meetings, Voter Assistance Advisory Committee (VAAC) meetings, and All Partners meetings. The “automated captions” option is enabled on the agency’s Zoom account; Zoom-generated captions are available to attendees, who can choose to turn on this feature during the meeting.

(Note: For VAAC meetings and All Partner meetings, the agency also provides human-typed Communication Access Realtime Translation (CART) services as an option to all attendees. For All Partners meeting attendees, the CART transcription is made available after the call.)

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Version effective March 2026

Training data is proprietary to Zoom.

Input Data

Speech at the CFB's public meetings.

Output Data

Text in a transcript and captions.

Vendor Information

Vendor Name(s): Zoom

Vendor Type(s): Procurement/Paid pilot

This is a platform service.

Department of City Planning (DCP)

Voice to Vision

Department: Community Planning and Civic Engagement

First Used: March 2025

Purpose Type: Information presentation

Autonomy: Informative

Population Impacted: Individuals (Voice to Vision shares individual comments shared during community planning); Geographic space; Property; Group, organization, or business

Computation Type: Data transformation

Frequency of Use: Voice to Vision is viewed 10+ times per day

Vendor(s): Massachusetts Institute of Technology (MIT): The Center for Constructive Communication

For more information: <https://dl.acm.org/doi/10.1145/3715070.3757228>

Tool Description

Voice to Vision is a repository of community input shared during neighborhood planning processes. Public comments were tagged and summarized by large language models and visualized online. This allows community members to see how their input shaped planning and policy decisions. As of December 2025, it has been used for the Jamaica

Neighborhood Plan (<https://v2v-jamaicaplan.ccc-mit.org/>) and the White Plains Road Neighborhood Plan (<https://v2v-white-plains-road-plan.ccc-mit.org/>).

Tool Purpose

Voice to Vision uses artificial intelligence to support several functions for tagging and summarizing comments made during neighborhood planning sessions with the general public. Locations mentioned in community comments are identified and geotagged using GPT-4.1-mini and GPT-5-mini. Community comments are also tagged based on common themes using GPT-4.1-mini and GPT-5-mini. Finally, summaries of common topic areas are drafted by Claude 3.5 Sonnet, which also provides citations back to the original comments. All tagging, summarization, and citation is reviewed by staff prior to publication.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

Comments from community engagement sessions related to the neighborhood plans.

Output Data

Potential locations mentioned in comments, themes and goals mentioned in comments, theme summaries with citations from comments.

Vendor Information

Vendor Name(s): Massachusetts Institute of Technology (MIT): The Center for Constructive Communication

Vendor Type(s): No-cost engagement

Voice to Vision was developed as a research project by MIT's Center for Constructive Communication in coordination with DCP.

Department of Correction (DOC)

NYC Jail Population Forecast

Department: Operations Research

First Used: January 2025

Purpose Type: Resource allocation

Computation Type: Forecasting

Autonomy: Informative

Frequency of Use: Weekly / Monthly

Population Impacted: Group, organization, or business

Vendor(s): None

For more information: <https://council.nyc.gov/budget/wp-content/uploads/sites/54/2025/09/DOC-MOCJ-FY25-TC-July-Report-Population-Forecast.pdf>

Tool Description

This is a two-year population projection for the New York City jail system based on historic trends from the past 12 years.

Tool Purpose

An older forecast model was created prior to this for internal informational and research purposes. However, in 2024, Terms and Conditions were added to DOC's budget requiring a more formal tool be developed and published for the jail system. The first of these reports was released in January 2025, and updates have been released every 6 months since. Internally, the Department has been using the tool for high-level planning and budget purposes.

Data Analyzed

Identifying information may be collected or disclosed in the training data and input data.

Training Data

Training data is from internal sources and includes historic aggregated jail admissions rates, NYPD daily arrest rates, daily average population counts, and information for incarcerated individuals including admission dates, historic changes in incarceration status (e.g., City Sentenced, Pretrial Detainee, State Ready), and length of stay.

Input Data

Inputs for inference include the latest aggregated population levels, admission rates, arrest rates, and information for incarcerated individuals including admission dates, historic changes in incarceration status, and length of stay.

Output Data

The system's output is a daily forecast for the expected population and the uncertainty around the daily forecast, for each day over the next two years. The uncertainty is estimated by the errors collected in a series of rigorous back-testing of the model and is shown as daily "prediction intervals" (i.e., the expected ranges for population) for each day over the same two-year period.

Pocketalk

Department: Division of Programs and Community Partnerships

First Used: January 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Supervised

Frequency of Use: Limited assessments (6 total)

Population Impacted: Individuals (DOC staff, people in custody, and visitors to DOC facilities)

Vendor(s): Pocketalk

For more information: <https://pocketalk.com/>

Tool Description

Pocketalk is designed to support real-time, two-way conversation translation across more than 90 languages. The Department of Correction has implemented a small-scale, limited pilot to assess the efficacy of the devices in a real-life environment. If successful, Pocketalk will be piloted across select facilities to supplement (not replace) qualified human interpretation services. Data captured through Pocketalk's use will assist the Department in monitoring usage trends, operational effectiveness, and potential expansion opportunities to enhance language accessibility across the agency.

Tool Purpose

Language translation for people in custody and visitors.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

Audio recordings of conversations.

Output Data

Written and spoken translation of the conversation into the specified language.

Vendor Information

Vendor Name(s): Pocketalk

Vendor Type(s): Off-the-shelf

Pocketalk is a two-way translation device. Pocketalk utilizes multiple engines to provide translations across 92+ languages, including localized dialects and slang.

Risk of Recidivism Tool

Department: Division of Programs and Community Partnerships

First Used: July 2025

Purpose Type: Resource allocation

Computation Type: Classification

Autonomy: Informative

Frequency of Use: Approximately weekly

Population Impacted: Individuals
(Individuals who are newly admitted to
DOC custody)

Vendor(s): None

Tool Description

The tool is designed to identify individuals at high risk of recidivism so they can receive additional services.

Tool Purpose

The purpose of the tool is to identify newly admitted individuals at high risk of recidivism within 12 months of discharge to the community so they can be offered targeted services that support successful reintegration. This initiative is funded by a Department of Justice grant that supports the hiring of two Discharge Planners to deliver these services, including weekly one-on-one case management, evidence-based curricula, and connections to community resources. The program is grounded in research

demonstrating that enhanced, targeted interventions for individuals at high risk of recidivism lead to improved outcomes.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Information about persons in custody discharged to the community in 2020 and 2021.

Input Data

Prior admissions, self-reported homelessness, self reported drug or alcohol misuse, whether the top charge is a misdemeanor, and age of first admission to DOC custody.

Output Data

The tool outputs a classification indicator of either “high” or “low” risk of recidivism. Program staff then provide additional support to those who are at high risk.

Department of Environmental Protection (DEP)

AI Cybersecurity Tool

Department: Bureau of Business Information Technology

First Used: July 2025

Purpose Type: Risk management

Computation Type: Classification

Autonomy: Monitored

Frequency of Use: Not specified

Population Impacted: Group, organization, or business

Vendor(s): Not disclosable

Tool Description

DEP uses AI to analyze cybersecurity data to identify potential threats quickly and coordinate an effective response.

Tool Purpose

The information produced by the tool is used to identify unusual or potentially cybersecurity-risk activity across our environment quickly. DEP uses the insights to prioritize investigations, streamline cybersecurity operations, and respond more effectively to potential cyber threats.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Computer and network logs, including DEP staff user IDs and computer information.

Output Data

Alerts, risk scores, unusual network patterns, potential cybersecurity issues, and recommended actions, including DEP staff user IDs and computer information.

Vendor Information

Vendor Name(s): Not disclosable

Vendor Type(s): Procurement/Paid pilot

Not disclosable due to cyber concerns and to prevent exposing details that could assist malicious actors.

FloodVision

Department: Bureau of Sustainability

First Used: September 2025

Purpose Type: Risk management

Autonomy: Fully autonomous

Population Impacted: Geographic space;
Property; Group, organization, or business

Computation Type: Scoring

Frequency of Use: Not specified

Vendor(s): ClearObject

Tool Description

FloodVision detects flooding in various areas of New York City by analyzing Department of Transportation (DOT) road camera footage and provides information about the flood to DEP to help inform data-driven decisions.

Tool Purpose

The FloodVision data is one of several data points that helps corroborate flooding hot spots in priority areas.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Images and video from DOT road cameras.

Input Data

Images and video from DOT road cameras.

Output Data

Whether flooding has been detected and the approximate depth of the flood.

Vendor Information

Vendor Name(s): ClearObject

Vendor Type(s): Professional services

ClearObject implemented FloodVision using Google Cloud Platform's AI cloud offerings.

Google Gemini

Department: Bureau of Business Information Technology

First Used: September 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Supervised

Frequency of Use: Once

Population Impacted: Individuals (App users)

Vendor(s): Google

Tool Description

DEP used Gemini to generate an image that was edited and used as the icon for the NYC Noise app.

Tool Purpose

DEP used Gemini to help in part of a process to create a mobile app icon.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

A text prompt.

Output Data

An image that was combined with other images to create the app icon.

Vendor Information

Vendor Name(s): Google

Vendor Type(s): Off-the-shelf

A free version of Google Gemini was used to create this graphic.

Microsoft Copilot for M365

Department: Public Affairs

First Used: July 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Multiple times per week

Population Impacted: Group, organization, or business; Individuals

Vendor(s): Microsoft

Tool Description

Microsoft Copilot for M365 is a tool that helps by drafting social media content, drafting scripts for video content, and providing translations of DEP content into other languages. This use was part of a pilot of Microsoft Copilot for M365 within Public Affairs, which began in July 2025 and is ongoing.

Tool Purpose

To assist in writing and presentation of agency objectives and public information for use on social media channels.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to Microsoft, with access to data in the DEP Microsoft tenant.

Input Data

Staff prompts Copilot using DEP factual information and a desired tone for the output.

Output Data

Appropriate content for social media or video scripts; translated materials.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): No-cost engagement

Microsoft provided the tool and some staff training.

Microsoft Copilot for M365

Department: Public Affairs

First Used: July 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Multiple times per week

Population Impacted: Group, organization, or business; Individuals

Vendor(s): Microsoft

Tool Description

Microsoft Copilot for M365 is a tool that helps by drafting social media content, drafting scripts for video content, and providing translations of DEP content into other languages. This use was part of a pilot of Microsoft Copilot for M365 within Public Affairs, which began in July 2025 and is ongoing.

Tool Purpose

To assist in writing and presentation of agency objectives and public information for use on social media channels.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to Microsoft, with access to data in the DEP Microsoft tenant.

Input Data

Staff prompts Copilot using DEP factual information and a desired tone for the output.

Output Data

Appropriate content for social media or video scripts; translated materials.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): No-cost engagement

Microsoft provided the tool and some staff training.

Mufflers Noise Violations

Department: Bureau of Environmental Compliance

First Used: January 2025

Purpose Type: Performance evaluation

Autonomy: Supervised

Population Impacted: Individuals (Owner of vehicle exceeding muffler noise limits); Geographic space

Version effective March 2026

Computation Type: Classification

Frequency of Use: Daily

Vendor(s): Intelligent Instruments

Tool Description

Noise monitoring cameras record loud noises along with their noise decibel levels. The recordings are analyzed automatically to determine whether the type of noise recorded is from a car muffler, honking, or other types of noise. The software also tries to use Optical Character Recognition to read the vehicle license plates.

Tool Purpose

The information received is used by DEP staff to help inform whether a noise summons should be issued and to whom. While the tool helps with the analysis, the staff is still reviewing the recordings manually to verify the information provided.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Video recordings from installed cameras on New York City roads.

Output Data

Vehicle plate number and noise type.

Vendor Information

Vendor Name(s): Intelligent Instruments

Vendor Type(s): Professional services

Intelligent Instruments provides the cameras that are installed in New York City roadways. Their software platform is doing the analysis on the recordings.

Department of Health and Mental Hygiene (DOHMH)

Bowtie2

Department: Disease Control – Public Health Laboratory

Version effective March 2026

First Used: June 2022

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: 1-2 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/BenLangmead/bowtie2>

Tool Description

Bowtie2 aligns sequencing data to a reference sequence using Burrows-Wheeler transformations. It is geared to use with Illumina sequencing data.

Tool Purpose

Bowtie2 is an intermediate step in the workflow to analyze COVID-19 variants in wastewater.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Laboratory-generated sequencing reads and an indexed organism reference genome.

Output Data

Genomic alignments indicating sequence variation relative to reference, supporting variant analysis and surveillance comparisons.

Burrows-Wheeler Aligner (BWA)

Department: Disease Control – Public Health Laboratory

First Used: July 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: 1-2 times per week

Version effective March 2026

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/lh3/bwa>

Tool Description

Burrows-Wheeler Aligner (BWA) aligns sequencing data to a reference sequence.

Tool Purpose

BWA aligns sequence data to reference using Burrows-Wheeler transformations. This tool is optimal for low-divergent genomic data and short read data, such as Illumina sequence data. This tool is used to predict the order in which the fragments generated by sequencers are pieced together to form a complete genomic sequence data. This tool is used for Legionella and PulseNet sequencing analyses.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Laboratory-generated sequencing reads and an organism-specific reference genome selected for public health investigation.

Output Data

Read-to-reference alignments showing genomic similarity and variation, used for variant detection and comparison.

ChoiceMaker (CM)

Department: Disease Control – Bureau of Investigations

First Used: June 2003

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: Daily

Population Impacted: Individuals
(Individuals of all ages with a record in the
Citywide Immunization Registry)

Vendor(s): HLN Consulting

For more information: <https://www.choicemaker.com/>

Tool Description

ChoiceMaker (CM) is a record-matching tool that identifies duplicate records belonging to the same individual.

Tool Purpose

CM is used by the Bureau of Investigations (BOI) and Healthy Homes to identify duplicate immunization and lead records. The outputs produced by CM are used in ongoing manual and automated deduplication processes (record merging).

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

The training data was developed by the vendor and DOHMH staff using DOHMH data. It includes DOHMH staff decisions on the likelihood of a match between 2 records.

Input Data

CM uses demographic data (e.g., names, date of birth, address, identifiers) and health event data (e.g., date and type of event) from BOI's Citywide Immunization Registry and Healthy Homes' LeadQuest registry in its evaluation.

Output Data

The program outputs a series of record pairs and a match probability for each pair.

Vendor Information

Vendor Name(s): HLN Consulting

Vendor Type(s): Professional services

A vendor was involved in the development of the program initially. CM is now available as an open-source program. The DOHMH implementation, which has been customized using the training data, is maintained by HLN Consulting.

Genome Analysis Toolkit (GATK)

Department: Disease Control - Public Health Laboratory

First Used: October 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Classification

Autonomy: Supervised

Frequency of Use: 3-4 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://gatk.broadinstitute.org/hc/en-us>

Tool Description

The Genome Analysis Toolkit (GATK) is a suite of tools for variant calling and filtering after sequence alignment. It uses naive Bayes to qualify aligned bases as sequence or erroneous data, which would be excluded from the final genomic sequence.

Tool Purpose

GATK is used to identify mutations and call upon differences from the reference, which is used to generate the predicted complete sequence.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Uses known variant site resources curated by public consortia for recalibration; no local model training performed.

Input Data

Reference genome and aligned sequencing data generated through public health laboratory workflows.

Output Data

Identified genomic variants relative to reference sequence, used for mutation tracking, resistance analysis, and outbreak comparison.

