Software Application Procurement Principles for Utah Government Entities

_____

Commission on Protecting Privacy and Preventing Discrimination

Office of the State Auditor
Issued February 1, 2021

_____

OFFICE OF THE

STATE AUDITOR

AUDIT LEADERSHIP:

John Dougall, State Auditor
The Commission on Protecting Privacy and Preventing Discrimination

OFFICE OF THE
STATE AUDITOR

Software Application Procurement Principles for Utah Government Entities

Commission on Protecting Privacy and Preventing Discrimination

Office of the State Auditor
Issued February 1, 2021

We recommend government entities apply the following principles as they procure commercial software applications or engage in development of custom software applications that use, gather or consolidate personally-identifiable information (PII) or other sensitive data. These principles are most suited to new or emerging technologies, such as artificial intelligence (AI) or machine learning (ML), that may not have a long history to draw upon for software application and vendor evaluation as well as for "startup" or young vendors that likewise may not have extensive history.

1. **Limit Sharing of Sensitive Data:** Government entities should fully understand their data. They should limit sharing of sensitive data (private data, PII, etc.) to the greatest extent possible to protect individual privacy and should not share more than is necessary to perform the required task. Data should be filtered and restricted within the government's systems before being transferred into the vendor's application. Wherever possible, a government entity should anonymize data, but government entities should recognize that sensitive data can be reconstructed from previously anonymized sources.

2. **Minimize Sensitive Data Collection and Accumulation**: A software application should collect no more sensitive data than it needs, and should retain that sensitive data no longer than it needs to accomplish the specified purpose.

3. **Validate Technology Claims - including Capability Review:** A vendor should clearly demonstrate the validity of their marketing claims. Example claims that warrant particular caution include
    a. Asserted use of AI or ML,
    b. Proposed use of disparate data sources, especially social media or integration of government and private sources, and
    c. Real-time capabilities, especially real-time data intelligence or analysis.

4. **Rely on Objective, Repeatable Metrics**: Vendors make various claims about the ability of their software applications to deliver value within a given accuracy or efficiency measure. Do not rely on anecdotes as validation of these claims. Government entities should invest in software applications where the value can be measured on an ongoing basis. A reputable vendor should include success criteria in any Request For Proposal (RFP) response, and these should include metrics that are easy to measure and compare across time and vendors. The RFP should also request that work to automate the gathering and reporting of these metrics be included in the project definition.

5. **Assess Threat Models**: The vendor should be able to enumerate the people, processes, and technological interfaces that constitute an attack or risk surface for their proposed software solution. These threats should be prioritized, and high-priority threats should have recommended mitigations. The vendor should have a vulnerability reporting process. A documented history of conducting and remediating penetration tests is a significant benefit.

6. **Perform In-Depth Review of AI/ML Algorithms:** All claims of AI or ML should be clearly validated and explained, including:
    a. AI algorithms used in the software application
    b. model training and evaluation procedures used
    c. model outputs utilized by product / feature
    d. source and composition of the training, validation and test datasets
    e. demonstration of having a license to use those datasets
    f. pre- and post-processing methods applied to data
    g. processes for model lifecycle management, including on-going evaluation

    The output of an AI-based software application may still have issues of bias or lack of fairness, even if the inputs and system are reasonably judged not to include such failings. The output of the software application should be monitored to ensure protection of privacy and avoidance of improper or unexpected bias.

7. **Demonstrate Privacy Compliance: Privacy Specific Items and Protection:** The vendor should demonstrate compliance with privacy regulations, such as CCPA or any similar laws enacted by Utah.
    a. The vendor should define what constitutes PII under the contract. A government entity's default definition may be overly narrow and may need adjustment for particular problems.
    b. The vendor should describe their anonymization process and how it protects against the use of secondary data to de-anonymize data. A governmental entity should evaluate the effectiveness of that process to mitigate de-anonymization in the context of the software application.

c. Specific certifications may be required for a specific application, but such certifications may still be insufficient to protect privacy.

8. **Review Steps Taken to Mitigate Discrimination**: Ensure that the vendor has considered the question of bias and discrimination within their software application and that the vendor has mechanisms, such as audit results, to demonstrate that their software application does not disproportionately affect various categories of individuals, particularly any federally protected class.
    a. For example, consider sources of data that may include implicit or historic bias (e.g., distribution of video cameras by region or neighborhood)
    b. For example, consider how model choice and training may introduce bias.
    c. Understand the interpretation of model output.
    d. For AI-based or ML-based software applications, determine whether the source of training and model data has been evaluated by subject matter experts.
    e. Entity should use best in class models for evaluation to prevent discrimination, particularly in the case of biometric analysis or facial recognition. As an example, the U.S. NIST provides evaluations of the accuracy of facial recognition based on demographic differences.

9. **Determine Ongoing Validation Procedures:** The government entity must have a plan to oversee the vendor and vendor's solution to ensure the protection of privacy and the prevention of discrimination, especially as new features/capabilities are included.

10. **Require Vendor to Obtain Consent of Individuals Contained Within Training Datasets:** Many biometric characteristics may be captured without an individual's knowledge or consent. Examples may include facial recognition or gait analysis. Ensure that a vendor has consent from the individuals whose biometric characteristics are used in training datasets.

11. **Vet Key Vendor Personnel:** Key vendor personnel may need background checks. The type of checks may vary depending upon the sensitivity of data that personnel have access to. These checks need to be validated to the procuring government entity.

12. **Evaluate Vendor Corporate Management and Vendor Solvency:** Evaluate the financing history of the vendor and their solvency to ensure they can carry out the contract. Placing code in escrow may be a compensating mechanism here.