NYC Use Guidance: Generative Artificial Intelligence

1.0 Introduction

Generative Artificial Intelligence ("GenAI") refers to any AI whose primary function is to generate content, in the form of code, text, images, and more. GenAI technologies are rapidly evolving and becoming increasingly embedded in everyday tools and workflows. They present opportunities for agencies of the City of New York ("City") to improve efficiency, enhance communication, and support public service delivery—such as by drafting content or summarizing information.

These technologies also raise critical concerns related to accuracy and reliability, data privacy and cybersecurity, bias and discrimination, environmental and worker impacts, and the potential erosion of public trust if used in ways that are opaque, unaccountable, or inappropriate for government contexts. New Agentic AI features can exacerbate these concerns or introduce new ones entirely.[1]

The City is committed to a responsible approach to emerging technologies. This Use Guidance provides a framework for the responsible use of GenAI by City personnel, aligned with the City's AI Principles, and applicable laws and policies.

1.1 What's New in this Guidance

In 2024, the City released Preliminary Use Guidance: Generative Artificial Intelligence to orient the City's approach on the use of this technology.[2] This Use Guidance updates that document, offering:

- A new format intended to lead agency stakeholders through key aspects of decision-making.
- An expanded exploration of the challenges GenAI use can raise, aligned to each of the City's AI Principles, as well as the recommended steps to address them. New content brings this Guidance up to date with recent developments of GenAI technology and its use, as well as with new developments in governance.
- An expanded glossary of terms to support common understanding and scope across agency work.

2.0 Purpose

The Office of Technology and Innovation ("OTI") is issuing this Use Guidance to support City agencies in the responsible exploration of GenAI technologies. This document aims to help agencies understand the appropriate considerations and guardrails for Gen AI usage broadly. Future supplements to this document will include more specialized guidance to help agencies understand how the Use Guidance can be applied in common, practical settings where GenAI systems are being used across sectors.

3.0 Authority

The Office of Technology and Innovation ("OTI") was formed under Mayoral Executive Order 3 of 2022 ("EO 3") in order to unify technology teams across government and centralize coordination around existing and emerging technologies. OTI serves as the City's central technology agency, leading "the development, coordination and implementation of the City's information technology, information security, information privacy and telecommunication matters."

4.0 Roles and Responsibilities

As with all technology products and tools, City personnel should access GenAI systems only when such access has been approved by responsible agency personnel, and as authorized by agency-specific and citywide requirements. The relevant personnel and requirements may vary across GenAI system, but will frequently include the following responsible agency personnel:

- Agency Chief Contracting Officer ("ACCO")
- Agency General Counsel
- Agency Chief Information Officer ("CIO") or Agency Chief Technology Officer ("CTO")
- Agency Chief Information Security Officer ("CISO")
- Agency Privacy Officer ("APO")
- Agency Algorithmic Tools Liaison[3]
- Various business or operational owners

5.0 Use Guidance

OTI has organized this Use Guidance around the City's AI Principles.[4]

Sections 5.1 through 5.5 address each of the City's five AI Principles and how they apply to GenAI systems. Each section contains four subsections:

- Guiding Questions: The starting point for understanding how GenAI use relates to the AI principles
- Why This Matters: An explanation of the related risk considerations
- What You Should Do: A summary of strategies to help preempt or plan to mitigate risk
- Example Scenarios: Additional illustrations for the points introduced above

All examples included herein are illustrative and are not intended as endorsements of specific use cases. Please consult both citywide and agency-specific policies before implementing new GenAI use cases. For any questions or inquiries about GenAI usage, please reach out to ai@oti.nyc.gov.

A summary of all Guiding Questions is provided in Appendix A for an at-a-glance view.

5.1 Validity and Reliability

Does this AI system effectively address the specific problem it's intended to solve, and can it perform reliably over time and across different contexts?

Guiding questions to ask

- **Suitability for Purpose:** Is GenAI the only feasible way to achieve the stated purpose? Can other predictive analytics or automation tools also achieve the stated purpose with similar performance? Given that GenAI outputs can change substantially even when given the same or similar prompts, how much variation in the output is acceptable for the stated purpose?
- **Output Verification:** What processes are needed to flag and review errors in the GenAI system's outputs? How much human work, subject matter expertise, and judgment are needed to check and address these errors? Does your agency have appropriate resources in place to do this work, now, and on an ongoing basis?
- **Error Impact and Remediation:** How often are errors expected to occur? If an error goes uncorrected, how severe are the potential consequences? Might they include health, safety, legal, or regulatory ramifications, for example? What processes are needed to remediate errors? How will your agency ensure those processes are effective?
- **Performance Monitoring:** How will the GenAI system's performance be monitored on an ongoing basis? What are the acceptable levels of performance? Is there a change management plan when metrics deviate from such acceptable levels?

