Office of Technology & Innovation

Agency Compliance Reporting of Algorithmic Tools

CALENDAR YEAR 2024

Overview

This summary communicates the results of New York City's fifth annual process for reporting on algorithmic tools. Under Local Law 35 ("LL 35")[1], the city continues its commitment to provide the public with a transparent view of these applications of agency data and technology. Pursuant to mayoral Executive Order 3 of 2022 ("EO 3")[2], the city's Office of Technology & Innovation ("OTI") manages this process, providing guidance to agencies and ensuring that agency materials are prepared for the public.

This report is also available as an open data set at https://data.cityofnewyork.us/d/jaw4-yuem.

Key Changes for 2024

Changes to the report have been made based on research done for the NYC AI Action Plan, which was published in October 2023.

There are six new and four amended questions meant to provide additional context around ownership of a tool within the agency and their purpose, impact, and computational type. A full list of reportable elements is available in the Appendix.

OTI updated its guidance to clarify that tools that create, transform, or edit information provided to the public should be considered to have a material public impact.

Twelve agencies, indicated in the Summary of Agency Reports, participated in the reporting process for the first time this year.

Summary of Agency Reports

The following table (continued over the next two pages) summarizes reporting from city agencies for 2024.

Agency	Tools Reported
Administration for Children's Services	4
Board of Corrections *	0
Board of Standards and Appeals *	0
Borough President of Manhattan *	0

[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

Agency	Tools Reported
Business Integrity Commission	0
Campaign Finance Board *	0
Civil Service Commission *	0
Civilian Complaint Review Board	0
Commission on Human Rights	0
Commission on Racial Equity *	1
Commission to Combat Police Corruption *	0
Conflicts of Interest Board	0
Department for the Aging	0
Department of Buildings	0
Department of City Planning	0
Department of Citywide Administrative Services	0
Department of Consumer and Worker Protection	0
Department of Correction	0
Department of Cultural Affairs	0
Department of Design and Construction	0
Department of Environmental Protection	1
Department of Finance	0
Department of Health and Mental Hygiene	16
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	3
Department of Transportation	1
Department of Veterans' Services	0
Department of Youth and Community Development	0
Economic Development Corporation	0
Equal Employment Practices Commission *	0
Financial Information Service Agency *	0
Fire Department	3
Landmarks Preservation Commission	0
Law Department	0
Mayor's Office	8
Mayor's Office of Contract Services	0
Mayor's Office of Criminal Justice **	0

Agency	Tools Reported
New York Police Department	4
NYC Emergency Management	0
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	5
Public Advocate *	0
Public Design Commission	0
School Construction Authority	0
Taxi & Limousine Commission	0
Total	55

* New to reporting this year.

** In 2024, the Mayor's Office of Criminal Justice became an independent agency. Previously, their reporting was included as part of the Mayor's Office.

Table of Contents

Administration for Children's Services (ACS) 1
Commission on Racial Equity (CORE) 6
Department of Environmental Protection (DEP) 7
Department of Health and Mental Hygiene (DOHMH) 8
Department of Investigation (DOI) 22
Department of Social Services (DSS) 23
Department of Transportation (DOT) 27
Fire Department (FDNY) 28
Mayor's Office (MO) 32
New York Police Department (NYPD) 40
NYC Public Schools (NYCPS) 45
Office of Chief Medical Examiner (OCME) 53
Office of Technology and Innovation (OTI) 54
Appendix 59

Administration for Children's Services (ACS)

Accelerated Safety Analysis Protocol Tool

Department: Division of Policy, Planning, and Analysis

First Used: May 2018
Updated in 2024: No
Computation Type: Ranking
Purpose Type: Performance evaluation
Population Type: Individuals
Identifying Information: Yes

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.

Data Analyzed

Training Data	ACS trained the model on ACS historic administrative data about closed investigations from April 2014 to April 2016. The training set included about 142,026 observations. The model was tested on closed investigations from April 2016 to April 2017 with 53,477 observations.
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.

Vendor Information

Vendor Name: N/A

Caseloads Projection

Department: Division of Policy, Planning, and Analysis

First Used: July 2024

Computation Type: Forecasting

Population Type: Geographic space

Updated in 2024: Tool created in CY2024

Purpose Type: Resource allocation

Identifying Information: No

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 was 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.

