Office of Technology & Innovation

CY2022

SUMMARY OF AGENCY COMPLIANCE REPORTING OF ALGORITHMIC TOOLS

Introduction:

This summary communicates the results of the City's third annual process for reporting on algorithmic tools. Pursuant to Mayoral Executive Order 3 of 2022 ("EO 3"), 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.[1]

Previously under the guidance of Mayoral Executive Order 50 of 2019 ("EO 50"), and now under Local Law 35 ("LL35"),[2] the City continues its commitment to providing the public with a transparent view of these applications of agency data and technology. Changes in place for reporting in 2022 and beyond are outlined below.

Key Changes for 2022:

Reporting for Calendar Year 2022 was completed using the definitions and requirements of LL 35 and related policies, included in Appendix A. LL 35 defines an algorithmic tool as:

> *Any technology or computerized process that is derived from machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis, that is used to make or assist in making decisions about and implementing policies that materially impact the rights, liberties, benefits, safety or interests of the public, including their access to available city services and resources for which they may be eligible. Such term includes, but is not limited to tools that analyze datasets to generate risk scores, make predictions about behavior, or develop classifications or categories that determine what resources are allocated to particular groups or individuals, but does not include tools used for basic computerized processes, such as calculators, spellcheck tools, autocorrect functions, spreadsheets, electronic communications, or any tool that relates only to internal management affairs such as ordering office supplies or processing*

[1] For the full text of EO 3 of 2022, see: https://www.nyc.gov/office-of-the-mayor/news/003-002/executive-order-3

[2] Executive Order 3 of 2022 moved responsibility for algorithmic tool management and policy to the newly created OTI, and Local Law 35, passed in 2021, mandates the reporting of algorithmic tools on a yearly basis. This 2022 report reflects the requirements of LL 35. For the full text of LL 35, see: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA29B34-9266-4B52-B438-A772D81B1CB5

payments, and does not materially affect the rights, liberties, benefits, safety or interests of the public.[3]

OTI provided guidance to assist agencies in understanding terms used in this new definition, compared to the terms and definitions under EO 50, and to illustrate tools that do and do not meet the definition in the law. Additionally, LL 35 partially modified the information to be reported for each tool compared to EO 50. OTI again provided agencies with guidance on responding to the law's reporting requirements. A summary of reportable information for 2022 is included in the Algorithmic Tool Directory below.

[3] *Ibid.*

SUMMARY OF AGENCY REPORTS

The following table on pages 3-4 summarizes the reporting results from City agencies for 2022.

Agency	Number of Tools Identified	Number of Tools Reported
Administration for Children's Services, ACS	2	2
Business Integrity Commission, BIC	0	0
Civilian Complaint Review Board, CCRB	0	0
Commission on Human Rights, CCHR	0	0
Conflicts of Interest Board, COIB	0	0
Department for the Aging, DFTA	0	0
Department of Buildings, DOB	0	0
Department of City Planning, DCP	0	0
Department of Citywide Administrative Services, DCAS	0	0
Department of Consumer and Worker Protection, DCWP (formerly Department of Consumer Affairs, DCA)	1	1
Department of Correction, DOC	1	1
Department of Cultural Affairs, DCLA	0	0
Department of Design and Construction, DDC	0	0
Department of Education, DOE	3	3
Department of Environmental Protection, DEP	0	0
Department of Finance, DOF	0	0
Department of Health and Mental Hygiene, DOHMH	9	9
Department of Housing Preservation and Development, HPD	0	0
Department of Investigation, DOI	1	1
Department of Parks & Recreation, DPR	0	0
Department of Probation, DOP	0	0
Department of Records and Information Services, DORIS	0	0
Department of Sanitation, DSNY	0	0
Department of Small Business Services, SBS	0	0
Department of Social Services, DSS	1	1
Department of Transportation, DOT	0	0
Department of Veterans' Services, DVS	0	0
Department of Youth and Community Development, DYCD	0	0
Economic Development Corporation, EDC	0	0
Fire Department of New York, FDNY	6	6
Landmarks Preservation Commission, LPC	0	0
Law Department	0	0
Mayor's Office*	3	3
New York City Housing Authority, NYCHA	0	0
New York City Police Department, NYPD	3	3
NYC Emergency Management, NYCEM	0	0
Office of Administrative Trials and Hearings, OATH	0	0

Office of Chief Medical Examiner, OCME	1	1
Office of Technology and Innovation	0	0
School Construction Authority, SCA	0	0
Taxi & Limousine Commission, TLC	0	0
TOTAL	**31**	**31**

*Note that the set of offices included in the Mayor's Office's reporting has been updated for 2022 reporting, based on changes in the City's organizational structure. Notably, the Civic Engagement Commission, which previously reported independently, now reports as part of the Mayor's Office.