Guppy

Department: Disease Control - Public Health Laboratory

First Used: June 2020

Updated in 2025: No

Purpose Type: Data management

Computation Type: Classification

Autonomy: Supervised

Frequency of Use: A few times per month

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): Oxford Nanopore Technologies

For more information: <https://nanoporetech.com/software/other/guppy>

Tool Description

Guppy converts electric signals to predict a nucleotide and enables filtering of low-quality calls.

Tool Purpose

This is a tool designed specifically for Oxford Nanopore Technology data. This is a neural network based basecaller: a tool that determines nucleotide bases of a genetic material and converts electric signals into strings to represent genomic data. In addition to basecalling, the tool also performs filtering of low-quality reads (a stretch of sequenced genetic material). This is the initial step that converts electric signals to fragments of sequence data, which can then be used for COVID-19 sequencing analysis.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Vendor-developed basecalling models are trained by Oxford Nanopore Technologies on proprietary nanopore signal datasets representing electrical signal-to-nucleotide mappings.

Input Data

Raw electrical signal data generated by Oxford Nanopore sequencing instruments from clinical or surveillance specimens.

Output Data

Basecalled nucleotide sequences with quality metrics representing specimen genetic content, used for downstream genomic surveillance and outbreak analyses.

Vendor Information

Vendor Name(s): Oxford Nanopore Technologies

Vendor Type(s): Off-the-shelf

Developed and maintains the tool.

HELIX

Department: Division of Information Technology

First Used: November 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Individuals (Agency staff)

Vendor(s): Microsoft

Tool Description

HELIX is a generative tool that allows agency staff to ask general questions based on data available in the OpenAI GPT-4o model. In addition, HELIX can also be grounded on approved documents allowing staff to ask questions specific to those documents.

Tool Purpose

The tool is used to generate responses to prompts entered by staff, requesting the generative AI tool to compose a text response to a text input. These outputs are potentially shared with the public.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Prompts provided by the users of the system.

Output Data

Version effective March 2026

The output data for the tool is the response generated by the OpenAI large language model. In addition, any HELIX configurations grounded on approved documents will generate responses based on those documents.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): Off-the-shelf

Helix was built using Microsoft Azure OpenAI.

Immunization Calculation Engine (ICE)

Department: Disease Control – Bureau of Immunization

First Used: 1997

Updated in 2025: Yes

Purpose Type: Information presentation

Computation Type: Forecasting

Autonomy: Monitored

Frequency of Use: Daily

Population Impacted: Individuals (Anyone who needs a vaccine)

Vendor(s): HLN Consulting

For more information: <https://cdsframework.atlassian.net/wiki/spaces/ICE/overview>

Tool Description

The Immunization Calculation Engine (ICE) is an immunization evaluation and forecasting system whose default immunization schedule supports all routine childhood, adolescent, and adult immunizations based on the recommendations of the Advisory Committee on Immunization Practices. ICE is free and open-source.

Tool Purpose

ICE is used by the Bureau of Immunization to evaluate a patient's immunization history and generate appropriate immunization recommendations.

Update Description

Vaccine recommendations were updated in CY2025.

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

N/A

Input Data

ICE uses demographic data (e.g., date of birth) and vaccination data (e.g., immunization date, vaccine group and type) in the evaluation process. Data used are stored in the Citywide Immunization Registry.

Output Data

The program returns recommendations on whether a patient has completed a vaccine series or is due for vaccines.

Vendor Information

Vendor Name(s): HLN Consulting

Vendor Type(s): Professional services

A vendor was involved in the development of the program and continues to be involved in program enhancements. ICE is also available as an open-source program. The DOHMH implementation is maintained by HLN Consulting.

Improving Foodborne Disease Outbreak Detection by Incorporating Complaints Identified in Social Media Data

Department: Bureau of Communicable Disease

First Used: November 2016

Updated in 2025: Yes

Purpose Type: Triage

Computation Type: Scoring

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Individuals (Public who dine at New York City restaurants and are Yelp users and New York City restaurants)

Vendor(s): Columbia University

For more information: <https://publichealth.cs.columbia.edu/>

Tool Description

Restaurant-associated foodborne disease outbreaks are often identified through complaints received via 311 non-emergency information system; however not all individuals report to 311. DOHMH in collaboration with Columbia University developed a text classifier program which monitors Yelp and Twitter data to identify complaints of foodborne illness which was supported by grants from the Alfred P. Sloan Foundation and the National Science Foundation. As of April 2023, the tool no longer uses data from Twitter (X) due to application programming interface (API) changes.

Tool Purpose

The model uses data from Yelp restaurant reviews and previously used Twitter data that was available on Twitter's publicly available API. Twitter (X) removed free access to their publicly available API in April 2023, so these data are no longer included in our analyses. The classifiers assign a "sick score" to each Yelp review indicating the likelihood that the review pertains to foodborne illness. The sick score is based on whether the review contains key words indicative of foodborne illness (e.g. "vomit"); the Yelp classifier also incorporates if the review indicates that multiple people became sick and if the review indicates a time between eating at a restaurant and illness onset (incubation period) that is consistent with foodborne illness. Each review with a sick score greater than or equal to a threshold value are reviewed and annotated by DOHMH foodborne disease epidemiology and environmental health staff to determine if the review was actually reporting foodborne illness possibly associated with a New York City restaurant; if yes, Yelp messages are sent to Yelp reviewers, requesting that they contact DOHMH. Data from annotations are used to improve classifier performance. Foodborne disease complaints identified through Yelp are combined with foodborne disease complaints reported to 311 to improve efficiency of outbreak detection.

Update Description

Since January 31, 2025, DOHMH has been unable to receive data for this tool. Columbia University and DOHMH staff have been unable to resolve the issue without additional resources.

Data Analyzed

Identifying information may be collected or disclosed in the training data and input data.

Training Data

Training data was used in the development of the Yelp classifiers. The training data consisted of restaurant reviews obtained from Yelp by Columbia University; a subset of these data were joined with annotations provided by DOHMH staff. The annotations of restaurant reviews focused on the following: 1) if the review indicated foodborne illness;

Version effective March 2026

2) if the incident occurred in the past 30 days; 3) if multiple people were sick; and 4) if the incubation period was consistent with foodborne illness. The training data is periodically updated (with annotations from DOHMH) to improve the classifiers.

Input Data

Yelp reviews of New York City restaurants are pulled from a privately available API provided by Yelp.

Output Data

The output data includes a “sick score” that the classifiers assign to each Yelp review indicating the likelihood that the review pertains to foodborne illness.

Vendor Information

Vendor Name(s): Columbia University

Vendor Type(s): No-cost engagement

DOHMH staff, including Bureau of Communicable Disease, Office of Environmental Investigations, and Division of Informatics and Information Technology & Telecommunications, and Columbia University are involved in making decisions about the tool. Columbia University Department of Computer Science professors and doctoral students maintain the classifier. The project was previously funded by the Alfred P. Sloan Grant, for which The Fund for Public Health in New York provided administrative support and grant management to DOHMH. This support and management ended at the completion of the grant in 2021.

IQTREE

Department: Disease Control - Public Health Laboratory

First Used: May 2020

Updated in 2025: No

Purpose Type: Data management

Computation Type: Clustering

Autonomy: Supervised

Frequency of Use: 2-3 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/iqtree/iqtree3>

Tool Description

IQTREE uses maximum-likelihood regression to create phylogenetic trees from genomes.

Tool Purpose

Produced phylogenetic trees are used to help rule in or out outbreaks of COVID-19 or other organisms.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Multiple sequence alignments derived from public health genomic datasets.

Output Data

Phylogenetic trees representing evolutionary relationships among isolates, used to assess transmission clusters and outbreak linkage.

kSNP4

Department: Disease Control - Public Health Laboratory

First Used: March 2022

Updated in 2025: No

Purpose Type: Data management

Computation Type: Clustering

Autonomy: Supervised

Frequency of Use: 2-3 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://pubmed.ncbi.nlm.nih.gov/37948764/>

Tool Description

kSNP4 uses multiple algorithms (e.g., maximum-likelihood, parsimony, neighbor-joining) to infer phylogenetic trees from genomes.

Version effective March 2026

Tool Purpose

Produced phylogenetic trees are used to help rule in or out outbreaks of bacteria.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Assembled genome sequences generated from laboratory sequencing data.

Output Data

Single-nucleotide polymorphisms-based relatedness results and phylogenetic trees summarizing genetic similarity among isolates, used for outbreak investigation.

Minimap2

Department: Disease Control - Public Health Laboratory

First Used: May 2020

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: 3-4 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/lh3/minimap2>

Tool Description

Minimap2 aligns sequencing data to a reference sequence.

Tool Purpose

Minimap2 uses optimal chaining scores to align sequencing data to reference genomes. This tool is faster and more optimal for long read sequences, such as Oxford Nanopore Technologies data. This tool is used to predict the order in which the fragments

generated by sequencers are pieced together to form a complete genomic sequence data. This tool is used for COVID-19 and monkeypox virus sequencing analyses.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Laboratory-generated sequencing reads and an appropriate reference genome.

Output Data

Sequence alignments identifying genomic similarities and differences, used for genomic characterization and surveillance.

Multiple Alignment using Fast Fourier Transform (MAFFT)

Department: Disease Control - Public Health Laboratory

First Used: January 2021

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: 3-4 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/GSLBiotech/mafft>

Tool Description

Multiple Alignment using Fast Fourier Transform (MAFFT) aligns multiple sequencing data.

Tool Purpose

MAFFT includes several algorithmic methods, including guided tree, scoring matrices, and sequence alignment algorithms to realign multiple genomic sequencing data. The

tool aligns sequences to help identify differences. This is used in all sequencing analysis prior to building a phylogenetic tree or distance tree.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Homologous nucleotide or protein sequences generated from laboratory sequencing or validated reference databases.

Output Data

Aligned sequences identifying conserved and variable genomic regions, used for phylogenetic and outbreak analyses.

Pangolin

Department: Disease Control - Public Health Laboratory

First Used: July 2021

Updated in 2025: Yes

Purpose Type: Data management

Computation Type: Clustering

Autonomy: Supervised

Frequency of Use: 3-4 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/cov-lineages/pangolin>

Tool Description

Pangolin assigns lineage names to SARS-CoV-2.

Tool Purpose

Pangolin uses a combination of several methods, including random forest tree, classification methods, and maximum parsimony to assign lineage names to SARS-CoV-2 genomic sequences to bin sequences that are more likely to be similar. This is a tool that designates a name based on a nomenclature for COVID-19 sequence data.

Version effective March 2026

Update Description

Updated to pangolin version 4.3.3.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Curated SARS-CoV-2 genome dataset with Pango lineage designations is maintained by the Pango Network and periodically updated.

Input Data

SARS-CoV-2 consensus genome sequences generated through public health sequencing workflows.

Output Data

Viral lineage assignments indicating genetic classification, used for variant tracking and epidemiological reporting.

PHYLOViZ

Department: Disease Control - Public Health Laboratory

First Used: October 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Clustering

Autonomy: Supervised

Frequency of Use: 3-4 times per year or as required

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://www.phyloviz.net/>

Tool Description

For representing the possible evolutionary relationships between strains, PHYLOViZ uses the goeBURST algorithm, a refinement of eBURST algorithm by Feil et al., and its expansion to generate a complete minimum spanning tree.

Tool Purpose

PHYLOViZ is used to generate the minimum spanning tree relationships.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Allelic profiles or phylogenetic outputs generated from sequencing-based typing analyses.

Output Data

Graphical relatedness networks identifying clustered isolates, used for outbreak visualization and surveillance reporting.

PulseNet 2.0

Department: Disease Control - Public Health Laboratory

First Used: September 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Clustering

Autonomy: Supervised

Frequency of Use: 4-5 times per week

Population Impacted: Individuals (Patients with culture-positive foodborne bacterial isolate); Biological sample

Vendor(s): Centers for Disease Control and Prevention, Booz Allen Hamilton

For more information: <https://stacks.cdc.gov/view/cdc/138367>

Tool Description

PulseNet 2.0 is a suite of tools used to align and analyze bacterial genomes.

Tool Purpose

PulseNet 2.0 is used to: 1. Assemble the bacterial genome (since the sequencing process involves fragmenting the bacterial DNA and then amplifying it into millions of pieces); 2. Identify the genus, species, and serotype of the bacterial isolate; 3. Perform quality control checks to ensure the sequence meets certain quality standards; 4. Perform core

and whole genome multi-locus sequence typing (a technique used to type bacteria based on their genetic code); 5. Perform cluster analysis for cases related to one another based upon case definitions recommended by the Centers for Disease Control and Prevention (CDC). This information is then communicated to partners including foodborne epidemiologists at the Bureau of Communicable Disease, who investigate all reported cases of foodborne disease, with those investigations potentially resulting in restaurant inspections, closures, and food recalls.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Standardized sequencing data submitted by public health laboratories.

Output Data

Nationally comparable relatedness and clustering results used for multi-jurisdictional outbreak detection and surveillance coordination.

Vendor Information

Vendor Name(s): Centers for Disease Control and Prevention, Booz Allen Hamilton

Vendor Type(s): Procurement/Paid pilot

Developed and maintains the tool.

Spades

Department: Disease Control - Public Health Laboratory

First Used: October 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: 2-3 times per week

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/ablab/spades>

Tool Description

Spades uses several algorithms to simplify genomic read data into de Bruijn graphs and finds overlaps to assemble genomes.

Tool Purpose

Spades is an intermediate step in the workflows of bacterial analyses.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Raw sequencing reads produced from clinical, environmental, or foodborne pathogen samples.

Output Data

Draft genome assemblies reconstructing organism genomes, used for typing, resistance detection, and surveillance analyses.

Vsearch

Department: Disease Control - Public Health Laboratory

First Used: June 2022

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Supervised

Frequency of Use: As required per week/month

Population Impacted: Individuals (Patients who tested positive for the targeted pathogens and whose samples were sequenced); Biological sample

Vendor(s): None

For more information: <https://github.com/torognes/vsearch>

Tool Description

Vsearch uses the Needleman-Wunsch algorithm to merge read pairs and align and dereplicate sequences to detect chimeric genomic sequences.

Tool Purpose

Vsearch is an intermediate step in the workflow to analyze COVID-19 variants in wastewater.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Nucleotide sequences from laboratory samples and, where applicable, curated reference databases.

Output Data

Similarity or clustering results identifying related sequences, used for organism identification or comparative genomic analysis.

Department of Investigation (DOI)

Facial Recognition Technology

Department: Compliance

First Used: March 2019

Purpose Type: Not specified

Autonomy: Informative

Population Impacted: Individuals

Updated in 2025: Yes

Computation Type: Matching

Frequency of Use: Not specified

Vendor(s): Not disclosable

Tool Description

The tool analyzes an uploaded image or video and searches and compares it with lawfully possessed images to generate a pool of possible matches. If possible matches are identified, a trained DOI examiner visually analyzes and evaluates potential matches

Version effective March 2026

to assess reliability of a match consistent with agency policy and applicable laws. A match serves as an investigative lead for additional investigative steps and does not constitute a positive identification.

Tool Purpose

Facial recognition is a digital technology that DOI uses to analyze uploaded images or videos of people and objects obtained during an investigation by comparison with lawfully possessed images. Facial recognition generates possible matches of an object or individual from this analysis and comparison. The purpose of the tool is to assist DOI investigations of matters within its jurisdiction including fraud and other criminal activity.

Update Description

Routine vendor updates.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Vendor uses publicly available open-source media data.

Input Data

Images or videos of people and object obtained during an investigation.

Output Data

Lawfully possessed images of possible matches, if found.

Vendor Information

Vendor Name(s): Not disclosable

Vendor Type(s): Off-the-shelf

The vendors provide ongoing technical assistance. Confidentiality agreements are in place with the vendors.

