San Francisco Generative AI Guidelines (2025)

For All City and County personnel, including employees, contractors, consultants, volunteers, and vendors working on behalf of the City

July 7th, 2025

Top 5 Guidelines for Using Generative AI

1. Whether generated by AI or a human, you are ultimately responsible for any content you use or share.
2. Copilot Chat is available to City employees as a secure option for most Generative AI tasks. Other secure enterprise tools may also be available to staff. We strongly discourage the use of Generative AI tools that have not been purchased or vetted by the City. If you must use public or consumer Generative AI tools, never enter sensitive, confidential, or City data, as these tools may store or reuse the information you provide.
3. Always review, edit, fact-check, validate, and/or test AI generated content, which isn't always accurate.
4. Always disclose AI use when it contributes to public-facing or sensitive work or required by regulations. Ensure the Generative AI tool is properly recorded in the City's 22J inventory and provide direct notice to impacted individuals.
5. Never use Generative AI to generate deepfakes—fake images or recordings--or other content that could be mistakenly interpreted by someone to be real.

1. Introduction

Enterprise Generative AI (GenAI) tools procured and licensed by the Department of Technology (DT) are now available for use by staff of the City and County of San Francisco (City), opening new opportunities to improve the effectiveness, efficiency, and responsiveness of City services for all San Franciscans.

Unlike other AI technologies used by the City—which support informed decisions based on input data—GenAI tools can generate new content based on patterns identified in large datasets, often in a matter of seconds. Examples include text, images, music, and code.

However, the use of GenAI technology by the City also poses unique risks to its workers, residents, and visitors. While AI-generated content can appear authoritative and polished, it can be inaccurate, biased, or misleading. GenAI use can also heighten the risk of privacy breaches, unauthorized data sharing, and cybersecurity threats. Furthermore, overreliance on GenAI for decisions that affect the public's rights or safety can reduce transparency, weaken accountability, and erode trust in government.

1.1. Purpose of the Guidelines

These guidelines are designed to help City staff use GenAI tools effectively and responsibly, while maintaining public trust, protecting resident data, and preserving the integrity of City systems.

Although GenAI tools provided by DT offer stronger privacy and security protections compared to public or consumer AI tools, they still pose significant risks. These include bias, misinformation, factual errors, hallucinations, copyright violations, and inconsistent outputs. Such risks are heightened when City employees rely on these tools without exercising the necessary human oversight. To mitigate these risks, all GenAI outputs should be carefully reviewed before use to ensure accuracy, fairness, and alignment with City values and policies. City staff must also adhere to existing data protection and governance requirements, and provide transparency to the public about when and how GenAI tools are being used, in compliance with the City AI Transparency Ordinance (Chapter 22J).

1.2. Scope of the Guidelines

With some important distinctions, **the guidelines outlined in Sections 2.1 through 2.3 of this document apply to both:**

- **Enterprise GenAI tools**—like ChatGPT Enterprise, Microsoft Copilot, Snowflake Cortex, Adobe apps and Express, and other approved systems—licensed and managed through the Department of Technology (DT). These tools:
    - Have been **procured and configured** for City use;
    - **Allow use of sensitive City data (and restrict any vendor use of City data for AI training).** This includes protected health information (PHI) and personally identifiable information (PII) when using tools that have Business Associate Agreements (BAAs) in place, such as Microsoft Copilot and Snowflake Cortex;
    - Have **passed cybersecurity risk assessments;** and
    - Support the City's data retention standards and obligations for public request for access

- **Public or consumer GenAI tools** — such as free online chatbots or apps not procured or managed by the City. These tools do not provide adequate privacy protections, and information shared with them is often used to train the underlying models, posing potential confidentiality risks.

While the use of public or consumer GenAI tools for City business is strongly discouraged, we recognize that such tools may still be used in limited circumstances. **If you want to use public or consumer GenAI tools, you must obtain prior departmental approval and follow the precautions outlined in Section 2.4.**

2. City Guidelines for Generative AI Use

The GenAI uses outlined below are grouped by risk level, each with corresponding mitigation strategies and disclosure requirements.