Data Analyzed

Training Data	The model was trained on caseloads from January 2021 to July 2023 and tested on caseloads from July 2023 to July 2024.
Input Data	Predictions are based on administrative data on investigations and Family Service Units' involvement were 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 was 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.

Vendor Information

Vendor Name: N/A

Housing Prioritization

Department: Division of Policy, Planning, and Analysis

First Used: April 2023
Computation Type: Ranking
Population Type: Individuals
Updated in 2024: No
Purpose Type: Resource allocation
Identifying Information: No

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

Training Data	ACS trained the model on ACS historic administrative data regarding preventive services started between 2014 and 2020. An 80/20 split of data to train on 80 percent and test on 20 percent ensuring that no family appears in both sets. The training set contains 140,242 observations between January 2014 and December 2020. The test set consisted of 34,508 observations between January 2014 and December 2020.
Input Data	Predictions are based on administrative data about prior and current child welfare involvement at the start of a case. This includes 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.

Vendor Information

Vendor Name: N/A

Prevention Score Card

Department: Division of Policy, Planning, and Analysis

First Used: September 2021
Computation Type: Ranking
Population Type: Individuals
Updated in 2024: No
Purpose Type: Performance evaluation
Identifying Information: Yes

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.

Data Analyzed

Training Data	ACS trained the model on ACS historic administrative data about closed investigations from July 2009 to June 2016. Training set included about 158,787 observations. The model was tested on closed investigations from July 2016 to June 2018 with 46,969 observations.
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.

Vendor Information

Vendor Name: N/A

Commission on Racial Equity (CORE)

Adobe Express

Department: Communications

First Used: July 2024

Updated in 2024: Tool created in CY2024

Computation Type: Data generation

Purpose Type: Information presentation

Population Type: Individuals (General public)

Identifying Information: No

Tool Description

Generates images in response to a verbal prompt.

Tool Purpose

Used to generate visual content to be adapted for agency use in social media.

Data Analyzed

Training Data	Training data is proprietary to Adobe Creative Cloud.
Input Data	Description of image sought.
Output Data	Image relevant to social media messaging.

Vendor Information

Vendor Name: N/A

Department of Environmental Protection (DEP)

Idling Complaints Program

Department: Bureau of Environmental Compliance

First Used: August 2022
Updated in 2024: No
Computation Type: Classification
Purpose Type: Performance evaluation
Population Type: Individuals
Identifying Information: No

Tool Description

A contractor helped create an AI tool that analyzes the audio and visual aspects of pictures and videos submitted by citizens of alleged car idling complaint occurrences that are in violation of New York City air pollution laws.

Tool Purpose

The analysis from the tool makes a recommendation to staff reviewers whether the submitted evidence support an occurrence of car idling in violation of New York City laws. The tool also provides a level of confidence in its recommendation. The tool does not make the review decision in the Idling Complaints system. It is still entirely up to the staff to decide whether to take the tool's recommendation or not.

Data Analyzed

Training Data	Videos and pictures of cars idling submitted by citizens, along with staff decisions on whether the picture/video constituted as an idling violation.
Input Data	Videos and pictures submitted by citizens through our web portal.
Output Data	Recommendation, confidence level, description of its decision from the tool.

Vendor Information

Vendor Name: Acuvate

Acuvate developed the AI tool that performs the automated analysis of the submitted evidence.

Department of Health and Mental Hygiene (DOHMH)

Bowtie2

Department: Disease Control – Public Health Laboratory

First Used: June 2022
Updated in 2024: No
Computation Type: Matching
Purpose Type: Data management
Population Type: Individuals; Biological sample
Identifying Information: No

Tool Description

Aligns sequencing data to a reference sequence. Bowtie2 aligns sequencing data to a reference 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

Training Data	N/A

Input Data	Sequence reads (fastq) for single or paired-end runs (sequence reads can be considered strings).
Output Data	Aligned reads in SAM format.

Vendor Information

Vendor Name: N/A

Burrows-Wheeler Aligner (BWA)

Department: Disease Control – Public Health Laboratory

First Used: July 2017
Computation Type: Matching
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

Aligns sequencing data to a reference sequence.

Tool Purpose

Burrows-Wheeler Aligner (BWA) is aligning 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

Training Data	N/A
Input Data	Sequence reads (fastq) for single or paired-end runs (sequence reads can be considered strings).
Output Data	Aligned reads in SAM format.