Why this matters

- The outputs from GenAI models often change substantially, even when given the exact same inputs or minor wording changes to text prompts.[5]
- GenAI models are trained on vast amounts of data. If the training data are incorrect, inconsistent or outdated, the outputs will reflect that. These issues tend to worsen over time.
- Risks of GenAI use can have real-world consequences based on how those errors translate into mistakes in applications like health, public safety and law.
- GenAI systems have some unique risks. For example:
  - Hallucination: GenAI systems can create answers that seem plausible but are incorrect, especially when relevant information is missing;
  - Context Rot: When processing very long inputs or conversations, GenAI models can forget information about what was input in the beginning, including "system prompts," i.e., instructions that set expectations about the outputs;
  - Sycophancy Bias: The outputs of GenAI systems are biased toward pleasing the user; and
  - Jailbreaking: Malicious inputs can be used on a GenAI system to bypass safeguards and produce unsafe outputs. This can occur without a user's knowledge, via "Indirect Prompt

Injection," when the malicious content is contained in documents retrieved and referenced to generate the response.

- These kinds of errors can be hard to fix, because it is often unclear where the problem occurred given the complexity of models used, and because user prompts may be imprecise. Fixes for specific issues can also introduce new errors where none previously existed.[6]
- Techniques to improve the correctness of outputs don't always work. For example, GenAI can cite the wrong documents, or cite material out of context. Systems with integrated Agentic AI capabilities can write or execute code to obtain the desired results, but these capabilities introduce new risks that the generated code is incorrect or has unintended consequences such as data loss.
- GenAI use often shifts the need for human labor from generating content to verifying correctness of outputs. This verification work, however, can be so onerous that it compromises expected productivity gains from integrating GenAI in the first place.

What you should do

- Clearly define the purpose of the GenAI system before use, and ensure its capabilities align with the task at hand.
- Rigorously test GenAI systems in your domain before using them more widely. For example, conduct small, controlled internal trials and audits to validate that the systems work as expected across a variety of potential use cases, including edge cases, before rolling out to a broader user base. Or consider conducting "red teaming" exercises: adversarial testing to identify a system's limitations and where things may go wrong.
- Have a process for validating outputs on an ongoing basis, and incorporate human-in-the-loop review from subject matter experts where relevant. Check outputs even when they look plausibly correct at first glance.
- Check any AI-Generated citations for accuracy, not just to ensure the source document is valid, but also that the referenced content is correct.
- Monitor system performance over time and re-evaluate its reliability as use cases evolve and as models get updated.
- Check that updates created to fix specific issues are effective and do not introduce new errors.
- Once you understand the limitations of the GenAI system you are using, help others (including coworkers, staff, or the audience) understand these limitations through documentation and discussion.

Example scenarios

- **Good Practice:** Use a GenAI system to draft a first-pass summary of public meeting notes for internal review, followed by staff revision and validation before release, including any dates, numbers, and stats provided by the system.

- **Poor Practice:** Use GenAI to draft talking points for a press conference on a policy announcement without subject-matter review.

5.2 Social Responsibility

Could the use of this AI system result in unfair or inequitable outcomes, and have you taken steps to identify and mitigate bias and harm—especially to vulnerable communities? Have public benefits been carefully weighed against cost and environmental impact; have impacts related to job quality and security, and workplace surveillance been considered?

Guiding questions to ask

- **Audience:** Who is the audience for the output being generated: personnel within your agency, other agencies, or the public?
- **Impact on Decision-Making:** Will the GenAI system's output influence policy, resource allocation, or services – or is it for reference and ideation only?
- **Bias and Equity:** How might the development and use of the GenAI system create, sustain, or magnify discriminatory biases? Are there possible biases in the training data that may reinforce stereotypes, for example? Could the intended use case itself exclude specific populations, limit accessibility, or have other equity impacts?
- **Cost and Environmental Impact:** Is this GenAI use impactful enough to outweigh its cost and energy demands? Does this use align with broader City sustainability goals?[7]
- **Worker Impact:** Could this use of GenAI substantively change or disrupt staff members' work or roles? Could it raise concerns about workplace surveillance?
- **Public Engagement:** Is engagement with relevant individuals and communities needed to fully understand and mitigate impacts?