As a general rule, City employees must always thoroughly review, edit, fact-check, validate, and/or test their output, as applicable. **You are ultimately responsible for any content you use or share.**

2.1. Low-Risk Use: Internal Efficiency Tasks Performed Using Enterprise Generative AI Tools

You may use City-procured AI tools for:

- Drafting internal emails, memos, or communications
- Creating summaries of meetings, documents, or reports
- Writing, editing, or debugging code
- Generating outlines or first drafts of internal materials
- Improving language access between the general public and City staff.

These uses help improve efficiency and reduce workload, but you remain the expert reviewer.

Safeguards and Responsibilities

Even when handling simple tasks, AI can make errors, omit context, or return outdated or biased information. There's no such thing as zero risk. To ensure responsible and safe use of AI tools:

- For summaries or memos, only use material you are already familiar with so you can spot issues. You should be able to independently evaluate the quality and correctness of the output.
- Always double-check factual claims, hyperlinks, and references to ensure content is backed by evidence.
- Always review and edit AI's output critically before using or sharing it.
- Only use GenAI for coding tasks if you already know the programming language. You must be able to review, debug, and test any AI-generated code.

Disclosure Requirements

Disclosure is not required when using AI for internal drafting or support. You remain fully responsible for any content you use or share, including any errors or omissions introduced by AI.

2.2. Medium- to High-Risk Use: Public-Facing or Sensitive Work Performed Using Enterprise Generative AI Tools

Use extra caution and follow additional steps when City-approved AI tools are used to perform tasks affecting public communication, services, or decisions such as:

- Drafting or translating public-facing content.
- Drafting interview questions and screening materials for hiring processes.
- Summarizing policy-related data.
- Supporting decisions related to services, enforcement, or eligibility.
- Contributing to documents that affect regulation or safety.

For these use-cases, AI can serve as a support tool, but it should never make final decisions that affect individuals or public outcomes.

Safeguards and Responsibilities

To ensure responsible and safe use of Enterprise AI tools used to perform public-facing or sensitive work:

- Only use GenAI if you have deep subject-matter expertise to review its output. This ensures you can spot errors, detect harmful implications, review suggestions critically, and make informed decisions based in AI-generated content.
- Review and edit AI-generated outputs to ensure they reflect the City's values, promote equity, and uphold ethical standards and digital accessibility standards including the use of inclusive language and design practices that serve residents with visual, cognitive, or motor impairments.
- Actively monitor for instances of bias and correct them manually.

Disclosure Requirements

When using Enterprise GenAI tools for public-facing or sensitive work, usage should be properly documented through the 22J process.

To promote transparency toward the public, individuals impacted by AI systems must receive a direct notice disclosing that GenAI substantially contributed to the work product. At minimum, this direct notice must include: a clear plain-language, multilingual statement that GenAI was used; the name and version of the tool used; a note indicating that the content was reviewed by City staff; and contact information for questions, appeals, or corrections.

> ***Direct Notice – Sample Template (please check with your department for specific requirements or approved language)***
>
> *This content was generated with the assistance of [Tool Name, Version] and reviewed by City staff for accuracy.*
> *For questions, appeals, or corrections, contact: [Email].*
>
> *View this notice in: [Español] · [中文] · [Tagalog]*
> *Or contact [email] for additional options.*

Depending on the type of content and its intended audience, other regulatory requirements related to disclosures or disclaimers, such as those under AB 3030, may also apply. Staff should always consult with their department for specific guidance on compliance with applicable laws and policies.

Furthermore, if you **paraphrase, quote, or incorporate any AI-generated content** into your own work (whether text, image, data, or other), you should **cite it appropriately, j**ust as you would any external source. Established citation guidelines, such as those from the MLA Style Center, recommend including the following details: a description of the **prompt; t**he **name and version** of the AI tool; the **developer/company; t**he **date of interaction; and a URL** to the tool or session, if applicable.