ALGORITHMIC TOOL DIRECTORY

As a result of the 2022 agency compliance reporting process, the following algorithmic tools were identified for public reporting. The directory that follows provides general information about these tools to facilitate transparency into the way agencies are leveraging relevant technologies for delivering services to New Yorkers.

For each of the tools reported, the directory provides:

- Name of the agency reporting the tool,
- Month and year in which the algorithmic tool began to be used, if known,
- Name and a brief description of the algorithmic tool,
- Purpose for which the agency is using the tool, and description of how the information received from the tool is used,
- Type of data collected or analyzed by the tool and the source of the data, and
- Whether a vendor or contractor was involved in the development or ongoing use of the tool, a description of such involvement, and the name of such vendor or contractor when feasible.

All information is provided directly by the reporting agency.

Agency: Administration for Children's Services	
Name of Tool Severe Harm Predictive model	**Date Tool Began to be Used** May 2018
Description of Tool Predictions of Severe Harm (identifying likelihood of substantiated allegations of physical or sex abuse within the next 24 months) are based on 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. If the Quality Assurance 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. No staff in Quality Assurance unit or in the investigative unit sees the	

scores. The model only supports the decision about which investigations are prioritized for review by the Quality Assurance unit.

Purpose of Tool and Description of Agency Use

The Quality Assurance 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 Quality Assurance 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 Quality Assurance review.

Type of Data Collected/Analyzed and Source of Data

Training data: ACS trained the model on ACS historic administrative data about closed investigations from January 2013 to March 2014. Training set included about 119,668 observations. The model was tested on closed investigations from April 2014 to December 2014 with 82,331 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 (SCR) call and time spent in foster care. Only ACS administrative data are used in the model.

Description of Vendor Involvement

N/A

Agency: Administration for Children's Services

Name of Tool	**Date Tool Began to be Used**
Repeat Maltreatment (RM) model	July 2017

Description of Tool

Predictions of Repeat Maltreatment (identifying the likelihood of being involved in a future indicated investigation within the next 24 months) are based on Machine Learning methodology and are calculated for all children receiving prevention services from ACS prevention service providers. The prediction is made with the assumption that the case is closed the day the model is run. 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 end of a prevention service. This model was formerly known as the Service Termination Conference (STC) model and was used by Preventive Services managers to identify whether or not a case should have ACS or provider-agency facilitation at service termination conference.

Purpose of Tool and Description of Agency Use

The RM model was initially used to prioritize ACS facilitation of prevention termination conferences, so that ACS could be certain that services had been provided in these cases. When a family is ready to exit ACS prevention services, an end of services conference is required (known as a "Service Termination Conference"). However, Service Termination Conferences are generally no longer facilitated by ACS and this prioritization is no longer necessary.

The RM model is also used to group prevention providers into quartiles to assess their performance in comparison to others in the same quartile, based on the service-need/risk levels of the families they've served during the previous year. This is a retrospective analysis for performance management.

Type of Data Collected/Analyzed and Source of Data

Training data: ACS trained the model on ACS historic administrative data about closed investigations from July 2009 to December 2015. Training set included about 136,982 observations. The model was tested on closed investigations from January 2016 to December 2017 with 48,771 observations.

Input data: Predictions are based on administrative data about prior and current child welfare involvement including SCR investigations and time spent in foster care. Only ACS administrative data are used in the model.

Description of Vendor Involvement

N/A

Agency: Department of Consumer and Worker Protection	
Name of Tool Route Automation	**Date Tool Began to be Used** July 2020

Description of Tool

Inspection Supervisor selects an inspector, enters a date and the number of businesses to be inspected, and the geographic area to be considered. The system identifies businesses in the selected area and assigns them to the route based on inspection priority until the number of businesses entered has been reached. Then the tool runs a Simulated Annealing Algorithm to optimize the order businesses appear on the route based on proximity and method of travel.

Purpose of Tool and Description of Agency Use

DCWP inspectors conduct inspections based on a route, or list of businesses to be inspected on a specific day, which must be pre-approved by their supervisor. The Route Automation tool generates a route for an inspector on a specific date based on configuration variables and geographic area. All routes generated by the tool still require supervisor review and approval.

Type of Data Collected/Analyzed and Source of Data

Training data: The tool is not trained in an AI or Machine Learning sense.

The tool makes decisions based on configuration tables, businesses and their licenses and inspection and violation history, and uses a Simulated Annealing Algorithm to optimize the order in which businesses appear on a route.

Input data: Inspection date, business category and address, licenses held (if any), last inspection date and type, violation history (if any), date of inspection request or license application/renewal (if applicable), Inspection Unit (of the inspector), and geographic area.

Description of Vendor Involvement

The tool was designed and built by PruTech, an outside vendor contracted to design and build DCWP's Automated Inspection Management System (AIMS) and its accompanying Mobile Enforcement platform. The tool is part of the AIMS system.

