City of Sugar Land's AI Guiding Principles

Scope

This guidance applies to all City departments, employees, and entities involved in the development, procurement, deployment, or use of AI technologies.

Purpose Statement

The purpose of this document is to establish a robust framework of guidelines and principles for the ethical and efficient utilization of artificial intelligence (AI) within the City of Sugar Land. Embracing the dynamic potential of AI to enhance services and processes, our goal is to integrate AI alongside other cutting-edge technologies, fostering the development of a smart and sustainable City and community. This creation of this guidance has largely been spurred on by the rapid and widespread deployment of generative AI technologies.

As artificial intelligence (AI), particularly generative AI, continues to evolve, it is anticipated that advancements and innovations will persist over time. Nonetheless, the guiding principles outlined herein are intended to remain steadfast and serve as a foundational framework for the City's utilization of AI. These principles are designed to ensure consistency, ethical standards, and reliability in the deployment and management of AI technologies within our community.

At the core of our strategy lies a steadfast commitment to aligning AI integration with established best practices and commitments, particularly in the realm of risk management, while prioritizing security, transparency, accountability, and adherence to other ethical principles. Recognizing the dual nature of AI, with its capacity for improving operational efficiency and social equity as well as posing risks if not appropriately governed, our guidelines prioritize responsible AI usage.

Guiding Principle 1: WE WILL BE PURPOSEFUL

Employees are encouraged to experiment with the application of AI through controlled trials or small-scale pilot projects to assess the feasibility, effectiveness and impact of integrating that AI technology into specific use cases. The engagement of AI should address City business or service issues, be purpose driven, and should align with providing a life even better than one can imagine.
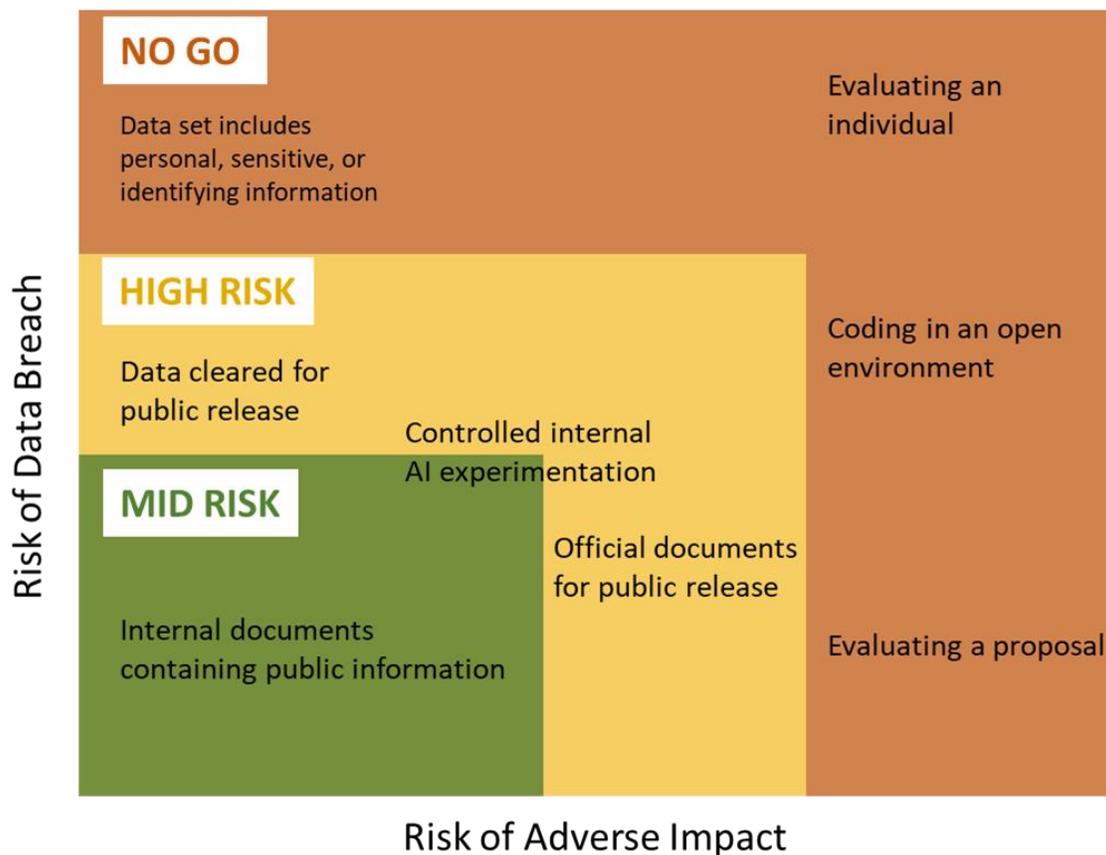
The City's use of AI is intended to be results-driven, helping to streamline bureaucratic processes, optimize resource allocation, and improve decision-making by analyzing large amounts of data. Employees are advised to consider ethical implications and ensure that biases are mitigated throughout the development and deployment of AI systems, thereby guaranteeing these systems serve the public good.

