



Ensuring Responsible AI at SailPoint

Core Principles for AI Development



Ensuring Responsible AI at SailPoint

January 2025

Introduction

At SailPoint, AI-driven capabilities are at the core of identity security. SailPoint's AI-powered solutions enhance decision-making, accelerate risk detection, and deliver seamless integrations. We leverage AI to deliver customized recommendations and actionable data insights tailored uniquely to each customer by deciphering intricate data patterns and user behaviors.

However, with the increasing reliance on AI, concerns about responsibility and trustworthiness have understandably surfaced. At SailPoint, we recognize these apprehensions and prioritize the following characteristics of trustworthy AI to help ensure that our products support the ethical adoption and use of AI.

1. Privacy & Security

As an identity security company, security is at our core. All of our products are developed and released in accordance with our strict cybersecurity certification requirements and rigorous security measures. Customer personal information is governed by our GDPR-compliant [privacy program](#). We use personal information only when necessary and avoid training models using personal data whenever possible. If personal information is required for the functionality of a feature, then the feature will be designed to use a dedicated model that is deployed only to the specific customer's tenant in order to minimize data usage and exposure risk. Additionally, our engineering and security teams use a set of both automated and manual tools to check for security vulnerabilities during the development process and in production. Automated unit test suites that include security components are executed multiple times daily as part of the continuous product builds.

For more about information security, privacy, and compliance at SailPoint, visit our Trust Center at <https://www.sailpoint.com/why-us/trust>.

2. Explainability, Interpretability, and Transparency

For users and operators of AI systems, the ability to make sense of, contextualize, and understand the impacts of the system are critical requirements for exercising appropriate oversight and effective AI deployment. Where most helpful, SailPoint's products include detailed insights about the factors that contribute to AI model outputs, such as recommendations. This helps the user to ensure that our recommendations are relevant to their unique organization. For example, within our Identity Outliers feature, we provide data points to the end user such as:

- Peer Access Similarity – How similar an identity's access is to the closest peer
- Rare Access – The percentage of the identity's entitlements held by less than 1% of the organization
- Entitlement Count – An identity's total number of entitlements
- Roles/Access Profiles with a Single Entitlement – The number of the identity's assigned roles and access profiles that contain only one entitlement

These data points allow users to understand, identify, and override any decisions which they may disagree with. These data points also provide the transparency that many of our customers require to trust the outputs of our ML models.

3. Fairness and Bias Management

Fairness in AI-driven recommendations is an important consideration. Our AI-based products leverage machine-learning algorithms. In cases where the products make recommendations impacting individuals, these algorithms are dedicated to a customer's instance and trained on that customer's data. This means that each recommendation made by our AI is based on an organization's own data, helping ensure that outputs resonate with the customer's unique, specific requirements. Customers can choose which attributes, or datapoints, to use to influence the model for their instance, based on relative appropriateness for that customer's organization. Because SailPoint governs security for enterprises, the relevant attributes about individuals are based on business information rather than on unique or sensitive personal information that could expose individuals to unfairly biased outcomes. Commonly chosen attributes are:

- job title
- department
- location
- cost center



This level of customization and transparency helps provide our customers with confidence that attributes historically linked to unfairness in social systems (e.g., race, gender, religious preference) are not being used to influence outcomes. In addition, SailPoint's data science team scrutinizes each algorithm prior to deployment in an effort to detect and rectify any potential biases via a series of readiness assessments and monitors deployed models to identify changes in performance and detect unfair outcomes. Our team of experts dedicates extensive effort to executing models and confirming equitable, unbiased recommendations, prior to making those models available in our product features.

4. Accountability

As AI tools take an ever-more active role in our world, concerns about the potential impact of AI and AI-driven decision-making on humans and our environments are rightfully top-of-mind. SailPoint's philosophy emphasizes using AI only when it makes sense to solve real customer problems; in turn, the intended purposes that our AI tools serve are focused and discrete, existing exclusively within the SailPoint ecosystem.

We also recognize that human judgment is irreplaceable. While our products (when used for their intended purposes) would not be expected to have negative impacts on individuals, we also ensure that our AI features do not make decisions automatically. Instead, a human reviewer is presented with recommendations and explanations and is empowered to determine what actions (if any) to take. Choosing to decline the suggestion is always a valid and available option.

5. Validity & Reliability

All SailPoint models are evaluated, tested, and monitored by SailPoint using our internally-managed AI Platform. Models deemed performant enough for release are deployed using the AI Platform after development testing is complete. Prior to deployment to a customer's tenant, models are trained and validated against that customer's data on a per-feature basis. We then monitor model performance using dashboards and alerts generated by our AI Platform to detect changes in performance or performance degradation. Metrics and performance indicators including accuracy, precision, recall, F1 score, probabilistic distance metrics, user feedback, KL-divergence, and drift are computed and continually monitored via



dashboard and alerts. Models are retrained on a regular schedule ranging from daily to monthly, or more frequently as necessary if drift is observed.

Summary

At SailPoint we are committed to the fairness, transparency, and trustworthiness of our AI offerings. We encourage continuous dialogue and feedback, further strengthening our stance of providing safe, beneficial, and equitable AI experiences, tailored uniquely to each customer.