Agency: Department of Correction	
Name of Tool Housing Unit Balancer (HUB)	**Date Tool Began to be Used** April 2017
Description of Tool The HUB is comprised of two functions: (1) a classification tool based on decision trees that determines an individual's propensity for violence, and (2) a housing area risk assessment, which utilizes advanced predictive analytics (i.e., neural networks) to determine optimal housing areas based on the classification scores of people in custody. The primary operational use of the HUB is for the classification score, which is used to track populations and optimize housing arrangements. The last day the department utilized the HUB system for classification purposes was January 31, 2022.	
Purpose of Tool and Description of Agency Use The Housing Unit Balancer (HUB) is used for informing housing decisions made by operational staff designed to produce less conflict in housing areas.	
Type of Data Collected/Analyzed and Source of Data Training data: Classification function: The training set comprised of over 60,000 records of people in custody and was retrieved in 2014. The test set was comprised of over 4,000 people in custody. These sets were used to create and validate the ultimate solution for the classification scoring of people in custody. Housing Area Risk Assessment function: The training set was incident data for 24 months (01/13-01/15) and the test set was incident data for 6 months (01/15-06/15). Input data: Classification function: (1) past institutional conduct, (2) top criminal charge, (3) security risk group affiliation, (4) age, (5) Brad H indicator. Housing Area Risk Assessment function: composition levels of (1) people in custody, (2) security risk group members, (3) security risk groups, (4) housing types, (5) age, (6) length of stay, and (7) duration of stay in housing areas.	
Description of Vendor Involvement The tool was created by McKinsey & Company and implemented in 2017	

Agency: Department of Education	
Name of Tool MySchools	**Date Tool Began to be Used** August 2018
Description of Tool 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

Purpose of Tool and Description of Agency Use

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 (3K, pre-K, Gifted & Talented, middle school, and high school). The tool encompasses a family-facing portal, a school-facing portal, and an administrative portal.

Type of Data Collected/Analyzed and Source of Data

Training data: The algorithm was already widely recognized for its advantages prior to adoption in New York City. The DOE consulted with a team of researchers at MIT 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).

Description of Vendor Involvement

We have a 5 year contract with the agency Blenderbox who designed the application and implemented the algorithmic matching functionality. The work is meant to transition to be run in-house, by the Division of Instructional and Information Technology (DIIT) within the Department of Education, by the end of the contract. The team at DIIT has already begun to takeover maintenance and development of the tool.

Agency: Department of Education

Name of Tool	**Date Tool Began to be Used**
NYCDOE APPR Measures of Student Learning (MOSL) Growth Model	September 2013

Description of Tool

The growth model uses a variety of student-level (assessment scores, English Language Learner, Disability, and Economic Disadvantage indicators), classroom-level (e.g. % Students With Disabilities), and school-level data (e.g. % English Language Learners, % 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 used to either 1) identify "peer groups" of students, from which student growth percentiles (SGPs) are determined, or 2) 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 MOSL Rating. The MOSL Rating is combined with the MOTP 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.

Purpose of Tool and Description of Agency Use

In accordance with New York state law and New York State Education Department (NYSED) regulations, the Department developed and maintains a "growth model" to produce Measures of Student Learning (MOSL) ratings for use in annual professional performance reviews (APPR) for teachers and principals. The MOSL ratings are combined with Measures of Teaching/Leadership Practice (MOTP/MOLP) ratings to produce an annual Overall Rating for each eligible educator.

Type of Data Collected/Analyzed and Source of Data

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

Input data: The growth model makes use of three types of data: (1) students’ end-of-year assessment scores, (2) enrollment and attendance records that link students to teachers and schools, and (3) historical academic and demographic information used to identify groups of similar students.

Description of Vendor Involvement

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

Agency: Department of Education

Name of Tool	**Date Tool Began to be Used**
NYCDOE APPR Measures of Teaching/Leadership Practice (MOTP/MOLP) Calculation	September 2013

Description of Tool

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 an MOTP/MOLP Rating. The MOTP/MOLP Rating is combined with the MOSL Rating to produce an Overall Rating for each eligible educator. 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.

Purpose of Tool and Description of Agency Use

In accordance with New York state law and New York State Education Department (NYSED) regulations, the Department developed and maintains databases and calculation rules to produce Measures of Teaching/Leadership Practice (MOTP/MOLP) ratings for use in annual professional performance reviews (APPR) for teachers and principals. The MOTP/MOLP ratings are combined with Measures of Student Learning (MOSL) ratings to produce an annual Overall Rating for each eligible educator.

Type of Data Collected/Analyzed and Source of Data

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.

Description of Vendor Involvement

N/A

Agency: Department of Health and Mental Hygiene	
Name of Tool Newborn Home Visiting Program Screening Tools	**Date Tool Began to be Used** PHQ-9 and GAD-7 screening tools were implemented by the program in September 2015. The breastfeeding assessment tool was implemented by the program in November 2014

Description of Tool

The NHVP database uses an algorithm based on screening tool results to determine breastfeeding and maternal depression referrals.