Department of Social Services (DSS)

Homebase Risk Assessment Questionnaire (RAQ)

Department: Human Resources Administration Homebase

First Used: June 2012

Updated in 2025: No

Purpose Type: Resource allocation

Computation Type: Scoring

Autonomy: Monitored

Frequency of Use: Daily

Population Impacted: Individuals
(Households seeking Homebase assistance)

Vendor(s): None

Tool Description

Homebase applicants answer screening questions about their current housing situation, history of disruptive experiences, shelter history, and other domains. Each of the answers is assigned a number of points, and applicants that reach a certain point threshold are eligible for deeper Homebase services, such as financial assistance and case management. Workers are able to override a limited number of model decisions with permission of a supervisor.

Tool Purpose

The Homebase program was created to prevent households from entering the Department of Homeless Services (DHS) shelter system. Since New York City has a range of antipoverty programs and the number of households entering shelter is small compared to the pool of New Yorkers who have an eviction filing each year, the Agency had to ensure that the households who most needed additional homelessness prevention services were being enrolled in Homebase programs. Research showed that staff were not accurately able to predict who would or would not enter the DHS shelter system and that using a risk assessment would provide a better way to match resources to the families who would benefit the most.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

The RAQ was developed based on analysis of data on Homebase enrollees from 2004 to 2008, conducted in conjunction with a team of academic researchers, to determine

predictive factors for those entering shelter. It was updated in 2023 based on analysis led by DSS researchers of 2013-2016 Homebase data.

Input Data

Factors include, among others: personal characteristics such as age and pregnancy; educational attainment and employment status; housing issues such as eviction, discord, a move in the past year; past and recent experience of homelessness.

Output Data

The tool produces a score that is used to assess eligibility for full versus brief Homebase services.

Maestro Automated Valuation System (MAVS)

Department: Accountability Office

First Used: January 2025

Purpose Type: Triage

Autonomy: Informative

Population Impacted: Individuals
(Medicaid clients who file personal injury lawsuits)

Computation Type: Scoring

Frequency of Use: Regularly

Vendor(s): Gainwell/Health Management Systems, Inc. via contract with NYS Office of Medicaid Inspector General (OMIG)

Tool Description

Maestro Automated Valuation System (MAVS) uses machine learning to identify Medicaid encounters that are related to a given injury for which a Medicaid client received funds from a personal injury lawsuit. The Medicaid program is legally entitled to place a lien to recoup these medicaid funds.

Tool Purpose

MAVS uses machine learning to identify Medicaid encounters that are related to a given injury for which a Medicaid client received funds from a personal injury lawsuit. The Medicaid program is legally entitled to place a lien to recoup these Medicaid funds. Under current practice, a DSS staff member reviews MAVS output before sending any notification to the client or attorneys in the case.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

The model was trained on MAESTRO data (the system storing the DSS medicaid lien data) and specifically on the determination decisions made manually by DSS staff members.

Input Data

Medicaid data warehouse encounter data and MAESTRO data on personal injury details.

Output Data

Each Medicaid encounter in the client's MAESTRO file receives a determination: "RELATED" or "UNRELATED."

Vendor Information

Vendor Name(s): Gainwell/Health Management Systems, Inc. via contract with NYS Office of Medicaid Inspector General (OMIG)

Vendor Type(s): Professional services

DSS and NYS OMIG meet with Gainwell regularly to discuss MAVS and the underlying MAESTRO system.

SmartVAN / TargetSmart

Department: Public Engagement Unit

First Used: November 2019

Purpose Type: Data management

Autonomy: Informative

Population Impacted: Individuals (New Yorkers potentially eligible for benefits)

Updated in 2025: No

Computation Type: Matching

Frequency of Use: Occasionally

Vendor(s): EveryAction and TargetSmart

Tool Description

The Mayor's Public Engagement Unit (PEU) uses SmartVAN to manage outreach across a range of projects. SmartVAN provides functionality to create lists of potential clients to contact, collect personal information and survey responses from clients, and conduct outreach via phone banks and canvassing. SmartVAN also contains a frequently updated commercial dataset, provided by TargetSmart, of New York City residents and their demographic, contact, and other information. PEU uses this preloaded data to create outreach lists when data on existing clients or from partner agencies is unavailable or insufficient to meet the scope of the outreach project.

Version effective March 2026

Tool Purpose

In 2025, PEU used TargetSmart data within SmartVAN on various outreach projects. For instance, PEU used it to pull lists of New Yorkers who recently turned 26 for outreach to offer help enrolling in health insurance.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Training data is part of vendor's proprietary processes.

Input Data

Input data is part of vendor's proprietary processes.

Output Data

The algorithmically-derived data that PEU accesses is the output of proprietary algorithmic processes developed and operated by TargetSmart. These algorithmic processes include matching multiple input datasets to determine residency, contact information, and demographics on New York City residents. SmartVAN also includes a number of algorithmically determined likelihood scores, including scores for the likelihood that a household contains children under 18, etc.

Vendor Information

Vendor Name(s): EveryAction and TargetSmart

Vendor Type(s): Procurement/Paid pilot, Professional services

EveryAction and TargetSmart jointly provide the SmartVAN product. EveryAction is the software provider. TargetSmart is the data provider. TargetSmart is the entity who applies algorithmic techniques. EveryAction provides access to this data through their platform.

Statara

Department: Public Engagement Unit

First Used: August 2025

Purpose Type: Data management

Computation Type: Matching

Autonomy: Informative

Frequency of Use: Occasionally

Population Impacted: Individuals (New Yorkers who may be eligible for certain benefits)

Vendor(s): Statara

For more information: <https://statara.com/about-us/>

Tool Description

Statara leverages algorithmic matching and scoring to provide a commercial dataset of New York City residents and their demographic, contact, and other information. The Department of Social Services (DSS)/the Mayor's Public Engagement Unit (PEU) uses this data to create outreach lists when data on existing clients or from partner agencies is unavailable or insufficient to meet the scope of the outreach project.

Tool Purpose

In 2025, PEU used Statara data to drive outreach on several projects. For example, PEU used Statara to create lists of New Yorkers potentially eligible for Senior Citizen Rent Increase Exemption using factors like age and whether they were likely not a homeowner. For these projects, PEU uses Statara-provided addresses and/or phone information to conduct outreach.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Training data is part of vendor's proprietary processes. The vendor receives no DSS/PEU data as part of training or input.

Input Data

Input data is part of vendor's proprietary processes. The vendor receives no DSS/PEU data as part of training or input.

Output Data

The algorithmically derived data that PEU accesses is the output of proprietary algorithmic processes developed and operated by Statara. These algorithmic processes include matching multiple input datasets to determine residency, contact information, and demographics on New York City residents. Statara also includes a number of algorithmically determined likelihood scores, including scores for the likelihood that a household contains children under 18, etc. PEU receives a regularly updated data file

from Statara for its use. PEU subsets and/or merges this file with other PEU data sources to drive outreach projects.

Vendor Information

Vendor Name(s): Statara

Vendor Type(s): Procurement/Paid pilot, Professional services

Statara provides us with an updated consumer data file for New York City residents on a monthly basis.

Tenant Harassment Risk Model (THRM)

Department: Public Engagement Unit

First Used: October 2025

Purpose Type: Resource allocation

Autonomy: Informative

Computation Type: Scoring

Frequency of Use: Informs weekly outreach starting in October

Population Impacted: Property; Individuals (Tenants may or may not receive outreach due to the score of their building)

Vendor(s): None

Tool Description

The Tenant Harassment Risk Model (THRM) was developed by Housing Preservation and Development (HPD) to identify buildings where tenants were more likely to experience harassment or other challenges with their landlord.

Tool Purpose

In collaboration with HPD, PEU uses the THRM model to identify buildings for proactive tenant outreach (door knocking) to identify individual issues as well as buildings that could benefit from tenant organizing through the HPD Partners in Preservation program.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Version effective March 2026

The THRM uses data from multiple sources for training and input. Much of the data is from NYC Open Data, as well as agency partners and third parties such as Property Shark. The data includes variables such as violations, 311 requests/complaints, building financials, eviction filings. The methodology did not involve segmenting training data. All residential buildings were used in the analysis to develop the weights.

Input Data

The THRM uses data from multiple sources for training and input. Much of the data is from NYC Open Data, as well as agency partners and third parties such as Property Shark. The data includes variables such as violations, 311 requests/complaints, building financials, eviction filings. All residential buildings are included in input data.

Output Data

The THRM combines the output of multiple linear regressions to generate component scores focused on Construction, Eviction, Financial Distress, and Physical Distress (all on a 0-100 score) as well as an overall Tenant Harassment Risk Score (0-100). PEU received scores for residential buildings with at least 11 units and in which at least 50 percent of units are rent regulated.

Department of Transportation (DOT)

Midtown in Motion / Adaptive Control Decision Support System

Department: Traffic Operations

First Used: July 2011

Updated in 2025: No

Purpose Type: Performance evaluation

Computation Type: Classification

Autonomy: Monitored

Frequency of Use: Continuous

Population Impacted: Geographic space; Individuals (Travelers in the Manhattan midtown core)

Vendor(s): KLD Engineering, P.C.

Tool Description

Midtown in Motion (MIM) is a program that measures traffic congestion in the midtown core of Manhattan (from 1st to 9th Avenues and from 57th to 34th Streets, inclusive), using sensor data captured within and beyond the zone. This congestion classification (light, moderate, moderate-heavy, heavy) is used by the Adaptive Control Decision

Support System (ACDSS) to choose the optimal signal timing along the avenues to reduce congestion and improve traffic flow for all modes of transportation.

Tool Purpose

The purpose of the tool is to improve congestion in the midtown core of Manhattan through responsive traffic signal management.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

Motor vehicle travel times captured by Radio Frequency Identification (RFID) sensors.

Output Data

Recommended traffic signal plan.

Vendor Information

Vendor Name(s): KLD Engineering, P.C.

Vendor Type(s): Procurement/Paid pilot

KLD Engineering, P.C. is the developer of ACDSS and currently handles the maintenance contract.

Department of Youth and Community Development (DYCD)

Adobe Creative Suite

Department: Communications

First Used: May 2025

Purpose Type: Information presentation

Autonomy: Informative

Computation Type: Data transformation

Frequency of Use: Weekly

Population Impacted: Individuals (People in New York City); Group, organization, or business

Vendor(s): Adobe

Tool Description

Adobe Creative Cloud (CC) is a comprehensive collection of cloud software applications focused on graphic, video, and photography tools.

Tool Purpose

The creative suite has implemented AI tools in its software. For Adobe Photoshop, you are able to use AI to spot-edit photos if you are trying to remove a logo on an article of clothing for example.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Photos, videos, graphics uploaded by a staff member to edit and transform into content for social media.

Output Data

Designed audio, visuals and graphics based on the designer who exports a finished asset.

Vendor Information

Vendor Name(s): Adobe

Vendor Type(s): Off-the-shelf

Adobe developed and maintains the tool.

Adobe Premier Pro

Department: Communications

First Used: April 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Weekly

Population Impacted: Individuals (People in New York City)

Vendor(s): Adobe

Tool Description

Adobe Premier Pro is a video editing software used to create post-event recap videos and highlight reels, including AI-generated captions.

Tool Purpose

Adobe Premier Pro uses AI to provide auto-subtitle captioning for videos.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Video files are uploaded to clip, cut, and edit into a final video.

Output Data

Exported video with music, subtitles, and professional editing.

Vendor Information

Vendor Name(s): Adobe

Vendor Type(s): Off-the-shelf

Adobe developed and maintains the tool.

ChatGPT

Department: Communications

First Used: April 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Weekly

Population Impacted: Individuals (People in New York City); Group, organization, or business

Vendor(s): OpenAI

Tool Description

ChatGPT is an advanced AI language model that can understand and generate text based on the input it receives. It can assist with a wide range of tasks, including summarizing text, rephrase text to a lower reading level, and assist in providing content ideas for social media.

Tool Purpose

DYCD's Communication Unit have used ChatGPT to complete the following tasks: assist in summarizing text to create more concise captions, helping brainstorm DYCD related content ideas for social media, and taking current text and rewriting it to any necessary reading level.

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

Training data is proprietary to OpenAI.

Input Data

DYCD staff enter text prompts related to creating and editing social media content.

Output Data

Text response to the user prompt, which is reviewed and adapted by staff prior to publication.

Vendor Information

Vendor Name(s): OpenAI

Vendor Type(s): Off-the-shelf

ChatGPT is a freely available tool created and maintained by OpenAI.

DYCD Virtual Assistant Chatbot

Department: Information Technology - Application Development

First Used: September 2020

Purpose Type: Information presentation

Autonomy: Fully autonomous

Population Impacted: Individuals (People in New York City)

Updated in 2025: Yes

Computation Type: Matching

Frequency of Use: Averaging about requests 2100 a day

Vendor(s): Microsoft

For more information: <https://www.nyc.gov/dycd>

Tool Description

The DYCD Virtual Assistant Chatbot provides support to the Community Connect team by providing predefined answers to questions asked to the Chatbot by the public. This tool enables the public to quickly get answers to the most commonly asked questions on DYCD programs while reducing the number of requests to the Community Connect team on basic questions. The tool allows for the Community Connect team spend more time supporting the public on more nuanced scenarios and complex questions.

Tool Purpose

Information received is used to understand public questions and public interest in programs offered to produce the matching predefined response relevant to the question.

Update Description

The DYCD Virtual Assistant Chatbot was upgraded in 2025 to use the Microsoft Azure Language, a cloud-based service that provides natural language processing features for understanding and analyzing text to help build intelligent applications. In particular, the chatbot uses the Custom Question Answering and Conventional Language Understanding components.

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

Staff worked to capture and maintain information regarding the programs in a knowledge base.

Input Data

Questions are input to chat bot from user in general text format.

Output Data

Version effective March 2026

The Conversational Language Understanding (CLU) model is then used to capture the users intent, which in turn is directed to our Custom Question Answering (CQA) model, that searches the knowledge base, ranks the possible answers and returns the best match.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): Procurement/Paid pilot

Microsoft Azure provides the software for implementing the chatbot.

OpusClip

Department: Communications

First Used: June 2024

Purpose Type: Information presentation

Autonomy: Informative

Population Impacted: Individuals (Any viewer watching the content. Our targeted audience is New Yorkers); Group, organization, or business

Updated in 2025: Yes

Computation Type: Ranking

Frequency of Use: Biweekly

Vendor(s): OpusClip

Tool Description

OpusClip uses their data and AI to analyze video in the context of current social and marketing trends. It detects the most compelling, shareable, and emotionally impactful moments within the video that is uploaded.

Tool Purpose

Primarily used for our DYCD Next Gen Podcast. DYCD will record the podcast and use OpusClip to find moments in the video that may bring the most engagement when we post on our social media platforms.

Update Description

Tool is a web platform and updates regularly.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Finalized video meant for public viewing.

Output Data

The platform analyzes an uploaded video and provides the time stamps and clips of the original video that have a good chance of creating engagement with the viewer to comment, share, and like. We can then export the selected clips.

Vendor Information

Vendor Name(s): OpusClip

Vendor Type(s): Off-the-shelf

OpusClip develops and maintains the product.

Fire Department (FDNY)

Emergency Medical Services (EMS) Hospital Suggestion Algorithm

Department: Management, Analysis, and Planning

First Used: December 2020

Updated in 2025: No

Purpose Type: Resource allocation

Computation Type: Ranking

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Individuals
(Patients); Geographic space

Vendor(s): None

For more information: <https://doi.org/10.1016/j.scs.2022.104104>

Tool Description

The Emergency Medical Services (EMS) Hospital Suggestion Algorithm is used to determine the closest, appropriate hospital to the incident location based on the medical needs of a patient requiring transport.

