AI Handbook

City of San José

Maintained and updated by:

City of San José's Information Technology Department: Digital Privacy Office

digitalprivacy@sanjoseca.gov

Last updated:

March 18, 2024

Purpose

The City of San José AI Handbook provides guidance on how to comply with the City of San José AI Policy. Refer to this handbook for all matters relating to City of San José AI usage, including purchasing, procurement, governance, incident response, and education.

This document covers the following topics:

1. **AI Policy:** The foundational document for the City of San José's approach to AI; refer to the "Artificial Intelligence (AI) Policy"
2. **AI Review:** Required for all new technology procurements and data initiatives that involve an AI system; see "AI Review"
3. **AI Governance:** The framework for managing and monitoring the full lifecycle of all City of San José AI systems; see "Artificial Intelligence (AI) Governance"
4. **Link to the Generative Guidelines:** Guidance for City of San José staff in using Generative AI.

This document will continue to be updated to provide the latest information on the City of San José's AI Policy and practices.

TABLE OF CONTENTS

Artificial Intelligence in City of San José

Background

As the capital of Silicon Valley, the City of San José has a history of using innovative solutions to provide impactful services to residents with limited public sector resources. In recent years, the City has explored the capabilities of artificial intelligence (AI) and machine learning systems to improve the delivery of City services and support data analytics, like with AI traffic cameras and Automated License Plate Readers. In response to the City's increased use of AI systems, the Digital Privacy Office (DPO) has established an AI review process for technology procurements that involve the use of AI.

This handbook provides detail into the City of San José's AI governance established by the Artificial Intelligence (AI) Policy. As the City of San José's AI governance practices and the AI industry mature, the City of San José's relevant policies and this handbook will continue to evolve.

Identifying AI

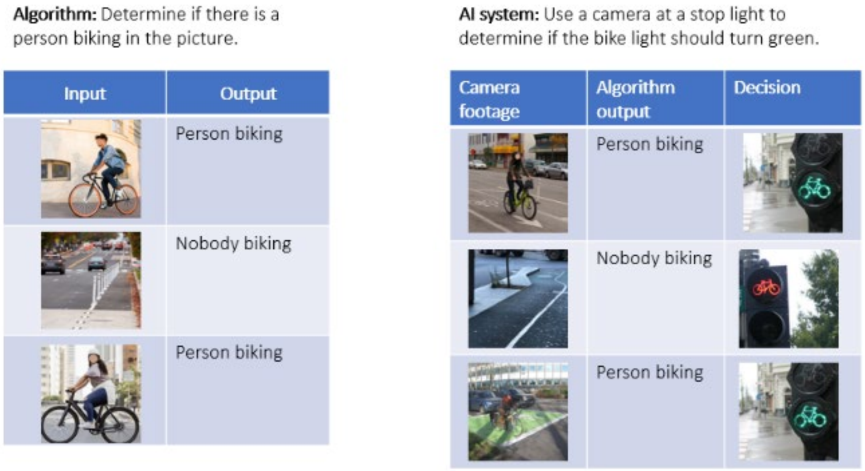
"AI" and "AI system" are defined in the AI Policy. In practice, the City of San José uses a checklist of questions to identify AI systems.[1] Procurement officers who process technology proposals may make particular use of this checklist. A technology may be considered an "AI system" if it elicits positive answers to any of the following questions:

- Does the technology use data to provide predictions, recommendations, insights, or decisions?
- Does the technology augment or replace human decision-making?
- Does the company use words such as "personalized", "tailored", and "adaptive" in its marketing?

Although the City of San José may identify a technology as an AI system, not all AI systems require a full AI Review. Learn more about the medium and high-risk AI systems that require full review in the section "AI Review Process: Step by Step, Step 2: Risk Analysis".

The figure below illustrates the relationship between an algorithm and an AI system. Consider the example of an AI system designed to change traffic lights based on the presence of a bicyclist. While the algorithm's sole function is to identify a bicyclist, the AI system changes the traffic light to accommodate the bicyclist identified by the algorithm.

[1] This checklist is inspired by The EdTech Equity Project's school procurement guide.



Algorithm: Determine if there is a person biking in the picture.

Input	Output
	Person biking
	Nobody biking
	Person biking

AI system: Use a camera at a stop light to determine if the bike light should turn green.

Camera footage	Algorithm output	Decision
	Person biking	
	Nobody biking	
	Person biking	

Figure 1: Example of an algorithm and an AI system. In this example the algorithm identifies if a bicyclist exists in the photo. That algorithm is then used in an AI system to turn a bike light green at an intersection.

AI Governance

Governance Structure

The governance of AI systems in the City of San José involves several actors, policies, and practices that work in coordination to ensure that the City of San José uses AI tools in a responsible manner. The diagram below explains the roles and responsibilities of actors in the City of San José who frequently participate in AI governance activities.

AI Policy Roles & Responsibilities

San José residents

- The City of San José is ultimately accountable to the residents it serves. For AI uses cases that present a high potential risk, the City of San José aims to involve the public in the process of reviewing and deploying the AI system. In general, the City of San José aims to be transparent about how it uses AI with the broader public.

Final Authority

- City Council

Escalated Authority

- The Chief Information Officer (CIO) approves AI-related policies set forth by the City Digital Privacy Officer (CDPO) and, if applicable, sends them to City Council for final approval.
- The Office of the City Attorney reviews AI-related policies from a legal perspective.
- The Purchasing & Risk Management Office reviews procurement requests. They may refer procurement requests to the CDPO if the procurement seems to involve some technology or AI system.

Reviewing Authority

- The CDPO is the main actor in AI governance activities and responsible for creating AI-related policies and guidance for the City of San José. The CDPO is responsible for ensuring all procurement requests that involve an AI system are given an AI review.
- The Chief Information Security Officer (CISO) is responsible for ensuring technology procurement requests are reviewed from a cybersecurity perspective.