Purpose of Tool and Description of Agency Use

Based on the results from the PHQ-9 and GAD-7 screening tools, a client is deemed eligible for a referral to the internal program social worker for follow up on maternal mental health services. In addition, a breastfeeding assessment tool also embedded within the NHVP database is used to make referrals to the program's lactation support services. Based on the client's score, there is a decision made around the intensity of lactation support services that the client will receive.

Type of Data Collected/Analyzed and Source of Data

Data comes from standard tools around depression (PHQ-9) and anxiety (GAD-7). The threshold for the depression screening is a score of 10 and above, and threshold for the anxiety screening tool is also a score of 10 or above. This triggers a referral with the client's consent to the program's social worker for services, unless there is an immediate need identified such as imminent harm to self or others. Data from the breastfeeding tool is obtained in a questionnaire completed through the lead home visitor or the International Board Certified Lactation Consultant (IBCLC's) observation around baby's readiness to feed, latching, parental comfort to feed the baby, observation of milk transfer from mother to baby and parental report of baby's stool/urine output frequency.

Description of Vendor Involvement

The Newborn Home Visiting Program Database is a DOHMH DITT Developed in-house system. The breastfeeding screening tool was developed within the program by a certified nurse.

Agency: Department of Health and Mental Hygiene	
Name of Tool	**Date Tool Began to be Used** November 2016

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

Description of Tool

Foodborne disease outbreaks are identified through many mechanisms. Restaurant associated outbreaks are often identified through complaints received via NYC's 311 non-emergent information system, however not all individuals report to 311. The New York City Department of Health and Mental Hygiene (NYC 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. These data are used in addition to complaint data received through NYC's 311 system to identify and respond to foodborne disease outbreaks.

Purpose of Tool and Description of Agency Use

The model uses data from Yelp restaurant reviews and Twitter data that is available on Twitter's publicly available API. The classifiers assign a "sick score" to each Yelp review or tweet indicating the likelihood that the review or tweet pertains to foodborne illness. The sick score is based on whether the review/tweet 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 and tweet 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/tweet was actually reporting foodborne illness possibly associated with a NYC restaurant; if yes, Yelp messages are sent to Yelp reviewers, requesting that they contact DOHMH, and a Twitter message with a survey link is tweeted back to Twitter users to confirm foodborne illness. Data from annotations are used to improve classifier performance. Foodborne disease complaints identified through Yelp and Twitter are combined with foodborne disease complaints reported to 311 to improve efficiency of outbreak detection.

Type of Data Collected/Analyzed and Source of Data

Training data was used in the development of both the Yelp and Twitter classifiers. Columbia University provided DOHMH with sample datasets for Yelp reviews and tweets, which were reviewed and annotated by DOHMH staff. For Yelp, training data was 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. For Twitter, training data was focused on if the tweet was indicating foodborne illness and if the incident occurred in NYC.

Description of Vendor Involvement

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.
Agency: Department of Health and Mental Hygiene
Name of Tool Pangolin Division of Disease Control **Date Tool Began to be Used** July 2021
Description of Tool 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.
Purpose of Tool and Description of Agency Use Assigns SARS-CoV-2 viral sequences obtained from clinical specimens to known lineages. Used to determine circulating lineages in NYC.
Type of Data Collected/Analyzed and Source of Data Nucleic acid sequences recovered from pathogen genomes using high throughput sequencing. The ML model is trained using the sequence data collected from around the world to generate variant nomenclature. Input data are designated with a variant name by maximum parsimony and random forest tree classification methods.
Description of Vendor Involvement N/A
Agency: Department of Health and Mental Hygiene
Name of Tool IQTREE Division of Disease Control **Date Tool Began to be Used** May 2020
Description of Tool IQTREE uses maximum-likelihood regression to create phylogenetic trees from genomes.
Purpose of Tool and Description of Agency Use Produces phylogenetic trees which are used to rule in or out SAR-CoV-2 sequences in outbreaks.
Type of Data Collected/Analyzed and Source of Data Assembled nucleic acid sequences recovered from pathogen genomes using high throughput sequencing as input data to this tool.
Description of Vendor Involvement N/A
Agency: Department of Health and Mental Hygiene
Name of Tool kSNP3 Division of Disease Control **Date Tool Began to be Used** March 2022
Description of Tool

kSNP3 can use multiple algorithms (maximum-likelihood, parsimony, neighbor-joining) to infer phylogenetic trees from genomes.

Purpose of Tool and Description of Agency Use

Produces phylogenetic trees which are used to rule in or out bacteria such as N. meningitidis in outbreaks.

Type of Data Collected/Analyzed and Source of Data

Assembled nucleic acid sequences recovered from pathogen genomes using high throughput sequencing as input data to this tool.