Vendor Information

Vendor Name: N/A

ChoiceMaker (CM)

Department: Disease Control – Bureau of Investigations

First Used: June 2003

Computation Type: Matching

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

Updated in 2024: No

Purpose Type: Data management

Identifying Information: Yes

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

Training Data	The CM model was trained on human decisions.
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: HLN Consulting

A vendor was involved in the development of the program initially. CM is now available as an open-source program. The DOHMH implementation is maintained by HLN Consulting.

Genome Analysis Toolkit (GATK)

Department: Disease Control - Public Health Laboratory

First Used: October 2017
Computation Type: Classification
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

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

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

Data Analyzed

Training Data Sets of known variant sites.
Input Data Fasta, uBam, SAM/BAM/CRAM, VCF.
Output Data BAM, TXT, VCF.

Vendor Information

Vendor Name: N/A

Guppy

Department: Disease Control - Public Health Laboratory

First Used: June 2020
Computation Type: Classification
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

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 (ONT) 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

Training Data	The default models within Guppy are trained on a mixture of native and amplified DNA/RNA, from multiple organisms including plant, animal, bacterial, and viral genomes.
Input Data	DNA/RNA strand passing through the nanopore. Raw data is stored as .fast5 files.
Output Data	.fast5, fastq, or BAM files.

Vendor Information

Vendor Name: Oxford Nanopore Technologies

Developed and maintains the tool.

Immunization Calculation Engine (ICE)

Department: Disease Control – Bureau of Immunization

First Used: 1997

Computation Type: Forecasting

Population Type: Individuals (Anyone who needs a vaccine)

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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 available through **https://cdsframework.atlassian.net/wiki/spaces/ICE/overview**.

Tool Purpose

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

Data Analyzed

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: HLN Consulting

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: Disease Control – Bureau of Communicable Disease

First Used: November 2016

Computation Type: Scoring

Population Type: Individuals (Public who dine at NYC restaurants and are Yelp users and NYC restaurants)

Updated in 2024: No

Purpose Type: Triage

Identifying Information: Yes

Tool Description

Restaurant-associated foodborne disease outbreaks are often identified through complaints received via NYC311 non-emergency information system; however not all individuals report to NYC311. 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 NYC311 to improve efficiency of outbreak detection.

Data Analyzed

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; 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: Columbia University

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
Computation Type: Clustering
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

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

Training Data	N/A
Input Data	FASTA, NEXUS, CLUSTALW, PHYLIP.
Output Data	Readable report, maximum likelihood tree in NEWICK format, log file for entire run.

Vendor Information

Vendor Name: N/A

kSNP4

Department: Disease Control - Public Health Laboratory

First Used: March 2022
Computation Type: Clustering
Population Type: Individuals; Biological sample
Updated in 2024: Yes
Purpose Type: Data management
Identifying Information: No

Tool Description

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

Tool Purpose

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

Update Description

Previously used kSNP3, updated this year to kSNP4.

Data Analyzed

Training Data	N/A
Input Data	Fasta.
Output Data	ML tree in NEWICK format, log & configuration files, Fasta file.

Vendor Information

Vendor Name: N/A

Multiple Alignment using Fast Fourier Transform (MAFFT)

Department: Disease Control - Public Health Laboratory

First Used: January 2021
Computation Type: Matching
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

Aligns multiple sequencing data.

Tool Purpose

Multiple Alignment using Fast Fourier Transform (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

Training Data	N/A
Input Data	Sequences can be in GCG, FASTA, EMBL (nucleotide only), GenBank, PIR, NBRF, PHYLIP or UniProtKB/Swiss-Prot (protein only) format.
Output Data	FASTA or Clustalw.

Vendor Information

Vendor Name: N/A

Minimap2

Department: Disease Control - Public Health Laboratory

First Used: May 2020
Computation Type: Matching
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

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

Training Data	N/A
Input Data	Sequence reads (fastq) for single or paired-end runs (sequence reads can be considered strings).
Output Data	Aligned reads in SAM format.

Vendor Information

Vendor Name: N/A

Pangolin

Department: Disease Control - Public Health Laboratory

First Used: July 2021
Computation Type: Clustering
Population Type: Individuals; Biological sample
Updated in 2024: Yes
Purpose Type: Data management
Identifying Information: No

Tool Description

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.