Department Stakeholders

- Any City of San José department may submit a procurement request for an AI system.
- Departments can propose pilots, ideas, or potential use cases of AI.
- Departments have a designated AI lead, responsible for interfacing with the CDPO to ensure department is following City of San José practices. The AI lead also raises concerns from the department around AI to the DPO.
- The IT Department's CDPO serves as a first line for reviewing all technology. They may conduct a threshold analysis, or a deeper AI impact assessment depending on their assessed risk level of the project.

AI Working Group (AIWG)

Led by DPO, employees from various departments in the AIWG discuss AI-related issues and projects in the City of San José. The AIWG is composed of department AI leads and potentially other department representatives.

AI Advisory Group

Led by DPO, external stakeholders advise the CDPO and the CIO on the policies and activities related to the City of San José's AI governance. The Advisory Group consists of external stakeholders, including AI experts from industry, academia, civil rights, and members of the public. The Advisory Group meets quarterly to discuss AI topics with the DPO. While the Advisory Group plays a critical role in informing the City of San José's decisions, decision-making power remains within the City of San José.

AI Review

Introduction

In addition to the City of San José's privacy, cyber, infrastructure, and data review protocols, the CDPO conducts an AI Review to ensure that the proposed AI system complies with the City of San José's Guiding Principles for AI systems, and relevant privacy policies (in consultation with the Office of the City Attorney). While not all steps of the City of San José's AI Review protocols may be necessary or appropriate for every project, the protocol provides the general review framework for any project.

It is important to note that throughout the product lifecycle of all AI systems, there should be a consistent and continuous review of the technology through a human-centric lens. Consistent review allows the City of San José to ensure that the AI system continues to provide value and protects against potential harms going undetected.

What needs an AI Review?

Not all AI systems, whether out-of-the-box or customized, require a full-fledged AI Review.[2] Simple rule-based systems, which rely on a series of hard-coded conditional rules to produce one of multiple pre-defined outputs, may not be subject to the AI Review process. In contrast, algorithm-based systems, which rely on complex logic to make predictions based on patterns in a set of training data, are more frequently subject to the AI Review process. As a general rule, AI systems should be reviewed on a regular basis (e.g., annually, quarterly, etc.).

Whether an AI system is subject to an AI Review also depends on the potential risk of the AI system in question. To classify the risk of an AI system, the CDPO conducts a risk threshold assessment before initiating an AI review of a given proposal (see the subsection "Step 2: Risk Analysis" within the section "AI Review Process: Step by Step" for more information on the risk threshold assessment).

[2] It is important to note that vendors contracted under professional, personal, or general consulting agreements may be using AI based tools on the City of San José's behalf. When evaluating professional services agreements, determine if the vendor may be using AI systems to generate reports, analyze data, or provide insights and request information about such systems. Depending on the circumstances, the City of San José may wish to also review such third-party AI systems in accordance with the same rules and policies described below.

See the chart below for examples of AI systems that do and do not require an AI Review. **These examples are not exhaustive, but rather are meant to guide and reinforce concepts.**

AI Review is **required**	AI Review is **optional**
1. Predictive policing system that impacts the deployment of City resources 2. Identity recognition based on one-to-many matching (e.g., license plate reader, facial recognition) 3. AI system that meets organizational thresholds that require an RFP according to City procurement standards 4. Automated decision-making system that automates a decision which traditionally requires human review 5. AI systems that impact or integrate with infrastructure systems consistent with City of San José services (e.g., translation service for 311)	1. General website recommendations (e.g., recommended videos on the City of San José's YouTube channel) 2. Personalized notification system for City of San José events 3. Strictly rule-based logic (e.g., an accounting software to calculate taxes owed, a motion-detection based alarm system)

Figure 2: An example list of AI systems that do and do not require an AI Review.

AI Review Framework: At a Glance

The AI Review Framework guides the City of San José in reviewing AI systems during the public procurement process. The Framework outlines the actions that from the early project proposal stage to final approval and ongoing monitoring of the AI system.

To understand how the AI Review process is adapted for Request for Proposals (RFP), refer to the section "Request for Proposals (RFP) AI Review Protocol".

