# Generative AI Guidelines

Last updated: April 24, 2025

## Executive Summary:

Generative Artificial Intelligence (AI) is a branch of AI technology that interacts with user-provided creativity to generate content—such as memos, letters, proposals, images, voice, and music— at the request of a user. Many organizations have banned Generative AI, while others allow unrestricted usage. The City of San José (City) recognizes the opportunity for a controlled, ethical, and responsible approach to use AI that acknowledges the benefits to efficiency while mitigating the risks around AI bias, privacy, and cybersecurity.

Previous iterations of the Generative AI Guidelines were released prior to the creation of a City-wide AI Policy. Recognizing the collaborative work between City staff, San José residents, the Digital Privacy Advisory Task Force, and countless partner government agencies in the GovAI Coalition, the City published its AI policy (City Policy Manual 1.7.12), governing these guidelines.

**At a baseline, users must follow these rules while using Generative AI for City work or using City resources, including direct services like ChatGPT, Zoom AI, and Copilot and extensions like Compose.ai:**

1. Information entered into Generative AI systems may be subject to a Public Records Act (PRA) request, may be viewable and usable by the company, and may be leaked unencrypted in a data breach. Unless using a City-issued Generative AI tool, users are prohibited from using any prompts that may include information not suitable for public release (such as confidential or personally identifiable information).
2. Review, revise, and fact check via multiple sources any content created from a Generative AI system as these systems can make mistakes. Users are responsible for the accuracy and appropriateness of material created with AI assistance.
3. Limit the environmental impact of your Generative AI usage by using search engines for general questions, using fewer prompts where feasible, and putting boundaries on

outputs when you only need a few bullet points or a short summary (e.g., instructing Generative AI to "respond in less than 100 words.").

4. Always cite images, video, and audio generated by AI. Cite other content generated entirely by AI. See more detail on how and when to cite in the "Citing Generative AI" section on page 10. Citing Generative AI maintains trust with our residents that content published by the City is reliable and overseen by human City staff.
5. Record at least one instance of your Generative AI use via the City's Generative AI Form. This informs improvements to staff training and AI guidance.
6. Create an account specifically for City use, if a Citywide staff account was not provided, to ensure public records are kept separate from personal records. Review "Getting started with Generative AI for City use" on page 8 for more details. If a user agrees to the terms and conditions of a system that the City does not have a formal agreement with, the user is responsible for complying with those terms and conditions.
7. Departments may provide additional rules around Generative AI. Consult your manager or department contact if there are additional department-specific rules.
8. Refer to this document quarterly, as guidance will change with the technology, laws, and industry best practices. Check the "Change Log" to identify changes. Bookmark the Information Technology Department Generative AI Guidelines webpage for easy access to the latest on generative AI guidelines at the City. You may also subscribe for updates to San José's Generative AI Guidelines.
9. Users are encouraged to participate in the City's established workgroups to help advance AI usage best practice in the City and enhance the Guidelines. Review "Joining AI Working Group" section on page 10.
10. Attend AI training sessions or review online training session conducted by ITD. Staff can express interest to join upcoming AI upskilling cohorts.

# Table of Contents

## Change Log

| Date | Content |
|---|---|
| July 20, 2023 | First Release |
| September 23, 2023 | New Sections:<br>1. **"Ownership of Generated Content."** Emphasized users' responsibility to verify, edit, and manage the generated content<br>2. **"Verify the Copyright of All Generated Content."** Advised users to verify content does not infringe copyright and, if unsure, to edit before using.<br>3. **"Understand the Terms and Conditions."** Emphasized need for users to review terms and conditions.<br>Edited Sections:<br>1. **"Create an Account Specific for City Usage."** Clarified that inputs and outputs of a Generative AI system may be subject to a PRA request.<br>2. **"Accuracy."** Noted that Generative AI companies do not guarantee the content they generate is accurate.<br>Miscellaneous Edits:<br>1. Clarified applicability of Public Records Acts requests;<br>2. Added acknowledgement of Generative AI extensions;<br>3. Reinforced the date limit of ChatGPT;<br>4. Refined names of principles to match the evolving terminology;<br>5. Updated link to usage reporting form and to the latest Generative AI Guidelines;<br>6. Updated acceptable uses of Generative AI for programming; |
| April 24, 2025 | Edited Sections:<br>1. **"Executive Summary" and other locations.** Added statement "Unless you have been approved by the Information Technology Department (ITD) and are using a City enterprise Generative AI tool, do not use any prompts that may include information not suitable for public release"<br>2. **"Record usage of Generative AI."** Clarified that reporting only needs to happen once per user per type of use to support the City in understanding value of Generative AI to City.<br>3. **"Cite Generative AI."** Clarified purpose of citation is about building trust with residents and when to cite. |

| | |
|---|---|
| | 4. **"Create an account specifically for City-related work."** Clarified work account clearly separates usage of Generative AI for work vs. personal in the event of a Public Records Act Request.<br><br>5. **"Appendix: High Risk and Prohibited Use Cases."** Moved "coding" to high-risk use case rather than prohibited use case.<br><br>6. **Adjusted language risk classifications to "Low Risk", "Mid Risk" and "High Risk" to align with** City AI Handbook<br><br>New Sections:<br><br>1. **"Limit the Environmental Impacts of Your Generative AI Usage."** Explains how to be thoughtful and efficient in using Generative AI to reduce environmental impact and improve quality of your usage |

## Definitions

**User**: Staff, contractors, interns/fellows, volunteers, or others using Generative AI for City work purposes.
**City**: The city government of San José located in California, United States of America.
**Generative AI**: A machine that automatically creates content such as text, audio, or image.
**Artificial Intelligence (AI)**: Machines doing tasks that typically require human intelligence.
**Machine Learning**: A type of AI in which computers use data to "learn" tasks through algorithms.
**Algorithm**: A set of steps or procedure designed to perform a task, such as mathematical operations (e.g., addition) or logical rules, to solve a problem or perform a task.

## Purpose of Guidelines

"Generative AI," such as ChatGPT, grew from a niche topic to a variety of publicly available tools with hundreds of millions of adopters in less than one year. Among other things, Generative AI presents an incredible opportunity for people to increase their efficiency and efficacy in work. Generative AI has also been used for several irresponsible applications including faking news headlines,[1] leaking personal information,[2] and enabling phishing cyber-attacks.[3]

The City is actively working to create policies and procedures around AI in general. This document serves as part of an evolving governance structure around responsible AI usage.

## Application of the Guidelines

This document applies to all use of Generative AI by a City staff member, contractor, volunteer, or other person while performing a role for the City (collectively "users"). This document does not apply to users of Generative AI for personal purposes or business purposes unassociated with the City.

Generative AI does not refer to algorithms that a person directly defines. For example, a spreadsheet a human created to calculates taxes owed based on income is not "Generative AI." A general rule is that if you cannot write the system's entire algorithm, either because you do not understand the math or because it would take years to write down, then it is probably AI.

[1] "ChatGPT is making up fake Guardian articles. Here's how we're responding," The Guardian
[2] "OpenAI says a bug leaked sensitive ChatGPT user data," Engadget
[3] "Council post: How ai is changing social engineering forever," Forbes

**Departments may provide additional rules on the usage of Generative AI.** Users should consult their manager if there are additional rules specific to their department.

## Principles for Using Generative AI

Usage of Generative AI shall follow the City's AI principles:

2. **Privacy: Submit information to Generative AI tools that is ready for public disclosure.** This includes any text, photos, videos, or voice recordings you share with the AI. Be mindful that the AI output may include unexpected personal information from another user and ensure removing any potential private information before publishing.
3. **Accuracy**: The City maintains trust with its residents and partners by providing accurate information. **Review and fact check all outputs you receive from a Generative AI**. Users should consult trustworthy sources to confirm that the facts and details in the AI-generated content are accurate. Trustworthy sources include official City documents and peer-reviewed journals. Consult your supervisor for other trustworthy sources (e.g., newspapers, blogs, or datasets). Be aware that many systems, like ChatGPT, may only use information up to a certain date (e.g., 2021 for ChatGPT) and cannot guarantee the content they generate is accurate.
4. **Transparency**: The user shall be clear when using Generative AI. **This can often include citing that you used AI in creating a product.** See how and when to cite Generative AI in the "Citing Generative AI" section under "Guidance while Using Generative AI."
5. **Equity:** AI system responses are based on patterns and relationships learned from large datasets derived from existing human knowledge, which may contain errors and is historically biased across race, sex, gender identity, ability, and many other factors. **Users of Generative AI need to be mindful that Generative AI may make assumptions based on past stereotypes and need to be corrected.** Establish guidelines to address equity as it relates to services in your department.
6. **Accountability**: **The person using AI is accountable for the content it generates.** Use Generative AI with a healthy dose of skepticism. The level of caution used should correspond to the risk level of the use case (see "Assessing Risk in Generative AI Use Cases"). It is always important to verify information provided by Generative AI.
7. **Beneficial:** User should be open to responsibly incorporating Generative AI into their work where it can **make services better, more just, and more efficient**. For example, a tool like ChatGPT can help users go from an outline to a draft Council memorandum quickly,

enabling them to focus more time on the analyses and findings that inform recommendations to Council.