Guiding Principle 2: WE WILL MANAGE RISK

The application of new technologies almost always comes with risk and unintended consequences. Before applying AI technologies, responsible parties should make a deliberate and reasonable attempt to identify potential risks, liabilities, likely consequences, severity of consequences, risk responses, and potential mitigations. AI users should determine with a high level of confidence which identified risks are acceptable and those that might render the application of AI unsuitable. When a user plans to apply AI technology to a data set for a specific purpose, process, or result, they should consider the potential consequences and risks of that effort escaping its original purpose.

Designing an all-encompassing, one size fits all, AI Risk Matrix evaluation tool would fail to fully protect Sugar Land's citizens and employees from the potential negative consequences of errant AI use. Therefore, it rests with the process owner, those who will employ AI for generative purposes, to consider potential risks and outcomes. The responsible party may wish to collaborate with other process owners, participants, and end users, to ask the "what if" questions necessary to identify risks and determine acceptability. An example AI Generative Risk Matrix is provided.

Example Risk Matrix – Generative AI

NO GO

Data set includes personal, sensitive, or identifying information

Evaluating an individual

HIGH RISK

Data cleared for public release

Coding in an open environment

Controlled internal AI experimentation

MID RISK

Official documents for public release

Internal documents containing public information

Evaluating a proposal

Risk of Data Breach (vertical axis)

Risk of Adverse Impact (horizontal axis)

Guiding Principle 3: WE WILL ENSURE CREDIBILITY, ACCOUNTABILITY, & TRANSPARENCY

As trailblazing public servants, the scrutiny that accompanies this brand of service should be accepted and embraced. A commitment to maintaining a standard of transparency in all efforts, including the use of AI technologies, should be upheld.

The City's AI efforts should be committed to credibility, reliability, sustainability, and verifiability. Ensuring the mathematical accuracy and scientific soundness of AI technologies for each specific application should be a priority. Industry standards for margins of error should be adhered to, validating the reliability and accuracy of AI products.

Staying informed about industry standards, best practices, and emerging research is expected of AI users and process owners. Relevant updates and advancements should be actively incorporated into the City's AI initiatives to keep AI use current, effective, and aligned with the latest developments in the field.

Accountability for the processes, results, and consequences of AI products and outcomes should be maintained. Transparency is empowered through discernible AI-generated content, with visible markers such as watermarks facilitating this. When content generated by AI is publicized, an effort to specify this to the public should be made.

Guiding Principle 4: WE WILL PROVIDE PRIVACY, SAFETY & SECURITY

In all applications of AI technology within the City, security, safety, and privacy should be prioritized. AI systems should uphold confidentiality, with stringent safeguards in place to prevent unauthorized access or use. Efforts should consistently focus on minimizing risks to individuals. Preservation of privacy is paramount, necessitating the protection of personally identifiable information (PII) and other sensitive data, including that which may jeopardize homeland security, from unauthorized disclosure, access, or manipulation.

To ensure adherence to these principles, consider the following actionable methods:

1. Understanding Data Usage: Prior to engaging with an AI system, users should be familiar with its data collection and utilization processes, and possible sale of data to third parties. Review associated privacy policies to grasp the scope of data collection, processing methods, and access protocols.

2. Minimal Disclosure of Personal Information: Users should limit the provision of personal information to only what is strictly necessary for effective use of an AI system. Users

should also exercise caution in sharing sensitive data, disclosing it only when indispensable.

3. Regular Permissions Review: Users should routinely assess permissions granted to AI programs on their devices. Authorize only those permissions indispensable for optimal functionality.

By adhering to these practices, the City can ensure that AI technologies operate within a framework that prioritizes security, safety, and privacy for all stakeholders.

Guiding Principle 5: WE WILL CULTIVATE EQUITY & INTEGRITY

The responsible development and deployment of AI necessitates a commitment to inclusivity, respect, fairness, and freedom from bias. Inclusivity is paramount; AI prompts should be constructed with all stakeholders in mind, regardless of background or identity. Inputs used for algorithms should undergo rigorous scrutiny to ensure that the output generated by AI will be fair and free from bias or stereotypes. Input data and prompts should not be simultaneously entered into an algorithm in a way that reinforces or suggests existing biases or stereotypes. Furthermore, prompts should not be written to generate a desired answer; instead, they should rely on value-neutral language so that the validity of the resulting responses can be confirmed for accuracy, fairness, and freedom from bias. Adherence to these principles ensures that AI systems benefit everyone and uphold accountability at the City.