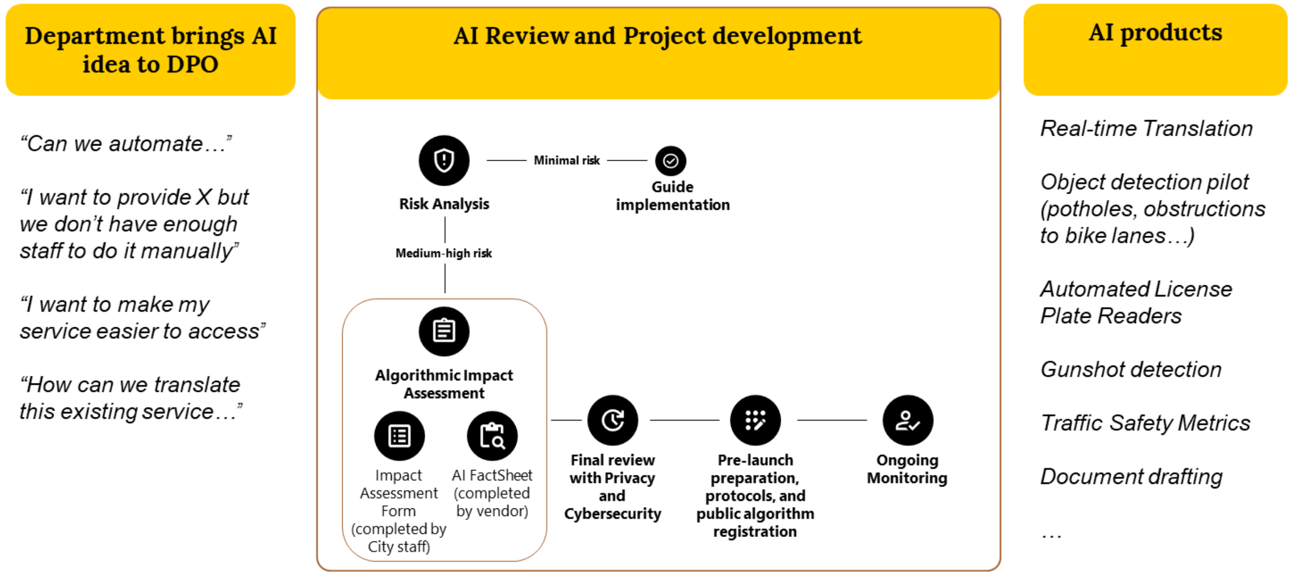

Department brings AI idea to DPO
"Can we automate…"
"I want to provide X but we don't have enough staff to do it manually"
"I want to make my service easier to access"
"How can we translate this existing service…"
AI Review and Project development
Risk Analysis
Minimal risk
Guide implementation
Medium-high risk
Algorithmic Impact Assessment
Impact Assessment Form (completed by City staff)
AI FactSheet (completed by vendor)
Final review with Privacy and Cybersecurity
Pre-launch preparation, protocols, and public algorithm registration
Ongoing Monitoring
AI products
Real-time Translation
Object detection pilot (potholes, obstructions to bike lanes…)
Automated License Plate Readers
Gunshot detection
Traffic Safety Metrics
Document drafting
…

Figure 3: AI Review Framework.

1. **Procurement Request or Existing Application Update:**

   Departments seeking to procure an AI system first engage the CDPO to discuss their technology proposal. Departments should provide existing material related to the project, including details on project purpose, data collected and specific uses and benefits of automating a process, recommendation, or decision.

2. **Risk Analysis:**

   The CDPO conducts an "AI Risk Threshold Analysis" to assess the risk of the proposed AI system and to determine if the project necessitates a full-fledged assessment. Step 2 below provides further guidance and examples of risk on a gradient scale from low-risk to high-risk.[3]

3. **Assessment:**

   The CDPO facilitates the following steps to evaluate the potential risks and benefits of the proposed AI system.

   The following steps should be adhered to in accordance with the City of San José's risk tolerance:

   a. **Impact Assessment Form:** The business-owning department(s) completes the required Impact Assessment Form, or equivalent.

[3] It is important for the City of San José to understand and document its risk tolerance for AI systems. Key considerations that may be helpful for assessing an institution's risk tolerance include:
- What experience does City of San José have with AI?
- What are City of San José's existing risk management frameworks and practices?
- Who are the key stakeholders involved in City of San José's AI strategy?
- Are there any specific regulations or policies that will influence City of San José's use of AI?

b. **AI FactSheet:** The vendor completes the required AI FactSheet. The CDPO may work with the vendor to obtain more details about the AI system. Equivalent information should be provided if the AI system is developed internally.
    i. **If request involves a Request for Proposals (RFP):** The vendor may not be determined at this time. Require potential vendor(s) to complete the AI FactSheet. See "Request for Proposals (RFP) AI Review Protocol" for guidance on RFP questions.

4. **Public Engagement:**

   If the proposed AI system is considered of particular public interest, the CDPO conducts in-person outreach, targeting communities with limited access to online comments (either due to language or internet access issues). Community feedback is then incorporated into the Data Usage Protocol.

   The City of San José should prioritize reducing barriers for public participation, particularly for those directly impacted by the AI system, especially historically marginalized or disadvantaged communities.

5. **Final Review:**
   Projects are reviewed by the CDPO and Cybersecurity Office. The CDPO provides assessment, approval/denial, and/or recommendations.

   At CIO's discretion, a review may rise to a relevant Council Committee.

6. **Pre-Launch Preparation:**
   a. **Data Usage Protocol:** Medium-risk and high-risk AI systems may necessitate a Data Usage Protocol to govern the collection, access, processing, and sharing of data around the AI system to ensure that the project complies with the City of San José's Digital Privacy Policy.
   b. **AI Inventory:** The approved project proposal is added to a publicly viewable online inventory of the AI systems deployed by the City of San José.[5] The Impact Assessment Form and AI FactSheet are made publicly available online. The DPO regularly updates the AI Inventory as new AI systems are adopted and archives old systems as they are phased out of use.
   c. **Training:** Users of the approved AI system are given training to properly deploy, operate, and maintain the technology. Training is often provided by the vendor or other third party.

7. **Ongoing Monitoring:** Departments report annual metrics defined in the Data Usage Protocol. Reports usually require metrics on data usage and project effectiveness. Public can comment on data usage and annual updates online at this link.

AI Review Framework: Step-by-Step

The previous section provided a high-level summary of the AI Review Framework. In this section, each of the seven steps that comprise the AI Review Framework are explained in greater detail.

Step 1: Proposal

An AI Review is triggered when a business-owning department in the City of San José department submits a procurement request for a technology involving an AI system. An AI Review can also be initiated when a vendor has made updates to a product's functionality or released a new version of the AI system.

An AI Review can be formally initiated through the DPO.

For a consultation on potential projects, to initiate an informal "AI Risk Threshold Analysis", or to ask any questions, contact the DPO directly at digitalprivacy@sanjoseca.gov. To see a list of AI systems currently used by the City of San José, refer to the online AI Inventory here.

Step 2: Risk Analysis

The City of San José will conduct a risk analysis on the proposal to determine is a full-fledged review is required for the system. The review will be overseen by CDPO.