## Getting Started with Generative AI for City Use

### Usage of Generative AI may be Subject to the Public Records Act

Any retained conversations relating to City work may be subject to public records requests and must comply with the City's retention policies. In addition, users will need to comply with the California Public Records Act and other applicable public records laws for all City usage of Generative AI. This means any prompts, outputs, or other information used in relation to a Generative AI tool may be released publicly. Unless you have been approved by the Information Technology Department (ITD) and are using a City enterprise Generative AI tool, do not use any prompts that may include information not suitable for public release.

### Create an Account Specifically for City-related Work

If you choose to use Generative AI for City-related work, you shall have an account for all Generative AI usage in your role at the City using a City email address. The purpose of this is to ensure proper retention of public records and avoid comingling of public and personal records. This account should not be used for any personal purpose. Users can use their City email address for City usage, or they can create a shared account using a different work email address. For example, the Digital Privacy Office might create a shared ChatGPT account using the digitalprivacy@sanjoseca.gov email address. Regardless of whether a shared or work email address is used to create an account, users should use a unique password for the service. **Like any other account which uses a City email address, the password should not be the same password used to log in to any City devices.** For example, if a data breach occurs on ChatGPT (which happened in March 2023)[4] and your password is stolen, a hacker should not be able to log into your laptop with that information.

If users use personal devices or accounts to conduct City work, the records generated may still be subject to search and disclosure. The records generated may include both the content users input and the content users receive from the Generative AI system.

### Understand the Terms and Conditions

If you choose to use Generative AI for City work and agree to the terms and conditions of a system without a City agreement in place, you are responsible for complying with those terms

[4] ChatGPT confirms data breach, raising security concerns. (2023, May 16). Security Intelligence.

and conditions. Contact ITD at digitalprivacy@sanjoseca.gov for a list of City-approved Generative AI tools. As of January 2025, the City has begun issuing ChatGPT Teams licenses.

## Check the Copyright of All Generated Content

Many tools, such as Microsoft Copilot[5] and ChatGPT[6], now offer some protections in the event of the AI producing copyrighted content. However, users should still check if the content they use from any Generative AI systems does not infringe on the copyright of other content. For example, City employees can check the copyright of image-based content by searching the generated image in a search engine like Google or Bing. This approach does not guarantee protection against copyright infringements. If users are uncertain if content violates copyright, they should either edit the content to be original or not use it.

Currently, AI generated content is not considered eligible for copyright.[7]

## Ownership of Generated Content

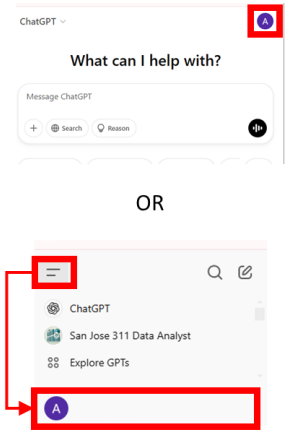
In most cases, the user owns the content they input into a Generative AI service and the information they receive as an output. The user can use the content at their discretion, in accordance with City policy and any terms and conditions agreed to. However, many Generative AI companies still retain the right to use both the input and output content for their own commercial purposes. For example, this could include a Generative AI company using City data to train their models or distributing City output data for marketing campaigns. This emphasizes the importance to opt-out of data collection and to only provide information the City is ready to make public should be entered into a Generative AI system.

## Opt out of data collection if possible

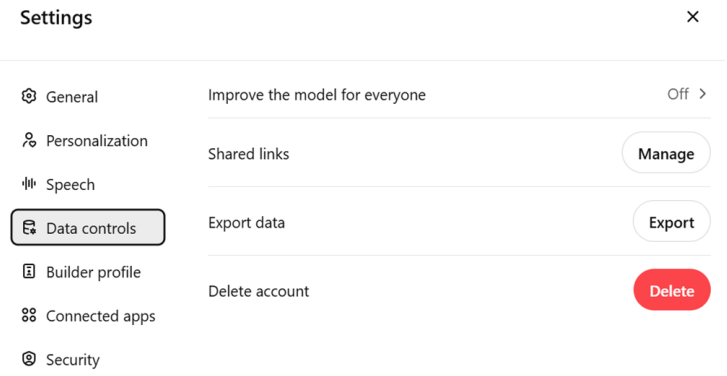
Some services offer an option to opt out of data collection. This means the generative AI system will not keep the data you provide, and it will not be used in the system's models. Opt out of data collection and model training whenever possible. For example, you can opt out of ChatGPT by going to "settings" → "data controls" → "Improve the model for everyone."

[5] https://techcommunity.microsoft.com/discussions/businessapplicationsforpartners/introducing-the-microsoft-copilot-copyright-commitment/3922303
[6] https://www.theguardian.com/technology/2023/nov/06/openai-chatgpt-customers-copyright-lawsuits
[7] https://www.reuters.com/legal/ai-generated-art-cannot-receive-copyrights-us-court-says-2023-08-21/

# Opting out of ChatGPT data collection (as of December 2024)




1. Click on your profile picture and select "Settings". Your profile will be on the top-right or you will click the top-left and find your profile on the bottom-left.




2. Go to "Data controls" and turn off "Improve the model for everyone"

## Joining AI Working Groups

The City is dedicated to providing practical guidance around AI that protects people from harm while providing the best services to residents. To accomplish this, the City has three engagement groups dedicated to informing AI use in the City:

8. City AI working group: City staff discuss AI policy, use cases, and guidelines. Users can learn more about AI in the City, discuss potential ideas in their departments, and flag any potential concerns.
9. Digital Privacy Advisory Taskforce: External Taskforce of experts around Digital Privacy and AI. The Taskforce advises and recommends on the City's digital privacy practices, including responsible AI.[8]
10. GovAI Coalition: The City of San José is collaborating with government agencies across the country to ensure that the AI systems we use serve all of our communities. The group collaborates on items including responsible AI governance, vendor accountability, and sharing use case experiences. If you are an agency interested in joining, you can do so at sanjoseca.gov/govai.

[8] Digital Privacy Advisory Taskforce webpage: https://www.sanjoseca.gov/your-government/departments-offices/information-technology/digital-privacy/digital-privacy-advisory-task-force

In addition to these three groups, the City holds opportunities for the public to provide feedback, including in-person sessions in San José, virtual sessions, and online at sanjoseca.gov/digitalprivacy. Members of the public can contact the Privacy and AI team directly at digitalprivacy@sanjoseca.gov.