Example: Quoting AI-Generated Text

When asked to describe the symbolism of the green light in The Great Gatsby, ChatGPT provided a summary about optimism, the unattainability of the American dream, greed, and covetousness. However, when further prompted to cite the source on which that summary was based, it noted that it lacked "the ability to conduct research or cite sources independently" but that it could "provide a list of scholarly sources related to the symbolism of the green light in The Great Gatsby" ("In 200 words").

Works-Cited-List Entry (MLA Style):

"In 200 words, describe the symbolism of the green light in *The Great Gatsby*" follow-up prompt to list sources. *ChatGPT*, 13 Feb. version, OpenAI, 9 Mar. 2023, https://chat.openai.com/chat.

If a GenAI tool references or summarizes secondary sources (e.g., articles, studies, or reports), you must verify the original material and cite the primary source directly, just as you would in traditional research.

2.3. Prohibited Uses

To protect public trust, safety, and ethical standards, do **not** use GenAI tools for any of the following:

- Relying on AI to create City official documents or make decisions without expert human review.
- Generating images, audio, or video that could be mistaken for real people (including public officials or members of the public).
- Creating "deepfakes" or impersonations of any person or official—even with disclaimers.
- Fabricating fictional survey respondents or public input for research or outreach purposes.
- Relying on AI to review legal or regulatory issues.

2.4. Using Public or Consumer Generative AI Tools: Additional Guidance

The use of public or consumer Generative AI tools in place of City-approved enterprise solutions is strongly discouraged. Any experimental use of non-approved Generative AI technologies must receive prior departmental approval.

To ensure the security of City systems and data and best serve the public when using public or consumer GenAI tools, follow this guidance in addition to the ones outlined for Enterprise GenAI Tools:

- **Never enter sensitive or protected data that cannot be fully released to the public, including personal information, health information, City data, and/or financial information.** This information can be viewed by the companies that make the tools and, in some cases, other members of the public. Once entered, this information becomes part of the public record. The handling and disclosure of sensitive information is already governed by several City policies, including but not limited to:

  - Charter Section 16.130, Privacy First Policy
  - Administrative Code Section 12M.2(a), Nondisclosure of Private Information
  - Campaign & Governmental Conduct Code section 3.228, Disclosure or Use of Confidential City Information
  - The "Computers and Data Information Systems" section of the Department of Human Resource's Employee Handbook (January 2012, page 48)
  - Please refer to Citywide Data Classification Standard for more specifics on data classification and department responsible roles

- **Never conceal the use of public or consumer GenAI tools** from your coworkers and remain transparent about when and how these tools are used in your work.

2.5. Guidance for Departmental IT Leaders

Departmental IT leaders have a responsibility to support right-sized GenAI uses that deliver the greatest public benefit. This begins with ensuring that AI projects are problem-led and not technology-led, by centering the specific needs and challenges faced by the department and the communities it services.

In parallel, Departmental IT leaders must ensure responsibility and transparency in how GenAI tools are used, especially when they could influence department decisions, erode the public's rights or safety, or affect access to critical services. This responsibility includes ensuring that their department complies with the transparency requirements set forth in Chapter 22J of the Administrative Code.

To support these goals, IT leaders should adhere to the following guidance:

- If your department is considering adopting or piloting a new GenAI tool, please reach out to the Emerging Technology Team at ai@sfgov.org early in the process to ensure alignment with citywide standards, policy requirements, and support.
- When purchasing technology that includes GenAI, collect the information required by 22J from vendors during contract execution or shortly thereafter. This includes the name of the technology and vendor; a description of its purpose, function, intended use, and operational context; training and generated data; an explanation of how the technology works; what it is optimizing for and its accuracy (preferably with numerical metrics); conditions affecting performance; testing for bias (including results); procedures for reporting issues or incidents; oversight mechanisms; and whether collected data may be used to train proprietary or third-party systems.
- Designate at least one 22J Lead responsible for:
  - Managing compliance with 22J.
  - Coordinating inventory submissions and acting as the point of contact for the Emerging Technology Team.
  - Determining whether each GenAI technology your department uses--or is planning to procure, borrow, or receive--qualifies for an exemption under 22J.
  - Submitting required information for non-exempt GenAI tools under 22J through LogicGate 22J or an MS form.