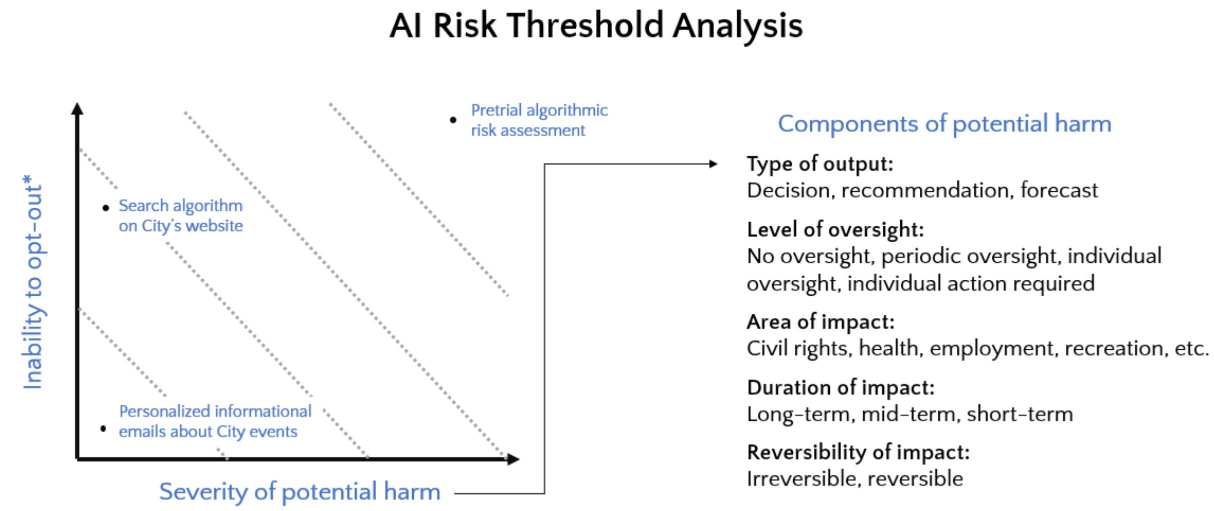
The City of San José aligns its approach to risk with the National Institute for Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (AI RMF). While the NIST AI RMF does not necessarily describe exactly *how* to evaluate AI risk, the AI RMF defines characteristics of trustworthy AI and provides important guidance for organizations that seek to build governance and risk management processes for AI systems.[4]

While there are many methods to evaluating AI risk, the approach outlined below is designed to be easily implemented by the City and its limited capacity. The AI Risk Threshold Analysis is intended to be a practical tool that balances real-world constraints with the need to adequately ascertain the risk of an AI system.

In the AI Risk Threshold Analysis model, the impacted individual's inability to opt-out of the use of the AI system and the severity of potential harm of the AI system are the two main factors for evaluating AI risk (see Figure 4).[5]

[4] NIST AI Risk Management Framework: https://www.nist.gov/itl/ai-risk-management-framework and example applications: https://airc.nist.gov/Usecases.

[5] The AI Risk Threshold Analysis model is partly inspired by the risk matrix found in AI Ethics Impact Group's "From Principles to Practice: An interdisciplinary framework to operationalise AI ethics". https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf


AI Risk Threshold Analysis
Pretrial algorithmic risk assessment
Search algorithm on City's website
Personalized informational emails about City events
Inability to opt-out*
Severity of potential harm
Components of potential harm
Type of output:
Decision, recommendation, forecast
Level of oversight:
No oversight, periodic oversight, individual oversight, individual action required
Area of impact:
Civil rights, health, employment, recreation, etc.
Duration of impact:
Long-term, mid-term, short-term
Reversibility of impact:
Irreversible, reversible

Consider: How severe is the potential harm of the AI system if it is inaccurate, includes harmful bias, is not privacy-respecting, etc.?

Figure 4: The risk of an AI system can be estimated by identifying the impacted individual's inability to opt-out of the use of the AI system and the severity of potential harm posed by the system.

The components of potential harm consist of the type of output, the level of oversight, the area of impact, the duration of impact, and the reversibility of the impact. In Figure 4, the risk level of each component of potential harm *decreases* from left to right. For example, for type of output, "decisions" are higher risk than "forecasts"; for level of oversight, "no oversight" is higher risk than "individual action required"; for area of impact, "impacts on civil rights" are higher risk than "recreational impacts".

Although the AI Risk Threshold Analysis provides a systematic method for evaluating AI risk, estimating the risk of an AI system will vary for each context. For example, the distinction between a short-term and mid-term impact can be unclear and context-dependent; should a duration of one month be considered a short-term or a mid-term impact? The answer will depend on the unique context in which the AI system is being deployed and on the values of the community (see "Step 4: Public Engagement"). Is an AI-powered chatbot that leverages ChatGPT to provide residents information about applying to social welfare programs a mid-risk or a high-risk system? Again, the answer is highly context-dependent and requires the contextual and subject matter expertise of the CDPO.

It is possible for an AI system to simultaneously possess characteristics typically associated with both mid-risk AI systems *and* characteristics typically associated with high-risk AI systems. Categorizing these complex sociotechnical manifestations of risk is not always straightforward and requires discretion in conducting the risk analysis. When the risk level of an AI system is unclear, a good rule of thumb is to default to the higher risk categorization.

While estimating AI risk is not clearcut and requires the discretion of the CDPO, it can nevertheless be helpful when triaging risk to reference characteristics typically associated with low-, mid-, or high-risk AI systems. The tables below feature characteristics that tend to be associated with low-, mid-, and high-risk AI systems, along with examples of each risk archetype commonly found in the public sector. The tables below are non-exhaustive and are intended to help in thinking about AI risk.

Low-risk AI Systems	
Description	**Low-risk:** AI system that involves an opt-in approach to a person being subject to the system. The system generates predictions but does not automate decision-making and involves anonymized information used to provide general improvements to the City of San José. Notice is provided upon collection if any personal information is involved. The effects or impacts of the system are reversible, and typically short-term.
Characteristics	• **Data Collection:** Notice is provided upon collection if any personal information is involved, as well as documentation to support the safe storage and handling of data. • **Inferential:** AI System provides analysis, insights, or predictions but these are for informational purposes only and are not tied to automated decision-making. • **Negligible impact on humans:** The AI system is used for non-critical tasks with no negative material impact for humans. • **Mundane applications:** The AI assists with routine tasks like text completion. • **Accuracy and Validity:** Accuracy and validity metrics for the AI system are known. • **Transparency:** Access to appropriate levels of information based on the stage of the AI system's lifecycle is provided and tailored to the role or knowledge of individuals interacting with the AI system. • **Explainable and interpretable:** The meaning of the AI system's output(s) is understood in the context of its designed functional purpose.
Examples of low-risk AI systems	• Software that generates a comprehensive profile of a client by aggregating inputted data. • A process that matches users to a basic administrative outcome such as time slots for appointments or next available client services specialist. • A system or tool that permits the operations of basic computer processes such as opening programs, sending electronic communications, autocorrecting, or using a calculator. • The use of generative AI for general research purposes, where the outputs are not included in public documents, policies, or decision-making frameworks.