## Guidance while using Generative AI

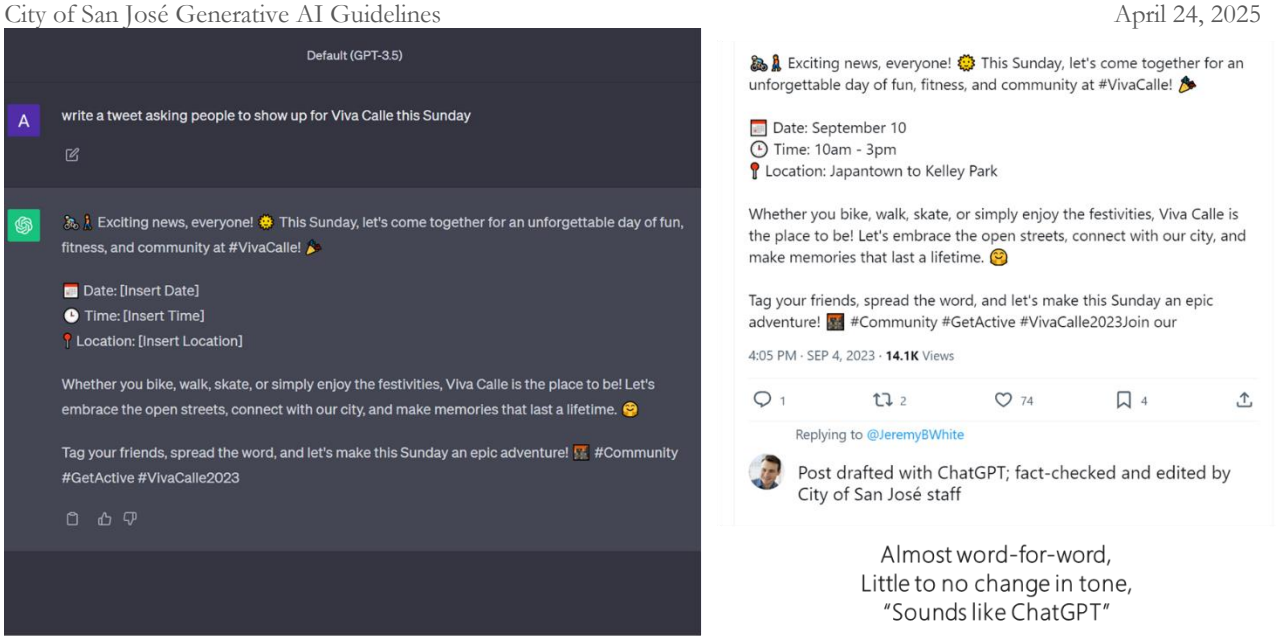
### Cite Use of Generative AI

**Citing Generative AI tools supports transparency and helps build trust between the City and its residents.** Residents expect communications from the City to be factual, credible, and verified by staff before releasing. While the City will use AI to provide better services more efficiently, it is essential that the City proactively communicate when material is from Generative AI but still vetted by the City.

Cite text from Generative AI in the following circumstances:

11. When the content is entirely AI generated from AI. This is common with AI-generated translations, AI-generated transcripts, and AI chatbots
12. When the tone of the content reads more like Generative AI the tone of the City. If the content is edited before release—as it should be—this will typically not apply.

For example, the following example twitter post uses the exact same language, emojis, and style as was generated by AI.




*A user asking ChatGPT to create a tweet, and ChatGPT producing a tweet. The user posted almost exactly what ChatGPT provided, so the user also cites ChatGPT and assures the reader that information was fact-checked and edited by City staff.*

Any statements used as fact must cite a credible source rather than the AI. Credible sources include official City documents and peer-reviewed journals. Consult your supervisor for other trustworthy sources (e.g., newspapers, blogs, or datasets). All images and videos must cite any AI used in their creation, even if the images are substantially edited after generation.

### How to Cite:

Generative AI can be cited as a footnote, endnote, header, or footer. If the content is completely AI generated with no opportunity for human review, such as a speech-to-text transcript or a public AI chatbot, then use a citation like the following:

"This text is generated by AI and may contain errors. Do not use it as an official source."

If the content has an opportunity for human review but still reads like Generative AI, use a citation like the following:

"Text drafted with Generative AI, fact-checked by City of San José staff. Sources for facts and figures are provided as they appear."

**Citations for images and video must be embedded into every frame of the image or video**. For support on how to do this, see the "Creating Images or Video" use case in the appendix or reach out to digitalprivacy@sanjoseca.gov.

## Record Usage of Generative AI

City staff are constantly discovering innovative ways to use Generative AI tools. To support knowledge sharing across the City, staff must report their usage of Generative AI at least once through this form: https://forms.office.com/g/3Znipym4k5. The form will take 1 minute. Responses are anonymous and will be used to inform future AI training and guidance.

Report usage when you use Generative AI in a way you have not used it before. For example, if you use Generative AI to draft grant applications, you do not need to submit the form each time you draft a grant application. Instead, you only need to submit it once for the overarching email drafting use case.

Additional guidance and advice around using Generative AI can be found in the Appendix.

## Limit the Environmental Impacts of Your Generative AI Use

The development and use of AI tools has impacts on the environment, especially Generative AI tools. For example, the amount of energy required for a single query to a large language model (e.g., single message to ChatGPT) is estimated to be anywhere between 6-10 times more than a traditional web search (0.3 Wh vs. 2.9 Wh).[9] In addition, prompting it for 10-50 queries consumes 500 milliliters of water.[10]

By 2030, an estimated 200 Tera Watt hours (TWh), or 1.6% of total US electricity consumption will come from AI.[12] This translates to more electricity consumption than all refrigerators in the US combined.[13]

[9] https://www.goldmansachs.com/pdfs/insights/pages/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf

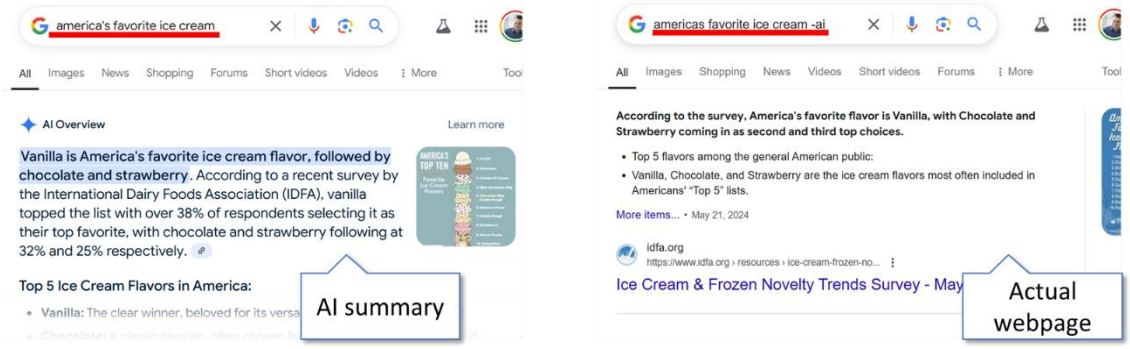
[10] Ren, Shaolei. "How much water does AI consume? The public deserves to know" (2023)

[12] Goldman Sachs estimates 8% of US electricity will come from data centers, and 20% of data center electricity will be used for AI. https://www.goldmansachs.com/pdfs/insights/pages/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf

[13] Based on U.S. Energy Information Administration's 2015 Residential Energy Consumption Survey, which estimates 303 Trillion British thermal units (Btu), or around 89 Tera Watt hours (TWh) of total US energy consumption from refrigerators

To fully address the significant environmental impacts of AI, thoughtful action across all levels of government and industry will be required. At the same time, there are individual actions that all users of Generative AI tools should take to limit their environmental impacts:

13. **Use a search engine for general questions.** Use typical search engines (e.g., Bing, Google, DuckDuckGo) that require significantly less energy for one-time questions and finding information. This approach will also help you find authoritative sources and avoid inaccuracies seen with Generative AI tools. Search engines often also use generative AI (e.g., Google and Bing), which you can turn off by adding "-ai" at the end of your search. See example below.