Description of Vendor Involvement

N/A

Agency: Department of Health and Mental Hygiene

Name of Tool	**Date Tool Began to be Used**
PHYLOViZ Division of Disease Control	October 2017

Description of Tool

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 (MST)

Purpose of Tool and Description of Agency Use

Used to generate the minimum spanning tree relationships which are used to rule in or out Legionella strains in outbreaks.

Type of Data Collected/Analyzed and Source of Data

Assembled nucleic acid sequences recovered from pathogen genomes using high throughput sequencing and qualitative data to represent sample and patient data.

Description of Vendor Involvement

N/A

Agency: Department of Health and Mental Hygiene

Name of Tool	**Date Tool Began to be Used**
GATK Division of Disease Control	October 2017

Description of Tool

A suite of tools for variant calling and filtering after sequence alignment. It uses naive Bayesian to qualify aligned bases as sequence or erroneous data, which would be excluded from the final genomic sequence.

Purpose of Tool and Description of Agency Use

Used to identify mutations and other differences in sequences in microbial genomes. Can be used to determine different characteristics of microbial genomes.

Type of Data Collected/Analyzed and Source of Data

Assembled nucleic acid sequences recovered from pathogen genomes using high throughput sequencing, and quantitative data that represent the quality of the values.

Description of Vendor Involvement

N/A

Agency: Department of Health and Mental Hygiene

Name of Tool	**Date Tool Began to be Used**
BioNumerics Division of Disease Control	September 2017

Description of Tool

BioNumerics is software used to store and analyze sequencing data from bacterial pathogens that are implicated in food outbreaks. In short, bacterial isolates derived from clinical and environmental sources are received by the NYC Public Health Laboratory where they are processed, tested, and ultimately sequenced to identify clusters of disease.

Purpose of Tool and Description of Agency Use

BioNumerics is used to 1) re-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 whole genome multi-locus sequence typing (wgMLST, 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 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.

Type of Data Collected/Analyzed and Source of Data

Nucleic acid sequences recovered from pathogen genomes using high throughput sequencing.

Description of Vendor Involvement

N/A

Agency: Department of Health and Mental Hygiene

Name of Tool	**Date Tool Began to be Used**
ChoiceMaker (CM) Division of Disease Control	Approx. June 2023

Description of Tool

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

Purpose of Tool and Description of Agency Use

CM is used by 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).

Type of Data Collected/Analyzed and Source of Data

CM uses demographic data (e.g., names, DOB, address, identifiers) and health event data (e.g., date and type of event) from BOI's Citywide Immunization Registry (CIR) and Healthy Homes' LeadQuest registry in its evaluation. The program outputs a series of record pairs and a match probability for each pair.

Description of Vendor Involvement

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.

Agency: Department of Investigation	
Name of Tool Facial Recognition Technology	**Date Tool Began to be Used** March 2019

Description of Tool

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.

Purpose of Tool and Description of Agency Use

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.

Type of Data Collected/Analyzed and Source of Data

Training data: Self-trained in system usage.

Input data: Images.

Description of Vendor Involvement

Out-of-the-box product. The vendor provides ongoing technical assistance. Confidentiality agreements are in place with the vendor.

Agency: Department of Social Services	
Name of Tool Homebase Risk Assessment Questionnaire (RAQ)	**Date Tool Began to be Used** June 2012

Description of Tool

Homebase applicants answer questions about their current housing situation, history of disruptive experiences, and shelter history. Each of the answers is assigned a number of points, and applicants that reach a certain point threshold are eligible for additional 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.

Purpose of Tool and Description of Agency Use

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 enrolled in public assistance or 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 much better way to match resources to the families who would benefit the most.

Type of Data Collected/Analyzed and Source of Data

Training data: The RAQ was developed based on analysis of data on Homebase enrollees from 2004 to 2008, conducted in conjunction with a team of academic researchers, to determine predictive factors for those entering shelter.

Input data: Personal characteristics such as age and pregnancy; educational attainment, employment status, benefits status; housing issues such as eviction, discord, number of moves, recent discharge from institutions; traumatic childhood experiences; past and recent experience of homelessness.

Description of Vendor Involvement

DHS contracted with researchers to evaluate years of Homebase administrative data to develop a risk assessment. 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

Agency: Fire Department of New York	
Name of Tool RBIS (Risk Based Inspection Program); ALARM (A Learning Approach to Risk Modeling)	**Date Tool Began to be Used** November 2019

Description of Tool

ALARM is a combined approach using machine learning and risk ratios to assess the risk of a building for structural fire ignition (probability) and civilian fire injury/death (impact). The machine learning algorithm takes incident data, housing characteristics, and 311 data and creates a probability of structural fire ignition. This is combined with a civilian injury or death risk ratio for the building which is based on building characteristics, incident data and nearby felony crimes to create a risk score (range is 1-9), with 1 being highest risk and 9 being lowest. Buildings are prioritized within each of the nine risk scores according to the residential population in each building.