Update Description

Pangolin is updated as new sequences are published, and as the virus evolves. Currently, it is version 4.3.

Data Analyzed

Training Data	Trained on a dataset of genomes that have been designated to Pango lineages using whole genome information.

Input Data	Fasta files.
Output Data	.csv file with taxon name and lineage assigned.

Vendor Information

Vendor Name: N/A

PHYLOViZ

Department: Disease Control - Public Health Laboratory

First Used: October 2017
Computation Type: Clustering
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

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

Used to generate the minimum spanning tree relationships.

Data Analyzed

Training Data	N/A
Input Data	txt, NEWICK, FASTA.
Output Data	Minimum spanning tree.

Vendor Information

Vendor Name: N/A

PulseNet 2.0

Department: Disease Control - Public Health Laboratory

First Used: September 2017
Updated in 2024: Yes

Computation Type: Clustering

Purpose Type: Data management

Population Type: Individuals; Biological sample

Identifying Information: No

Tool Description

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.

Update Description

This tool was put on the cloud and accessed through CDC Secure Access Management Services site.

Data Analyzed

Training Data	N/A
Input Data	Fastq.
Output Data	txt, Excel.

Vendor Information

Vendor Name: Centers for Disease Control and Prevention

Developed and maintains the tool.

Spades

Department: Disease Control - Public Health Laboratory

First Used: October 2017
Computation Type: Matching
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

Tool Description

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

Tool Purpose

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

Data Analyzed

Training Data	N/A
Input Data	Fastq.
Output Data	Fastas and other files for corrected reads; scaffolds, contigs, paths in GFA format; fastg assembly graph.

Vendor Information

Vendor Name: N/A

Vsearch

Department: Disease Control - Public Health Laboratory

First Used: June 2022
Computation Type: Matching
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: No

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

Training Data	N/A
Input Data	Sequence reads (fastq, Fasta) for single or paired-end runs (sequence reads can be considered strings).
Output Data	FASTA, FASTQ, tables, alignments, SAM.

Vendor Information

Vendor Name: N/A

Department of Investigation (DOI)

Facial Recognition Technology

Department: Not specified

First Used: March 2019
Computation Type: Matching
Population Type: Individuals
Updated in 2024: Yes
Purpose Type: Not specified
Identifying Information: Yes

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 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

General system updates by vendor to improve quality of search results and to fix bugs.

Data Analyzed

Training Data	Vendor uses publicly available open-source media data.
Input Data	Images.
Output Data	Images.

Vendor Information

Vendor Name: Not disclosable

Out-of-the-box products. 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: Not specified

First Used: June 2012

Computation Type: Scoring

Population Type: Individuals (Households seeking Homebase assistance)

Updated in 2024: No

Purpose Type: Resource allocation

Identifying Information: Yes

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 DHS shelter system. Since NYC 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

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.

Vendor Information

Vendor Name: Multiple researchers

DHS contracted with researchers to evaluate years of Homebase administrative data to develop a risk assessment. The DSS research team then led an updated analysis that led to tool revisions. The published research papers are listed below:

https://ajph.aphapublications.org/doi/10.2105/AJPH.2013.301468

https://www.journals.uchicago.edu/doi/abs/10.1086/686466?mobileUi=0&journalCode=ssr

https://www.tandfonline.com/doi/abs/10.1080/10511482.2022.2077801

NYC Enterprise Data Solutions Service (EDS)

Department: Public Engagement Unit

First Used: March 2024
Computation Type: Matching
Population Type: Individuals
Updated in 2024: Tool created in CY2024
Purpose Type: Data management
Identifying Information: Yes

Tool Description

The NYC Enterprise Data Solutions Service (EDS) automatically matches and links person records from multiple source systems, regardless of whether they share a common identifier such as a social security number. EDS supports cross-agency data integration and interoperability, particularly across health and human service agencies. EDS works by processing business rules that standardize name and geocode address data before applying a series of deterministic and probabilistic matching rules that generate detailed or summarized reports of record linkages.