*Example of a Google search with and without the phrase "-ai" included*

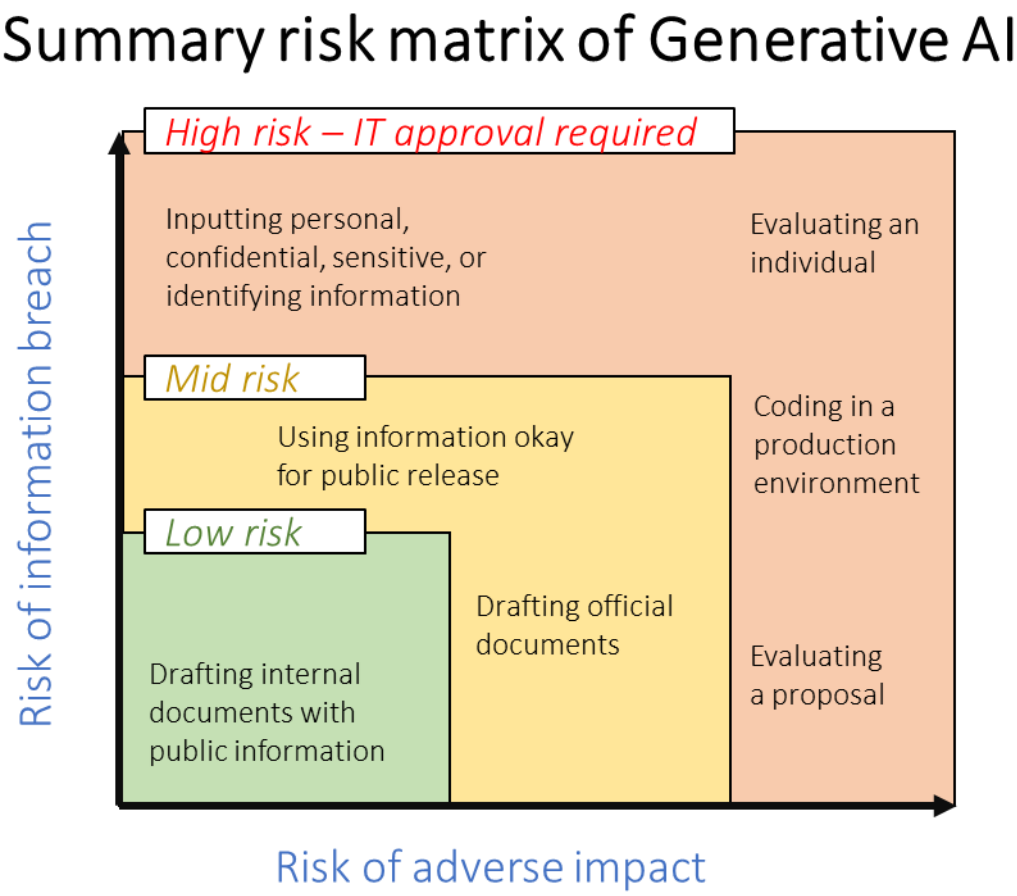
14. **Be thoughtful in your use of Generative AI tools**. Generative AI functionality is often turned on by default in many tools even when it is not useful. For example, many AI meeting tools take notes at every meeting on your calendar by default, even if you are not interested in the meeting or do not attend the meeting. Not only does this produce unnecessary environmental impacts, but it also risks creating inaccurate records of meetings that cannot be verified for accuracy.
15. **Minimize the number of prompts.** Try to use fewer prompts to accomplish your goals. For example, rather than drafting a memo in ten prompts, see if you can combine individual prompts to reduce the total number of queries to five while still generating a sufficient memo draft. Giving highly specific, detailed prompts will not only improve your responses, but it can also reduce the number of prompts you need to complete your task.
16. **Minimize the output expected from the AI**. Be specific around the output that you want and the length of output you would like. Generative AI tools can often provide more output than is needed to achieve the task at hand. Add guidance in your prompts like "respond in less than 100 words" or "respond in five bullet points" to make the output not only easier to understand, but also less costly to the environment.

## Assessing Risk in Generative AI Use Cases

The risk presented by Generative AI tools varies by use case, with the risk spectrum ranging from mid-risk to high-risk to intolerable risk.

Generative AI risk is determined by two key factors:

17. Risk of information breach: the potential harm if the information exchanged with a Generative AI is released to an unintended audience. This can include entering personally identifiable information, sensitive records, or confidential business information into a Generative AI platform. Additionally, any information entered into a Generative AI platform may be subject to the California Public Records Act and other similar laws. If you would not share the information in a public forum, do not share it with a Generative AI platform.
18. Risk of adverse impact: the potential harm of using the output for a decision, task, or service. This impact can be different for different populations and should be considered from an equity lens, such as adverse impacts to people of a certain race, age, gender identity, or disability status. Not only can AI be biased, but it can also provide false information. In general, if Generative AI is used in relation to City processes that can alter an individual or community member's rights, freedoms, or access to services, it should be thoroughly reviewed by multiple users before any document is finalized or action is taken.




*Summary of assessing risk in Generative AI use cases.*

## When Engaging in Mid-risk Use Cases

**Keep in mind the tone and specific language in the AI output**. Generative AI is trained on a global context and may not use the vocabulary or tone consistent with the City and its values. Simple examples include replacing "citizen" with "resident" in documents, and capitalizing "City" when referring to the City of San José. These documents, like any others, require thorough review before moving from draft to final product.

**Cite verifiable sources for all facts and figures** (past memos, newspapers, research papers, etc.). ChatGPT or other Generative AI platforms are not definitive sources. Facts should be accompanied by links or citations to sources that the general public could find, such as news articles or research papers. ChatGPT and other AI can fabricate sources if asked, so do not rely on them for finding citations either. Find sources directly and confirm they are legitimate before using.

Unless you have been approved by ITD or using a City enterprise Generative AI tool, do not use any prompts that may include information not suitable for public release. This includes information such as: draft RFP requirements not yet public; vendor transactions; procurement approvals; content that could propagate discrimination, harassment, or biased viewpoints (e.g., stereotypical language); or internal City decisions or communications containing privileged or classified information (e.g., legal case files, attorney work product, non-disclosure agreement information, City bank account information, protected health-related information about individuals, law enforcement investigations, information that could lead to security breaches, etc.).

Additional details on risk can be found in the Appendix.

## Concluding Thoughts

Generative AI presents users an opportunity to work better, faster, and smarter. However, because the technology and the laws surrounding it are evolving and present unknown risks, its adoption comes with ethical considerations. Remember the fundamental rules when using any Generative AI:

19. Review and follow City policies and guidelines.
20. Never submit personal or confidential information into a Generative AI.
21. Review, revise, test, and fact check any output from a Generative AI.
22. Be transparent when content was drafted using Generative AI.
23. Return to this document often, as guidance on usage will change rapidly.

By keeping the above guidance in mind when using generative AI tools, we can ensure the safe and responsible use of AI by the employees of the City, and we can protect privacy, security, and maintain public trust. If you or your department has any questions, comments, or concerns around using Generative AI, please contact your team at digitalprivacy@sanjoseca.gov. The Privacy office can provide users trainings, set up AI evaluations, and help your team do the best with Generative AI.

# Appendix

## A Definition of Generative Artificial Intelligence

Generative Artificial Intelligence, commonly referred to as “Generative AI” or “GenAI,” is an “automated system” used to generate “content.”

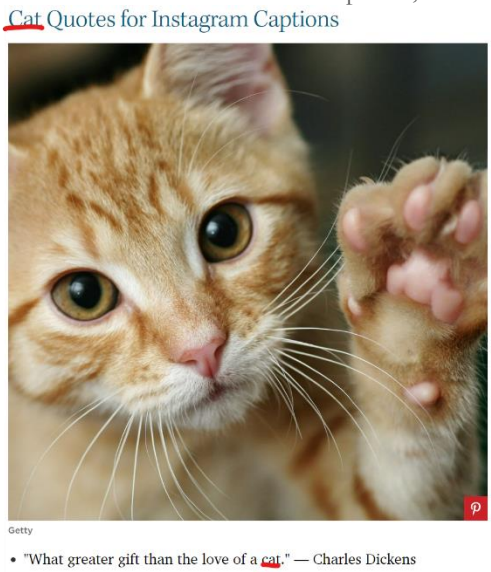
An "automated system" is any system, software, or process that uses computation as part of a system to generate outputs, outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities.[15]

“Content” includes text, emails, presentations, images, video, audio, architectural documents, diagrams, and other forms of media.

Generative AI uses massive datasets to generate content that someone would want given a prompt (see definition of “prompt” below). For example, ChatGPT has collected data on millions of webpages to identify sentence patterns that commonly come next after someone types a phrase. Online information is paired with human training where algorithm developers manually judge and correct the output of the system. For example, it may have required a combination of millions of webpages and a human developer to train ChatGPT that “Jack fell down, and broke his crown” should be completed by “and Jill came tumbling after.”

[15] Definition derived from the United States White House Office of Science and Technology Policy AI Bill of Rights (October 2022) and the National Institute of Standards and Technology Artificial Intelligence Risk Framework (January 2023)

Billions of images are shared online every day, along with hundreds of thousands of hours of video[16] and countless text posts. Much of this information is connected to other information on the internet. For example, pictures of cats are often connected with captions that have the word "cat" in them. These connections allow a computer to, after millions of connections, "learn" what a cat looks like. Eventually, a computer can create an image of a cat based on all the previous images it has seen.