Purpose of Tool and Description of Agency Use

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 BINs), as a part of the Risk-Based Inspection Program.

Type of Data Collected/Analyzed and Source of Data

Training data: In order to create the models, the team utilized a 5-year incident dataset and reserved 99% of the data to train the probability model and 80% of the data to train the impact model.

Input data: The ALARM risk score utilizes data from our fire and EMS dispatch system, building characteristic data, 311 calls, felony crimes, census data and civilian injury data.

Description of Vendor Involvement

ALARM was built in-house by a team of analysts from the Management, Analysis and Planning Bureau.

Agency: Fire Department of New York

Name of Tool	**Date Tool Began to be Used**
EMS Hospital Suggestion Algorithm	March 2007

Description of Tool

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."

Purpose of Tool and Description of Agency Use

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

Type of Data Collected/Analyzed and Source of Data

Input data: The EMS Hospital Suggestion algorithm relies on telematics data from the Department of Citywide Administrative Services city-owned vehicles collected between 2015 and mid-2016 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.

Description of Vendor Involvement

This algorithm and the resulting output file that is used in our EMS CAD system to suggest hospitals was provided by a vendor, until September 2020. The Department currently creates this file using a new algorithm, developed in-house by the Bureau of Management Analysis and Planning in conjunction with engineers from Columbia University.

Agency: Fire Department of New York

Name of Tool EMS Unit Suggestion Algorithm	**Date Tool Began to be Used** March 2007
Description of Tool The algorithm computes a list of geographic atoms in order of closest to furthest in 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.	
Purpose of Tool and Description of Agency Use The EMS Unit Suggestion Algorithm is used to determine which order of geographic regions (known as atoms) to search in order for the EMSCAD system to select an appropriate EMS unit for dispatch to an incident.	
Type of Data Collected/Analyzed and Source of Data Input data: The EMS Unit Suggestion algorithm relies on historical FDNY CAD trip time data which is used to calibrate a network analysis model which derives atom-to-atom transport times.	
Description of Vendor Involvement This algorithm and the resulting output file that is used in our EMS CAD system to suggest atom order for unit search is currently provided by a vendor, Deccan International.	
Agency: Fire Department of New York	
Name of Tool EMS Hospital Load Balancing Algorithm	**Date Tool Began to be Used** January 2021
Description of Tool The algorithm requires three data inputs: the estimated travel time from any ATOM to any hospital, the number of available beds for every hospital and the estimated number of transports that will occur the following day at every hospital. The algorithm first determines if any hospital is expected to receive more patients than available beds. If overload is expected, the algorithm reallocates the necessary ATOMs such that no hospital is overloaded, and the reallocation of any ATOM is done so with minimal additional travel time. The optimized output - known as a pattern - is directly input into the EMS CAD system for use in the following day.	
Purpose of Tool and Description of Agency Use The hospital load balancing algorithm is designed to optimize hospital transports in a way that proactively avoids hospitals from being congested with too many patients, while at the same time minimize the total travel times as much as possible. The outputs of the algorithm are used in the EMS Computer Aided Dispatch (EMS CAD) system to provide EMS crews with an optimal hospital to transport patients.	

Type of Data Collected/Analyzed and Source of Data

Training data: This is an optimization model and was not "trained".

Input data: There are three input data sources. First is an ATOM-to-hospital travel time matrix. This matrix is computed by FDNY based on historical travel time information. Second, the algorithm consumes hospital bed availability data from the HERDS dataset generated by the New York State Department of Health and provided by NYC DOHMH. Finally, we estimate the number of transports that will occur at each hospital for the next day by using historical data and computing a 3-day moving average.

Description of Vendor Involvement

The tool was developed internally at FDNY in partnership with Columbia University's Industrial Engineering and Operations Research Department.

Agency: Fire Department of New York

Name of Tool	**Date Tool Began to be Used**
EMS Ambulance Scheduling Tool	June 2021

Description of Tool

The tool requires the average number of medical emergencies per hour by dispatch area and the ambulance schedule of each dispatch area. Based on this information, the algorithm will optimize tour start times to maximize the minimum difference between the supply of ambulances and the demand for an ambulance.

Purpose of Tool and Description of Agency Use

The purpose of the tool is to match the supply of EMS ambulances to the demand for ambulances (medical emergencies) over a 24-hour period for each EMS dispatch area. The tool uses an existing ambulance schedule for each dispatch area and optimizes their start times in order to match the demand for an ambulance. The tool supports FDNY EMS in developing an ambulance schedule citywide.

Type of Data Collected/Analyzed and Source of Data

Training data: This is an optimization model and was not "trained". The model optimizes based on the hourly average number of incidents in a dispatch area.