    - Notifying DT of any updates if GenAI tools are decommissioned, replaced or added to your department's inventory.
    - Supporting annual compliance reviews by the City Controller.
- Encourage staff to participate in GenAI–related training offered by the City to understand risks, use cases, and functionalities of specific Enterprise GenAI tools.
- If a Gen-AI feature becomes available to staff without your department's prior approval, contact the vendor to learn about privacy and security implications, and deactivate the functionality if needed. If the tool is licensed and managed by DT, report the issue immediately to the Emerging Technology Team at ai@sfgov.org.

3. Data Protection Requirements

The use of City data in Enterprise AI tools is subject to the following restrictions:

- For **Copilot Chat** and **Snowflake**, you can use **Level 4 data and below** (Levels 1–4).
- For **ChatGPT Enterprise**, you can use **Level 3 data and below** (Levels 1–3).
- **Only use PHI (Protected Health Information) in tools that have a BAA (Business Associate Agreement) in place, such as Copilot Chat and Snowflake subject to your department's approval**.
- To verify which types of department-specific data are permitted for use with City-approved tools, always check with your Department.
- **Do not enter any sensitive or protected data including personal information, health information, and/or financial information into public or consumer AI tools** not provisioned or approved for City use.

Content entered into or generated by GenAI tools may constitute public records and may be subject to disclosure under the **California Public Records Act (CPRA)**, as well as the **City and County of San Francisco's public access and records retention requirements**, including the **Sunshine Ordinance**. For Copilot Chat, user inputs and AI-generated content are retained for 30 days, consistent with the retention period for Microsoft Teams. For ChatGPT Enterprise, content is retained for 90 days.

4. Data Governance

Generative AI tools depend on accurate, clean, and well-organized data. If City data is outdated, inconsistent, or poorly maintained, AI outputs may be inaccurate or misleading.

Every department has a role in:

- Keeping records and metadata up to date
- Using standardized formats per the City's Data Management Policy
- Ensuring data is managed according to City policies

Strong data governance supports reliable and fair AI usage.

5. Background

5.1. Current Legislative and Regulatory Landscape

Since the release of the City's first GenAI guidelines in December 2023, the national regulatory landscape has shifted significantly. President Biden's 2023 Executive Order, which prioritized AI safety, security, and responsible use, was rescinded in early 2025 and replaced with a lighter-touch, pro-innovation policy that emphasizes deregulation and national competitiveness over public safeguards. At the state level, Governor Newsom reaffirmed California's leadership in responsible AI governance with the release of a *Report on Frontier AI Policy* in June 2025. The report highlights growing risks--like AI-generated scams, disinformation, and threats involving dangerous materials--and calls for strong, practical safeguards. These include enhanced transparency, independent evaluations, whistleblower protections, and systems to track harmful incidents. Built on a "trust but verify" approach, the report argues that thoughtful regulation supports, rather than hinders, responsible innovation.5.2. San Francisco: the Home of AI

San Francisco is at the center of the AI boom, with many of the world's most influential AI companies--including OpenAI, Anthropic, Databricks--as well as a growing ecosystem of startups and research labs headquartered in the City. As the home of AI, San Francisco is also on the front lines of addressing its real-world impacts on residents, workers, and public services. Given the scale of innovation underway, and the potential consequences of inaction, the City has both a responsibility and an opportunity to lead in the civic use of AI through responsible and accountable governance.