AI systems apply this same approach to music, books, poems, voices, videos, and anything else created on the internet.

## Prompts and Generative AI

Generative AI relies on a user (e.g., a person) to "prompt" the AI to generate content. "Prompts" are any direction provided by a user. Examples of Generative AI include:

1. Creating text based on a prompt
2. Creating a picture or video based on a prompt
3. Making an audio file of a famous person saying something they did not say
4. Creating a movie scene based on a text prompt and pictures of the characters

Examples of prompts include:

1. **Text prompt to generate text content.** For example: "Tell me a story about three people becoming friends despite their differences"
24. **Text prompt to generate picture/video content.** For example: "Draw a cow with long hair and an ornate bell"
25. **Voice and text prompt to generate audio content.** For example: [Upload a recording of Tim Cook] "Say 'I'll just warn you now, I don't know how to use a computer' in the voice provided."

[16] (QUT), Q.U.of T. (2022) *3.2 billion images and 720,000 hours of video are shared online daily. can you sort real from fake?*, *QUT*. Queensland University of Technology

26. **Image and text prompt to generate picture/video:** "Re-draw this bear with cleaner lines and give the bear a crown. Then show a clip of the bear running."

## Details for Understanding Generative AI Risk

### Understanding "Risk of Information Breach"

**General rule**: If the information exchanged with a Generative AI system would be harmful to a person or community if made public, it is a high or intolerable risk. Services like ChatGPT have been compromised in the past and leaked personal information.[17] Until private applications with higher security are deployed in the City, all information exchanged with Generative AI has a reasonable risk of being compromised.

**Low-risk** information includes non-identifying and non-confidential information. For example, a simple email response or instructive documents often contain only general information that would not present any risk if made public.

**Mid-risk** information includes personally identifiable information (e.g., full name, birth date, email address) and confidential business information that may have larger implications to City processes. **Until a private application is deployed with security measures approved by the Cybersecurity Office, no high-risk information shall be provided to a Generative AI system.**

**High-risk** information includes highly sensitive and identifying information. This includes data such as credit card numbers, bank account information, social security numbers, and other information that requires rigorous security measures and compliance standards before being processed.

### Understanding "Risk of Adverse Impact"

**General rule**: If you are using Generative AI in relation to City processes that can alter an individual or community's rights, freedoms, or access to services, it is at least high risk and should be thoroughly reviewed before any document is finalized or action is taken. Additionally, any action that could reasonably lead to the City engaging in legal infringements on intellectual property are prohibited.

[17] "OpenAI says a bug leaked sensitive ChatGPT user data" https://www.engadget.com/openai-says-a-bug-leaked-sensitive-chatgpt-user-data-165439848.html

Low-risk impact includes tasks associated with drafting internal messages, internal documentation, and idea generation. These tasks can be sped up with the support of Generative AI but require many more steps before reaching a public impact.

Mid-risk impact includes tasks associated with official City documents or messaging. It also includes uses that require substantial editing and review before usage. These tasks require thorough review at the time of generation before using in any work context. Special care should be taken when a task may impact individuals differently across factors such as race, age, gender identity and disability (e.g., a memo about tree canopy inequity in neighborhoods).

High-risk impact includes tasks that undermine trust in the City through false statements or news; deny people due process such as in resource allocation, job evaluations, and purchasing decisions; or expose the City to substantial security or legal risks. Generative AI does not have reasoning behind the content it produces and cannot justify a decision.

## Examples of Generative AI Use Cases by Risk Level

### Examples of Low-risk Use Cases

#### 1. Drafting messages to staff and trusted partners

Generative AI tools can help users draft emails or other messages to staff and trusted partners. ChatGPT is a tool commonly used for this purpose. You can prompt ChatGPT to provide formal sounding language from general framing of the message. You can also have AI draft emails in different tones by asking for a different tone.

Additional Guidance:

27. You may be inclined to use ChatGPT to help with email replies. Do not copy your current email thread into ChatGPT. The email was sent to select people and may be confidential.
28. Be mindful about the purpose of the email, and if it is appropriate to use Generative AI for drafting it. For example, Vanderbilt University received heavy backlash for using ChatGPT to draft an email in response to a school shooting.[18]

Example:

[18] Korn, J. (2023) *Vanderbilt University apologizes for using CHATGPT to write mass-shooting email | CNN business*, *CNN*. Available at: https://www.cnn.com/2023/02/22/tech/vanderbilt-chatgpt-shooting-email/index.html

29. Prompt ChatGPT with the following: "I am the lead product manager for housing technology initiatives. We interview users to gather product requirements, prioritize features, and work with software developers on implementation. Draft me an email asking the software developers how long the housing database will take to implement, and what risks to implementation they see."
30. Carefully read through the email, perform final fact-check and other edits to the draft email. Manually add in personal information or internal confidential details before sending.
31. Cite at the end of the message "*Some of this content was drafted using ChatGPT. All facts, figures, and statements were reviewed by the sender to be accurate."*
32. If someone replies to your email asking for what you would like to see in the database, you can return to ChatGPT and prompt it with the following: "I want the database to be easily understood by our field staff. Draft this request to the developers." Read the draft, fact-check, and manually add information as needed. Cite ChatGPT at the end of the message as done previously.

### 2. Framing written content not intended for official release

Generative AI can be useful for creating an outline or structure for your written content. This can include an outline for a cover letter, long-form writing, project documentation, or speaker notes for a presentation. When the written content is not intended for official public release, it presents less risk than official City publications (like memos or policies). ChatGPT is the most common tool for this use case. You can write a few key points you would like to detail, any themes you want present, the kind of voice you would like, and how long you would like the content. Remember that information you input into a Generative AI system may be subject to a Public Records Act request.

Additional Guidance: Unless you have a Generative AI trained to your context—a feature likely not available for another year—the tool will provide generic language that does not apply to the City. For example, ChatGPT may use the word "citizens" rather than "residents" when referring to the people we serve because it is not used to San José's specific circumstances. As always, make sure to review, revise, and fact-check any output from Generative AI.

Example prompt steps:

33. Prompt ChatGPT with the following: "I am writing an instruction manual for how City staff should add content to the City's website. Draft an outline that can be posted on our

intranet. It should have a section about how to create lists, add hyperlinks, and show pictures using a content management system. Draft in a formal tone but make the text clear and approachable."

34. Review, revise, and fact-check. Manually enter any confidential or private information into the draft or final version.
35. Cite that you used ChatGPT in the drafting process. See how to cite Generative AI in the "Citing Generative AI" section under "Guidance while Using Generative AI,"

### 3. Learning from a document

You may copy a large public document into a Generative AI tool and ask the AI questions about the document. ChatGPT is the most common tool for this use case.

Additional Guidance: The document or information you paste into the Generative tool AI should already be public information.

Example prompt steps:

1. Ask ChatGPT to "Summarize the following document. Let me know if there is any mention of California, cities, or San José."
2. Copy the text from a public document and paste it into ChatGPT. An example document would be the text from this news article: https://www.fiercewireless.com/wireless/san-jose-plans-smart-city-infrastructure-verizon-and-at-t. Copy text and paste into ChatGPT
3. You do not need to cite ChatGPT unless you quote specific text outputted from ChatGPT in future written content.

### 4. Brief list of other mid-risk use cases

36. Helping you come up with a name for your team. For example, "give me ten names for a team focused on AI and Privacy that uses the acronym 'SAFE'." Be sure the name is not already used by another team in the City.
37. Learning about a new topic in a way that you can understand. For example, "explain quantum mechanics to me like I'm five." Verify anything you learn from ChatGPT before applying the knowledge in a City context.
38. Helping you find the right word for a concept. For example, "What is the word for the second-to-last episode in series." Once the AI provides the word, search the word on Google (or elsewhere) to confirm it means what you think it means.

## Examples of Mid-risk Use Cases

### 1. Drafting memos and or other public-facing City documents

Generative AI tools can help users draft memos and other public-facing documents more efficiently. Because the content is meant for public release, it is treated as high-risk and should **be reviewed and edited multiple times before release**. ChatGPT is the tool most used for this purpose.

Additional guidance: The City expects users to produce their own research that informs memos, such as information related to policy changes and program changes. Memos, press releases, and other publications also have their own City-specific formats and standards to follow. Consult your supervisor to make sure your memo follows City standards.