Input data: There are two input data sources. First, the supply of ambulances which includes the start and end times of each ambulance tour in the city, as well as the starting location of the ambulance. The second input is the demand, which is the number of medical emergencies per hour per dispatch area.

Description of Vendor Involvement

The tool was developed internally at FDNY in partnership with Columbia University's Industrial Engineering and Operations Research Department.

Agency: Fire Department of New York

Name of Tool	**Date Tool Began to be Used**
EMD Schedule Optimization Tool	June 2021

Description of Tool

The algorithm requires two datasets. First, the tool requires the average number of medical calls per hour for a 24-hour period. Second, the tool requires a user to specify the number of call takers assigned to each tour. Based on these two inputs, the tool provides a projection of supply (call takers) versus demand (medical calls). Additionally, the tool can take the total number of available staff and optimally allocate them across tours to maximize the minimum difference between supply and demand. Based on these outputs, EMD officers can identify times during the day when call taker utilization is high and reallocate staff to accommodate.

Purpose of Tool and Description of Agency Use

The purpose of the tool is to provide Emergency Medical Dispatchers (EMD) staff a tool to optimally allocate call takers during a 24-hour period. The tool uses an expected number of incoming calls and the number of personnel scheduled to work in order to allocate the call takers to different shifts such that the supply of call takers exceeds the demand for call takers.

Type of Data Collected/Analyzed and Source of Data

Training data: This is an optimization model and was not "trained" using training data. The algorithm relies on actual historical data to determine average hourly medical calls.

Input data: The tool requires an hourly count of medical calls arriving during a 24-hour period. Additional "data" requirements are input from the user depending on user-driven scenarios. For example, a user could specify five 8-hour tours per day (at different start times) rather than existing four tours (2 8-hour tours and 2 12-hour tours).

Description of Vendor Involvement

The tool was developed internally at FDNY in partnership with Columbia University's Industrial Engineering and Operations Research Department.

Agency: Mayor's Office	
Name of Tool Scorecard Blockface Sampling Algorithm - MO - Operations	**Date Tool Began to be Used** March 2022

Description of Tool

The Scorecard program sends inspectors across New York City to rate street and sidewalk cleanliness. The sampling algorithm creates a monthly list of blocks for inspectors to visit and rate.

Purpose of Tool and Description of Agency Use

The primary goal of the algorithm is to produce a sample of blockfaces that is statistically sound and geographically representative. This list is used to rate street and sidewalk cleanliness citywide, as well as by borough and DSNY district.

Type of Data Collected/Analyzed and Source of Data

The blockface sample is selected from the Pavement Edge File, which is part of the NYC Planimetric Database managed by the Office of Technology and Innovation. Sampling is weighted towards blockfaces in high-density areas and includes extra sampling of blockfaces in Business Improvement Districts (BIDs). It also takes into account the linear miles of street within a DSNY District.

Description of Vendor Involvement

The sampling algorithm was developed by the former Mayor's Office of the Chief Technology Officer in partnership with the Mayor's Office of Operations.

Agency: Mayor's Office

Name of Tool	**Date Tool Began to be Used**
Methodology for Poll Site Language Assistance - MO - CEC	November 2020

Description of Tool

Since no dataset is currently available that reliably captures the number of limited English proficient (LEP) registered voters for all program languages, the 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 NYCBOE 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.

Purpose of Tool and Description of Agency Use

This is a methodology for determining how the New York City Civic Engagement Commission (CEC) will provide interpretation services at poll sites for limited English proficient voters. The methodology explains how the NYCCEC will identify the languages and locations in which interpretation services will be offered during the November 2020 election and beyond. These services supplement the interpretation assistance provided by NYC Board of Elections in several languages. Under the Charter, the NYCEC can only provide interpretation services in a language if: (1) it is a designated citywide language; or (2) 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.

Type of Data Collected/Analyzed and Source of Data

Input data: For citywide estimates, this methodology uses current data from the American Community Survey (ACS) 2016-2020 5-year estimates. This methodology also uses the American Community Survey Census Tract 2016-2020 5-year Public Use Microdata Samples for poll site level analysis; this is the most current and accurate data available on 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.

Description of Vendor Involvement

N/A

Agency: Mayor's Office

Name of Tool	**Date Tool Began to be Used**
SmartVAN / TargetSmart - MO - PEU	November 2019

Description of Tool

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.

Purpose of Tool and Description of Agency Use

In 2022, PEU has 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 in SmartVAN used to conduct text and phone outreach to help New Yorkers access the Affordable Connectivity Program (ACP) in specific zip codes. In cases like these, TargetSmart's determination of who lives in which zip codes affects whether New Yorkers receive PEU outreach. Additionally, the algorithm that TargetSmart uses to match phone numbers to individuals and determine if they are mobile phones or not determines the type of outreach that New Yorkers receive.

Type of Data Collected/Analyzed and Source of Data

The algorithmically-derived data that PEU accesses via SmartVAN 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.