Why this matters

- Without proper safeguards, AI-Generated Content can be exclusionary or worsen representational harms.
    - GenAI models are trained on vast datasets, and training data often contains low-quality data scraped from the internet or other sources that may not be properly checked for accuracy, bias, or toxicity.
    - Beyond data, the way the core problem or desirable outputs are defined, or how performance is measured can also lead to bias, encoding assumptions about whose interests are relevant and whose interests can be ignored.
- Modern GenAI systems are costly to use and consume significant energy, water, and computing resources. Such demands can have indirect consequences on New Yorkers, like higher energy prices or the construction of electricity transmission infrastructure or data centers that could impact marginalized communities.

- While the integration of GenAI systems can make certain tasks faster and easier, it can also change or displace workers' tasks or roles, require reskilling or changes to internal procedures, and disrupt team dynamics or job quality.
- A socially responsible approach to GenAI requires active efforts to test for bias, engage diverse perspectives, prioritize equity and inclusion throughout the system's design and deployment, and carefully weigh monetary, environmental, and worker impacts against other public benefits a system may offer.

What should you do

- Understand how your agency's overall mission is aligned with citywide policies and goals for social responsibility, and how responsible GenAI adoption can support those efforts.
- Bias and Equity
    - Consider how GenAI outputs describe, or may be received by, different populations, especially marginalized populations.
    - Review GenAI outputs for possible stereotypes or other representational harms, especially when used in public-facing materials.
    - When feasible, consult impacted communities or subject-matter experts to identify risks and mitigation strategies.
- Environmental Impact
    - Understand what resources are needed to use the GenAI system, and also what resources were used to build the GenAI system. Consider resource use in the supply chain, e.g., during data collection or manufacture of computer hardware.
    - Seek information from vendors about their energy use and mitigation strategies (e.g., green cloud commitments, detailed energy usage reporting, data center efficiency metrics, and renewable energy commitments) and incorporate sustainability into procurement criteria.
    - Consider less costly alternatives such as smaller models (sometimes called small language models), or non-AI solutions altogether (see Reliability and Validity).
- Worker Impact
    - Engage workers in decision-making on whether and how to integrate any GenAI systems in their workflows.
    - Train workers on any new system, and prepare them for any workplace disruption its integration may pose.
    - Notify workers of any new data collection employed by new systems, and limit such data collection wherever possible.
    - Consult with appropriate human resources and labor relations teams on any changes to job function or description, as required.

Example scenarios

- **Good Practice**: Run a small pilot of a GenAI-assisted informational chatbot for City services and gather feedback from diverse residents before scaling.
- **Poor Practice**: Use GenAI to auto-generate service descriptions without testing for inclusive language or accessibility across reading levels.
- **Poor Practice**: Use a large model to look up basic information that could easily be found through less resource-intensive tools.
- **Poor Practice:** Deploy a new GenAI system to your agency that affects the efficiency or capacity of agency workers without consulting them in a pilot first or soliciting feedback on the plan.

5.3 Information Privacy

Does the system collect, use, or share identifying information, and if so, does it comply with all relevant privacy laws and policies while protecting individual data rights?

Guiding questions to ask

- **Public Trust:** Has the agency articulated a specific, legally permissible purpose and obtained approval from the APO for the use of identifying information in the GenAI system? If this GenAI system will involve the use of identifying information obtained from the public, will the public be informed?
- **Data Minimization:** Will the minimum identifying information necessary be used for the stated purpose?
- **Accountability:** Are you compliant with relevant information privacy policies, both citywide and agency-specific, and will additional policies be created to support the use of the GenAI system for its articulated purpose?

Why this matters

- All City agency use of GenAI systems must comply with applicable privacy laws and policies, including the Identifying Information Law (NYC Admin. Code §§ 23-1201 to 23-1205),[7] the Citywide Privacy Protection Policies and Protocols (CPPPP),[8] and any agency-specific privacy policies.

What you should do

- Follow citywide and agency privacy policies.
- Consult with your APO on questions relating to GenAI use involving identifying information, including whether a Privacy Impact Assessment (PIA)[9] is appropriate and consideration of contextual integrity in the contemplated use of identifying information (see CPPPP § 4.3; Guidance for Assessing Contextual Integrity). [10]

- Use only the minimum identifying information necessary to complete your task (see CPPPP § 2.0; Privacy Principles - Data Minimization).[11]

Example scenarios

Do: Use an enterprise GenAI system involving the collection of identifying information which has been approved by your agency privacy officer.

5.4 Cybersecurity

Is this AI system designed and deployed in a way that protects the City, City employees, and the public against cybersecurity threats, and does it align with citywide cybersecurity policies and standards?