Example prompt steps:

1. Provide context around the memo but only provide public details: "We are building an encampment management work order system at the City to better coordinate services and have just completed the detailed design. We will be presenting to the City council on the latest update."
2. Then request ChatGPT to draft a memo: "Draft a memorandum with the following sections and key points: Introduction: (add bullet points), Human-Centered Design work (add bullet points), Requirements (add bullet points), Next-steps (add bullet points).
3. After initial memo draft, prompt ChatGPT to "Draft a conclusion summarizing all the prior sections."
39. Manually add in any non-public information to the draft memo produced by ChatGPT.
40. Carefully read through memo, perform fact-check and other edits to memo to maintain a tone consistent with City documents.
41. Cite verifiable sources (past memos, newspapers, research papers, etc.) for all facts and figures in the memo.
42. If required, cite that you used ChatGPT in the drafting of the memo. See how and when to cite Generative AI in the "Citing Generative AI" section under "Guidance while Using Generative AI."

## 2. Writing an RFP/RFB/Vendor relations

Generative AI tools can help users draft RFP/RFBs more efficiently. Because the content is meant for public release, it is a high-risk use case. ChatGPT is the tool most used for this purpose.

Additional guidance: RFPs and other publications have their own City-specific formats and standards to follow. Consult your supervisor and purchasing business partner to make sure your memo follows City standards. Guidance can be found on the City intranet, including the "Strategic Procurement Guidelines for RFP and Contract Requests To Finance-Purchasing." Take special care not to provide ChatGPT information that is not meant to be public yet. For example, if the specific requirements of the RFP are not meant to be public yet, do not input them into your prompts.

Example prompt steps:

43. Provide context around the procurement document without providing non-public details: "We are procuring an encampment management work order system at the City to better coordinate services."
44. After ChatGPT responds, ask to draft the procurement document: "Draft an RFP with the following sections and key points: Introduction: (add bullet points), Scope of Work (add bullet points), Requirements (add bullet points), Cost Breakdown (add bullet points)."
45. Manually add in any non-public information to draft document produced by ChatGPT.
46. Carefully read through the document, perform fact-check and other edits.
47. If required, cite that you used ChatGPT in the drafting process. See how and when to cite Generative AI in the "Citing Generative AI" section under "Guidance while Using Generative AI."

## 3. Writing advertisements, social media posts

Generative AI tools can help users draft promotional material. Because the content is meant for public release, it is a high-risk use case. Chat GPT is the tool most used for this purpose.

Additional guidance: None, refer to AI principles and guidance for high-risk use cases.

Example prompt steps:

48. Provide ChatGPT with details around the needed post and audience, for example: "Draft a cute tweet of less than 240 characters that reminds families that tomorrow is Walk and Roll to school day."
49. Review output, edit to make personal to San José and relevant department or office, and post.

## 4. Writing job postings or job descriptions

If you provide a Generative AI with a list of qualities you want and a role title, it can help you draft a formal-sounding job description. Because the content is meant for public release, and job requirements can have a substantial impact on who applies, it is a high-risk use case.

Additional guidance:

The City expects users to follow existing standards on the format and content of job postings based on classifications. Consult your Human Resources (HR) business partner or your Department's HR representative for information on job classifications and postings.

Additionally, be mindful of the language used in the requirements, responsibilities, and tone used in the job posting. Check if the job description seems to use language stereotypically associated with a specific race or gender. Use gender-neutral language: Avoid using gender-specific pronouns (he, she) and job titles (fireman, firewoman). Instead, opt for inclusive terms such as "they" and "fire officer." Remove gender-coded words: Avoid using adjectives that may be associated with a specific gender, such as "aggressive" or "nurturing." Use neutral descriptors, like "results-driven" or "collaborative."[19]

Example prompt steps:

50. Provide ChatGPT with the previous **public** posting for an Analyst I position and request that ChatGPT "Draft a similar job description, but with a focus on using information to inform park capital projects."
51. Manually add in any non-public information to draft document produced by ChatGPT.
52. Carefully read through the document and edit for a more neutral tone, perform fact-check and other edits.

[19] Krakovetskyi, O. (2023, March 22). Eliminating Bias in Job Descriptions with ChatGPT – The DevRain Tech Blog - Medium. *Medium*. https://medium.com/devrain/eliminating-bias-in-job-descriptions-with-chatgpt-72b92ebc7911

## 5. Creating images or video

Some Generative AI tools such as Stable Diffusion and Dall-E can create images or video clips based on text prompts. The City needs to maintain its legitimacy as a trustworthy source when using video and images, which requires substantial precautions whenever using AI-generated visual content.

Additional Guidance:

53. **Use only for illustrative purposes.** For historical events, use real images rather than generated. For example, if you want a picture of a giraffe wearing a suit and tie for your presentation, generate it. If you are proposing a new visual diagram or abstract concept, you can also generate it. If you want a picture of the Mayor at City Hall, find an actual picture.
54. **Require a citation embedded into the image or video at all times**. Images and videos can easily be taken out of their original context and misinterpreted as reality. To prevent a news article or other secondary source from treating an image as fact, all images and frames of a video must specify that they were generated using an AI system. The citation shall be included in the image itself and cannot be removed without editing or cropping the image.

Example Use Case:

1. Provide a prompt: “drawing of falcon and its chicks on top of a skyscraper”
2. Choose your image
3. Embed the citation into the image: “Image generated by DALLE-2”

55. Add alt text into the image that clearly states the image was generated by an AI system

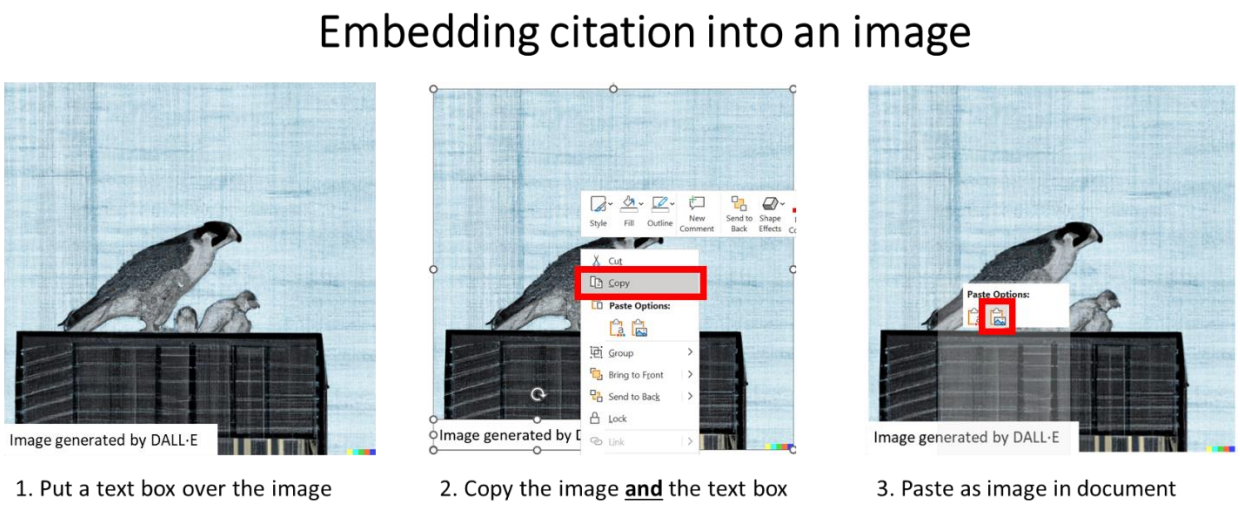



*Figure 1: Three steps of embedding a text box into an image in order to cite the Generative AI system used for the image.*

## 6. Language Translation

Large Language Models like ChatGPT have been demonstrated to translate text better than the previous generation of AI translation systems. Past systems typically translated word-for-word (e.g., "hello" into "hola") while Large Language Models can translate the meaning of a phrase, which often performs better with improved understanding of context.[20] The City has developed a custom translation assistant[21] trained on existing translation pairs provided by the City and partners in the GovAI Coalition.[22]

**Translations can still be wrong, and whenever possible should be confirmed by a fluent speaker of both languages.** Consult the City Manager's Language Access Unit or your language access manager for more guidance on translations.

Suggested use: Provide a document, paragraph, or phrase and ask to translate into another language.