Medium-risk AI Systems	
Description	**Medium-risk**: AI system that involves identifiable information to provide targeted government services desired by the data owner with periodic oversight. Notice is provided at time of collection and often requires written consent. The system may have a short- to medium-term impact on quality of life factors.
Characteristics	• **Data Collection:** Notice is provided upon collection of sensitive personal information, but it is unknown how data is stored and handled. • **Opt-out consent:** Users are automatically enrolled unless they actively choose to opt out. • **Accuracy and Validity:** Accuracy and validity metrics for the AI system are mostly known. • **Transparency:** Access to appropriate levels of information based on the stage of the AI Systems lifecycle is somewhat provided and tailored to general roles of individuals interacting with the AI system. • **Explainable and interpretable:** To a limited extent, the output(s) can be related back to inputs and model assumptions. • **Human Oversight:** The AI system offers suggestions, insights or predictions, but humans retain final decision-making power. • **Economic impact:** Has minor impact on workforce and economic opportunity. • **Periodic oversight:** Human monitoring occurs at intervals, not continuously. • **Moderate impact:** The AI system may temporarily impact quality of life, but has minimal long-term risks, such as economic, legal, or reputational consequences.
Examples of mid-risk AI systems	• Recruiting software that recommends relevant job openings to candidates based on their skills and experience, and final hiring decisions remain with human recruiters. • An AI model that evaluates loan applications based on financial data, but human loan officers make the final approval or denial decisions. • Marketing and advertising software that tailors ads to users' interests based on online behavior, while users retain control over their data and opt-out options. • A tutoring software that tailors learning based on student data, however the student or the student’s parents can choose the learning pathway and material difficulty.

High-risk AI Systems	
Description	**High-risk**: AI system that is potentially rights-impacting or safety-impacting within areas such as: critical infrastructure, biometrics, legal representation, and highly sensitive personal information traditionally kept hidden, like Social Security Numbers, credit card numbers, etc. The high-risk system may automate decision-making, have significant material impact on quality of life, and be subject to minimal or no human oversight.
Characteristics	• **Compulsory**: Users have no opt-out option and are automatically subjected to the AI system. • **Data Collection:** No notice or documentation is provided regarding collection, storage, and handling of personal data. • **Automated Decisions:** The AI directly drives decisions with minimal or no human input. • **Minimal Oversight:** Human monitoring is either absent or very infrequent. • **Significant Impact:** Decisions made by the AI system can profoundly affect quality of life, including: ○ Employment: Job opportunities, hiring or firing decisions, salary recommendations. ○ Healthcare: Diagnosis, treatment plans, eligibility for care. ○ Criminal Justice: Risk assessment, bail determinations, sentencing recommendations. ○ Public Safety: Conflict resolution, suspect selection, ticketing and fines, resource allocation. ○ Economic: Automated social services benefit distribution, audits. ○ Sensitive Data: The AI system likely processes highly sensitive personal data with significant consequences for misuse. ○ Financial: Automated spending which has the potential to violate liquidity and use-of-funds requirements, automated loan acceptance/denial. ○ Infrastructure: The reduction of internet bandwidth or power being inconsistent with the needs of the public.
Examples of high-risk AI systems	• Generative AI that creates realistic, yet fabricated, videos or audio recordings, posing risks for misinformation and reputational harm. • A system that estimates individual risk factors for insurance, credit, employment, or healthcare

	without human oversight in decision-making, potentially leading to bias and discrimination. • An autonomous weapons system that can choose and engage targets without human intervention. • A biometric and facial recognition software that identifies individuals in real-time based on facial features, raising privacy and potential bias concerns. • A system that recommends criminal sentences based on offender data and prior sentencing outcomes from historical case data.

Step 3: Assessment

During the assessment stage, the business-owning department(s) fills out the required Algorithmic Impact Assessment Form for the CDPO and the vendor completes the AI FactSheet. Higher-risk projects require a Data Usage Protocol to guide how the project complies with the City of San José's Digital Privacy Policy. Lower-risk projects do not need to undergo a full-fledged review.

Impact Assessment Form

The Impact Assessment Form is completed by staff from the business-owning department procuring the system.[6] The Form consists of questions that are intended to capture information including, but not limited to:

- Project objective:
  - Please clearly describe the project use case, the current process, and the desired outcome.
  - Which Department is owning this system?
  - Who in the Department is responsible for this system?
  - Why does your department choose automation as an approach to this problem? What other approaches to solving this problem were considered (if any) and what led to choosing automation?
- Vendor details:
  - Will the AI system be designed, developed, deployed, or maintained by vendors or third parties?
  - How can the CIty test the AI system before it is put into use?
- Transparency:
  - How do individuals receive a notice in advance of interacting with the AI system? (For example, if a user is interacting with a chatbot, the system lets the user know they are talking to a chat bot instead of a human.)
  - How can third-party auditors easily view the AI system's data to perform evaluations?
  - How could AI system operators or residents know if the system outputs an error? What ability will they have to correct or appeal an error?

[6] See an example of a completed Impact Assessment Form from the City of San Jose here: https://www.sanjoseca.gov/home/showpublisheddocument/94187/638107653163800000.