Tool Purpose

The algorithm computes a list of hospitals in order of closest to furthest in time for each medical condition category as currently established. (For example, there is a list of hospitals computed in order of closest in time for all hospitals that accept General Emergency Department patients and for all hospitals that accept special conditions, such as burns.) Depending on the medical needs category of the patient, the algorithm produces a pre-determined list of hospitals which is based on the location of the patient and then made available to the crew as a list of “closest, most appropriate hospitals.”

Data Analyzed

No identifying information is collected or disclosed.

Training Data

The EMS Hospital Suggestion algorithm relies on automatic vehicle location data from ambulances transporting patients to hospitals between 2018 and 2019 to calibrate a network analysis model that derives incident to hospital transport times. The order of suggested hospitals are then compared with five years of historical EMS hospital transport data from before the COVID-19 pandemic (2015-2019) to validate and correct the network model.

Input Data

The inputs for the algorithm include the location and medical call type of the patient.

Output Data

The algorithm outputs the closest, most appropriate hospitals.

Emergency Medical Services (EMS) Unit Suggestion Algorithm

Department: Management Analysis and Planning

First Used: March 2007

Updated in 2025: Yes

Purpose Type: Resource allocation

Computation Type: Ranking

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Geographic space; Individuals (individuals calling 911 for emergency medical services)

Vendor(s): Deccan International

Tool Description

The Emergency Medical Services (EMS) Unit Suggestion Algorithm is used to determine which order of geographic regions (known as atoms) to search in order for the EMSCAD system to select an appropriate EMS unit for dispatch to an incident.

Tool Purpose

The algorithm computes a list of geographic regions (known as atoms) in order of closest to furthest in travel time for each atom in the city. This list of ordered atoms is the output of an algorithm that relies on a calibrated network model to derive travel time estimates. The output is an excel file which is converted into an EMSCAD-compatible file and loaded into the system for real-time unit selection capabilities. Starting in September 2025, The file was generated and implemented as a 24/7 source file, meaning the recommended search order is not currently varying by time of day. In September 2025, the model was updated to accept time of day as an input.

Update Description

Starting in September 2025, input data includes time of day to make computations. The output of this process now includes outputs for different times of day and day of week.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

The EMS Unit Suggestion algorithm relies on historical EMSCAD trip time data which is used to calibrate a network analysis model which derives atom-to-atom transport times.

Input Data

The input for the algorithm is a geographic location, time of day (as of September 2025), and historical emergency unit response time data.

Output Data

The algorithm outputs a recommended EMS unit for dispatch.

Vendor Information

Vendor Name(s): Deccan International

Vendor Type(s): Professional services

This algorithm and the resulting output file that is used in our EMSCAD system to suggest atom order for unit search is currently provided by a vendor, Deccan International.

Facial Recognition Technology

Department: Bureau of Fire Investigations

First Used: December 2022

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Classification

Autonomy: Informative

Frequency of Use: Not specified

Population Impacted: Individuals

Vendor(s): Not disclosable

Tool Description

The tool analyzes an uploaded image or video and compares the image to lawfully possessed images to generate a pool of possible matches. Any matches serve only as investigative leads and Bureau of Fire Investigations personnel conduct an additional investigation. The match alone is not treated as a positive identification.

Tool Purpose

The purpose of the tool is to assist Fire Department investigations into arson and other criminal activities that fall within the agency's jurisdiction.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Publicly available open-source social media data.

Input Data

Images of individuals of interest and images from social media or the internet.

Output Data

The tool generates possible image matches of an object or individual.

Vendor Information

Vendor Name(s): Not disclosable

Vendor Type(s): Off-the-shelf

Vendor provides ongoing technical support. Confidentiality agreements are in place with the vendor.

Fire Brush Fire Detection Camera

Department: Bureau of Fire Investigations

First Used: September 2025

Purpose Type: Risk management

Computation Type: Classification

Autonomy: Informative

Frequency of Use: Not specified

Population Impacted: Property;
Geographic space

Vendor(s): Claro AI

Tool Description

This tool is an early warning detection system that uses solar-powered cameras that send data to a secure AI server. The server analyzes the images using artificial intelligence to detect signs of smoke or fire.

Tool Purpose

If smoke or fire is detected on the camera, an alert is sent to the Bureau of Fire Investigations and the Fire Department Operation Center to view the camera and determine if an emergency response is required.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Reference images of smoke and fire-related objects.

Input Data

Images of treetops from FDNY cameras in City parks or other City-owned locations.

Output Data

Email alert and images.

Vendor Information

Vendor Name(s): Claro AI
Version effective March 2026

Vendor Type(s): Off-the-shelf, Procurement/Paid pilot

Vendor provides ongoing technical support.

RBIS (Risk Based Inspection Program); ALARM (A Learning Approach to Risk Modeling)

Department: Management, Analysis, and Planning

First Used: November 2019

Updated in 2025: No

Purpose Type: Risk management

Computation Type: Scoring

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Property;
Individuals (Civilian fire injuries/fatalities)

Vendor(s): None

Tool Description

A Learning Approach to Risk Modeling (ALARM) creates risk scores for each building in the city. These scores are used to schedule our Fire Operations building inspections within the inspectable population of buildings in the city (~330,000 Building Identification Numbers), as a part of the Risk-Based Inspection Program (RBIS).

Tool Purpose

ALARM is a combined approach using machine learning and risk ratios to assess the risk of a building for structural fire ignition (probability) and civilian fire injury/death (impact). The machine learning algorithm takes incident data, housing characteristics, and 311 data and creates a probability of structural fire ignition. This is combined with a civilian injury or death risk ratio for the building which is based on building characteristics, incident data and nearby felony crimes to create a risk score (range is one to nine), with one being highest risk and nine being lowest risk. Buildings are prioritized within each of the nine risk scores according to the residential population in each building.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Each month the team uses a five-year incident dataset and reserves 99 percent of the data to train the ignition model and 80 percent of the data to train the impact model.

Input Data

The ALARM risk score utilizes data from our Fire and Emergency Medical Services dispatch systems, building characteristic data, 311 calls, felony crimes, census data and civilian injury data.

Output Data

The tool outputs a risk score from one (highest risk) to nine (lowest risk).

Mayor's Office (MO)

Adobe Photoshop

Department: Mayor's Office of Media and Entertainment

First Used: January 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Matching

Autonomy: Informative

Frequency of Use: Almost every day

Population Impacted: Individuals (Viewers who watch the city's television channels; some individuals who appear in Mayor's Office of Media and Entertainment's television content)

Vendor(s): Adobe

For more information: <https://www.adobe.com/products/photoshop.html>

Tool Description

NYC Media uses Adobe Photoshop to edit images. Adobe Photoshop uses generative AI to allow users to edit images without manual work.

Tool Purpose

NYC Media uses Adobe Photoshop to make slight edits to some images that appear in some content that is produced in-house and broadcast on the city's television edits. For example, to comply with Federal Communications Commission regulations for non-commercial educational stations, we may use an AI tool to blur a company's logo on a t-shirt. As another example, we may use the AI tool to add visual interest, for example, to add legs in a picture that is cropped at the waist.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

According to Adobe's website, "generative AI Image models were trained on licensed content, such as Adobe Stock, and public domain content where copyright has expired."

Input Data

Images copyrighted by the City of New York or licensed to the City of New York pursuant to an agreement that authorizes edits and, if involving images of people, content that is covered by a written consent form.

Output Data

Visual content that is broadcast on the City's television stations.

Vendor Information

Vendor Name(s): Adobe

Vendor Type(s): Off-the-shelf, Procurement/Paid pilot

Adobe regularly updates the Photoshop software.

Adobe Premiere Pro

Department: Mayor's Office of Media and Entertainment

First Used: 2021

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Monitored

Frequency of Use: About every day

Population Impacted: Individuals
(Individuals who watch content on NYC Media's television channels; individuals who speak in content on NYC Media's television channels)

Vendor(s): Adobe

For more information: <https://www.adobe.com/products/premiere.html>

Tool Description

NYC Media uses Adobe Premiere Pro to edit video content broadcast on the city's television stations. Within Adobe Premiere Pro, we use AI-powered tools to help generate closed captions of some video content that is edited in-house prior to broadcast.

Tool Purpose

We use AI-powered tools to help a human editor generate closed captions of some video content that is edited in-house prior to broadcast on the city's television channels.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

According to Adobe's website, "Speech to Text is powered by a combination of Adobe proprietary technology — including Adobe Sensei machine learning— and third-party technologies."

Input Data

Spoken words in video programs.

Output Data

Closed captions.

Vendor Information

Vendor Name(s): Adobe

Vendor Type(s): Off-the-shelf, Procurement/Paid pilot

Adobe provides regular software updates.

AI Transcription on Teams

Department: Mayor's Office to End Domestic and Gender Based Violence

First Used: April 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Monitored

Frequency of Use: Daily

Population Impacted: Individuals
(Participants in the meeting might provide their name and business affiliations)

Vendor(s): Microsoft

Tool Description

Microsoft Teams has a built-in feature that uses AI to create a transcript of the meeting.

Tool Purpose

The tool provided a transcript of a meeting which was then reviewed by Mayor's Office to End Domestic and Gender Based Violence staff for accuracy. The transcript was then emailed to meeting participants.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to Microsoft.

Input Data

The words spoken during a meeting are captured by the tool.

Output Data

The tool provides text of the words spoken during the meeting.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): Off-the-shelf

Microsoft provides this tool as part of Teams.

AppTek OmniCaption 300 Closed Captioning Appliance

Department: Mayor's Office of Media and Entertainment

First Used: November 2022

Purpose Type: Information presentation

Autonomy: Fully autonomous

Population Impacted: Individuals (People who watch live City Council and mayoral content televised on NYC Gov and other content televised on NYC World; people who appear in the content that is televised)

Updated in 2025: No

Computation Type: Data transformation

Frequency of Use: Almost daily

Vendor(s): AppTek

For more information: <https://www.apptek.ai/>

Tool Description

The AppTek OmniCaption 300 closed captioning appliance uses AI-enabled automatic speech recognition to create closed captions of live television content.

Tool Purpose

The Mayor's Office of Media and Entertainment uses the AppTek Omni 300 closed captioning appliance to provide closed captioning of live-broadcasted events (e.g., City Council hearings) and content that is cablecast on NYC World.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

According to AppTek's website, the "OmniCaption 300 closed captioning appliance was developed for and trained on broadcast news, sports, weather and other programming."

Input Data

The input data are words spoken by people during live broadcasts of public hearings, meetings, and events and content on NYC World.

Output Data

Closed captions that reflect the written text of the input data (spoken words).

Vendor Information

Vendor Name(s): AppTek

Vendor Type(s): Off-the-shelf, Procurement/Paid pilot

AppTek provides support for the Omni 300 closed captioning system.

Canva

Department: Mayor's Office of Equity and Racial Justice, NYC Office of Talent and Workforce Development

First Used: January 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Monthly

Population Impacted: Individuals
(Audience for generated materials includes internal and external stakeholders, partner organizations and the general public)

Vendor(s): Canva

Tool Description

Canva is a cloud-based online graphic design platform for creating professional visuals and documents, such as presentations, websites, and similar products.

Tool Purpose

The purpose for each Mayor's Office that uses Canva AI is listed below.

For the Mayor's Office of Equity and Racial Justice, Canva's AI feature is used to generate visual material for social media communication.

For NYC Office of Talent and Workforce Development, Canva's AI feature is used to receive suggestions for designing graphics for invitations, flyers, and meeting and presentation decks for events and presentations. The NYC Talent team has limited graphic design expertise on staff and the tool enables more streamlined designs for promotional materials.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to Canva.

Input Data

Description of image sought.

Output Data

Editable graphic and text in response to user design.

Vendor Information

Vendor Name(s): Canva

Vendor Type(s): Off-the-shelf

Canva provides a free version of their software to the public, which both Mayor's Office's use. Neither hold a contract with Canva.

ChatGPT

Department: Mayor’s Office to End Domestic and Gender Based Violence, Office of Community Mental Health, Office of Talent and Workforce Development, and Mayor’s Office of Nonprofit Services

First Used: March 2024

Updated in 2025: Yes

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Daily

Population Impacted: Individuals (Public audiences, website and social media users, internal and external stakeholders, including groups and individuals)

Vendor(s): OpenAI

Tool Description

ChatGPT is used by Mayor’s Offices as a supplemental tool to support a variety of operational, administrative, and communications functions. ChatGPT is an advanced AI language model that can understand and generate human-like text based on the input it receives. The tool assists staff in improving the clarity, accuracy, and accessibility of written materials, conducting background research, and streamlining workflows. Its use is tailored by each Mayor’s Office to meet programmatic needs.

Tool Purpose

The purpose for each Mayor’s Office that uses ChatGPT is listed below.

The Mayor’s Office to End Domestic and Gender Based Violence (ENDGBV) staff have used ChatGPT to complete the following tasks: assist in gathering information for literature reviews for public facing ENDGBV reports, draft and edit public e-mails, and assist in creating potential job interview questions based on content in the job description.

The Office of Community Mental Health (OCMH) uses the tool in WordPress to write alt text for images on their website, making it accessible to people using assistive technology. The alt text for images are written with inclusive, neutral, and bias-free language. OCMH writes alt text for images of actual humans without the use of ChatGPT to avoid bias and non-consensual use of people’s information. ChatGPT is mostly used in this way to describe images with a lot of text.

The Office of Talent and Workforce Development uses ChatGPT monthly to make text content more streamlined, grammatically correct, and concise; some of this text content is shared with the public or external stakeholders.

The Mayor's Office of Nonprofit Service's uses it to edit content for their social media and Capacity Building trainings for nonprofits.

Update Description

ChatGPT is updated routinely by OpenAI. In addition to model updates, additional offices within the Mayor's Office have begun to use the product in CY2025: the Office of Community Mental Health began use in August 2025, the Office of Talent and Workforce Development began use in January 2025, and the Mayor's Office of Nonprofit Services began use in January 2025.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to OpenAI.

Input Data

Text prompts were provided to ChatGPT.

Output Data

ChatGPT provided text and image responses to the text prompts.

Vendor Information

Vendor Name(s): OpenAI

Vendor Type(s): Off-the-shelf

OpenAI created and maintains ChatGPT.

GoTranscript

Department: Deputy Mayor for Communications

First Used: 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Multiple times a week

Population Impacted: Individuals (General public); Other (Mayor's remarks and speeches)

Vendor(s): GoTranscript

For more information: <https://gotranscript.com/transcription-services>

Tool Description

GoTranscript is an audio and video transcription service.

Tool Purpose

The Mayor's research team submits the audio for the Mayor's remarks and press conferences to the GoTranscript AI transcript service, which returns a first draft. The research team then does a second and third read, correcting any errors and formatting it properly before sending it to be published as a public transcript.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Audio recordings of the mayor's remarks.

Output Data

Draft transcript for the audio or video recordings of the Mayor's remarks.

Vendor Information

Vendor Name(s): GoTranscript

Vendor Type(s): Procurement/Paid pilot

Purchase use of services.