[20] Cms, O.- *Official site of the State of New Jersey, Department of Labor and Workforce Development | New Jersey & USDR, with support from Google.org, release LLM training materials to build AI translation assistants for public services*. Available at: https://www.nj.gov/labor/lwdhome/press/2024/20241205_aitransition.shtml.
[21] Assistant URL: https://chatgpt.com/g/g-muc4N5C0u-translation-gpt. Built in partnership with US Digital Response (USDR), Kansas City, MO, and other partners in the GovAI Coalition. See the full dataset of translation data on San José's Open Data Portal: https://data.sanjoseca.gov/dataset/agencies-translation-pairs/resource/05a81ebf-5a95-463d-ae7b-1efeddd9ad25
[22] GovAI Coalition: Sanjoseca.gov/govai

Additional guidance: If your translation is not verified by a human, clearly cite the translation as AI Generated and may contain mistakes. Should be cited in the target language. For example, if you are translating text from English into Spanish, the citation should be present in Spanish. Your citation should clearly state "This translation was generated by artificial intelligence and may contain errors or inaccuracies."

Example prompt steps:

56. Tell the AI what you generally want to communicate in another language
57. Ask the AI to translate that text into the target language
58. Wherever possible, confirm the translation with a fluent speaker in both languages – consult the City's Language Access Unit
59. Suboptimal options for confirming the translation include opening a new chat with ChatGPT and asking it to translate the text back into the original language
60. If the translation was not verified by a human, include at the beginning of the document that "This translation was generated by artificial intelligence and may contain errors or inaccuracies." Ensure this language is translated into the language of the published document.

### 7. Creating presentation slides

If you provide Generative AI with some public information, it can create a presentation for you. It is being integrated into existing applications like PowerPoint. Currently there is no clear leader in Generative AI for presentations, but a few examples are beautiful.ai and gamma.app. Presentations are automatically high risk because they go beyond text into images, which can present false information if the audience believes the image is real.

Suggested use: Provide a public document, or an outline with public information.

Additional guidance: Similar guidance to other public-facing documents, but with the additional requirement to cite all AI Generated images clearly on the image.

Example prompt steps:

61. Provide the AI with an article or outline from publicly available information. For example, an article about the evolution of cats.[23]

[23] Article used in example was Vsadmin. (2022, April 8). *The Evolution of Cats*. Killarney Cat Hospital. https://www.killarneycat.com/the-evolution-cats/

62. Generate the presentation
63. Fact check all content, review for tone and language
64. Cite Generative AI images

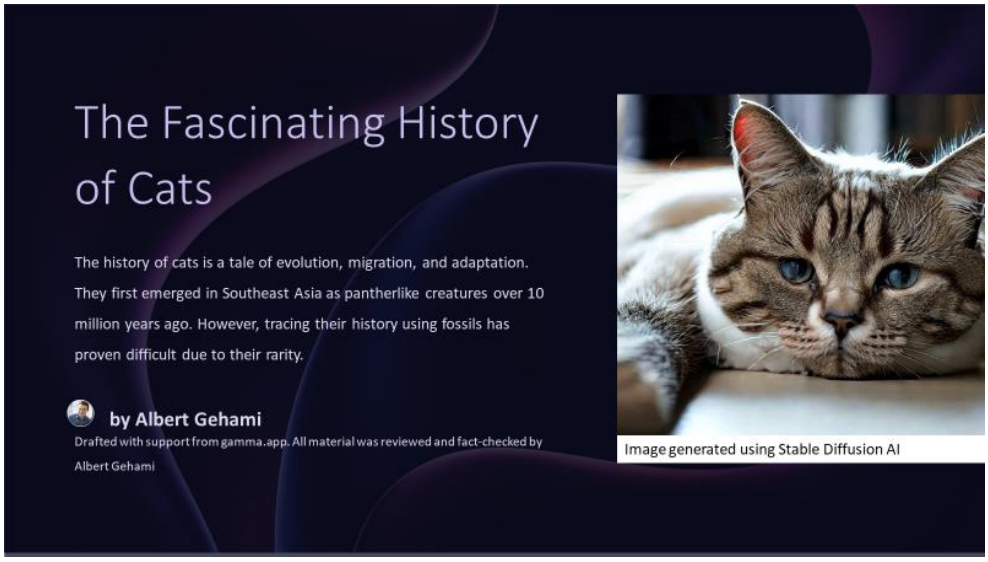



*Figure 2: Image of a generated presentation with support from the AI tool gamma.app. The image on the right was generated using Stable Diffusion AI and was cited as such*

## 8. Programming

AI tools can be useful for helping write, edit, and fix code. It serves as an additional tool in partnership with websites and internal documentation in generating code that operates in the City. Most common Generative AI tools (e.g., ChatGPT, Claude, and Copilot) provide coding support. Other tools (e.g., GitHub Copilot, Tabnine) are built specifically to support programming.

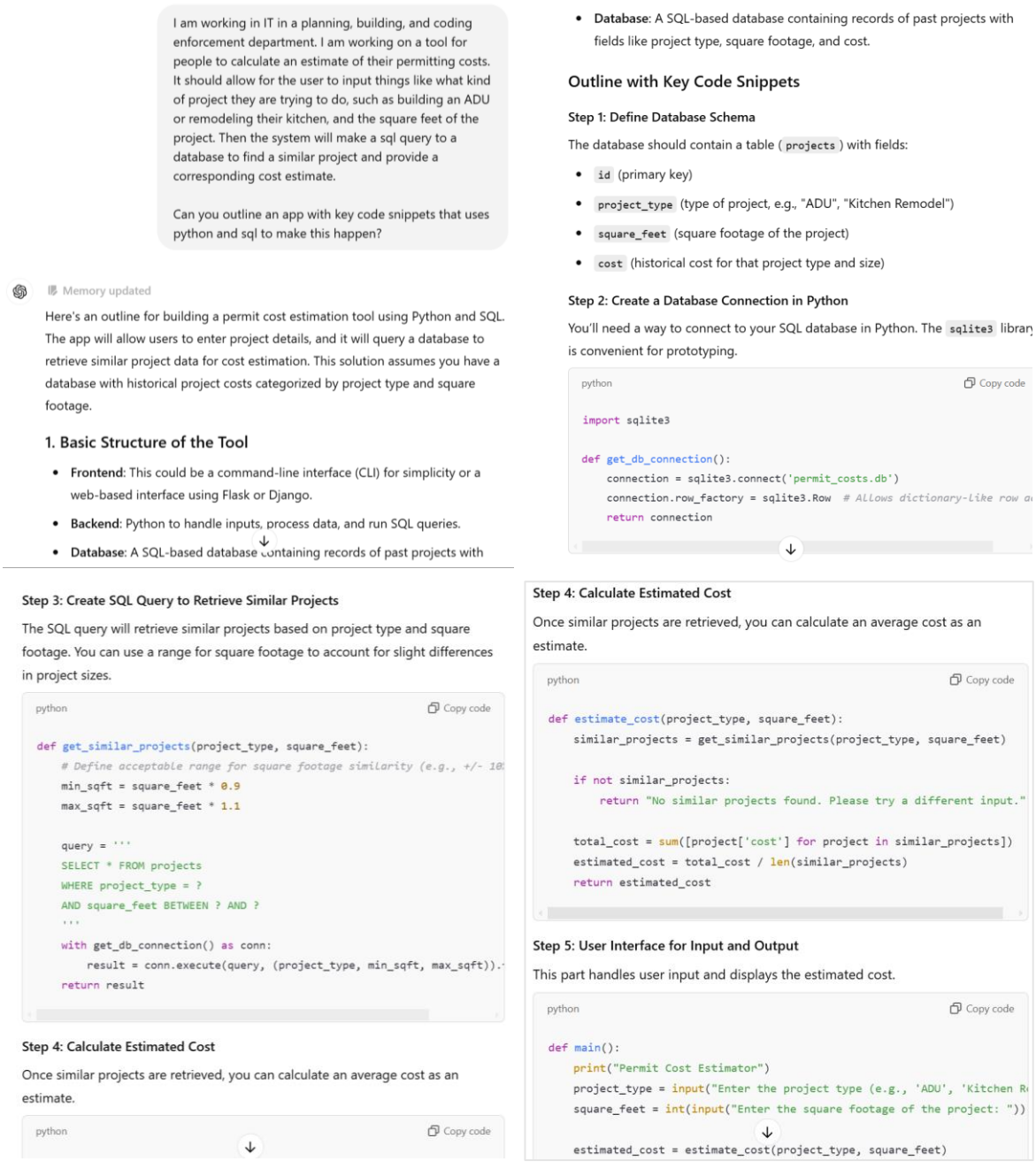
Suggested use: Create a draft framework for a new coding task or to upload an existing code segment for debugging.