Guiding questions to ask

- **Tool Authorization and Licensing:** Are you using a GenAI system that has been reviewed, licensed, and approved for agency use—rather than a free GenAI product or unvetted version?
- **Account and Access Security:** Are you using a City-managed or enterprise account with permissions and security controls that follows relevant City or agency policies?
- **Compliance**: Are you compliant with relevant cybersecurity policies?
- **Data Governance**: Have you made sure that input data being used in your GenAI system has been appropriately classified?[12]
- **Controls for AI System Protection:** Have you implemented controls to ensure the safety and integrity of AI systems? At the systems level, is there logging and control to prevent internal theft or unauthorized access? Are there safeguards against reverse engineering? At the model level, are there controls against risks like adversarial inputs and data poisoning?

Why this matters

- GenAI systems not designed for enterprise use often lack essential protections such as encryption, data governance, and administrative controls.
- Using personal accounts is prohibited by citywide cybersecurity policies.
- Improper account use increases the risk of unauthorized access, data leaks, and other cyber threats.
- If security controls are not in place, risks to intellectual property and data confidentiality may not be mitigated.

What you must do

- Use pre-approved City-managed accounts for all GenAI-related work.
- Comply with citywide cybersecurity policies.

- Confirm that all accounts comply with agency security policies, including Multi-Factor Authentication ("MFA"), log collection, and password complexity requirements.
- Ensure that proper logging and audit capabilities are in place to ensure that all activity in AI systems comply with citywide cybersecurity mandates.

Example scenarios

- **Good Practice:** Use a licensed account provisioned by your agency.
- **Prohibited:** Use a free AI image generator logged in through a personal account to create visuals for a City project that hasn't been released publicly.
- **Prohibited:** Logging into free web-based AI systems and inputting Sensitive and Restricted City Data.

5.5 Trust and Transparency

Would the public understand how this AI system is being used, and have you taken steps to ensure transparency, explainability, and the option for human oversight, appeal, or fallbacks where appropriate?

Guiding questions to ask

- **Output Attribution:** Have you clearly marked or disclosed that content was generated or modified by GenAI?
- **GenAI Use Disclosure:** Have you disclosed the use of this GenAI system under current reporting requirements (e.g., Local Law 35 of 2022)?[13]
- **Intellectual Property:** Does the incorporation of AI-Generated Content into City materials create copyright or other intellectual property risks?
- **Public Outreach and Human Fallbacks:** Does the relevant GenAI use pose risks to the City's reputation or relationship with the public? Is public outreach needed to support clear understanding of benefits and risks? Is there a need to provide a human "fallback" option for those who prefer not to engage with GenAI?

Why this matters

- Transparency is essential to preserving public trust in government. Without clear communication, the public may not understand how AI-Generated Content is produced, where human judgment was applied, or how to question AI-assisted outputs.
- Not disclosing AI use can lead to public confusion, diminished credibility, and legal risks—especially if outputs appear authoritative but contain errors or were produced without adequate oversight.

- Lack of attribution or unclear authorship can raise questions about intellectual property, accountability, or the legitimacy of content.[14]
- GenAI use can spark mistrust from the public, especially when goals, implementation details, and outcomes are not clear to those who may be impacted. Public engagement can be a critical mechanism to establish an accurate and nuanced understanding of public concerns and impacts, and to foster trust with New Yorkers.
- Making human alternatives or fallback mechanisms available can help ensure those who prefer not to use GenAI systems have trusted avenues to access relevant information and services.

What you should do

- Disclose when content is generated or assisted by GenAI.
- Follow agency protocols for reporting the use of GenAI systems, including pilot or experimental use cases.
- Ensure human review and accountability is in place, especially for content that affects public rights, services, or perceptions.
- Understand how copyright laws and City policies apply to GenAI-assisted content, and ensure adequate human contribution where required.
- Evaluate whether public engagement and/or human fallback mechanisms may be needed for your GenAI implementation, and work to integrate them.[15]

Example scenarios

- **Good Practice**: Note in a public newsletter that GenAI was used to generate initial drafts, with final edits by City personnel.
- **Required**: Report your use of a GenAI summarization tool as part of annual LL35 reporting, and any other tracking process, as required.
- **Poor Practice**: Release a policy explainer written by GenAI without clarifying authorship or verifying its legal accuracy.
- **Poor Practice:** Release a GenAI-based chatbot without engaging prospective users to understand needs, concerns, and impacts.
- **Poor Practice**: Publish GenAI-generated FAQs on a public website without staff review, attribution, or disclosure.

