AI Risk Management Framework

City of San José

Executive Summary

The National Institute of Standards (NIST) released version 1.0 of its AI Risk Management Framework (RMF) in 2023, NIST's voluntary framework for AI governance. In reviewing San José's existing AI governance against the NIST AI RMF, we identified gaps in our existing approach. To improve our existing governance, we recommend implementing the following components:

- Formal citywide AI policy that establishes key aspects of AI governance
- Education for department staff, including general AI education and training on City practices
- Formal mechanisms to solicit feedback from system end-users
- Comprehensive evaluation procedures for AI systems in the field

These components are part of our Equity through Data and Privacy 2023-2024 strategic plan. In the coming months we will identify any necessary budget asks to further address these gaps in 2024-2025.

Context

1. The City of San Jose uses AI to serve a population of 971,233 residents (July 2022).
2. AI has the potential to significantly impact residents' lives. Trust in the City's AI programs will enable greater feedback channels for community initiatives.
3. Since federal and state legislation lag behind AI adoption, the City plans to lead AI governance efforts that may be replicated in other municipalities and state agencies.
4. The AI Risk Management Framework (RMF) is helpful in identifying areas for growth in the City's current AI program and provides a model for standardizing AI policies.

Current state of San José AI Governance

The City's current AI governance process consists of the following components:

People:

1. Centralized AI governance owner in Information Technology Department
2. AI Governance Working Group
3. Advisory Taskforce in the form of the Digital Privacy Advisory Taskforce

Processes:

1. Required procurement review for all technology purchases
2. Initial risk analysis of any proposal
3. Impact assessment for mid and high-risk proposals
4. Data Usage Protocols that define how an AI system can be used
5. Public Algorithm Register for approved AI systems that impact the public
6. Ongoing monitoring of high-risk systems to ensure continued compliance and performance

More details on the City's AI Review Framework can be found online.

Through the review, we identified four key gaps in our current governance structure, discussed in the "results and impact" section.

NIST AI RMF Review Process

AI Risk Management Framework
Measure
Map
Govern
Manage

The AI RMF allows us to identify areas for growth in our AI risk management approach. It defines AI governance into four categories:

1. Govern: A culture of risk management is cultivated and present
2. Manage: Risks are prioritized and acted upon based on a projected impact
3. Map: Context is recognized and risks related to context are identified
4. Measure: Identified risks are assessed, analyzed, or tracked

Under each category are several components comprised of measurable subcomponents. We assessed our existing AI governance against the AI RMF's 72 subcomponents, and graded ourselves from one to four.

1. New, no action by the City yet
2. Setting up or in early progress
3. Fully established and operational
4. Established and constantly improving

NIST emphasizes that the goal is not to reach perfection in every subcomponent. It is a voluntary resource and each organization will have specific components most essential for it. For the sake of completeness in our initial assessment, we reviewed every subcomponent.

Results

The subcomponents range from core risk management including cybersecurity and legal protections, to novel risk management approaches specific to AI. Below are the averaged scores for each category:

1. Govern: 2.3
2. Manage: 2.3
3. Map: 2.8
4. Measure: 1.7

The low "Measure" score is partially due to many subcomponents focusing on the technical development process of AI models. Most AI systems in the City are procured rather than developed internally, but this may change in future years.

Identified Gaps

Checking our current AI governance processes against the AI RMF helped us identify multiple areas of growth, particularly in education and testing/measurement. After reviewing the AI RMF, we identified the need for:

- Formal citywide AI policy that establishes key aspects of AI governance, including:
  - Roles and responsibilities
  - Executive sponsorship
  - AI principles
  - Purpose and vision for AI in the City

- Education across City departments, including general AI education and specific training on our City's AI governance practices
- Formal mechanisms to solicit feedback from system end-users
    - Expectation for departments to raise concerns
    - Reporting mechanism for departments and administrators into policy
- Comprehensive evaluation and procedures for:
    - Robustness testing
    - System drift
    - Stress testing
    - Documenting system errors
    - Decommissioning AI systems

Next Steps

To close the gaps identified through this analysis, we will:

1. Draft and propose a citywide AI policy
2. Provide regular education for analyst and mid-level staff to learn how to use AI in accordance with the City's practices
3. Increase usage of avenues for feedback from system end-users, such as our Generative AI usage form
4. Explore partnership with a third party on an improved evaluation procedure for active AI systems

Feedback on the AI RMF – to share with NIST

NIST is seeking feedback on version 1.0 of their AI RMF. This section is to provide that feedback.

The AI RMF is a great tool for agencies to use as the field works toward standardized AI governance practices. It is a robust and comprehensive list of metrics for organizations to review their progress. The City used the AI RMF as a supplementary tool to "sanity check" our own risk management processes and identify areas for growth in our current AI program.

For many practitioners, the AI RMF may be somewhat redundant and overwhelming. The very long list of standards could feasibly be condensed into a shorter set that is more digestible for practitioners "on the ground" to reference and implement on a daily basis. As the City checked our own AI governance practices against the AI RMF, we found the process to be overly tedious and time-consuming. This is partly because many of the metrics in the AI RMF feel extremely redundant. Although metrics like MEASURE 1.1 and MEASURE 2.2 are slightly different, this difference is too subtle to make a meaningful difference to practitioners, and we found ourselves assigning out virtually the same rating and comments for redundant metrics.

In addition, a few metrics in the AI RMF seemed unreasonable or unattainable, at least in the current landscape of AI governance. These metrics were often too vague, abstract, and wordy. For example, MAP 2.3 details a long list of abstract items that should be documented or established throughout the lifecycle of the AI system. Similarly, MAP 5.1 itemizes a longlist of technical mechanisms and tests. For both of these metrics, only a very mature program would be able to fulfill them and understanding what the metrics are even asking for could be difficult.