Tool Purpose

In 2024, the Mayor's Public Engagement Unit (PEU) moved the operations of its Rent Freeze team (focused on enrollment in Senior Citizen Rent Increase Exemption, Disability Rent Increase Exemption, Senior Citizen Homeowners' Exemption, and Disabled Homeowners' Exemption programs) to a new case/client management system within our custom Salesforce instance. Prior to migrating data from the prior system (based in our EveryAction tool), PEU needed to identify duplicate clients within EveryAction and between EveryAction and Salesforce. The primary reason for duplication is that we use Salesforce to support a number of other teams that might have interacted with Rent Freeze clients as well. We used the EDS tool developed by our partners at NYC Opportunity and maintained by OTI to identify client records that were highly likely to be the same client. PEU developed its own process to combine outreach and case history information for the records identified as related to the same client to create a single record in PEU's Salesforce system across programs. Using the EDS tool helped us provide better service to our clients by providing our specialists with richer information about the client's interaction with the Rent Freeze, Tenant Helpline, Tenant Support Unit, and GetCovered teams.

Data Analyzed

Training Data	N/A
Input Data	Client contact information, including name, phone numbers, email addresses, home addresses.
Output Data	The same client information but with common identifier numbers.

Vendor Information

Vendor Name: N/A

SmartVAN / TargetSmart

Department: Public Engagement Unit

First Used: November 2019
Computation Type: Matching
Population Type: Individuals
Updated in 2024: Yes
Purpose Type: Data management
Identifying Information: Yes

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.

Tool Purpose

In 2023, PEU had the TargetSmart data within SmartVAN on a number of projects. PEU frequently uses the data to create lists of residents who live within certain zip codes that PEU wants to target for outreach. For example, PEU created lists using TargetSmart data to conduct door-knocking and phone banking outreach to New Yorkers identified as potentially eligible for the DOF Rent Freeze program based on TargetSmart data. In cases like these, TargetSmart's determination of who lives in which zip codes as well as estimated income affects whether New Yorkers receive PEU outreach. Additionally, the algorithm that TargetSmart uses to match phone numbers to individuals impacts the type of outreach that New Yorkers receive.

Update Description

We get regular updated lists of New Yorkers from the vendor directly into the EveryAction platform we use for outreach.

Data Analyzed

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: EveryAction and TargetSmart

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.

Department of Transportation (DOT)

Midtown in Motion / Adaptive Control Decision Support System

Department: Traffic Operations

First Used: July 2011

Computation Type: Classification

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

Updated in 2024: Yes

Purpose Type: Performance evaluation

Identifying Information: No

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.

Update Description

Ongoing maintenance.

Data Analyzed

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: KLD Engineering, P.C.

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

Fire Department (FDNY)

Emergency Medical Services (EMS) Hospital Suggestion Algorithm

Department: Management Analysis and Planning

First Used: March 2021

Computation Type: Ranking

Population Type: Individuals (Patients); Geographic space; Group, organization, or business

Updated in 2024: No

Purpose Type: Resource allocation

Identifying Information: No

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

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.

Vendor Information

Vendor Name: N/A

Emergency Medical Services (EMS) Unit Suggestion Algorithm

Department: Management Analysis and Planning

First Used: March 2007

Updated in 2024: No

Computation Type: Ranking

Purpose Type: Resource allocation

Population Type: Geographic space; Individuals; Group, organization, or business

Identifying Information: No

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 EMS computer-aided design (CAD) 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. The file is generated and implemented as a 24/7 source file, meaning, the recommended search order is not currently varying by time of day. The department is intending to implement time-of day search orders in the near future.

Data Analyzed

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.
Output Data	The algorithm outputs a recommended EMS unit for dispatch.

Vendor Information

Vendor Name: Deccan International

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.

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

Department: Management Analysis and Planning

First Used: November 2019

Computation Type: Scoring

Population Type: Property; Individuals (Civilian fire injuries/fatalities); Group, organization, or business

Updated in 2024: Yes

Purpose Type: Risk management

Identifying Information: No

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 NYC311 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.

Update Description

The fire ignition and injury/death models are recalculated monthly with fresh training data to create updated variable weights.

Data Analyzed

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, NYC311 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).

Vendor Information

Vendor Name: N/A

Mayor's Office (MO)

Adobe Photoshop

Department: Mayor's Office of Media and Entertainment

First Used: January 2024

Computation Type: Data generation

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

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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

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: Adobe

Adobe regularly updates the Photoshop software.