Description of Vendor Involvement

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.

Agency: New York City Police Department	
Name of Tool Facial Recognition Technology	**Date Tool Began to be Used** October 2011
Description of Tool Tool which may help investigators identify unknown subjects in law enforcement investigations.	
Purpose of Tool and Description of Agency Use Facial recognition is a digital technology that NYPD uses to compare images obtained during investigations with lawfully possessed arrest 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.	
Type of Data Collected/Analyzed and Source of 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. The facial recognition software will generate a pool of possible match candidates for review by trained Facial Identification Section investigators. • Training data is proprietary to the vendor.	
Description of Vendor Involvement Software developed and maintained by Dataworks	

Agency: New York City Police Department	
Name of Tool ShotSpotter	**Date Tool Began to be Used** March 2015
Description of Tool Provides acoustic gunshot detection to assist with emergency call response	
Purpose of Tool and Description of Agency Use 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.	
Type of Data Collected/Analyzed and Source of Data •Specialized software analyzes audio signals for potential gunshots, 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.
• Training data is proprietary to the vendor.

Description of Vendor Involvement

Software developed and maintained by ShotSpotter

Agency: New York City Police Department

Name of Tool	**Date Tool Began to be Used**
Patternizr	December 2016

Description of Tool

Aids crime analysis in detection of potential crime patterns.

Purpose of Tool and Description of Agency Use

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.

Type of Data Collected/Analyzed and Source of 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.
•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.

Description of Vendor Involvement

The tool was developed by data scientists and analysts at NYPD. Contractors and NYPD personnel integrated it into the Domain Awareness System. Personnel in Crime Control Strategies work with the Information Technology Bureau to maintain the tool.

Agency: Office of Chief Medical Examiner

Name of Tool	**Date Tool Began to be Used**
STRMix	January 2017

Description of Tool

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

software builds millions of conceptual DNA profiles. 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.
Purpose of Tool and Description of Agency Use 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.
Type of Data Collected/Analyzed and Source of Data Training data: Training data was not used in the sense of AI software. The 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.
Description of Vendor Involvement The software has been developed by New Zealand Crown Institute of Environmental Science and Research (ESR) with Forensic Science South Australia. The developer assisted the NYC OCME analyze and interpret our data during the validation of the software.

Appendix A – Local Law 35

File #: Int 1806-2019, **Version:** A

Int. No. 1806-A

By Council Members Koo, Lander, Ayala, Kallos, Cabán, Rosenthal, Louis, Dinowitz, Gennaro and Rose

A Local Law to amend the administrative code of the city of New York, in relation to reporting on algorithmic tools used by city agencies

Be it enacted by the Council as follows:

Section 1. Subchapter 1 of chapter 1 of title 3 of the administrative code of the city of New York is amended by adding a new section 3-119.5 to read as follows:

§ 3-119.5 Annual reporting on algorithmic tools. a. For purposes of this section, the term "algorithmic tool" means any technology or computerized process that is derived from machine learning, artificial intelligence, predictive analytics, or other similar methods of data analysis, that is used to make or assist in making decisions about and implementing policies that materially impact the rights, liberties, benefits, safety or interests of the public, including their access to available city services and resources for which they may be eligible. Such term includes, but is not limited to tools that analyze datasets to generate risk scores, make predictions about behavior, or develop classifications or categories that determine what resources are allocated to particular groups or individuals, but does not include tools used for basic computerized processes, such as calculators, spellcheck tools, autocorrect functions, spreadsheets, electronic communications, or any tool that relates only to internal management affairs such as ordering office supplies or processing payments, and does not materially affect the rights, liberties, benefits, safety or interests of the public.

b. Each agency shall report to the mayor's office of operations, or any other office or agency designated by the mayor, no later than December 31 of every year, every algorithmic tool that the agency has used one or more times during the prior calendar year.

c. Each agency shall provide the following information about each algorithmic tool reported pursuant to subdivision b of this section:

1. The name or commercial name, and a brief description of such algorithmic tool;

2. The purpose for which the agency is using such an algorithmic tool;

3. The type of data collected or analyzed by the algorithmic tool and the source of such data;

4. A description of how the information received from such algorithmic tool is used;

5. Whether a vendor or contractor was involved in the development or ongoing use of the algorithmic tool, a description of such involvement, and the name of such vendor or contractor when feasible; and

6. The month and year in which such algorithmic tool began to be used, if known.

d. The mayor's office of operations, or any other office or agency designated by the mayor, shall compile the information received pursuant to subdivisions b and c of this section and report it to the mayor and the speaker of the council, disaggregated by agency, no later than March 31 of every year.

e. No agency shall disclose any information pursuant to this section where such disclosure would violate local, state, or federal law, or endanger the safety of the public, or interfere with an active agency investigation.

§ 2. This local law takes effect immediately.

SJ/IB
LS #11068
12/7/21 10:00 pm