Methodology for Poll Site Language Assistance

Department: Civic Engagement Commission

First Used: November 2020

Updated in 2025: Yes

Purpose Type: Resource allocation

Computation Type: Ranking

Autonomy: Informative

Frequency of Use: Twice

Population Impacted: Individuals

Vendor(s): None

Tool Description

Since no dataset is currently available that reliably captures the number of limited English proficient (LEP) registered voters for all program languages, the Civic Engagement Commission (CEC) uses the percentage of LEP citizens of voting age (CVALEP) as a substitute or proxy measure of need. CEC ranks the program-eligible languages in order of magnitude of CVALEP and distributes poll sites to each language based on its ranking (excluding CVALEP persons that speak languages served by the NYC Board of Elections in certain New York City counties). The number of poll sites that will receive services in any given language will depend on each language's share of the total CVALEP in the population eligible to be served. For example, according to U.S. Census data, approximately 207,926 New Yorkers are CVALEP and speak a language that is served by this program. This proportionality approach allows CEC to balance goals of including diverse language communities as well as fair access to the total number of eligible voters within each language community. The program provides interpreters in program-eligible languages at poll sites based on U.S. Census data showing concentrations of CVALEP individuals who speak these languages and reside around each poll site. For each language, poll sites are chosen in descending order of concentration of CVALEP, until the language's share is met. This process is repeated for each language, thereby including the poll sites with the highest concentration of CVALEP for each program-eligible language until that language's share is met, and the total number of poll sites for which resources are allocated is reached. It may be possible, based on analysis of data, to reassign poll sites to languages with greater need; however, each language will receive a minimum of at least one poll site. Models used included the Thiessen polygon method to create a Voronoi diagram to determine CVALEP estimates.

Tool Purpose

This is a methodology for determining how the CEC will provide interpretation services at poll sites for LEP voters. The methodology explains how the CEC will identify the languages and locations in which interpretation services will be offered during the November 2025 election and beyond. These services supplement the interpretation assistance provided by NYC Board of Elections in several languages. Under the Charter, the CEC can only provide interpretation services in a language if it is a designated citywide language or it is spoken by a greater number of LEP New Yorkers than the lowest ranked designated citywide language and at least one poll site has a significant concentration of speakers of such language with LEP. This methodology ensures service for all languages that are eligible under the Charter.

Update Description

Data is now based on the more recent census (American Community Survey 2019-2023 five-year estimates).

Data Analyzed

No identifying information is collected or disclosed.

Training Data

N/A

Input Data

For citywide estimates, this methodology uses current data from the American Community Survey 2019-2023 five-year estimates. This methodology also uses the American Community Survey Census Tract 2019-2023 five-year Public Use Microdata Samples for poll site level analysis, which tracks resident New Yorkers at the neighborhood level. In addition, the methodology uses data from the Board of Elections on the location of election districts and poll sites.

Output Data

The algorithm estimates the number of citizens of voting age with Limited English Proficiency for each program-eligible language who could report to each polling site.

Microsoft Copilot

Department: NYC Office of Talent and Workforce Development, Mayor's Office of Nonprofit Services

First Used: January 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Weekly

Population Impacted: Individuals (Internal and external stakeholders, including the general public (e.g., MONS social media followers and nonprofit partners))

Vendor(s): Microsoft

Tool Description

Microsoft 365 Copilot is an AI-powered tool that helps with work tasks by responding to user prompts with AI-generated information.

Tool Purpose

Beginning in January 2025, the NYC Office of Talent and Workforce Development (NYC Talent) team uses Microsoft Copilot for notetaking by using the speech to text feature if

the staff member is unable to type in order to multi-task. Other Copilot features like predictive text in email are also periodically used when corresponding with external parties.

Beginning in September 2025, the Mayor's Office of Nonprofit Services (MONS) uses Microsoft Copilot by prompting it to edit content into different formats for their social media and capacity building trainings for nonprofits.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

For NYC Talent: Text or speech. For MONS: Training content (based on professional experience) and drafted social media content.

Output Data

For NYC Talent: Text. For MONS: Edited training and social media content.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): Off-the-shelf

Off-the-shelf product by Microsoft accessed via paid licenses.

Panelot

Department: Civic Engagement Commission

First Used: December 2025

Purpose Type: Resource allocation

Autonomy: Supervised

Population Impacted: Individuals (New Yorkers who applied to participate in The People's Money's Borough Assemblies)

Computation Type: Ranking

Frequency of Use: Once

Vendor(s): Panelot

For more information: <https://panelot.org/>

Tool Description

Panelot (panel selection by lot) is a not-for-profit system for randomly selecting citizen panels in a way that is representative of the population and fair to volunteers.

Tool Purpose

The Civic Engagement Commission (CEC) uses Panelot for the selection of participants in the Borough Assembly phase of our annual Participatory Budgeting process, The People's Money. Panelot receives as input the composition of the pool of volunteers, the desired panel size, and quotas meant to ensure a representative outcome (for example, in a panel of 100 people the number of women might be required to be between 47 and 53). Panelot outputs a list of quota-compliant panels, each with an assigned probability. The list of panels and their probabilities are chosen to be leximin optimal: the output first maximizes the selection probability of any volunteer, then maximizes the second lowest selection probability, then maximizes the third lowest, and so on. The CEC contacts applicants from the output lists. If an applicant is no longer available to participate, the CEC contacts backup volunteers from the same list.

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

N/A

Input Data

Input data consists of an anonymized list of applicants, each with a unique ID that allows them to be referenced with the internal CEC list of applicants. The input data only contains information relevant to reaching quotas. The fields are: type of housing, education, language, income, age, and zip code.

Output Data

Output data consists of a list of potential participants as well as back-ups for some quota categories.

Vendor Information

Vendor Name(s): Panelot

Vendor Type(s): Off-the-shelf

Panelot created and maintains the tool. Vendor is pro-bono.

Zencity

Department: Mayor's Office of Civic Engagement

First Used: July 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Classification

Autonomy: Informative

Frequency of Use: 3-4 times

Population Impacted: Group, organization, or business

Vendor(s): Zencity

Tool Description

Zencity is a survey and large language model (LLM) platform to get community feedback. Zencity has AI features that help write content and come up with research questions to mine media sentiment on specific topics.

Tool Purpose

Civic Engagement holds the city contract with Zencity. Part of this contract is using the platform's AI to help with content writing for additions to webpages/ information distribution and assists with compiling research on public sentiment from specific topics.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Zencity is a company with a proprietary model for their LLM. The sentiment analysis pulls from media sources from news to X posts.

Input Data

Media articles.

Output Data

Sentiment analysis and survey takeaways made public via reports or press releases.

Vendor Information

Vendor Name(s): Zencity

Vendor Type(s): Procurement/Paid pilot, Professional services

The vendor provided a training on how to use the platform and offers additional assistance. Zencity Organic is the AI function.

Zoom

Department: Mayor's Office of Media and Entertainment

First Used: 2020

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Occasional use

Population Impacted: Individuals (People who participate in rulemaking hearings and webinars; people who read transcripts of those hearings and webinars)

Vendor(s): Zoom

For more information: <https://www.zoom.com/>

Tool Description

Zoom is a virtual meeting platform; Zoom has an auto closed-caption function that uses AI.

Tool Purpose

The Mayor's Office of Media and Entertainment (MOME) uses Zoom for public hearings on rulemaking and for public webinars. MOME uses Zoom's auto transcript function and captioning function. (Note: MOME provides American Sign Language and human-typed Communication Access Realtime Translation services as a reasonable accommodation upon request.) If MOME publishes a transcript after the Zoom meeting, a human reviews and corrects it.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to Zoom. According to Zoom's website, "Zoom does not use any customer audio, video, chat, screen sharing, attachments, or other communications-like customer content (such as poll results, whiteboard, and reactions) to train Zoom's or its third-party artificial intelligence models."

Input Data

Speech at MOME's rulemaking hearings and agency webinars.

Output Data

Text in a transcript and captions.

Vendor Information

Vendor Name(s): Zoom

Vendor Type(s): Off-the-shelf, Procurement/Paid pilot

Regular updates to the application.

Mayor's Office of Contract Services (MOCS)

AI Transcription on Webex

Department: Mayor's Office of Contract Services Learning & Development

First Used: July 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Biweekly

Population Impacted: Individuals
(Participants might provide their name
and business affiliations)

Vendor(s): Cisco

Tool Description

WebEx has a built-in feature that uses AI to create a transcript of public meetings, which are shared with participants after the meeting.

Tool Purpose

MOCS uses the transcription for vendor trainings that are hosted on WebEx. It's a public link that is shared with all who register for or attend the training. A link to the recording, including transcription, is shared with those who request it.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to Cisco.

Input Data

The words spoken during a meeting are captured by the tool.

Output Data

The tool provides text of the words spoken during the meeting.

Vendor Information

Vendor Name(s): Cisco

Vendor Type(s): Off-the-shelf

Cisco provides this tool as part of Webex.

New York Police Department (NYPD)

Facial Recognition Technology

Department: NYPD

First Used: October 2011

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Informative

Frequency of Use: Not specified

Population Impacted: Individuals (General population)

Vendor(s): DataWorks

Tool Description

Facial recognition is a digital technology which may help investigators identify unknown subjects in law enforcement investigations.

Tool Purpose

Facial recognition is a digital technology that NYPD uses to compare images obtained during investigations with lawfully possessed arrest and parole photos. The tool analyzes an uploaded image, known as a probe image, and searches and compares against the image repository. The purpose of the tool is to enhance law enforcement's ability to investigate criminal activity as well as identify deceased persons and missing persons. When used in combination with human analysis and additional investigation, facial recognition technology is a valuable tool in solving crimes and increasing public safety.

Version effective March 2026

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

If NYPD investigators obtain a still image depicting a face of an unknown individual during an investigation, the image can be submitted for facial recognition analysis in accordance with NYPD facial recognition policy. Known as a probe image, NYPD facial recognition software compares the image to a controlled and limited group of lawfully obtained photos called the photo repository.

Output Data

The facial recognition software will generate a pool of possible match candidates for review by trained Facial Identification Section investigators.

Vendor Information

Vendor Name(s): DataWorks

Vendor Type(s): Procurement/Paid pilot, Professional services

Software developed and maintained by vendor.

Patternizr

Department: NYPD

First Used: December 2016

Purpose Type: Data management

Autonomy: Informative

Population Impacted: Individuals;
Property; Geographic space; Other (Crime
classification)

Updated in 2025: Yes

Computation Type: Matching

Frequency of Use: Not specified

Vendor(s): None

Tool Description

Patternizr aids crime analysis in detection of potential crime patterns.

Tool Purpose

Patternizr compares features of crimes and finds ones that are similar and may be part of a crime pattern. Analysts will look at the candidate crimes and suggest the formation of crime patterns to a pattern identification module. If a pattern is formed, detectives often consolidate the investigative efforts (e.g., one detective investigates all the crimes in the pattern.) The report filters non-normal trends into a spreadsheet and displays year-over-year counts of crimes that have non-normal trends. The tool requires a human user to evaluate the output data to see if complaints identified as similar are, in fact, connected to a pattern.

Update Description

Routine maintenance.

Data Analyzed

Identifying information may be collected or disclosed in the training data and input data.

Training Data

Separate models were trained for each of three different crime types (burglaries, robberies, and grand larcenies). These crime types have a sufficient corpus of prior manually identified patterns for use as training examples. This corpus consists of approximately 10,000 patterns between 2006 and 2015 from each crime type. A portion of this corpus includes complaint records where the same individual was arrested for multiple crimes of the same type within a span of two days.

Input Data

The input data is a candidate crime and its features. A complaint describes details of the crime, including the date and time (which can be a range if the precise time of occurrence is unknown), location, crime subcategory, modus operandi, and suspect information. This information is used to calculate the five types of crime-to-crime similarities used as features by Patternizr: location, date-time, categorical, suspect and unstructured text.

Output Data

Probability that a complaint is connected to a pattern.

ShotSpotter

Department: NYPD

First Used: March 2015

Updated in 2025: Yes

Purpose Type: Data management

Computation Type: Classification

Autonomy: Informative

Frequency of Use: Not specified

Population Impacted: Geographic space

Vendor(s): ShotSpotter

Tool Description

ShotSpotter provides acoustic gunshot detection to assist with emergency call response.

Tool Purpose

Provides acoustic gunshot detection to assist with emergency call response. The tool supports patrol operations in alerting units to potential gunfire and enhances investigations involving firearms.

Update Description

Routine maintenance.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

Specialized software analyzes audio signals for potential gunshots.

Output Data

The tool determines the location of the sound source, and once classified as potential gunfire sends the incident to acoustic experts for additional analysis. Notifications are sent for confirmed gunfire. ShotSpotter activations may result in evidence collection that can enhance case investigations. Problematic locations identified through alerts may require additional resource deployment and/or investigations.

Vendor Information

Vendor Name(s): ShotSpotter

Vendor Type(s): Procurement/Paid pilot, Professional services

Software developed and maintained by vendor.

NYC Emergency Management (NYCEM)

Microsoft Copilot for M365

Department: External Affairs

First Used: July 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Occasionally

Population Impacted: Individuals (General public)

Vendor(s): Microsoft

Tool Description

Microsoft Copilot for M365 is a generative AI tool.

Tool Purpose

NYCEM used Copilot to simplify preparedness language for public consumption and to improve accessibility using the plain language guidelines.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

NYCEM staff input preparedness recommendations that had been previously drafted.

Output Data

The program output simplified preparedness recommendations, taking into account plain language guidelines.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): No-cost engagement

Microsoft provides and updates the software and offered office hours to discuss the tool.

NYC Public Schools (NYCPS)

Algebra Teaching Assistant

Department: Division of Instructional and Information Technology

First Used: May 2023

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Daily

Population Impacted: Individuals
(Teachers, students)

Vendor(s): Microsoft

Tool Description

The Algebra Teaching Assistant uses the Division of Instructional and Information Technology AI platform to access specific algebra-focused content to provide responses to prompts related to algebra.

Tool Purpose

The tool is used to generate responses to prompts entered by a student or teacher, requesting the generative AI tool to compose a text response to a text input.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

The large language model (LLM) has been trained exclusively on curriculum from Illustrative Math.

Input Data

Prompts provided by the users of the system.

Output Data

The output data for the Teaching Assistant is the response generated by specifically developed LLM using the Illustrative Math curriculum.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): No-cost engagement

Microsoft provided technical guidance for their emerging generative AI technology and built some small modules of code for the specific Teaching Assistant use cases.

Annual Professional Performance Review Measures of Student Learning (MOSL) Growth Model

Department: Division of Instructional and Information Technology

First Used: September 2013

Updated in 2025: No

Purpose Type: Performance evaluation

Computation Type: Scoring

Autonomy: Supervised

Frequency of Use: Daily

Population Impacted: Individuals
(Teachers)

Vendor(s): Education Analytics

For more information: <https://tools.nycenet.edu/resources/awa-guide/>

Tool Description

The growth model uses a variety of student-level (assessment scores, English language learner, disability, and economic disadvantage indicators), classroom-level (e.g. percent students with disabilities), and school-level data (e.g. percent English language learners, percent students with disability, average prior achievement, school type) to estimate/predict a student's score on one of many possible course-culminating assessments. These predicted scores are either used to identify "peer groups" of students, from which student growth percentiles (SGPs) are determined, or compared to actual scores to determine student credit values. These units (SGPs or credit values) are then weight-averaged to generate an educator-level result - the Measures of Student Learning (MOSL) rating. The MOSL rating is combined with the Measures of Teaching/Leadership Practice (MOTP/MOLP) rating to produce an Overall Rating. Per state law 3012-d, annual ratings "shall be a significant factor in HR decisions." This is often implemented by making ratings a qualifying/disqualifying element in decision-making concerning employment, tenure, salary, and other professional opportunities.

Tool Purpose

In accordance with New York State law and New York State Education Department regulations, NYCPS developed and maintains a "growth model" to produce MOSL ratings for use in annual professional performance reviews for teachers and principals. The MOSL ratings are combined with MOTP/MOLP ratings to produce an annual Overall Rating for each eligible educator.

Data Analyzed

Identifying information may be collected or disclosed in the training data and input data.

