Before the
National Institute of Standards and Technology
Gaithersburg, MD 20899

In the Matter of	)	
Request for Information Related to NIST's	)	
Assignments Under Sections 4.1, 4.5 and 11 of the	)	
Executive Order Concerning Artificial Intelligence	)	Docket No.: NIST-2023-0009

COMMENTS OF THE CITY OF NEW YORK

The City of New York ("City"), through its Office of Technology and Innovation ("OTI"), submits this response to the National Institute of Standards and Technology's ("NIST") *Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence*.[1] As NIST seeks to carry out its responsibilities under Executive Order 14110 ("EO 14110"), the City welcomes the opportunity to communicate its relevant priorities and concerns as well as to emphasize the importance of the federal government's leadership in emerging areas of AI governance and other AI-related initiatives.

In 2023, the City issued the New York City Artificial Intelligence Action Plan ("AI Action Plan"), leveraging the 2022 consolidation of City technology and data functions under a unified Office of Technology and Innovation to outline a set of new, impactful commitments to support responsible use of AI across city government.[2] This pioneering work builds on earlier City efforts to address the incredible opportunities and unique impacts that AI has in the local government context, including the publication of an AI Primer and Strategy in 2021,[3] which broadly explored the meaning and implications of AI for New York City and outlined key opportunities and challenges for the City's ongoing work, and a first-in-the-nation directory of the City's algorithmic tools, in publication since 2020.[4]

While the City pursues the development of policies and other governance practices as committed through the AI Action Plan, it values the role that federal policymaking and regulation plays both

---

[1] NIST, *Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial* Intelligence, Federal Register Vol 88, No. 244, Docket 231218-0309, https://www.govinfo.gov/content/pkg/FR-2023-12-21/pdf/2023-28232.pdf

[2] City of New York, Executive Order 3 of 2022, which consolidated the City's technology and data teams under the newly-created Office of Technology and Innovation can be found at https://www.nyc.gov/assets/home/downloads/pdf/executive-orders/2022/eo-3.pdf ; the City's AI Action Plan is available at https://www.nyc.gov/assets/oti/downloads/pdf/reports/artificial-intelligence-action-plan.pdf

[3] The City's 2021 AI Strategy and appended AI Primer can be found at https://a860-gpp.nyc.gov/concern/nyc_government_publications/nv9355378?locale=en

[4] New York City Local Law 35 of 2022 can be found at https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA 29B34-9266-4B52-B438-A772D81B1CB5 ; the City's algorithmic tools reporting outputs from 2020 to 2022 can be found at https://www.nyc.gov/content/oti/pages/reports

in providing local governments with models for implementable governance frameworks, and in promoting emerging standards that may allow for greater harmonization of practices across different levels of government. NIST's AI Risk Management Framework ("AI RMF") is a key reference for the City, given the framework's intentional applicability across a range of organization types. In parallel, the City also heavily leverages NIST's Cybersecurity Framework and associated security control publications to perform security reviews, with a keen eye to FedRAMP's cloud accreditation process, to maximize the ability to reuse technologies that have already been accredited.

The City strongly believes that local governments face unique challenges when implementing responsible AI practices, often managing frontline delivery of government services to their residents with finite fiscal and human capital resources, and through complex organizational structures and administrative processes. Further, many cities—and the City of New York in particular—are home to diverse populations, including economically and socially disadvantaged groups, with a wide array of needs and vulnerabilities that cannot be addressed with a one-size-fits-all approach to service delivery. With limited resources and a focus on promoting equity and opportunity for residents, local governments may experience both operational and programmatic hurdles to implementing and governing AI solutions in ways that help to ensure responsible and fair use while reducing risk for residents. Generally, the City encourages NIST to continue promoting multi-sectoral perspectives into its ongoing development of guidance, standards, and best practices. More specifically, the City strongly emphasizes the need to view local governments as one of the most important categories of adopters of federal policymaking models and beneficiaries of emerging standards and best practices, in their parallel roles of service delivery, policymaking, and regulatory enforcement. Residents are often the most deeply and directly impacted by thoughtful implementation of federal guidance at the municipal level.

The sections that follow enumerate the City's comments for selected sections of the Request for Information.