Adobe Premiere Pro

Department: Mayor's Office of Media and Entertainment

First Used: 2021

Updated in 2024: No

Computation Type: Data generation

Purpose Type: Information presentation

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

Identifying Information: Yes

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

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: Adobe

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 2024: Tool created in CY2024

Computation Type: Data generation

Purpose Type: Information presentation

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

Identifying Information: Yes

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

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: Microsoft

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

Computation Type: Data generation

Population Type: 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 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

Tool Description

The tool uses AI-enabled automatic speech recognition to create closed captions of live television content.

Tool Purpose

The Mayor's Office of Media and Entertainment is using 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

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: AppTek

AppTek provides support for the Omni 300 closed captioning system.

ChatGPT

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

First Used: March 2024
Computation Type: Data generation
Population Type: Individuals
Updated in 2024: No
Purpose Type: Information presentation
Identifying Information: No

Tool Description

ChatGPT is an advanced AI language model that can understand and generate human-like text based on the input it receives. It can assist with a wide range of tasks, including answering questions, providing recommendations, and engaging in meaningful conversations.

Tool Purpose

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 and assist in creating potential job interview questions based on content in the job description.

Data Analyzed

Training Data	Training data is proprietary to OpenAI.
Input Data	Text prompts were provided to ChatGPT.
Output Data	ChatGPT provided text responses to the text prompts.

Vendor Information

Vendor Name: N/A

Methodology for Poll Site Language Assistance

Department: Civic Engagement Commission

First Used: November 2020
Computation Type: Ranking
Population Type: Individuals
Updated in 2024: Yes
Purpose Type: Resource allocation
Identifying Information: No

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 2024 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

We are now using 2017-2021 data from the American Community Survey 5-year data. The differences between new and previous data are not statistically significant, therefore the distribution of services will not be affected based on these differences.

Data Analyzed

Training Data	N/A
Input Data	For citywide estimates, this methodology uses current data from the American Community Survey 2017-2021 5-year estimates. This methodology also uses the American Community Survey Census Tract 2017-2021 5-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.

Vendor Information

Vendor Name: N/A

StratifySelect

Department: Civic Engagement Commission

First Used: November 2022

Computation Type: Matching

Population Type: Individuals (New York City residents (all boroughs))

Updated in 2024: No

Purpose Type: Resource allocation

Identifying Information: Yes

Tool Description

StratifySelect is used to select a group of people from a pool of applicants such that the selected group matches target demographics and is as randomized as possible given the need to match targeted demographics. Traditional stratified sampling can create cases in which some individuals have a near-zero chance of being selected. This method uses a new technique of explicitly computing a maximally fair output distribution and then sampling from that distribution to select the final panel, achieving a fairer

distribution of probabilities per applicant while maintaining fidelity to the demographics of the borough. The tool compares applicant demographic data to the data from the American Community Survey to achieve a representative sample. See this paper in Nature for details on the methodology: **https://www.nature.com/articles/s41586-021-03788-6**

Tool Purpose

The group selected by StratifySelect will be invited to participate in the CEC's Borough Assemblies. Each group contains a prioritized list including "backup options" in case some of the invitees decline or are unable to participate.

Data Analyzed

Training Data	N/A
Input Data	Fully anonymized data from interested people (age, gender, race, and Hispanic identity, borough, zip code) is used as input. We emphasize that no identifying data is stored, shared and/or transmitted during the entire process.
Output Data	The output is a subset of the same data such that the output group is both randomly selected and representative of each borough in these categories, as defined by public census data.

Vendor Information

Vendor Name: N/A

Zoom

Department: Mayor's Office of Media and Entertainment

First Used: 2020

Computation Type: Data generation

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

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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. We use Zoom's auto transcript function and captioning function. (Note: We provide American Sign Language and human-typed Communication Access Realtime Translation services as a reasonable accommodation upon request.) If we publish a transcript after the Zoom meeting, a human reviews and corrects it.

Data Analyzed

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: Zoom

Regular updates to the application.

New York Police Department (NYPD)

Evolv Express Weapons Detection System

Department: Not specified

First Used: July 2024

Updated in 2024: Tool created in CY2024

Computation Type: Classification

Purpose Type: Data management

Population Type: Individuals (General population); Property; Geographic space

Identifying Information: Yes

Tool Description

Electromagnetic weapons detection system.

Tool Purpose

Electromagnetic weapons detection devices emit ultra-low frequency, electromagnetic pulses (similar to those used in retail loss prevention) that pass through objects moving through the system. Sensors process the relayed information and the system uses this data to determine if it detects a potential firearm. The system is equipped with video cameras that are part of a real-time image-aided alert system that will indicate the presence of a firearm to monitoring personnel.