Training Data

The growth model process is employed in both retrospective and prospective ways. In the retrospective version, the results are determined entirely within-sample. In the prospective version, the coefficients of the model are estimated on multiple prior years of data.

Input Data

The growth model makes use of three types of data: students' end-of-year assessment scores, enrollment and attendance records that link students to teachers and schools, and historical academic and demographic information used to identify groups of similar students.

Output Data

The model outputs an estimate of a student's score on a course-culminating assessment.

Vendor Information

Vendor Name(s): Education Analytics

Vendor Type(s): Professional services

Education Analytics provides technical assistance and quality assurance for the growth model.

Annual Professional Performance Review Measures of Teaching/Leadership Practice (MOTP/MOLP) Calculation

Department: Division of Instructional and Information Technology

First Used: October 2013

Updated in 2025: No

Purpose Type: Performance evaluation

Computation Type: Scoring

Autonomy: Supervised

Frequency of Use: Daily

Population Impacted: Individuals
(Principals, assistant principals, teachers)

Vendor(s): None

For more information: <https://tools.nycenet.edu/resources/awa-guide/>

Tool Description

Throughout a school year, evaluators observe teachers/principals multiple times and use a rubric to provide a numerical rating on one or more rubric components. These rubric component scores are then weight-averaged according to collectively bargained rules to produce a Measure of Teaching/Leadership Practice (MOTP/MOLP) Rating. The MOTP/MOLP rating is combined with the Measures of Student Learning (MOSL) rating to produce an Overall Rating for each eligible educator. Per state education law 3012-d, annual ratings “shall be a significant factor in HR decisions.” This is often implemented by making ratings a qualifying/disqualifying element in decision-making concerning employment, tenure, salary, and other professional opportunities.

Tool Purpose

In accordance with New York State law and New York State Education Department regulations, NYCPS developed and maintains databases and calculation rules to produce MOTP/MOLP ratings for use in annual professional performance reviews for teachers and principals. The MOTP/MOLP ratings are combined with MOSL ratings to produce an annual Overall Rating for each eligible educator.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Pilot data prior to program launch was used to inform the weights assigned to various rubric components. However, the weights are ultimately determined via collective bargaining.

Input Data

Rubric component numerical ratings.

Output Data

The model outputs a score for teachers and principals.

Eureka! Chatbot

Department: Division of Instructional and Information Technology

First Used: August 2023

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Monitored

Frequency of Use: Daily

Population Impacted: Individuals (NYCPS staff and parents)

Vendor(s): Nagarro and Microsoft

For more information: <https://www.schools.nyc.gov/learning/digital-learning/applications-and-platforms/supporthub>

Tool Description

The Azure Cognitive Services technology and chatbot (internally branded as “Eureka!”) was configured and deployed in August 2023 to be the first response to calls to the NYCPS IT Service Desk. It accesses scripts to handle four common reasons for a user to call or contact the service desk: Password Reset, Create a Ticket, Ticket Status, Request for Information. The chatbot accesses pre-defined scripts to respond to user voice or text input. The user’s request is either serviced, completed and closed by the chatbot, or the user is given the option (at any time) to connect to a live agent.

Tool Purpose

The tool is used to respond to common IT service desk requests: Password Reset, Create a Ticket, Ticket Status, Request for Information. Users can access the tool by phone, by computer through the NYCPS Support Hub application, and from links from Microsoft Teams and other NYCPS systems, such as TeachHub.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Pre-defined scripts designed to respond to four common requests to the IT Service Desk.

Input Data

A voice call or text-based chat session initiated by a user and responded to by the Eureka! chatbot before being handled by a human Service Desk agent.

Output Data

The chatbot generates responses to user-entered prompts based on the training data or forwards the inquiry to a human Service Desk agent. Since its launch in August 2022, the chatbot has handled an average of 1,500 calls and 300 web-based inquiries each day. Approximately 30 percent of the voice calls and 10 percent of the web-based queries have been handled completely by Eureka! without being forwarded to a human Service Desk agent.

Vendor Information

Vendor Name(s): Nagarro and Microsoft

Vendor Type(s): Professional services

Developed by an IT services vendor (Nagarro) using Microsoft Cognitive services.

MySchools - Match

Department: Division of Instructional and Information Technology

First Used: August 2018

Updated in 2025: No

Purpose Type: Resource allocation

Computation Type: Matching

Autonomy: Monitored

Frequency of Use: Daily

Population Impacted: Individuals
(Students)

Vendor(s): Blenderbox

For more information:

https://enrollmentsupport.schools.nyc/app/answers/detail/a_id/3652

Tool Description

The tool utilizes the Gale-Shapley deferred acceptance algorithm to match applicants to schools. This algorithm has been in existence for many years, used internationally for various purposes. It's most common use is in the National Resident Matching Program for medical school students.

Deferred acceptance works as an iterative series of steps: students and programs are tentatively matched in each step, but nothing is finalized until the algorithm terminates (hence the deferred).

1. Each student "proposes" to their first choice;
 - Programs assign seats to students one at a time;
 - When all seats are filled, programs may reject previously accepted students in favor of new applications from students they prefer (e.g., students with a better lottery number);
 - Remaining students are rejected;
2. Students rejected in the last step "propose" to the next choice on their list;
3. The algorithm terminates when all students are matched or have proposed to all the programs they listed.

Tool Purpose

MySchools is an application used to house online school directories, collect application choices, and run the admissions matching algorithm that is used for all centralized admissions processes (3-K, pre-K, Gifted & Talented, middle school, and high school). The tool encompasses a family-facing portal, a school-facing portal, and an administrative portal.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

N/A

Input Data

Student biographical information (e.g., home address, poverty status, home language), student academic information (e.g., course grades, state test scores), and student school records (e.g., sending school).

Output Data

The algorithm outputs a school match for each student.

Vendor Information

Vendor Name(s): Blenderbox

Vendor Type(s): Professional services

We have a five-year contract with the agency Blenderbox who designed the application and implemented the algorithmic matching functionality. The work is meant to transition to be run in-house, by the Division of Instructional and Information Technology (DIIT) within NYCPS, by the end of the contract. The team at DIIT has already begun to takeover maintenance and development of the tool.

MySchools - Probability of Acceptance

Department: Division of Instructional and Information Technology

First Used: September 2024

Updated in 2025: Yes

Purpose Type: Information presentation

Computation Type: Scoring

Autonomy: Fully autonomous

Frequency of Use: Daily

Population Impacted: Individuals
(Students)

Vendor(s): Researchers from
Massachusetts Institute of Technology
(MIT)

For more information:

https://enrollmentsupport.schools.nyc/app/answers/detail/a_id/3652

Tool Description

This feature, added to MySchools in 2024, determines a probability of acceptance at a specific school for a future high school student. This is calculated and displayed as a student is selecting schools to apply to in the MySchools application.

Tool Purpose

Information is presented to students and parents to help them decide what high schools to apply to.

Update Description

As part of a limited pilot program, some eight grade students were notified that they had a high likelihood of acceptance to a high-performing school. This was done to encourage these students who may not be considering those schools to apply to attend high schools at those schools.

Data Analyzed

Identifying information may be collected or disclosed in the training data and input data.

Training Data

Model was developed by researchers affiliated with the Massachusetts Institute of Technology (MIT) and trained on information about New York City public high schools. Students will see an icon indicating whether they have a “high,” “medium,” or “low” chance of receiving an offer, based on the applicant’s admissions characteristics like district or borough, grades, priority group, and the school’s admissions method, such as whether the admission is open or screened.

Input Data

Student high school selections and student records.

Output Data

Probability of acceptance for the student to a specific high school, indicated as “high”, “medium” or “low.”

Vendor Information

Vendor Name(s): Researchers from Massachusetts Institute of Technology (MIT)

Vendor Type(s): Professional services

MIT developed the tool and the Division of Instructional and Information Technology integrated it into the MySchools system.

Open Gen AI and Teaching Assistant Tool

Department: Division of Instructional and Information Technology

First Used: May 2023

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Supervised

Frequency of Use: Daily

Population Impacted: Individuals
(Teachers, students)

Vendor(s): Microsoft

Tool Description

The generative AI system using large language models was a system custom-built by the Division of Instructional and Information Technology using advanced Microsoft technologies to create a set of generative AI tools. "Open Gen AI" accesses a large language model (currently OpenAI's GPT 3.5) to provide responses to a broad range of prompts.

Tool Purpose

The tool is used to generate responses to prompts entered by a student or teacher, requesting the generative AI tool to compose a text response to a text input.

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

ChatGPT training data is proprietary to OpenAI.

Input Data

Prompts provided by the users of the system.

Output Data

The output data for the Open Gen AI tool is the response generated by the ChatGPT large language model.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): Procurement/Paid pilot

Microsoft provided technical guidance for their emerging generative AI technology and built some small module of code for the specific NYCPS Gen AI and Teaching Assistant use cases.

Office of Chief Medical Examiner (OCME)

STRmix

Department: Department of Forensic Biology

First Used: January 2017

Updated in 2025: No

Purpose Type: Data management

Computation Type: Matching

Autonomy: Informative

Frequency of Use: Not specified

Population Impacted: Individuals (Those suspected of crime, arrested, and put on trial); Biological sample

Vendor(s): NicheVision Forensics, LLC

Tool Description

STRmix™ combines sophisticated biological modelling and standard mathematical processes to interpret a wide range of complex DNA profiles. Using well-established statistical methods, the software builds millions of conceptual DNA profiles.

Tool Purpose

STRmix is a forensic DNA analysis software program that uses a probabilistic genotyping algorithm to interpret complex DNA profiles, such as those from mixed samples that contain DNA from multiple contributors.

STRmix™ models any types of allelic and stutter peak heights as well as drop-in and drop-out behavior. It does this rapidly, accessing evidential information previously out of reach with traditional methods. STRmix™ is supported by comprehensive empirical studies with its mathematics readily accessible to DNA analysts, so results are easily explained in court.

Version effective March 2026

Data Analyzed

Identifying information may be collected or disclosed in the input data.

Training Data

Training data was not used in the sense of AI software. OCME performed thousands of tests using the software to validate it for optimum use with our current laboratory standard operating procedures and genetic analyzers.

Input Data

Forensic DNA profiles from crime scenes as well as the DNA profiles from victims and suspects of crimes.

Output Data

The output is a deconvolution of genotype probability distribution that lists all of the accepted genotype sets and their associated weights. These weights can take any value from zero to one.

Vendor Information

Vendor Name(s): NicheVision Forensics, LLC

Vendor Type(s): Professional services

The software has been developed by New Zealand Crown Institute of Environmental Science and Research with Forensic Science South Australia. The developer assisted OCME in analyzing and interpreting our data during the validation of the software.

Office of Technology and Innovation (OTI)

311 Mobile App AI Smart Select

Department: Applications

First Used: December 2025

Purpose Type: Information presentation

Computation Type: Classification

Autonomy: Informative

Frequency of Use: Continuous

Population Impacted: Individuals (Users of the 311 mobile app)

Vendor(s): Microsoft, IBM

Tool Description

311 Smart Select allows users of the 311 mobile app to upload a photo of the problem they would like to report, and then provides suggestions of matching service request types. Users can select the results that best describes their issue, and then will be directed to a pre-filled mobile service request form or knowledge article.

Tool Purpose

The purpose of the tool is to provide users with the most relevant service request type for their problem. Users may encounter problems in the city but not know the precise language to register their complaint effectively. This tool ensures that 311 users are easily and accurately able to report issues, and that the city can triage and address concerns effectively.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

The training data is proprietary to Microsoft.

Input Data

Images uploaded by the user. Images are virus scanned and stored to comply with Freedom of Information Law requirements.

Output Data

Recommended service request types.

Vendor Information

Vendor Name(s): Microsoft, IBM

Vendor Type(s): Off-the-shelf, Professional services

This feature is implemented using Microsoft Azure AI Foundry and Microsoft Azure AI Search. IBM is the professional services vendor who supported the development of the application.

Fusus CORE Elite AI

Department: Public Safety and Emergency Management

First Used: January 2025

Purpose Type: Risk management

Autonomy: Monitored

Population Impacted: Property;
Geographic space; Individuals (Individuals
in areas where cameras are placed.)

Computation Type: Classification

Frequency of Use: Continuous

Vendor(s): Fusus by Axon

For more information: <https://www.axon.com/products/axon-fusus>

Tool Description

Fusus CORE Elite AI enables real-time object detection, object classification, and alerting when connected to camera systems.

Tool Purpose

Fusus CORE Elite AI enhances operational situational awareness by detecting objects, people, or motion-based events in real time, such as a person or a vehicle entering a restricted area. The tool can be configured to identify specific detection parameters and create alert profiles for those parameters, allowing automated notifications to be sent to designated monitoring staff. Alerts are currently distributed to email groups monitored by OTI's Technology Operations Center and Emergency Management personnel.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Live or recorded video streams from connected cameras, which are stored locally and subject to a limited retention period.

Output Data

Detection alerts and labeled video metadata.

Vendor Information

Vendor Name(s): Fusus by Axon

Vendor Type(s): Off-the-shelf, Professional services

This tool is part of the FUSUS video management system.

Version effective March 2026

Google Translate

Department: Web Operations

First Used: 2013

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Continuous

Population Impacted: Individuals (Users of nyc.gov)

Vendor(s): Google

Tool Description

Google Translate enables machine translation of nyc.gov and subpages into over 100 languages.

Tool Purpose

The Google Translate widget is used to make information on nyc.gov and its subpages more accessible to New Yorkers with limited English proficiency. It translates content into the 10 languages required under the language access law (Local Law 30 of 2017) and over 100 others.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to Google.

Input Data

Web content on nyc.gov and its subpages, written in English.

Output Data

Translated web content into the selected language.

Vendor Information

Vendor Name(s): Google

Vendor Type(s): Off-the-shelf

Google provides this service for free to organizations.

LLM Generation of Draft List of City Agencies and Governance Organizations

Department: Office of Data Analytics

First Used: September 2025

Purpose Type: Data management

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Once

Population Impacted: Group, organization, or business

Vendor(s): Claude by Anthropic

For more information: https://data.cityofnewyork.us/City-Government/NYC-Agencies-and-Governance-Organizations/t3jq-9nkf/about_data

Tool Description

This algorithmic tool was used to populate an initial draft of and manage updates to the NYC Agencies and Governance Organizations dataset: a standardized, analysis-ready, machine-readable reference of New York City agencies and other organizations with New York City-specific governance functions.

Tool Purpose

Large language models were employed in the creation of the NYC Agencies and Governance Organizations dataset to help institute acceptance rules for entity inclusion and naming conventions and implement a pipeline: ingest, normalize, de-dupe/record-link, validate, and publish, with Unicode/encoding fixes, name parsing, and auditable transformations at each step. Structured human quality assurance (peer review, change logs, and schema validation) preceded scripted exports to NYC Open Data, which then feeds the NYC.gov Agency Directory.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Internal and external sources on New York City agencies and governance organizations including Citywide Org Chart; Mayor's Office of Operations' list, as well as lists from the

Department of Records and Information Services and the Department for Citywide Administrative Services.

Output Data

A single, analysis-ready table with preferred names, common alternates, leadership, and reporting lines powers both public discovery (NYC.gov Agency Directory) and internal analytics.

Vendor Information

Vendor Name(s): Claude by Anthropic

Vendor Type(s): Off-the-shelf

An off-the-shelf product was used to develop the tool.

Microsoft Copilot Pilot Program

Department: Research & Collaboration

First Used: June 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Daily, over the course of the testing period

Population Impacted: Individuals (City staff who participated in the pilot)

Vendor(s): Microsoft

For more information: <https://www.microsoft.com/en-us/microsoft-365-copilot>

Tool Description

From June to August of 2025, select staff at OTI tested Microsoft Copilot for Office 365 and Copilot Chat through a pilot program to determine whether the applications provided business users meaningful benefits, as well as how the applications would need to be managed and secured.