I. *Section 1.a.(1). Developing a companion resource to the AI [RMF…] for generative AI.*

The City identifies several challenges in implementing AI RMF core functions for generative AI tools. As noted above, the City encourages NIST to develop the companion to the AI RMF with local governments in mind as potential adopters of the framework. Accordingly, the companion will be most effective to the extent that its implementation is achievable within the scale, organizational complexity, and resource constraints common across local governments. In addition, the City raises the following considerations in line with the topics identified under this section by NIST in the Request for Information:

Risks and harms of generative AI, including challenges in mapping, measuring, and managing trustworthiness characteristics as defined in the AI RMF, as well as harms related to repression, interference with democratic processes and institutions, gender-based violence, and human rights abuses

New York City government is both large and complex, comprising almost 150 agencies and offices of varying sizes, mandates, and capacities to use and govern AI. The City is undertaking a range of steps under its AI Action Plan to build capacity across these units and establish central

policy and guidance. However, this organizational complexity complicates efforts to effectively map AI use, as a precursor to mapping AI risk, within City agency operations. As noted, the City has implemented pioneering efforts to identify and publicly report the use of "algorithmic tools" across its agencies since 2020. This work requires a range of procedures to support agency-level awareness and reporting on an ongoing basis.[5] These procedures must account for varied levels of maturity with the subject matter, internal agency structures and procedures already in place, and domain-specific conventions, constraints, and regulations, among other factors. Complicating efforts further, generative AI solutions are increasingly being "invisibly" integrated into enterprise software that is already in use by City stakeholders, and are more widely available to be leveraged by City staff within external consumer tools. These developments will likely amplify the challenges of supporting robust awareness and reporting of AI tools in a complex organizational context, and require ongoing attention and adaptation.

Additionally, while some AI solutions used by City agencies are developed internally, a substantial number are procured. Some of these systems are used off-the-shelf as built by vendors, while others are customized for compliance or data management, or otherwise fine-tuned to agency-specific requirements. Procurement of these tools often occurs at the agency level (though agencies may make use of larger citywide or state contracts). Generally, the reliance on procurement indicates a need to focus governance efforts at the procurement stage of an AI system's lifecycle, but such attention is challenging when procurements occur in high volume or through a wide array of procurement vehicles. Accordingly, the City is undertaking new steps through its AI Action Plan to both streamline AI contracting, where appropriate, and establish new standards, terms, or guidance for AI procurement across the City.

The growing integration of "invisible" AI functionality into existing products and platforms, where purchasers and users may not be aware of an AI component of the system being procured, or where AI functionality is introduced through updates without notification or modification to contracting terms, creates visibility challenges. Similarly, the use of external tools that the City does not procure at all also presents challenges for implementing governance measures through procurement. For example, a video uploaded by the City to an external platform like YouTube may be captioned or translated using generative AI. Additionally, third party services may use AI tools on behalf of vendors, subcontractors, or members of the public, impacting City service delivery in a way that is challenging to include in an AI risk management plan overseen by the City.

In summary, the City anticipates several challenges in performing the RMF's Map and Govern functions at the scale of city government: while procurement should be a key focus of local governments, they should also be cognizant of encounters with AI that may slip in unnoticed through unforeseen product upgrades or interfaces with external platforms.

The types of professions, skills, and disciplinary expertise organizations need to effectively govern generative AI

The City's efforts to establish and implement new governance measures for the use of AI, including generative AI, will require building skills and capacity in a wide variety of teams and

[5] *Ibid.*

roles. Integrating new measures in this area will require new knowledge and skills within procurement, legal, technical, privacy, and business teams to ensure the responsibility for assessing and mitigating AI risk is incorporated throughout. The City's AI Action Plan emphasizes this holistic impact and includes commitments to understand the scope of this need, what resources exist or are needed to support this knowledge-building, and to then launch initial efforts. NIST could also endeavor to layer AI skills into existing workforce development frameworks, including the NICE framework for cybersecurity workforce practices.[6] The City would appreciate any guidance NIST may be able to provide to support these efforts.

Forms of transparency and documentation…that are more or less helpful for various risk management purposes…in the context of generative AI models, and best practices to ensure such information is shared as needed along the generative AI lifecycle and supply chain

Please see comments to Section 1.a.(1), above.