- Equity:
    - What individuals and communities will interact with the AI system? For example, is the algorithm used on the general population (technology used in many public areas) or a specific group (e.g., children in a school program, a single neighborhood)?
    - How likely is it that the AI system impacts children under the age of 18?
    - Does this use case, and the information/decisions generated by the AI system, impact an individual's right or freedoms (e.g., if the AI system helps determine if a suspect can be put on bail or must remain in jail)?
    - Does this use case, and the information/decisions generated by the AI system, impact an individual's economic status (e.g., if the AI system helps determine if an individual can apply to affordable housing)?
    - Does this use case, and the information/decisions generated by the AI system, impact an individual's health, healthcare, well-being (e.g., if the AI system helps determine an individual's likeliness for colon cancer)?
    - Do decisions from the AI system impact the environment? (e.g., potential impact to carbon emissions, high tech waste)?
    - What issues could arise if the AI system is inaccurate?
- Human oversight
    - Please describe the level of autonomy of the AI system.
        - System operates automatically with no human intervention
        - System operates automatically with occasional retrospective reviews by humans
        - System operates automatically with opportunity for human to override any individual action
        - System produces recommendations but cannot act without human intervention
    - If there is human intervention in the AI system, is it by the vendor, City department/office, or both?
    - Please list the City roles/divisions that will be "touching" the system, or managing the deployment and use of the AI system.
    - How does the department provide training and resources to personnel to help them develop the skills they need to effectively operate the AI system?
    - In the event that the AI system does not work or is deemed to be inaccurate, what back-up measures are in place to continue providing services? In other words, can the City continue to provide the service without the AI system, and how would it do that?
- Accessibility
    - Have you considered how the AI system integrates and interacts with commonly used assistive technologies (e.g., screen readers, voice recognition software, etc.)?
    - Have you considered how users of diverse abilities will interact with the user interface?
    - How will feedback be collected by individuals with disabilities regarding the system?

    - How will feedback from individuals with disabilities be implemented into the system?
    - Where will accessibility features and resources be documented and readily available?
    - Will there be specialized training for individuals with accessibility needs?
- Liability
    - Who in City of San José will ultimately be accountable and responsible if the system fails to operate as intended?
    - Who in City of San José has the authority to stop or limit the AI system's use?
    - If a vendor fails to meet contractual obligations, what are the alternative options that exist to ensure there is no loss of service? (Note that this ties in to the City's AI Incident Response Plan.)

AI FactSheet

The AI FactSheet is completed by the vendor of the AI system and captures basic facts about the AI system. The AI FactSheet enables the CDPO to better understand the technical details of the AI system and ultimately assess the risks and benefits it presents. The AI FactSheet is intended to capture information including, but not limited to:

- Training data
- Testing data
- Input and outputs
- Performance metrics
- Optimal conditions
- Poor conditions
- Bias

The CDPO will work with the vendor as needed to obtain necessary technical details about the AI system.

Step 4: Public Engagement

If the proposed AI system presents a significant potential risk or is of particular public interest, the CDPO conducts in-person outreach, targeting communities with limited access to online comments (either due to language or internet access issues). Community feedback is then incorporated into the Data Usage Protocol.

The City of San José should prioritize reducing barriers for public participation, particularly for those directly impacted by the AI system, especially historically marginalized or disadvantaged communities.

Public engagement can occur online, in-person, and in the built environment. Below are examples of formats for public engagement:

- Online
    - Interactive website portal: Create a dedicated website where citizens can submit feedback on specific AI initiatives, participate in surveys, and engage in discussions.

    - Social media town halls: Host live Q&A sessions on social media platforms like Facebook, Instagram, or X where experts discuss AI and answer public questions.
    - Online forums and communities: Create dedicated online forums or communities, possibly on sites like LinkedIn or Reddit, focused on AI policy and development.
- In-person
    - Public workshops and town halls: Organize in-person events where citizens can learn about AI, hear from experts, and discuss their concerns and suggestions. In addition, create interactive activities like brainstorming sessions to gain feedback.
    - Community outreach programs: Partner with local organizations, libraries, and community centers to host AI education and feedback sessions.
    - Citizen advisory boards: Establish a diverse advisory board of citizens to provide ongoing feedback on AI development and policy.
    - Focus groups and interviews: Conduct targeted focus groups and interviews with specific demographics or stakeholders to gather in-depth feedback on specific AI applications or concerns.
- Built environment
    - Interactive kiosks and installations: Install interactive kiosks in public spaces like libraries, parks, or government buildings where citizens can learn about AI and provide feedback through surveys, polls, or open-ended questions.
    - "Living labs" for testing AI applications: Designate specific areas or neighborhoods as "living labs" where citizens can experience and provide feedback on prototype AI applications in real-world settings.
- Additional considerations
    - Accessibility and inclusivity: Ensure all feedback channels are accessible to people with disabilities and diverse backgrounds. Use multiple languages, alternative formats, and assistive technologies.
    - Transparency and communication: Clearly communicate the purpose of collecting feedback, how it will be used, and how citizens can stay informed about the process.
    - Data privacy and security: Implement data security measures to protect citizen privacy and ensure feedback is handled ethically and responsibly.

Step 5: Review

After the necessary documentation and public engagement has been completed, the CDPO and Cybersecurity Office reviews the proposal and provides final approval or rejection of the proposal.

Depending on the level of risk presented by the proposal, this step may involve gaining approval from City Council. The review may require input from the Office of the City Attorney or other relevant departments on a case-by-case basis. Based on the review by CDPO, in some instances, the proposal may need to be revised before being approved.

Step 6: Pre-Launch Preparation

Following approval, relevant documentation will be published on the digital privacy webpage, including:

- Data Usage Protocol: An electronic copy of the Data Usage Protocol is added to the digital privacy webpage.
- AI Inventory: If the project features an AI system, the approved project proposal is added to a publicly viewable online register of the AI systems deployed by the City of San José.[8] The Impact Assessment Form and AI FactSheet are also made published online. The CDPO regularly updates the AI Inventory as new AI systems are adopted and archives old systems as they are phased out of use.