Tool Purpose

Microsoft Copilot for Office 365 and Copilot Chat were tested to better understand the products and their potential value for City staff. Testers were encouraged to use the applications as part of their routine work flows, completing specific tasks that ranged from legal analysis of proposed legislation, reviewing stakeholder feedback for policy documents, data analysis on survey results, to developing content for social media. While

most test cases were experimental and did not have a public impact, some outputs may have been used in public communications.

Data Analyzed

Identifying information may be collected or disclosed in the training, input, and output data.

Training Data

Training data is proprietary to Microsoft. For Copilot for Microsoft 365, the model was also tuned using information about the files and structure of the OTI tenant, known as the “knowledge graph.”

Input Data

Users could input text and attach files.

Output Data

The tool output text, images, and documents, depending on the user prompt.

Vendor Information

Vendor Name(s): Microsoft

Vendor Type(s): No-cost engagement, Off-the-shelf

Microsoft provided 50 trial licenses for testing Microsoft Copilot for Office 365 and made Copilot Chat available to 40 users (Copilot Chat is free for Microsoft users but turned off in the OTI-hosted tenant).

MyCity Chatbot

Department: Applications

First Used: September 2023

Purpose Type: Information presentation

Autonomy: Fully autonomous

Population Impacted: Individuals
(Primarily City residents, visitors, and business owners)

Updated in 2025: Yes

Computation Type: Ranking

Frequency of Use: Continuous

Vendor(s): Microsoft, EY

For more information: <https://chat.nyc.gov>

Tool Description

The NYC MyCity chatbot is a beta AI-powered chatbot that provides information and access to services for residents and businesses in New York City. The chatbot was decommissioned in February 2026.

Tool Purpose

The NYC MyCity chatbot is a beta AI-powered chatbot that provides information and access to services for residents and businesses in New York City. It's currently focused on two main areas: Business Services and 311 information on City services. Basic information on MyCity is also included, but is a much smaller aspect of the index. The chatbot provides information on starting or operating a business in New York City, answers questions about permits, licenses, regulations, and other business requirements, and connects users with relevant resources and support services. It also offers information on various city services and benefits, especially those covered by 311 knowledge articles, and helps users find resources related to childcare, career, and other areas. The chatbot is using Microsoft's Azure AI technology and OpenAI's ChatGPT 4-o large language model.

Update Description

The chatbot content was expanded to include 311 knowledge articles.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

Training data is proprietary to the vendor.

Input Data

Text queries are input by the user on the MyCity portal.

Output Data

The tool produces text responses with references and links primarily based on information from Business Services and 311 knowledge articles.

Vendor Information

Vendor Name(s): Microsoft, EY

Vendor Type(s): Procurement/Paid pilot, Professional services

Microsoft provides cloud-based ChatGPT services, and EY is the professional services vendor.

Omnichannel Language Translation

Department: 311

First Used: January 2024

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Informative

Frequency of Use: Continuous

Population Impacted: Individuals
(Customers who contact 311 via text/SMS)

Vendor(s): Microsoft

Tool Description

The Omnichannel Language Translation tool delivers multi-language capability for the 311 text/SMS channel. The tool supports the 10 designated citywide languages to enable 311 agents to interact with customers in their language.

Tool Purpose

The algorithmic tool converts the customer's text inquiry in their chosen language into English, allowing the text agent to understand, research and reply to the inquiry. The tool converts the agent's English language response to the customer's chosen language among the 10 designated citywide languages.

Data Analyzed

No identifying information is collected or disclosed.

Training Data

The training data is proprietary to Microsoft.

Input Data

Text/SMS inquiries from customers via 311-NYC.

Output Data

Responses to customer inquiries in the language the customer used.

Vendor Information

Vendor Name(s): Microsoft

Version effective March 2026

Vendor Type(s): Procurement/Paid pilot

Omnichannel is part of the Microsoft suite available to OTI as part of the Dynamics customer relationship management platform. Microsoft supported the design, development, and testing of the tool preparation and deployment.

Zoom Automated Captions

Department: Office of Data Analytics

First Used: March 2021

Updated in 2025: No

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Captions were produced for 45 events from Open Data Week 2025

Population Impacted: Individuals
(Captions are provided to those who choose to watch the videos online.)

Vendor(s): BetaNYC, Zoom

For more information: <https://www.youtube.com/@NYCOpenDataWeek>

Tool Description

Creates virtual closed captioning/live transcription during Zoom meetings.

Tool Purpose

Captions are provided to attendees of Zoom meetings held by the NYC Open Data team at the Office of Data Analytics in conjunction with the civic tech non-profit BetaNYC under the Open Data Week and Open Data Ambassador initiatives. The full transcription of the event is then added to the meeting recordings, which are uploaded on YouTube. The purpose of the captions, both for the live event and the recording, is to improve meeting accessibility.

Data Analyzed

Identifying information may be collected or disclosed in the input and output data.

Training Data

Training data is proprietary to Zoom.

Input Data

Live audio from Zoom meeting.

Output Data

Captions/transcript of meeting.

Vendor Information

Vendor Name(s): BetaNYC, Zoom

Vendor Type(s): Off-the-shelf, Professional services

BetaNYC is our collaborator on the Open Data Week and Open Data Ambassador initiatives. They own and operate the Zoom account that is used for meetings under these initiatives, have access to the transcription files, and use these when editing and uploading video recordings onto YouTube.

Office of the Comptroller (Comptroller)

Google Translate

Department: Communications

First Used: January 2014

Updated in 2025: Yes

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Fully autonomous

Frequency of Use: Daily

Population Impacted: Individuals

Vendor(s): Google

For more information: <https://translate.google.com/>

Tool Description

Google Translate enables machine translation of Comptroller's website, its subpages, and printed materials to over 100 languages.

Tool Purpose

The Google Translate widget is used to make information on Comptroller's website, its subpages, and printed materials more accessible to New Yorkers with limited English proficiency. It translates content into 10 languages required under the language access law (Local Law 30 of 2017) and over 100 others.

Update Description

Google Translate was updated by the vendor (Google), and our agency continued using the tool under the updated version.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Web content on Comptroller's website and its subpages; content for printed materials written in English.

Output Data

Translated content into the selected language.

Vendor Information

Vendor Name(s): Google

Vendor Type(s): Off-the-shelf

Google Translate is a publicly available product.

Mirage

Department: Communications

First Used: January 2025

Purpose Type: Information presentation

Computation Type: Data transformation

Autonomy: Monitored

Frequency of Use: Once a week

Population Impacted: Individuals (Social media users who follow office socials)

Vendor(s): Mirage

For more information: <https://mirage.app/>

Tool Description

Mirage is a captioning tool used for social media videos. This tool assists staff to accurately add captions to the video visuals created by the Comptroller's office for social media publishing.

Tool Purpose

This tool adds captions to the videos created by the Comptroller's office for social media.

Data Analyzed

Identifying information may be collected or disclosed in the input data and output data.

Training Data

Training data is proprietary to the vendor.

Input Data

Video taken for social media.

Output Data

Video output with captions added to the input video.

Vendor Information

Vendor Name(s): Mirage

Vendor Type(s): Off-the-shelf

Vendor created product is used for captioning.

Appendix

How to Identify an Algorithmic Tool Qualifying Criteria

The definition of algorithmic tool from LL 35 can be broken out into three major criteria:

1. **Data Analysis.** Any technology or computerized process that is derived from machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis, that is used to make or assist in making decisions about and implementing policies that materially impact the rights, liberties, benefits, safety or interests of the public, including their access to available city services and resources for which they may be eligible.
2. **Decision-Making.** Any technology or computerized process that is derived from machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis, that is used to make or assist in making decisions about and implementing policies that materially impact the rights, liberties, benefits, safety or interests of the public, including their access to available city services and resources for which they may be eligible.
3. **Material Impact.** Any technology or computerized process that is derived from machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis, that is used to make or assist in making decisions about and implementing policies that materially impact the rights, liberties, benefits, safety or interests of the public, including their access to available city services and resources for which they may be eligible.

A tool or system must meet **all three** of these criteria to be defined as an algorithmic tool per LL 35. If a tool or system meets only two or fewer criteria, it will not qualify.

Examples

The examples included below represent both qualifying and non-qualifying tools and systems. The table can be used to help understand what characteristics may or may not cause a tool or system to qualify.

Example	Criterion			Qualify?	Explanation
	1	2	3		
Agency X creates an automated web application form for a specific benefit delivered by the agency. The criteria to be eligible for the benefit were established by the state law that funds the program.		✓	✓	No	Although the system assists in decision-making per Criterion 2, and the effects of those decisions are material per Criterion 3, the use of pre-determined criteria does not meet the definition of data analysis in Criterion 1.
Agency X's leadership routinely consults a dashboard of agency key performance indicators. The performance of these indicators leads to agency decision-makers to undertake reallocations of staff and other resources to different city geographies.		✓	✓	No	The dashboard allows leadership to make data-informed decisions per Criterion 2, and the outcomes of those decisions may materially impact the allocation of city services per Criterion 3, but the production and display of KPIs does not meet the definition of data analysis in Criterion 1.
Agency X develops a risk calculator that applies a score to each of its clients. The inputs for the score were determined through a machine-learning algorithm. The scores are used by the agency to determine eligibility for a specific service offered by the agency.	✓	✓	✓	Yes	Machine Learning approaches are explicitly considered in Criterion 1, and the outputs of the calculator are operationalized for decision-making per Criterion 2, and those decisions are materially impacting members of the public per Criterion 3.

Agency X's research division uses a variety of statistical regression techniques to analyze historical patterns in data to determine which client characteristics are most associated with adverse outcomes. The research, including variable coefficients, are shared with agency leadership for review.	✓		✓	No	Statistical regression is considered data analysis under Criterion 1, and the adverse outcomes considered in the analysis are material per Criterion 3, but the analysis is not being operationalized for agency decision-making as required by Criterion 2.
Agency X uses a logistic regression to analyze likelihood of clients' success in completing an agency program. The agency then creates an index derived from the model and applies that index to new, incoming clients, prioritizing those with a higher likelihood for success.	✓	✓	✓	Yes	Statistical regression is considered data analysis under Criterion 1, the agency is using the analysis regularly for its decision-making per Criterion 2, and the prioritization for an agency service is material per Criterion 3.
Agency X develops a predictive model to determine which vehicles in its fleet are likely to break down first. The agency runs the model monthly on the vehicles within the fleet based on current maintenance data, and generates a list of vehicles to undergo maintenance that month.	✓	✓		No	Predictive modeling is explicitly considered Criterion 1, and the agency is actively using the model's outputs for decision-making per Criterion 2, but neither the model's outputs nor outcomes have a material impact per Criterion 3.

Agency X procures and deploys a chatbot, where residents can make an inquiry online and either receive a computer-generated answer or be directed to a representative.	✓	✓	✓	Yes	Chatbots, an application of a natural language processing, would qualify for Criterion 1 as an example of Artificial Intelligence. The tool is being used to support agency decision-making per Criterion 2, and generating information that is shared with the public is considered a material impact per Criterion 3.
Agency X uses a statistical model to determine neighborhood characteristics associated with positive outcomes for residents. The agency uses the model to select 10 neighborhoods per quarter to receive mobile agency services.	✓	✓	✓	Yes	Statistical regression is considered data analysis under Criterion 1, the agency is using the analysis regularly for its decision-making per Criterion 2, and the selection of neighborhoods to receive agency services has a material effect on neighborhood residents per Criterion 3.

Reportable Elements

1. **Agency, Department.** The name of the agency and the department or office who used the algorithmic tool.
2. **Tool name.** The name or commercial name of the tool.
3. **Date of first use.** The month and year in which the tool began to be used, if known.
4. **Updated tool.** If used before the current year, whether the tool has been updated in the past year.
5. **Purpose type.** The type of purpose of the tool. See XXXX for full definitions.
 - a. Data management
 - b. Information presentation
 - c. Performance evaluation
 - d. Resource allocation
 - e. Risk management
 - f. Triage
6. **Computation type.** Type of computation used by the tool. See XXXX for full definitions.
 - a. Classification
 - b. Clustering
 - c. Data generation
 - d. Forecasting
 - e. Matching
 - f. Ranking
 - g. Sampling
 - h. Scoring
7. **Autonomy type.** Level of autonomy the tool has. See XXXX for full definitions.
 - a. Informative
 - b. Supervised
 - c. Monitored
 - d. Fully autonomous
8. **Frequency.** How frequently the tool was used in the calendar year.
9. **Population type [all that apply].** Population(s) are reviewed, assessed, or directly affected by the tool:
 - a. Individuals (please describe)
 - b. Group, organization, or business
 - c. Geographic space
 - d. Property
 - e. Biological sample

- f. Other (please describe)
- 10. **Website.** A website that provides more information about the tool.
- 11. **Tool description.** A brief description of the tool.
- 12. **Purpose description.** A description of how the information received from the algorithmic tool is used, including the purpose for which the agency is using the algorithmic tool.
- 13. **Update description.** If updated in the past year, how the tool was updated (e.g., fresh training data, variable weights).
- 14. **Identifying Information [all that apply].** Whether the following sets of data analyze, collect, or disclose “identifying information” as defined under New York City’s Identifying Information Law (Local Laws 245 and 247 of 2017):
 - a. Training data
 - b. Input data
 - c. Output data
 - d. No identifying information is analyzed, collected, or disclosed
- 15. **Training data, Input data, Output data.** The categories of data or the variables used in developing or operating the tool. Describe the sources of the data discussed.
- 16. **Vendor name.** If applicable, name of the vendor or contractor involved in the development or ongoing use of the tool.
- 17. **Vendor type [all that apply].** If applicable, type of vendor or contractor engagement:
 - a. Procurement/Paid pilot: A vendor is paid to use their product
 - b. No-cost engagement: A vendor is not paid for the short-term use of their product
 - c. Off-the-shelf: A vendor-created product is available to the public
 - d. Professional services: A vendor supports the implementation of a product
- 18. **Vendor involvement.** If applicable, a description of the vendor or contractor involvement.

Purpose Type

This appendix provides definitions and examples of use cases that may be associated with a given Purpose Type. Agencies should use this appendix to help populate the “Purpose Type” field for an algorithmic tool that has been identified for reporting.

Data Management systems are those providing access to data, storage or monitoring the storage of data, and controlling input-output operations within a data processing system.¹ Data management systems may perform one or more of tasks such as:

- deduplication of redundant entries,
- data cleaning by identifying anomalous entries,
- data validation ensuring that values are within expected ranges,
- imputing missing data,
- redaction of any sensitive or private data,
- data fusion to combine data records from multiple data sources, and
- disambiguating different entities with the same name.

Examples: Differentiating between voters that share the same name, removing instrument readings that are unphysical, flagging log entries of events with a negative time duration, combining records across different databases to provide a complete case history.

Performance Evaluation systems are those that judge whether, or measure of the extent to which, the outputs of a system fulfill some predefined requirements. These requirements may be set by compliance, legal, contractual, regulatory, or other procedural and organizational needs.²

Examples: Evaluating the number of cases correctly handled to measure staff productivity, monitoring an air quality sensor for anomalously hazardous pollution instrument readings, sensing traffic light failures to report the mean time between failure, and monitoring customer service interactions to determine the fraction of requests which exceed the response time specified in a service-level agreement.

Resource Allocation systems perform assignment and distribution of available resources. Resources may be physical, virtual, or human.