7.0 Changelog

Version	Change description	Authors	Date
1.0	Inaugural version	Alex Foard, Jiahao Chen and Renata Gerecke, Strategic Initiatives Division	03/04/2024
2.0	Guidance has been substantially elaborated and reformatted for utility, including integration, for each subsection of the Use Guidance, of: Guiding Questions, Why This Matters, What You Should Do, and Example Scenarios. Definitions have been updated to align with new content, and current City policy.	Hande Güven and Jessica Wang, Aspen Policy Academy; Alex Foard, Jiahao Chen, A. Kathryn Hohman, Soyoung Claire Park, Dean Labowitz, OTI	12/30/2025

8.0 Definitions/Glossary

Adversarial Input - Maliciously crafted Data intended as input to an AI system at Use Time to degrade the performance of the AI system, either by worsening the quality of its outputs, or by triggering vulnerabilities such as revealing sensitive training data.

Agency Privacy Officer - Staff designated by an Agency Head to act as its privacy officer whose responsibilities include Agency compliance with the Identifying Information Law and Citywide privacy policies.[16]

Agentic AI - An AI system with a level of autonomy that is Supervised or higher. In modern usage, Agentic AI refers to Generative AI systems that use AI-Generated Content to interact with other IT systems.[17] These actions are typically defined in terms of "tools," reusable computer instructions to perform standardized tasks such as web search, database access, or executing generated computer code; and "skills," which are reusable text describing how to use those tools that can be pasted into input prompts.

Artificial Intelligence ("AI") - A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.[18]
Autonomy - The extent to which the Operator is involved in acting on the AI system's outputs.

AI-Generated Content - Any content produced by GenAI.

Build Time - The part of the lifecycle of an AI system referring to the development of an AI model. At Build Time, i) a model is created or changed, and ii) the environment is not influenced by the AI.

City - The City of New York.

City Data - Data characterizing the City or its behavior; or Data owned, created, generated, stored, or maintained by, at the direction of, or for the benefit of the City; or any copies or derivatives of such Data.

Context Rot - The degraded performance of Generative AI that results from overly long inputs, resulting in Hallucinations or the production of unsafe content that bypasses internal safeguards.[19]

Data - Any information or representation(s) of information, knowledge, facts, ideas, concepts, or similar including any, texts, instructions, documents, databases, diagrams, graphics, drawings, images, sounds, or biometrics, that are communicated, created, generated, stored (in temporary or permanent form), filed, produced or reproduced, processed, or transmitted, in any form or media.

Data Poisoning - Manipulation of training data at Build Time to degrade the performance of the AI system at Use Time, either by worsening the quality of its outputs, or by triggering vulnerabilities such as revealing sensitive training data.

Fully Autonomous (level of Autonomy) - The AI system acts on its outputs without any Operator intervention.

Generative Artificial Intelligence ("Generative AI," "GenAI") - Any AI system whose primary function is to generate content, which can take the form of code, text, images, and more.[20]

Hallucination - AI-Generated Content that is plausibly correct at first glance, but is actually factually incorrect upon further inspection.[21]

Indirect Prompt Injection - a form of Adversarial Input or Data Poisoning of an Agentic AI system where data sources accessible to the Agentic AI system are manipulated with the intent that, when combined with a user-provided text prompt, results in degraded performance of the Generative AI system, either by worsening the quality of its outputs, or by triggering vulnerabilities such as revealing sensitive training data.[22]

Identifying Information - Any information obtained by or on behalf of the City that may be used on its own or with other information to identify or locate an individual.[23]

Information Technology ("IT") - Any services, equipment, or interconnected system(s) or subsystem(s) of equipment, that are used in the automatic acquisition, storage, analysis, evaluation, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information by the Agency.

Jailbreaking - The provision of malicious inputs into Generative AI systems, with the intent of bypassing safeguards to produce unsafe output.

Machine Learning - The study of computer algorithms that improve automatically through data, a subcategory of Artificial Intelligence.[24]

Monitored (level of Autonomy) - The AI system acts on its outputs unless the Operator overrides the action.

Multi-Factor Authentication ("MFA") - An authentication system that requires more than one distinct type of authentication factor for successful authentication. MFA can be performed using a multi-factor authenticator or by combining single-factor authenticators that provide different types of factors.[25]

Operator - The entity who, at Use Time, provides the input data to the computational model, oversees the execution of the computational model, and obtains its outputs.

Purpose - The role the AI system is intended to play in City government.

Restricted Data - Data with the highest level of sensitivity, whereof which the unauthorized disclosure, alteration, or destruction of such Data could be expected to have a severe or catastrophic adverse effect on the City's operations, organizational assets, or individuals.

Sensitive Data - Data intended only for internal City Agency use, where unauthorized disclosure, alteration, or destruction of such Data could be expected to have a serious adverse effect on the City's operations, organizational assets, or individuals or if such Data is only intended for internal use.