Efficacy, validity, and long-term stability of watermarking techniques and content authentication tools for provenance of materials, including in derivative work

Please see comments to Section 2.a below.

II. Section 1.a.(2). *Creating guidance and benchmarks for evaluating and auditing AI capabilities, with a focus on capabilities and limitations through which AI could be used to cause harm.*

The City is deeply interested in the development of guidance and other best practices to guide effective evaluation of AI systems, and wishes to convey several high-level considerations as NIST undertakes this work.

First, many of AI's potential harms, being sociotechnical in nature, are high-level and abstract, and therefore also difficult to define at a level that is unambiguous enough to permit effective testing at the operational level of AI systems. The City would appreciate NIST's guidance in how to define operational metrics for sociotechnical harms, in addition to the dimensions of bias, privacy, and transparency as discussed in the RMF, to include local government contexts, such as accessibility and degree of access to democratic participation.

Second, privacy should be a foundational consideration within guidance and benchmarks for evaluating and auditing AI capabilities, particularly to support and enable municipal utilization in balance with privacy protection and practice. Municipalities can have particularly complex regulatory landscapes from a privacy perspective with data elements subject to various federal, state, and local laws depending on their nature and the context in which they came into the City's possession. For example, on the local level, the City and its agencies operate in accordance with local privacy law, known as the *Identifying Information Law*, when collecting, disclosing, and

[6] See NIST Special Publication 800-181, Workforce Framework for Cybersecurity (NICE Framework), at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf and NICCS Cyber Career Pathways tool, at https://niccs.cisa.gov/workforce-development/cyber-career-pathways-tool

using identifying information in their operations. The *Identifying Information Law* maintains an expansive definition that is subject to its requirements and protections and would have to be considered in any City use of developed NIST guidance and benchmarks to the extent identifying information is a dependency on utilization.

Third, the City would also appreciate NIST's guidance on auditing AI capabilities within the lens of cybersecurity auditing and attestation. This would include updates of existing security control frameworks and language, in addition to recommendations on implementation of continuous monitoring guardrails for AI systems.

Finally, to ensure continuous improvement, there should be robust mechanisms established with clear roles and responsibilities for maintenance and updates of any guidance or best practices produced for the purposes of evaluating or auditing AI systems. The City would recommend NIST develop such mechanisms to allow for rapid adaptation to address emerging challenges and advancements in AI technology.

Availability of, gap analysis of, and proposals for metrics, benchmarks, protocols, and methods for measuring AI systems' functionality, capabilities, limitations, safety, security, privacy, effectiveness, suitability, equity, and trustworthiness.

Evaluation of AI systems is more difficult when the outputs of those systems are more complex data types. Those evaluation and measurement challenges will be of even greater concern as generative AI tools become more widespread. In the traditional discriminative mode, AI systems produce numerical or small categorical outputs whose correctness is relatively easier to define and measure, e.g., through the well-known measures of accuracy, precision, etc. associated with receiver operating characteristics. In contrast, generative AI systems produce very rich outputs whose correctness is much more nebulous. For example, an organization may wish to use generative AI to pictorially illustrate recommended safety procedures. When preparing such content, the end user will need to evaluate how accurately the generated images depict the steps described in text as a primary goal. At the same time, users may wish to consider many other secondary objectives as well, e.g. whether people and residences depicted in the illustrations accurately represent the diverse communities residing within the City, to consider the potential risk for representational harm.

The current evaluation practices for generative AI tools essentially measure the similarity of output to singular model answers which are assumed to be perfectly correct, collapsing the nuances of output to a single numerical dimension. It is instructive to contrast this practice with the classroom evaluation procedures used in social science and humanities settings, in which examination rubrics often specify multiple dimensions of evaluation, such as grammatical cohesion, clarity of argument, consistency of writing style, and so on, when evaluating student essays. For such situations, correct behavior often requires defining large sets of acceptable solutions, which themselves may be further refined by various degrees of acceptability along different desiderata.

In the absence of universal standards or protocols for addressing these evaluation and measurement challenges, the City has exercised caution with the release of any generative AI tooling to guard against inaccessible services, negative impacts on safety or rights, or inaccurate

communications, but welcomes NIST's ongoing development of adaptable standards and best practices.