Prior to implementation, relevant parties (e.g., staff in business-owning department) are given training to properly deploy, operate, and maintain the technology. Training is often provided by the vendor or other third party.

Step 7: Ongoing Monitoring

The Data Usage Protocol for a given technology proposal requires that the business-owning department of the AI system submits an Annual Usage Report. The report is typically 1-2 pages, drafted by the applicable department(s) and details:

1. Project summary
2. Required performance metrics as defined in the Data Usage Protocol (e.g., accuracy, effectiveness, cost)
3. Future plans for the technology initiative (e.g., project expansion, shift in usage)

Examples of past Annual Usage Reports can be found here.

Protocol for a Request for Proposals

Prior to issuing a Request for Proposals (RFP) or similar procurement process, departments should review the proposed solution to discern if it includes an AI component. Contact the CDPO directly at digitalprivacy@sanjoseca.gov, file a procurement request, or reach out to your department's privacy and AI representative to trigger an AI Review.

RFPs should incorporate questions that demand transparency around how the AI model works, clarify what protocols for human oversight are in place, and confirm that there are mechanisms for user review.

In consideration of the City of San José's public records obligations and transparency commitments, all vendors subject to the use of the protocols outlined in this document through the RFP process should be preemptively informed of City of San José's intended or required disclosure practices around bid documentation.

Use of the RFP Protocol

Below is an illustrative example of assessment questions and a scoring model of how to evaluate answers to these questions for an AI system. Recommended point values on a scale of zero to five are also provided for each category. Each section of assessment questions includes a "non-technical" question that asks the vendor to provide an answer in plain language that can be easily understood by a non-technical audience.

Any scoring methodology associated with evaluation of RFP bids is determined independently by City of San José in alignment of City of San José's stated principles and priorities. Please use this as a guide or rubric for best practices. It is the vendor's responsibility to be responsive to these questions, and it is the SME's responsibility to ensure that the vendor provides meaningful answers to the assessment questions.

Assessment Questions

1. System Overview

- Brief summary of the AI system.
- Purpose of the AI system, the intended use case, and users.
- Relevant context to the technology and maturity of the vendor.
- What is the policy on data collection, storage, and distribution?
- Training materials and implementation plan.
- Non-technical: Can you provide a non-technical overview of how your AI system operates and its key functionalities?

2. Data Training and Model Description

- How was the AI system trained, and what data was used?
- How often is data added to the training set?
- What data was used to test system performance?
- What conditions has the system been tested under?
- Provide a general description of the model(s) used.

- Non-technical: In layman's terms, can you explain how your AI system learns information and what kind of data it has been trained on?

3. System Operations

- How often are the models updated for users?
- Will the user have a choice in moving to the updated model or staying on the current model?
- Where is prompt and output data stored? Is this information used for future model versions?
- Do operators require specific education or certification to use the system?
- Non-technical: From a user's perspective, can you clarify where data is stored, as well as the process is for receiving updates or choosing to stay with a current model version?

4. Performance Evaluation

- How was the accuracy and effectiveness of the system measured?
- What metrics were used, and why?
- What is the range of accuracy of the AI system, and how does it vary depending on the data?
- What is the system optimizing for and under what constraints?
- Non-technical: In simple terms, how well does your AI system perform, and what aspects do its performance metrics prioritize?

5. Ethical Considerations

- What biases does the tool exhibit, and how does it handle that bias?
- Does the vendor report bias or justify why no bias would be present?
- How does the tool prevent or reduce harm to the end user?
- Non-technical: How do you ensure that your AI system treats all individuals and groups fairly, without any unintended biases?

6. System Reliability

- How does the AI system handle outliers?
- Do overwritten decisions feed back into the system to help calibrate it in the future?
- What conditions does the model perform best under?
- What conditions does the model perform poorly under?
- What are the limitations of the AI system?
- What expertise does this AI system require for operation, debugging, modification, and troubleshooting.
- Non-technical: Can you explain, in non-technical terms, how your AI system deals with unusual cases or incorrect predictions?

7. Interpretability and Explanation

- How does the AI system explain its predictions?

- Are the outcomes of the AI system understandable by subject matter experts, users, impacted individuals, and others?
- Non-technical: Can you share examples or scenarios illustrating how the AI system communicates its predictions in a way that is easy to understand for non-experts?

8. Monitoring and Correction

- How is the AI tool monitored to identify any problems in usage?
- Can outputs (recommendations, predictions, etc.) be overwritten by a human?
- Do overwritten outputs help calibrate the system in the future?
- Non-technical: For end-users, how can they be involved in monitoring and correcting any issues with the AI system?

9. Studies and Transparency

- Have the vendors or an independent party conducted a study on the bias, accuracy, or disparate impact of the system?
- If yes, can the City of San José review the study?
- Include methodology and results.
- Is the data used to train the system representative of the communities it covers?
- Non-technical: Can you provide examples of studies conducted to ensure fairness and accuracy, and how transparent are these studies?

10. User Interaction and Feedback

- How can the City of San José and its partners flag issues related to bias, discrimination, or poor performance of the AI system?
- How is the AI tool made accessible to people with disabilities?
- Has it been assessed against any usability standards, and if so, what was the result?
- What other human factors, if any, were considered for usability and accessibility of the system?
- Non-technical: How can users easily provide feedback on any issues they encounter with the AI system, and what measures have been taken to ensure accessibility for all users?

Scoring Model

1. System Overview

- 5 (Excellent): A comprehensive, non-technical overview that effectively communicates the AI system's purpose and functionality.
- 4 (Good): A clear and concise summary, providing a basic understanding of the AI system.
- 3 (Average): A brief overview but lacks clarity in conveying the system's purpose.
- 2 (Below Average): Limited information that does not effectively convey the system's purpose.
- 1 (Poor): No overview or insufficient information to understand the context.