While the City will continue to track federal and state developments, it is also charting its own course under Mayor Lurie's leadership by embracing intentional, practical AI adoption grounded in ethical standards and public trust—one that delivers more effective, efficient, and responsive services to all San Franciscans.

5.3. Development of Guidelines

The Emerging Technology Team developed these updated GenAI-focused guidelines in close coordination with the City's AI Advisory Committee, a staff working group that provides guidance on the adoption, governance, and ethical use of emerging technologies in the City. **6. Contact & Versioning**

The AI Advisory Committee will regularly update the City Guidelines for Generative AI Use to reflect new law, regulations, lessons learned from application, and developments in GenAI technology. Check these Guidelines regularly for updates, and bookmark to stay informed.

For questions or help with tool selection, training opportunities, or policy interpretation please check with the Emerging Technology Team at ai@sfgov.org.

This report includes content drafted with support from ChatGPT Enterprise (GPT-4o, OpenAI, June 2025). All content was reviewed and finalized by the Emerging Technology Team.

Glossary

Algorithms: are a set of rules that a machine follows to generate an outcome or a decision.

Artificial Intelligence (AI): refers to a group of technologies that can perform complex cognitive tasks like recognizing and classifying images or powering autonomous vehicles. Many AI systems are built using machine learning models. For a task like image recognition, the model learns pixel patterns from a large dataset of existing images and uses these patterns to recognize and classify new images.

Auditability for AI: AI where the outputs are explainable, monitored and validated on a regular basis.

Bard: is a conversational Gen AI chatbot built by Google

Black box models: are those where you cannot effectively determine how or why a model produced a specific result.

Chatbots: are computer programs that simulate conversations. Chatbots have been around for a few decades. Basic chatbots (without Gen AI) use ML to understand human prompts and provide more-or-less scripted answers that can guide users through a process. Gen AI chatbots can provide more human-like, conversational answers.

ChatGPT: is a conversational Gen AI chatbot built by OpenAI

Dall-e: is a Gen AI application that can generate images based on text prompts

Discriminative AI: In contrast to Gen AI, Discriminative AI models do not generate new content but can be used to predict quantities (for example, predicting home prices) or to assign group membership (for example, classifying images).

Enterprise Generative AI Tool: City-approved GenAI tool procured and managed by the Department of Technology or another City department. These tools are configured for City use, support sensitive data (with BAAs where applicable), meet cybersecurity standards, and follow strict privacy, legal, and data retention requirements.

Generative AI (Gen AI): refers to a group of technologies that can generate new content based on a user provided prompt. Many are powered by LLMs.

Large language models (LLMs): are a type of machine learning model trained using large amounts of text data. These models learn nuanced patterns and structure of language. This allows the model to understand a user generated prompt and provide a text response that is coherent. The responses are based on predicting the most likely word in a sequence of words and as a result, the answers are not always contextually correct. The training datasets used to build these models can contain gender, racial, political and other biases. Since the models have learnt from biased data, their outputs can reflect these biases. Generative AI applications are built using these LLMs.

Machine Learning (ML): is a method for learning the rules of an algorithm based on existing data.

Machine learning model: is an algorithm that is built by learning patterns in existing data. For example, a machine learning model to predict house prices is constructed by learning from historical data on home prices. The model may learn that price increases with square footage, changes by neighborhood, and depends on the year of construction.

Microsoft Copilot: an AI-powered assistant integrated across Microsoft 365 applications that helps users draft content, summarize emails, analyze data, and automate tasks.

Model validation: methods to determine whether the outputs generated by a machine learning model are unbiased and accurate.

Public or consumer Generative AI Tool: free or third-party Generative AI tool not managed by the City.

Snowflake Cortex: a suite of built-in Generative AI and machine learning capabilities within the Snowflake data platform. Cortex allows users to securely analyze, summarize, and generate insights from data using large language models, all within the City's existing Snowflake environment.

Training data: The dataset that is used by a machine learning model to learn the rules.