Software Regression - A change to system component that introduces additional defects that adversely affects functionality, reliability or performance.[26]

Supervised (level of Autonomy) - The AI system can act on its outputs, but requires explicit permission from the Operator to do so.

Sycophancy Bias - The phenomenon in which models tend to output answers that aim to please the user, leading to confirmation bias. [27]

Use Time - The part of the lifecycle of an AI system referring to the deployment of an AI model to serve its Purpose. At Use Time, i) the model is not changed, and ii) the environment is influenced by the AI.

Appendix A – Summary of Guiding Questions:

NYC AI Principle 1: Validity and Reliability	
Suitability for Purpose	Is GenAI the only feasible way to achieve the stated purpose? Can other predictive analytics or automation tools also achieve the stated purpose with similar performance? Given that GenAI outputs can change substantially even when given the same or similar inputs, how much variation in the output is acceptable for the stated purpose?
Output Verification	What processes are needed to flag and review errors in the GenAI system's outputs? How much human work, subject matter expertise, and judgment are needed to check and address these errors? Does your agency have the relevant personnel in place to do this work, now, and on an ongoing basis?
Error Impact and Remediation	How often are errors expected to occur? If an error is not corrected, how severe are the potential consequences? Might they include health, safety, legal, or regulatory ramifications, for example? What processes are needed to remediate errors? How will your agency ensure those processes are effective?
Performance Monitoring	How will the GenAI system's performance be monitored on an ongoing basis? What are the acceptable levels of performance? Is there a change management plan when metrics deviate from such acceptable levels?
NYC AI Principle 2: Social Responsibility	
Audience	Who is the audience for the output being generated: personnel within your agency, other agencies, or the public?
Impact on Decision-Making	Will the GenAI system's output influence policy, resource allocation, or services – or is it only for reference or ideation?
Bias and Equity	How might the development and use of the GenAI system create, sustain, or magnify discriminatory biases? Are there possible biases in the training data that may reinforce stereotypes, for example? Could the intended use case itself exclude specific populations, limit accessibility, or have other equity impacts?
Cost and Environmental Impact	Is this GenAI use impactful enough to outweigh its cost and energy demands? Does this use align with broader City sustainability goals?[28]

Worker Impact	Could this use of GenAI substantively change or disrupt staff members' work or roles? Could it raise concerns about workplace surveillance?
Public Engagement	Is engagement with relevant individuals and communities needed to fully understand and mitigate impacts?
NYC AI Principle 3: Information Privacy[29]	
Public Trust	Has the agency articulated a specific, legally permissible purpose and obtained APO approval for the use of identifying information in the GenAI system? If this GenAI system will involve the use of Identifying Information obtained from the public, will the public be informed?
Data Minimization	Will the minimum Identifying Information necessary be used for the stated purpose?
Accountability	Are you compliant with relevant information privacy policies, both citywide and agency-specific, and will additional policies be created to support the use of the GenAI system for its articulated purpose?
NYC AI Principle 4: Cybersecurity[30]	
Tool Authorization and Licensing	Are you using a GenAI system that has been reviewed, licensed, and approved for agency use—rather than a free GenAI product or unvetted version?
Account and Access Security	Are you using a City-managed or enterprise account with permissions and security controls that follows relevant City or agency policies?
Compliance	Are you compliant with relevant cybersecurity policies?
Data Governance	Have you made sure that input data being used in your GenAI system has been appropriately classified?[31]
Controls for AI System Protection	Have you implemented controls to ensure the safety and integrity of AI systems? At the systems level, is there logging and control to prevent internal theft or unauthorized access? Are there safeguards against reverse engineering? At the model level, are there controls against risks like Adversarial Inputs and Data Poisoning?
NYC AI Principle 5: Trust and Transparency	
Output Attribution	Have you clearly marked or disclosed that content was generated or

	modified by GenAI?
GenAI Use Disclosure	Have you disclosed the use of this GenAI system under current reporting requirements (e.g., Local Law 35 of 2022)?
Intellectual Property	Does the incorporation of AI-Generated Content into City materials create copyright or other intellectual property risks?
Public Outreach and Human Fallbacks	Does the relevant use of GenAI pose risks to the City's reputation or relationship with the public? Is public outreach needed to support clear understanding of benefits and risks? Is there a need to provide a human "fallback" option for those who prefer not to engage with GenAI?

Notes

[1] Recent GenAI developments have moved beyond content generation towards addressing specific user needs. Actions may include integrations with other systems (e.g. databases, code execution environments) that would allow GenAI to retrieve relevant documents or run complex calculations to generate better output. These enhanced systems are referred to as "Agentic AI." There is a full definition of this term in Section 8.0.