2. Data Training and Model Description

- 5 (Excellent): Detailed information on training data, model architecture, and transparency on model usage.
- 4 (Good): Clear explanation of the training process and model description.
- 3 (Average): Basic information on training data and model without much detail.
- 2 (Below Average): Limited information on training data and model description.
- 1 (Poor): No information or inadequate details regarding training data and model.

3. System Operations

- 5 (Excellent): Regular model updates with clear communication and user-friendly options.
- 4 (Good): Frequent updates with communication on changes, providing user choice.
- 3 (Average): Regular updates but lacking clear communication or user choice.
- 2 (Below Average): Infrequent updates with unclear communication and no user choice.
- 1 (Poor): No updates or communication about the system's status.

4. Performance Evaluation

- 5 (Excellent): Comprehensive metrics, clear justifications, and superior performance compared to other vendors.
- 4 (Good): Well-explained metrics with justified performance in line with industry standards.
- 3 (Average): Basic metrics explanation with average performance.
- 2 (Below Average): Limited metric explanation with subpar performance.
- 1 (Poor): No metric explanation or poor performance without justification.

5. Ethical Considerations

- 5 (Excellent): Thorough identification and handling of biases with transparent reporting.
- 4 (Good): Clear recognition and handling of biases with transparency.
- 3 (Average): Basic acknowledgment of biases without much transparency.
- 2 (Below Average): Limited recognition or handling of biases.
- 1 (Poor): No acknowledgment or handling of biases.

6. System Reliability

- 5 (Excellent): Robust handling of outliers, effective calibration, and adaptability to corrections.
- 4 (Good): Efficient handling of outliers and adaptability to corrections.
- 3 (Average): Adequate handling of outliers with some adaptability to corrections.
- 2 (Below Average): Limited handling of outliers and minimal adaptability to corrections.
- 1 (Poor): No handling of outliers or adaptability to corrections.

7. Interpretability and Explanation

- 5 (Excellent): Clear and understandable explanations that cater to both experts and general users.
- 4 (Good): Comprehensible explanations for predictions.
- 3 (Average): Basic explanations that may lack clarity.
- 2 (Below Average): Limited explanations that are often unclear.
- 1 (Poor): No explanations provided or entirely incomprehensible.

8. Monitoring and Correction

- 5 (Excellent): Robust monitoring, efficient correction mechanisms, and clear user involvement.
- 4 (Good): Efficient monitoring and correction mechanisms with user involvement.
- 3 (Average): Adequate monitoring with basic correction mechanisms and user involvement.
- 2 (Below Average): Limited monitoring and correction mechanisms with minimal user involvement.
- 1 (Poor): No monitoring or correction mechanisms, and no user involvement.

9. Studies and Transparency

- 5 (Excellent): Independent studies, transparent methodologies, and representative training data.
- 4 (Good): Third-party studies with transparent methodologies.
- 3 (Average): Some transparency in studies and methodologies.
- 2 (Below Average): Limited transparency in studies and methodologies.
- 1 (Poor): No studies or transparency in methodologies.

10. User Interaction and Feedback

- 5 (Excellent): User-friendly feedback mechanisms, high accessibility, and positive usability assessment.
- 4 (Good): Effective feedback mechanisms with good accessibility.
- 3 (Average): Adequate feedback mechanisms and accessibility features.
- 2 (Below Average): Limited feedback mechanisms and basic accessibility.
- 1 (Poor): No feedback mechanisms or accessibility features.

Use-case specific questions

Departments may want to include additional questions based on the specific use case for the AI system. Below are a few examples of use-case specific questions. The lists below should be used as a starting point for ideating questions that would be helpful to ask a vendor.

Use Case: Translation-Based AI Systems

- Please provide a list of all languages your solution supports for live speech translation, beyond the mandatory languages specified in Attachment A. Additionally, describe how often you introduce new languages to your platform and outline the process for these additions.

- Does your solution offer text-to-speech functionality or other methods to accommodate American Sign Language speakers in both in-person and virtual meetings? If so, please describe the features and functionalities that enable this accommodation, including any limitations or requirements for optimal performance.
- How do you measure the performance of your solution in terms of translation, speech-to-text, and text-to-speech for English, Spanish, Tagalog, Mandarin, and Vietnamese? Please specify the metrics used (e.g., BLEU score, human ratings, etc.) and the conditions under which these measurements were taken. Based on these metrics, what is the performance of your solution for the specified languages?
- Please provide/describe any third-party evaluations or benchmarks that have been conducted on your solution's translation capabilities. Include details on the evaluation process, criteria, results, and any subsequent improvements made to the solution based on these evaluations.
- What is your solution's average time delay between speech input and translation output?
- How does the solution handle rapid speech or overlapping conversations?
- How can the solution adapt to idiomatic expressions, cultural references, or local slang?

Use Case: ADA Compliance and Accessibility Considerations

- How well does your solution adapt to blurring, obstruction, poor lighting or any conditions that may lead to misclassification?
- Does your product offer built-in accessibility features for users with visual, auditory, motor, or cognitive disabilities? Please describe these features in detail.
- Has your product been tested for compatibility with common assistive technologies like screen readers, voice recognition software, and alternative input devices?
- Has your product undergone accessibility testing with users who have disabilities? If so, please share the testing methodology and key findings.
- How can users with disabilities provide feedback about the product's accessibility? Are there dedicated channels or support options for such concerns?

Generative AI Guidelines

Generative Artificial Intelligence (AI) is a branch of AI technology that can generate content—such as written word, presentations, images, video, voice, and music— at the request of a user. The City recognizes the opportunity for a controlled and responsible approach that acknowledges the benefits to efficiency while minimizing the risks around AI bias, privacy, and cybersecurity.

The City follows a set of Generative AI Guidelines, which are maintained on the City's IT website at https://www.sanjoseca.gov/your-government/departments-offices/information-technology/itd-generative-ai-guideline.