Data Analyzed

Training Data	Training data is proprietary to the vendor.
Input Data	Individuals walk through two towers which emit ultra-low frequency, electromagnetic pulses that pass through objects on the individual's person. Sensors process the information generated and the system uses this data to determine if it detects a potential firearm.
Output Data	If a potential firearm is detected, the system will capture a still image and an approximately three-second video of the individual moving through the system. The system will alert monitoring personnel that a potential firearm has been detected and wirelessly transmit the still image and video to a tablet being monitored by personnel. A cube will appear on both the still image and video clip, indicating the location of the potential firearm being worn or carried by the individual. The location of a cube is discerned by the system based on the electromagnetic data processed by the system sensors.

Vendor Information

Vendor Name: EVOLV

Software developed and maintained by vendor.

Facial Recognition Technology

Department: Not specified

First Used: October 2011
Computation Type: Matching
Population Type: Individuals (General population)
Updated in 2024: No
Purpose Type: Data management
Identifying Information: Yes

Tool Description

Tool 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.

Data Analyzed

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: Dataworks

Software developed and maintained by vendor.

Patternizr

Department: Not specified

First Used: December 2016

Computation Type: Matching

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

Updated in 2024: Yes

Purpose Type: Data management

Identifying Information: No

Tool Description

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

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.

Vendor Information

Vendor Name: N/A

ShotSpotter

Department: Not specified

First Used: March 2015
Computation Type: Classification
Population Type: Geographic space
Updated in 2024: Yes
Purpose Type: Data management
Identifying Information: No

Tool Description

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

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: Shotspotter

Software developed and maintained by vendor.

NYC Public Schools (NYCPS)

Algebra Teaching Assistant

Department: Division of Instructional and Information Technology

First Used: May 2023

Computation Type: Data generation

Population Type: Individuals (Teachers, students)

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

Tool Description

The Algebra Teaching Assistant uses the Division of Instructional and Information Technology (DIIT) AI platform to accesses 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

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: Microsoft

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
Computation Type: Scoring
Population Type: Individuals (Teachers)
Updated in 2024: No
Purpose Type: Performance evaluation
Identifying Information: No

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

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: Education Analytics

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

Computation Type: Scoring

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

Updated in 2024: No

Purpose Type: Performance evaluation

Identifying Information: No

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

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.

Vendor Information

Vendor Name: N/A

Eureka! Chatbot

Department: Division of Instructional and Information Technology

First Used: August 2023

Computation Type: Data generation

Population Type: Individuals (NYCPS staff and parents)

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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

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: Nagarro and Microsoft

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

MySchools - Match

Department: Division of Instructional and Information Technology

First Used: August 2018
Computation Type: Matching
Population Type: Individuals (Students)
Updated in 2024: No
Purpose Type: Resource allocation
Identifying Information: Yes

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. Perhaps most common is its use 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

Training Data	The algorithm was already widely recognized for its advantages prior to adoption in New York City. NYCPS consulted with a team of researchers at Massachusetts Institute of Technology who had been closely involved in its initial creation when we adopted it.
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: Blenderbox

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 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

Computation Type: Scoring

Population Type: Individuals (Students)

Updated in 2024: Tool created in CY2024

Purpose Type: Information presentation

Identifying Information: Yes

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.

Data Analyzed

Training Data	Model was developed by researchers affiliated with the Massachusetts Institute of Technology (MIT) and trained on information about NYC 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: Researchers from MIT

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

Computation Type: Data generation

Population Type: Individuals (Teachers, students)

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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

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: Microsoft

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
Computation Type: Not specified
Population Type: Individuals; Biological sample
Updated in 2024: No
Purpose Type: Data management
Identifying Information: Yes

Tool Description

STRmix™ combines sophisticated biological modeling 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. It grades them against the evidential sample, finding the combinations that best explain the profile. A range of likelihood ratio options are provided for subsequent comparisons to reference profiles. Using a Markov Chain Monte Carlo engine, 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.

Tool Purpose

STRMix is a probabilistic genotyping tool that is used to analyze mixtures of DNA profiles to help associate the crime scene evidence to potential victims or suspects of crimes.