[2] Version 1.0 of this Use Guidance, New York City Office of Technology & Innovation, Preliminary Use Guidance: Generative AI, March4, 2024, is available at https://www.nyc.gov/assets/oti/downloads/pdf/about/preliminary-use-guidance-general-artificial-intelligence.pdf

[3] Local Law 35 of 2022, Administrative Code of the City of New York, § 3-119.5, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA29B34-9266-4B52-B438-++A772D81B1CB5.

[4] New York City Office of Technology & Innovation, Artificial Intelligence: Principles & Definitions (v1.0), March 5, 2024, https://www.nyc.gov/assets/oti/downloads/pdf/about/artificial-intelligence-principles-definitions.pdf

[5] The terms of art are "nondeterminism" and "non-robustness" respectively. Reducing the sampling temperature to zero in modern GenAI systems reduces, but does not eliminate, these issues.

[6] This phenomenon is called a Software Regression.

[7] NYC Identifying Information Law (NYC Admin. Code §§ 23-1201 to 23-1205), https://codelibrary.amlegal.com/codes/newyorkcity/latest/NYCadmin/0-0-0-202942 (accessed December 4, 2025).

[8] New York City Office of Technology & Innovation, Citywide Privacy Protection Policies and Protocols (CPPPP) (v4.1), June 13, 2025, available to City personnel at https://cityshare.nycnet/html/informationprivacy/downloads/pdf/2025/2025_Citywide_Privacy_Protection_Policies_and_Protocols.pdf (accessed December 4, 2025).

[9] New York City Office of Technology & Innovation Privacy Impact Assessment (PIA), 2025, available to City personnel at https://cityshare.nycnet/html/informationprivacy/downloads/pdf/2025_Privacy_Impact_Assessment.pdf (accessed December 4, 2025).

[10] New York City Office of Technology & Innovation, (a) Citywide Privacy Protection Policies and Protocols (CPPPP) (v4.1), June 13, 2025, available to City personnel at https://cityshare.nycnet/html/informationprivacy/downloads/pdf/2025/2025_Citywide_Privacy_Protection_Policies_and_Protocols.pdf (accessed on December 4, 2025) and (b) Guidance for Assessing Contextual Integrity, available to City personnel at https://cityshare.nycnet/html/informationprivacy/downloads/pdf/2025_guidance_for_assesing_contextual_integrity.pdf (accessed on December 4, 2025).

[11] New York City Office of Technology & Innovation, Citywide Privacy Protection Policies and Protocols (CPPPP) (v4.1), June 13, 2025, available to City personnel at https://cityshare.nycnet/html/informationprivacy/downloads/pdf/2025/2025_Citywide_Privacy_Protection_Policies_and_Protocols.pdf (accessed on December 4, 2025).

[12] New York City Office of Technology & Innovation, (a) Citywide Data Classification Policy (No. P-03-PR-DS), August 21, 2023, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/P-01-PR-DS_Citywide_Data_Classification_Policy-Non-Restricted.pdf; and (b) Citywide Data Classification Standard (No. S-03-PR-DS), August 21, 2023, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/S-01-PR-DS_Citywide_Data_Classification_Standard-Sensitive.pdf

[13] Local Law 35 of 2022, Administrative Code of the City of New York, § 3-119.5, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA29B34-9266-4B52-B438-++A772D81B1CB5.

[14] The topic of copyright and AI is complex and rapidly evolving, as analyzed in the US Copyright Office's three-part report. https://www.copyright.gov/ai/ (accessed November 28, 2025).