Caution: AI tools may provide code that is outdated, which may use syntax that does not work with the latest version and/or have security vulnerabilities that were patched in following versions. The code may also have general issues that require debugging. Continue to follow best practices in coding, such as reviewing, validating, and testing (e.g., run unit testing, integration testing, peer code reviews, system testing, etc. to catch bugs, security vulnerabilities, and logical flaws. Always check for security vulnerabilities such as SQL injection, cross-site scripting (XSS) to identify any potential vulnerabilities in AI generated code.) any code in a development environment before it can impact real data and systems.

Example prompt steps:

65. Provide the AI with an outline of what you want to create
66. Request an outline that explains the code
67. Check all output, review for outdated code, and test in a development environment

See screenshots below of an example chat:



I am working in IT in a planning, building, and coding enforcement department. I am working on a tool for people to calculate an estimate of their permitting costs. It should allow for the user to input things like what kind of project they are trying to do, such as building an ADU or remodeling their kitchen, and the square feet of the project. Then the system will make a sql query to a database to find a similar project and provide a corresponding cost estimate.

Can you outline an app with key code snippets that uses python and sql to make this happen?

Memory updated

Here's an outline for building a permit cost estimation tool using Python and SQL. The app will allow users to enter project details, and it will query a database to retrieve similar project data for cost estimation. This solution assumes you have a database with historical project costs categorized by project type and square footage.

### 1. Basic Structure of the Tool

- **Frontend**: This could be a command-line interface (CLI) for simplicity or a web-based interface using Flask or Django.
- **Backend**: Python to handle inputs, process data, and run SQL queries.
- **Database**: A SQL-based database containing records of past projects with fields like project type, square footage, and cost.

### Outline with Key Code Snippets

**Step 1: Define Database Schema**

The database should contain a table (`projects`) with fields:

- `id` (primary key)
- `project_type` (type of project, e.g., "ADU", "Kitchen Remodel")
- `square_feet` (square footage of the project)
- `cost` (historical cost for that project type and size)

**Step 2: Create a Database Connection in Python**

You'll need a way to connect to your SQL database in Python. The `sqlite3` library is convenient for prototyping.

```python
import sqlite3

def get_db_connection():
    connection = sqlite3.connect('permit_costs.db')
    connection.row_factory = sqlite3.Row  # Allows dictionary-like row a
    return connection
```

**Step 3: Create SQL Query to Retrieve Similar Projects**

The SQL query will retrieve similar projects based on project type and square footage. You can use a range for square footage to account for slight differences in project sizes.

```python
def get_similar_projects(project_type, square_feet):
    # Define acceptable range for square footage similarity (e.g., +/- 10
    min_sqft = square_feet * 0.9
    max_sqft = square_feet * 1.1

    query = '''
    SELECT * FROM projects
    WHERE project_type = ?
    AND square_feet BETWEEN ? AND ?
    '''
    with get_db_connection() as conn:
        result = conn.execute(query, (project_type, min_sqft, max_sqft)).
    return result
```

**Step 4: Calculate Estimated Cost**

Once similar projects are retrieved, you can calculate an average cost as an estimate.

```python
def estimate_cost(project_type, square_feet):
    similar_projects = get_similar_projects(project_type, square_feet)

    if not similar_projects:
        return "No similar projects found. Please try a different input."

    total_cost = sum([project['cost'] for project in similar_projects])
    estimated_cost = total_cost / len(similar_projects)
    return estimated_cost
```

**Step 5: User Interface for Input and Output**

This part handles user input and displays the estimated cost.

```python
def main():
    print("Permit Cost Estimator")
    project_type = input("Enter the project type (e.g., 'ADU', 'Kitchen R
    square_feet = int(input("Enter the square footage of the project: "))

    estimated_cost = estimate_cost(project_type, square_feet)
```

*Figure 6: Images of using AI to outline and generate code for a database function and front end app. Generated using ChatGPT*

### 9. Brief list of other mid-risk use cases

Remember to follow the AI principles and general guidance for the mid-risk use cases below:

68. Creating diagrams. For example, "Create a flow chart of a tree turning into wood pulp and then into paper." Replace pictures before publishing.
69. Drafting papers. For example, "Here is my outline for my research paper, and my findings, draft a complete paper."

## Examples of High-risk Use Cases – prohibited unless approved by ITD

Below are examples of use cases that are prohibited for staff usage. If your department would like to request an exception, you can complete the AI Review form and contact digitalprivacy@sanjoseca.gov.[24]

### 10. Evaluations and Decisions

Why it requires ITD approval: Evaluating job applicants using AI has led to countless scandals of biased application reviews.[25] This evaluation issue also extends to other areas such as evaluating proposals or an existing employee.[26] AI-based evaluations expose the City to public protest across many key City functions such as hiring and purchasing.

**Additionally, Generative AI shall not be used to determine highly sensitive decisions such as an individual's health plan, cost of bail, conviction of a crime, grades, or admissions to a program.**

What can you do with Generative AI: AI can help flag key words and identify phrases within a document (see the mid-risk use case). It can also help you transform your bullet points in explaining your evaluation into complete sentences. However, the actual evaluation must be made by a person, and any AI generated content must be reviewed by a human before being shared.

[24] Link to AI review form: https://forms.office.com/Pages/ResponsePage.aspx?id=4DvjD0Jhlk-bjXgX1cJhOQHtWWB-PGxHmT9VcPkWpgRUMUVVTDRMRTIzODVVWVlDWUdTSzIxTTRJQi4u
[25] Vallance, B. C. (2022, October 13). AI tools fail to reduce recruitment bias - study. *BBC News*. https://www.bbc.com/news/technology-63228466
[26] In general, people react worse to negative evaluations from AI than they do to negative evaluations from people. Lopez, Alberto, and Ricardo Garza. "Consumer bias against evaluations received by artificial intelligence: the mediation effect of lack of transparency anxiety." *Journal of Research in Interactive Marketing* (2023).

Why it requires ITD approval: Replicating a person's voice with AI in any City document or recording would undermine the trust of staff and the residents. Potential legal concerns also exist regarding replicating a person's voice. Unless there is a specific application that is navigated with IT and the Privacy Officer, do not generate audio through AI.

## Additional Guidance around Generative AI

### Be Aware of Targeted Cyber Attacks Using Generative AI

Although City staff are already familiar with handling cyber risks like phishing and malware, the advent of generative AI introduces heightened cybersecurity risks as the attacks can be more complex and personalized. Cyber threat actors may use generative AI in their attacks in the following ways:

- **Writing AI-powered, personalized phishing emails**: With the help of generative AI, phishing emails no longer have the tell-tale signs of a scam—such as poor spelling, bad grammar, and lack of context. Plus, with AI like ChatGPT, threat actors can launch phishing attacks at unprecedented speed and scale.
- **Generating deep fake data**: Since AI can create convincing imitations of human activities—like writing, speech, and static and video images. Generative AI can be used in fraudulent activities such as identity theft, financial fraud, and disinformation.
- **Cracking CAPTCHAs and password guessing**: Used by sites and networks to comb out bots seeking unauthorized access, CAPTCHA can now be bypassed by hackers. By utilizing AI, they can also fulfill other repetitive tasks such as password guessing and brute-force attacks.

### Detecting Generative AI

Software developers are building tools, like GPTZero, GPT Radar, and Originality.ai, designed to detect if a body of writing was created by a generative AI tool. These tools are in early stages of development and their detection accuracy rate may not always be accurate and should be used with caution. For example, there have been numerous incidents of instructors using ChatGPT detection tools falsely accusing students of plagiarism, and endangering their grades and even diplomas.

Despite the limited accuracy of these tools, they allow residents to check if City documents were generated by AI regardless of whether users cite their usage or not. To build trust with residents,

users need to be proactive in communicating its usage of AI. Residents finding out on their own can cause reputation harm to the City.