Further, human input or feedback is often mentioned as a central element of developing and evaluating AI systems, including generative AI. However, the specific expertise needed in each case can vary quite widely and may not be in place within the unit deploying such tools. This may be particularly the case in a large local government, where individual agencies and teams can have widely varying technical and other capabilities. When encouraging the inclusion of human feedback, it is important to be clear about what kinds of experience the relevant staff should have and the desired nature of the requested feedback. Not all aspects of AI development and evaluation are suitable for all kinds of feedback. Some topics require specific technical, linguistic, or cultural understanding. Moreover, the outputs of human feedback can often be difficult to act on, as two reasonable humans could provide conflicting feedback or provide recommendations outside of the scope of what is feasible. The City has committed to a range of steps that will strengthen engagement with diverse stakeholders across its AI work, including varied public engagement efforts, establishment of an Advisory Network populated by AI experts in a variety of roles, industries, and sectors, and convening a Steering Committee composed of City leaders working on AI and adjacent topics at their agencies. Guidance from NIST on different methods of incorporating human input, and what kinds of input are necessary for different stages of developing and evaluating AI systems will be critically important to this work.

Risks arising from AI value chains in which one developer further refines a model developed by another, especially in safety- and rights-affecting systems

The need to manage the cost of AI risk management is a key concern for the City. Because the City mostly procures AI systems rather than building them in-house, and most systems are procured individually, its governance will be largely focused at the deployment level. However, at the scale of the City, where multiple deployments of the same systems may be required to serve different agencies' needs and domain-specific requirements, the City anticipates that the monetary cost and administrative burden to have every deployment tested de novo would be prohibitive. The City is in need of guidance of how to reconcile the needs for AI risk management with the need to reduce the cost of doing so. Specifically, NIST could provide valuable direction on how to control cost through appropriate decomposition of testing and evaluation to suitable units of analysis, be they model, system or deployment.

Since the practice of red teaming is already well-established in cybersecurity, the City is eager to learn if any best practices from software testing can be adapted to AI evaluation. For example, a key design paradigm in software testing is modularity, which drives the organization of software testing efforts into categories of unit tests, functional tests, integration tests, system tests, and so on. Similarly, the City would be interested to learn if a cost-effective methodology can be developed where evaluations and findings at the model level are transferable across multiple systems that use the same model; and similarly, how findings at the system level are transferable across multiple deployments of the same system. At the control level, there are existing resources that drive adoption of security control inheritance, such as the inheritance of controls from cloud service providers within the various "as a service" offerings (infrastructure, platform, software). The City recommends the development of similar security control overlays for AI models

integrated into cloud offerings, so that the customers using those services understand the customer responsibility versus the provider responsibility. The evolution of technology, such as generative AI, requires the expansion of the shared responsibility model to incorporate these new concepts.

Negative effects of system interaction and tool use, including […] Impacts on equity, including such issues as accessibility and human rights

As a large and diverse city with a mandate to drive equity and inclusion for New Yorkers in the delivery of City services, and the broader life of the city, the potential risks for AI to exacerbate social inequalities are of great concern. This concern is a primary driver behind the City's efforts outlined in the AI Action Plan, including particularly its commitments to establish a robust governance framework, and integrate meaningful public engagement across the City's AI efforts.

Generative AI creates new opportunities to address inequalities, but also presents new equity risks. For example, there is promising new work on machine translation of text and videos into sign language transcription, which has the potential for improving the accessibility of City communications. At the same time, machine translation tools are probabilistic in nature and therefore unable to guarantee perfect translation accuracy, and are likely to miss nuances like emotional affect and context of discourse that are challenging even for human translators to convey accurately. Such use cases exemplify how generative AI's potential to deliver material positive change must be coupled with suitable quality assurance, as discussed above. It is also worth emphasizing that while generative AI may be integrated on the promise of reducing the need for skilled labor that is in short supply, such tools often require the retention of human skills and expertise in order to function effectively and equitably. The City welcomes guidance from NIST on how it can strengthen its ongoing efforts to measure and mitigate the equity impacts of AI broadly.

The City has sought to address equity and inclusion in the context of broadband access and digital equity as these issues are a central priority for Mayor Adams' administration. While the City has launched a range of cornerstone efforts to address these needs,[7] there remain considerable gaps in broadband access and adoption nationally. From this perspective, we encourage NIST to account for not only the equity risks of AI at the product level, but also to consider where a lack of access to systems may lead to inequitable outcomes in the use of City AI tools.