Data Analyzed

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: NicheVision Forensics, LLC

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)

Google Translate

Department: Web Operations

First Used: 2013

Computation Type: Data generation

Population Type: Individuals (Users of nyc.gov)

Updated in 2024: No

Purpose Type: Information presentation

Identifying Information: Yes

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

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: N/A

MyCity Chatbot

Department: Applications

First Used: September 2023

Computation Type: Data generation

Population Type: Individuals; Group, organization, or business

Updated in 2024: Yes

Purpose Type: Information presentation

Identifying Information: No

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.

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 MyCity Basics. 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, 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 (LLM).

Update Description

Security updates and LLM upgraded to use ChatGPT 4-o.

Data Analyzed

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 based on information from Business Services and MyCity Basics.

Vendor Information

Vendor Name: Microsoft, EY

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

NYC311 AI Voice Pilot

Department: NYC311

First Used: March 2024
Computation Type: Data generation
Population Type: Individuals
Updated in 2024: Tool created in CY2024
Purpose Type: Information presentation
Identifying Information: No

Tool Description

NYC311 AI Voice pilot is a large language model (LLM)-powered voice call solution that provided information and access to services for residents, businesses, and visitors.

Tool Purpose

The NYC311 AI Voice was a pilot program that enabled NYC311 to test a generative AI voice application with NYC311 customers, for 21 days. On a daily basis, customers can reach the NYC311 call center, by dialing 311, 212-NEW-YORK, or 211. The pilot was kept on a small scale, available only to customers who dialed 211. The NYC311 AI Voice provided information to residents, businesses, and visitors on a wide range of inquiries, as well as providing updates on city programs, events, and notifications using NYC311's content.

Data Analyzed

Training Data	Training data is proprietary to the vendor.
Input Data	The customer made voice inquiries when contacting NYC311.
Output Data	NYC311 AI Voice provided the customer with information/responses from the NYC311 Content Application Programming Interface.

Vendor Information

Vendor Name: Microsoft, Nuance

Microsoft provided an LLM-powered solution to aid in handling customers who call NYC311 for information and services. The focus was handling voice interactions.

Omnichannel Language Translation

Department: NYC311

First Used: January 2024

Updated in 2024: Tool created in CY2024

Computation Type: Data generation

Purpose Type: Information presentation

Population Type: Individuals (Customers who contact NYC311 via text/SMS)

Identifying Information: No

Tool Description

The Omnichannel Language Translation tool delivers multi-language capability for the NYC311 text/SMS channel. The tool supports the 10 designated citywide languages to enable NYC311 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

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: Microsoft

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 2024: No

Computation Type: Data generation

Purpose Type: Information presentation

Population Type: Individuals (Attendees of Open Data Week and Open Data Ambassadors meetings who have spoken during the meeting or whose names have been mentioned during the meeting)

Identifying Information: Yes

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

Training Data	Training data is proprietary to Zoom.
Input Data	Live audio from Zoom meeting.
Output Data	VTT format file including captions/transcript of meeting.

Vendor Information

Vendor Name: BetaNYC

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.

Appendix

Reportable Elements

1. **Agency, Department.** The name of the agency and the department or office who used the tool.
2. **Tool name.** The name or commercial name of the algorithmic tool.
3. **Tool description.** A brief description of the algorithmic tool.
4. **Updated tool.** If reported previously, whether the tool has been updated in the past year.
5. **Update description.** If updated in the past year, how the tool was updated (e.g., fresh training data, variable weights).
6. **Date of first use.** The month and year in which the algorithmic tool began to be used, if known.
7. **Purpose type.** The type of purpose of the tool:
   a. Data management
   b. Information presentation
   c. Performance evaluation
   d. Resource allocation
   e. Risk management
   f. Triage
8. **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.
9. **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.
10. **Vendor name.** If applicable, name of the vendor or contractor involved in the development or ongoing use of the tool.
11. **Vendor involvement.** If applicable, a description of the vendor or contractor involvement.
12. **Identifying Information.** Whether the tool collects or analyzes "identifying information" as defined under New York City's Identifying Information Law (Local Laws 245 and 247 of 2017).
13. **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)

14. **Computation type.** Type of computation used by the tool:
    a. Classification
    b. Clustering
    c. Data generation
    d. Forecasting
    e. Matching
    f. Ranking
    g. Sampling
    h. Scoring