Examples: Access to seats in a public benefits program with finite capacity, assignment of teachers to classrooms, dispatch of response vehicle to emergency site.

¹ Adapted from § 3.6.6.2 Data Management, Information technology — *Metadata Registries Interoperability and Bindings (MDR-IB) — Part 1: Framework, common vocabulary, and common provisions for conformance*, ISO/IEC JTC 1/SC 32, ISO/IEC 20944-1:2013(en).

² Adapted from “performance evaluation”, Technical Committee ISO/TC 176/SC 2, *Quality management systems — Requirements*, ISO 9001:2015(en), § 9.

Risk Management systems coordinate activities to direct and control an organization regarding Risk.³ Risk itself is the effect of uncertainty on objectives, i.e., how a deficiency of knowledge has a consequence on a result to be achieved.⁴

Risk management can be broken down into several stages of a lifecycle⁵:

1. Risk Identification,⁶
2. Risk Analysis,⁷
3. Risk Evaluation,⁸
4. Risk Treatment,⁹
5. monitoring and review, and
6. recording and reporting

The first three stages are also collectively known as Risk Assessment.¹⁰ Risks can come in many forms and categories, for example: financial risks like credit risk, cybersecurity risks like data breach, public safety risks like the occurrence of a natural disaster, or legal risks like contractual liability.

Examples: Identifying students at high risk of failure, timing traffic lights to regulate the flow of traffic through a busy road, provisioning of backup generators ahead of an anticipated storm event.

³ Adapted from “risk management” a.k.a. “organizational risk”, ISO/IEC 31000:2018(en), *Risk management — Guidelines*, § 3.2.

⁴ Adapted from Technical Committee ISO/TC 262, *Risk management — Vocabulary*, ISO 31073:2022(en), <https://www.iso.org/standard/79637.html>, § 3.3.1 “risk”.

⁵ ISO/IEC 31000:2018(en) as cited above, § 6.

⁶ The process of finding, recognizing and describing Risks. From ISO 31073:2022(en) as cited above, § 3.3.9 “risk identification”.

⁷ The process to comprehend the nature of Risk and to determine the level of Risk. Risk Analysis provides the basis for Risk Evaluation and decisions about Risk Treatment. From ISO 31073:2022(en), as cited above, § 3.3.15 “risk analysis”.

⁸ The process of comparing the results of Risk Analysis with Risk criteria to determine whether the risk is acceptable or tolerable. From ISO 31073:2022(en), as cited above, § 3.3.25 “risk evaluation”.

⁹ The process to modify Risk. Risk Treatment can involve: avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk; taking or increasing risk in order to pursue an opportunity; removing the risk source; changing the likelihood; changing the consequences; sharing the risk with another party or parties [including contracts and risk financing]; and retaining the risk by informed decision. Risk Treatments that deal with negative consequences are sometimes referred to as “risk mitigation”, “risk elimination”, “risk prevention” and “risk reduction”. Risk Treatment can create new risks or modify existing risks. From ISO 31073:2022(en), as cited above, § 3.3.32 “risk treatment”.

¹⁰ The overall process of Risk Identification, Risk Analysis, and Risk Evaluation. From ISO 31073:2022(en), as cited above, § 3.3.8 “risk assessment”.

Triage systems perform initial assessments of incoming cases to determine their importance and assignment of priorities for follow-up action.¹¹ They apply one or more criteria to determine the most important items that require attention through a follow-up action.

Examples of criteria: Financial cost, proximity of a vehicle to a response location, urgency of request, severity of risk.

Examples of follow-up action: Dispatch of a vehicle, opening an investigation, approval of staff roster, or other activities that can be categorized as Resource Allocation or Performance Evaluation.

Examples of Triage: Determining the magnitude of response to an emergency incident, assessing the degree of need in a social services case.

¹¹ Adapted from Technical Committee ISO/TC 46/SC 10, *Information and documentation — Emergency preparedness and response*, ISO 21110:2019(en), § 3.17 “triage”; and Technical Committee ISO/TC 309, *Whistleblowing management systems — Guidelines*, ISO 37002:2021(en), § 3.12 “triage”.
Version effective March 2026

Computation Type

This appendix provides definitions and examples of the categories of computation in algorithmic tools. For Artificial Intelligence systems, the computation refers specifically to the computational model contained within them. Agencies should use this appendix to help them populate the “Computation Type” response for an algorithmic tool that has been identified for reporting.

The assignment of Computation Type is determined by identifying the Data Types of the inputs and outputs. A Data Type is an attribute of data which describes the possible values the data may take, as well as the operations that are allowed on that data. The following Data Types are used in this appendix:

- **Collection** describes multiple elements, whose order of appearance is irrelevant.
- **Nominal** describes two or more categories which cannot be compared to determine which one is greater, or whose ordering is not relevant to the use case.
- **Numerical** can be manipulated with arithmetic operations.
- **Ordinal** describes two or more categories, such that every pair of categorical labels can be compared to determine which one is greater.
- **Sequence** describes multiple elements, whose number may change each time a function is executed, and whose order of appearance is relevant.
- **Tree** describes multiple elements where some elements, but not all, can be ordered.

Classification. A model whose output Data Type is Nominal or Ordinal. Classification used to categorize input data into predefined categories.

Note: Classification models are not to be confused with Scoring models, whose outputs are Numerical. Classification models output a category. A Scoring model may be used inside a Classification model to generate a number; the Classification model can then take the numerical output, compare it to a threshold, and finally determine the category to assign based on whether the threshold was met. Examples 3 and 13 illustrate the difference.

Example 1: A system performs optical character recognition (OCR) to scan invoices received from vendors and issue payments automatically. Based on the format, layout and text identified within the document, the system labels whether or not each invoice contains enough documentation to pay the vendor. If the documentation is deemed acceptable, the payment is issued, otherwise the document is rejected, and no payment is issued. The output of this system is a Nominal Data Type, which can take the possible values of “accepted” or “rejected”.

Example 2: A system contains a model which computes a probability that an individual will be at risk of homelessness within the next 24 months. If the probability exceeds 20%,

the individual is considered high risk; if the probability is between 5% and 20%, the individual is considered medium risk; otherwise, the individual is considered low risk. The numerical probability is not relevant for the Triage purpose; only the low, medium or high-risk category matters for subsequent processing.

Example 3: A traffic monitoring system contains three components: i) a computer vision component which identifies moving vehicles on a road; ii) a component which measures the speed of each vehicle; and iii) a component which determines if a vehicle is speeding and issues a moving violation ticket. The output type of the third component is an Ordinal Data Type, which takes the possible values “issue ticket” or “no action”, and is the final output of the entire system.

Example 4: A clustering analysis is performed to identify common patterns of network traffic usage in a network logging and monitoring system. The resulting clustering analysis is then used to identify anomalous traffic patterns which may be evidence of a cybersecurity incident. The output of the anomaly detection model is a risk score, to which a threshold is then applied to determine if the anomaly should be reported. In this situation, a Clustering computation is performed to create the anomaly detector. However, the anomaly detector itself, has a computation type of Classification.

Clustering. A function which takes as input a Collection, and assigns each element of the Collection a label whose Data Type is a Nominal, Ordinal, or Tree. Clustering identifies common subgroups of data based on mutual similarity. Each data point is assigned a label which defines the group it belongs to. The group labels are usually unordered and hence constitute a Nominal Data Type, but occasionally can be ordered and constitute an Ordinal Data Type. Hierarchical Clustering is a special type of Clustering which learns groups with nested structures; the labels then define a Tree.

Examples of clustering: grouping of documents into common themes, grouping vendors by common product offering to identify potential competition opportunities, grouping sensors by their typical measurement patterns to identify potential malfunctions, identifying common segments of behavior in pedestrians to identify potential bottlenecks in road design, assigning phylogenetic strains in a population of microorganisms.

Forecasting. A function which takes as input a Sequence whose elements are of some Data Type, and outputs either a value of that Data Type or another Sequence whose elements are of that Data Type.

Example 5: A water consumption system collects historical water usage data, measured on an hourly basis, for the past 90 days, and uses a model to predict water usage for the next 24 hours. The computation in the model takes as input a Sequence of Numeric Data Type, whose values are the historical hourly water consumption rate, and outputs a Sequence of Numeric Data Type, whose values are the predicted hourly water consumption rate for the next 24 hours.

Example 6: An elevator control system collects historical passenger traffic in each elevator, measured in 5 second intervals, over the past 24 hours. When an elevator call button is pressed, the system uses a Predictive Model to forecast how many passengers will board on that floor. This computation takes as input a Sequence of Numeric Data Type, whose values are the number of passengers in the elevator in each interval, and outputs a Numeric Data Type, whose value is the predicted passenger load when the elevator stops at the called floor.

Other examples: estimating the number of riders on the subway over the next 30 days or predicting the demand for emergency services over a holiday weekend. Matching. A function which takes as input two Collections, and pairs each element of the first Collection with an element from the second Collection. Matching uses an algorithm to produce the best possible match, often in the presence of constraints.

Note. Matching models are not to be confused with Ranking models (see below). The former produces pairs between two sets of inputs, whereas the latter produces an ordered list selected from one set of inputs. Matching models may use Ranking models internally to generate the best candidates to pair with another item.

Example 7: A rostering system assigns workers based on the number of workstations available for a given work shift. Due to project requirements, there are preferences to situate some workers together and others apart. In addition, workplace accommodations result in requirements for some workers to only use specific workstations. Some workers are on vacation, and some workstations are malfunctioning and unavailable for the current work shift. The inputs to the rostering system are two Collections of Nominal Data Type, the first being the set of available workers and the second being the set of available workstations. The output of the rostering system is a Collection of pairs, each pair consisting of a worker assigned to a workstation.

Other examples: assigning students to seats in schools, scheduling the dispatch of emergency vehicles to incident locations, or residents to available apartments.

Ranking. A function which takes as input a Collection and another input and produces a Sequence whose elements are selected from the Collection.

Note: Ranking models are not to be confused with Matching models. The former produces an ordered list selected from one set of inputs, whereas the latter produces pairs between two sets of inputs. Ranking models may be used inside Matching models to generate the best candidates to pair with another item.

Note: Ranking models are not to be confused with Sampling models (see below). The former produces an ordered list of outputs, whereas the latter produces an unordered set of outputs. If the order matters, it is Ranking; otherwise, it is Sampling. Examples 10 and 11 illustrate the difference.

Example 8: A traffic plate reading system is connected to a database of registered vehicle license plate numbers. When given an image of a license plate, the system performs OCR and returns the top ten matches to known plate numbers in the database. The inputs are of type Image and Collection of Nominal Data Type, representing the picture of the license plate and the database of license plate numbers respectively. The outputs are of type Sequence of Nominal Data Type, being a sorted list of license plate numbers ranked by the likelihood of a correct match within the system.

Example 9: A chatbot on a website invites visitors to describe the issue they are having and recommends several follow-up actions that may help resolve their issue. A component of the chatbot system employs natural language understanding to infer the intent of the free-form text input by the user, and returns the top three responses from a predefined database of permitted responses that best match the intent of the input. This component takes as inputs Text and a Collection, the latter being the list of permitted responses, and produces as output a Sequence that is the top three responses that best match the inferred intent.

Example 10: An inspector needs to inspect apartments in a given neighborhood, but has limited time and can only afford to inspect a few apartments. The system that produces the roster takes as input a Collection of all apartments to be inspected, and produces a Sequence of apartments ranked by the likelihood of having issues to be fixed. Contrast with Example 11 below.

Other examples: recommending products to purchase from a vendor catalog based on a description of the desired item, returning best matches from a database of known suspects matching an image from a security camera. Sampling. A function which takes as input a Collection, and produces a Collection whose elements are selected from the Collection.

Note: Sampling models are not to be confused with Ranking models. The former produces unordered set of outputs, whereas the latter produces an ordered list of outputs. If the order matters, it is Ranking; otherwise, it is Sampling. Examples 10 and 11 illustrate the difference.

Example 11: An inspector needs to inspect all apartments in a given neighborhood over the course of a year. At the beginning of each week, they produce a roster of apartments to be inspected. The order of inspection of the apartments does not matter; the only requirement that all the selected apartments must be inspected by the end of the week. The system that produces the roster takes as input a Collection of all apartments that have yet to be inspected, and produces a Collection of apartments to be inspected in a given week. Contrast with Example 10 above.

Scoring. A function whose output Data Type is Numerical.

Note: Scoring models are not to be confused with Classification models, whose outputs are strictly of Nominal or Ordinal Data Type. Scoring models output a number, whereas

Classification models output a category. A Scoring model may be used inside a Classification model to generate a number; the Classification model can then take the numerical output, compare it to a threshold, and finally determine the category to assign based on whether the threshold was met. Examples 3 and 13 illustrate the difference.

Example 12: A system contains a model which computes a score for measuring an employee's productivity. The productivity score is used directly to determine the discretionary bonus for that employee.

Example 13: A traffic monitoring system contains two components: i) a computer vision component which identifies moving vehicles on a road and labels each vehicle in every video frame; and ii) a component which, given relevant video frames for each vehicle, computes the speed of that vehicle. The output type of the second component is a Numeric Data Type, which is the speed of the vehicle. The output of the second component is also the final output from the entire system.

Data Transformation. is any computation which does not fit into any of the other categories.

Examples: Real-time captioning systems which take video input and output closed captions in the form of text, meeting transcription systems which take audio recordings of meetings and output textual minutes, illustration systems which take text narratives and generate a sequence of images to accompany the text.

Autonomy

This appendix provides definitions and examples of the levels of autonomy in algorithmic tools. Agencies should use this appendix to help them answer "What is the level of autonomy for this tool?" for an algorithmic tool that has been identified for reporting.

Autonomy describes the degree of control that an algorithmic tool has to inform or make decisions on behalf of city staff. The level of Autonomy may depend on the context of both the relevant business process and the capabilities of the algorithmic tool.¹²

Informative. The tool does not act on its output. The employee uses or disregards the tool's outputs at will.

Example. A score that captures a resident's need of supportive services is provided to a case manager as they determine eligibility for a new program. This score is one of several pieces of information the case manager may choose to use (or not use) to determine whether a resident is eligible.

¹² OECD Framework for the Classification of AI Systems, OECD Digital Economy Papers No. 323, Feb. 22, 2022, OECD Publishing: Paris, p. 53 "Action autonomy level".
Version effective March 2026

Example. An agency deploys a customer service chatbot to receive and validate materials for an application. City staff must manually review all information received from the chatbot behind the scenes. Supervised. The tool can act on its outputs, but requires explicit permission from the employee to do so.

Example. A score that captures a resident's need of supportive services is provided to a case manager as they determine eligibility for a new program. The tool provides the case manager the option to lock in the computed score or substitute it with a rating of their choice at will before making determinations about who can enter the program.

Example. A tool produces a recommended inspection route that must be approved by a manager before being sent to an inspector.

Monitored. The tool acts on its outputs unless the employee overrides the action.

Example. An algorithmic tool provides a score of a resident's need for supportive services. A manager is able to reject the tool's computed rating and substitute it with a rating of their choice at will, but must document the reason for overriding the tool's score.

Example. A tool creates a light signaling pattern based on observed traffic patterns, which is automatically implemented but can be interrupted or changed by an employee at any time.

Fully autonomous. The tool acts on its outputs without any employee intervention.

Example. An algorithmic tool provides a score of a resident's need for supportive services and, if the resident meets a threshold, automatically enrolls them in a program.

Example. An agency deploys a customer service chatbot for automatically approving applications for permits. The chatbot receives documents supporting the permit application and issues a determination automatically, without review by city staff.

Example. An agency deploys a customer service chatbot for answering questions about a permitting process. It is able to do so without any input from city staff.