Applicability of testing paradigms for AI system functionality, effectiveness, safety, and trustworthiness including security, and transparency, including paradigms for comparing AI systems against each other, baseline system performance, and existing practice

The City would value guidance from NIST on how to weigh trade-offs in addressing different areas of risk, and particularly how to manage the varying needs and perspectives of different stakeholders in this regard. As noted, the City has committed to a broad array of engagement efforts in its AI Action Plan, with a range of stakeholder groups, recognizing the value and importance of gathering input and expertise from professional experts across sectors, as well as

[7] Information about a range of initiatives announced since 2022 can be found at https://www.nyc.gov/content/oti/pages/press-releases

New York City's diverse communities and populations. In practice, the City recognizes that consideration of any given AI tool or use case will require a delicate balancing of a range of perspectives and needs, in addition to compliance with any intersecting procedures, policies, or laws that may apply. For example, accountability to the public would generally favor greater transparency around city processes and decision-making, but in some cases, this may conflict with other stakeholder needs, or applicable policy or law, such as preserving the human rights and privacy of vulnerable individuals, particularly in sensitive contexts like receiving public assistance and safety. NIST's guidance would be invaluable to inform the Map function of identifying AI risks which may be in tension with each other, and reconciling any conflicts that may arise downstream in the Manage function.

III. Section 1. b. *Establish guidelines (except for AI used as a component of a national security system), including appropriate procedures and processes, to enable developers of AI, especially of dual-use foundation models, to conduct AI red-teaming tests to enable deployment of safe, secure, and trustworthy systems.*

As new generative AI functionality begins to become embedded in traditional applications, the City has begun to develop a process to assess the security of these applications. The City uses a threat-informed testing methodology that combines traditional web application with generative AI security assessment techniques to ensure an effective and comprehensive assessment. The City believes that the lessons learned from the approach could be valuable nationwide.

Nevertheless, as NIST works to develops its own guidelines, the City would recommend that any assessment framework be aligned to an industry accepted standard, e.g. the Open Worldwide Application Security Project (OWASP) Top 10 for LLM Applications, and the Cybersecurity Framework under NIST 800-53.[8]

The City's experience has been that aligning assessments with an industry standard can be beneficial for increasing security and trust, ensuring standardization and consistency that allow for ease of benchmarking and comparison, reducing redundancies, facilitating knowledge sharing and expertise, and enabling compliance with existing regulations and standards.

However, the City recommends that any standards developed should grant organizations the flexibility of adopting an assessment framework that is most appropriate for their specific use cases and is customizable to account for the unique needs and security context of each application.

[8] See OWASP Top 10 for LLM Applications, v1.1, https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-2023-v1_1.pdf and NIST SP 800-53 Rev. 5, Security and Privacy Controls for Information Systems and Organizations, at https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

IV. Section 2. a. *Reducing the risk of synthetic content: Content and tracking its provenance; techniques for labeling synthetic content, such as using watermarking; detecting synthetic content*

It is vital for the City of New York to be able to communicate effectively and broadly with the public, and for the public to verify and validate that communications they receive, especially in circumstances of heightened vulnerability, are genuinely from the City of New York. To this end, the City has a vested interest in the authentication of content and, by extension, the labelling and detection of synthetic content.

One proposed solution for addressing content authentication is the use of watermarking, a technology which would imperceptibly alter an image or text to either identify that it was synthetically created or, ideally, identify its provenance. In either case, the City is not aware of any watermarking tools that are sufficiently mature to be recommended for production use. While some generative AI products may produce a watermark, it is not impossible for this watermark to be spoofed or removed by a bad actor.

Additionally, existing watermarking technologies are generally designed for a specific modality, such as text or images, and the methods do not appear to be transferrable across multiple modalities. This poses a particular problem for the City of New York, which serves a large and diverse population with a wide range of needs to access City information and services. This requires the City to operate in many modalities, languages, and technology platforms; if watermarking is not universal across these barriers, their impact is diluted and may require additional, complex governance procedures.

Content authentication mechanisms can be designed through the packaging of content with suitable metadata, such as digital certificates. The City would benefit from research and best practices on methods for cross-platform and multilingual content authentication.