[15] New York City Office of Technology & Innovation, AI Public Engagement and Participation Guidance, v1.0, December 15, 2025.
[16] New York City Office of Technology & Innovation, Citywide Cybersecurity Program Glossary (v1.3), March 7, 2025.
[17] Anthropic, What is the Model Context Protocol (MCP)?, https://modelcontextprotocol.io/docs/getting-started/intro (accessed November 28, 2025).
[18] National Artificial Intelligence Initiative Act of 2020, 15 U.S.C. ch. 119 § 9401(3), https://uscode.house.gov/view.xhtml?path=/prelim@title15/chapter119&edition=prelim
[19] The term was coined by the user Workaccount2 on the Hacker News forum on June 18, 2025, in the thread titled "Is there a half-life for the success rates of AI agents?", https://news.ycombinator.com/item?id=44310054 (accessed November 28, 2025). This concept was developed in Simon Willison's Weblog, June 29, 2025, https://simonwillison.net/2025/Jun/29/how-to-fix-your-context/ (accessed November 28, 2025); and Kelly Hong, Anton Troynikov and Jeff Huber, Chroma Technical Report, Context Rot: How Increasing Input Tokens Impacts LLM Performance, https://research.trychroma.com/context-rot (accessed November 28, 2025).
[20] Helen Toner, What Are Generative AI, Large Language Models, and Foundation Models?, Center for Security and Emerging Technology, May 12, 2023, https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/; archived at https://web.archive.org/web/20250822172509/https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/.
[21] The term "hallucination" in the AI context first appeared in the academic literature on machine translation: see Phillipp Koehn and Rebecca Knowles, Six Challenges for Neural Machine Translation, in Proceedings of the First Workshop on Neural Machine Translation, pages 28–39, Vancouver. Association for Computational Linguistics. https://aclweb.org/anthology/W17-3204. The modern usage of "hallucination" to refer to plausibly correct errors in generated text more generally can be traced back to Joshua Maynez, Shashi Narayan, Bernd Bohnet and Ryan McDonald, On Faithfulness and Factuality in Abstractive Summarization, in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1906–1919, Online. Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.173.
[22] Kai Greshake, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz, Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. In Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISec '23). Association for Computing Machinery, New York, NY, USA, 26 November 2023, pages 79–90. https://doi.org/10.1145/3605764.3623985
[23] Local Law 247 of 2017, Administrative Code of the City of New York, § 23-1201, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3022112&GUID=A5F00A64-0D35-4B82-BD51-9F46A6A72E10.
[24] National Institute of Standards and Technology (NIST), The Language of Trustworthy AI: An In-Depth Glossary of Terms, March 29, 2023, doi:10.6028/NIST.AI.100-3.
[25] National Institute for Standards and Technology (NIST), Special Publication 800-63B-4, Digital Identity Guidelines: Authentication and Authenticator Management, July 31, 2025, https://doi.org/10.6028/NIST.SP.800-63b-4.
[26] Adapted from ISO/IEC/IEEE 24765:2017, Systems and software engineering — Vocabulary, §3.3371, "regression testing." https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:24765:ed-2:v1:en:term:3.3371
[27] Mrinank Sharma, Meg Tong, Tomasz Korbak, et al., "Towards Understanding Sycophancy in Language Models," October 23, 2023, available at https://www.anthropic.com/research/towards-understanding-sycophancy-in-language-models (accessed on December 4, 2025).
[28] Mayor's Office of Climate & Environmental Justice, PlaNYC: Getting Sustainability Done, April 2023, https://www.nyc.gov/content/climate/pages/reports-and-publications/planyc.
[29] New York City Office of Technology & Innovation, (a) Citywide Privacy Protection Policies and Protocols (v4.1), June 13, 2025, https://www.nyc.gov/assets/oti/downloads/pdf/reports/cpo/2025%20Citywide%20Privacy%20Protection%20Policies%20and%20Protocols_web.pdf; and (b)

Agency Privacy Officer Toolkit (v3.0), January 28, 2025, https://www.nyc.gov/assets/oti/downloads/pdf/reports/cpo/2025%20Agency%20Privacy%20Officer%20Toolkit%20_web.pdf.

[30] New York City Office of Technology & Innovation, (a) Citywide Application Security Policy (No. P-AS-01), November 1, 2018, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/P-AS-01-Citywide-Application-Security-Policy.pdf; (b) Citywide Application Security Policy and Standard (No. S-AS-01), November 1, 2018, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/S-AS-01-Citywide-Application-Security-Standard.pdf; (c) Citywide Cybersecurity Program (No. D-ID-GV-01, v1.0), October 23, 2019, available to City personnel at https://cityshare.nycnet/html/cityshare/downloads/it_wireless/info_security_policies/Non-Restricted-CSP-Citywide-Cybersecurity-Program.pdf; and (d) Citywide Cybersecurity for the Usage and Development of AI Systems Policy (No. P-08-PR-DS, v1.0), January 27, 2025, available to City personnel https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/P-08-PR-DS-Citywide-Cybersecurity-Usage-and-Development-of-Artificial-Intelligence-Systems-Policy-Non-Restricted.pdf.

[31] New York City Office of Technology & Innovation, (a) Citywide Data Classification Policy (No. P-03-PR-DS), August 21, 2023, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/P-01-PR-DS_Citywide_Data_Classification_Policy-Non-Restricted.pdf; and (b) Citywide Data Classification Standard (No. S-03-PR-DS), August 21, 2023, available to City personnel at https://cityshare.nycnet/intranets/cityshare_home/assets/cityshare_home/downloads/pdf/it-telecom/information-security-policies/S-01-PR-DS_Citywide_Data_Classification_Standard-Sensitive.pdf.