Auth0 Bot Detection

Model Card

Okta Model Cards are intended to provide information about models leveraged by Okta in Okta's product offerings and include information on the intended use cases, limitations, training, and evaluation of models. Model cards are not intended to be technical reports and are provided for informational purposes only. Model cards may be updated from time-to-time.

Model Card: Auth0 Bot Detection

Overview

- **Product/Feature Name:** Auth0 Bot Detection
- **Description**: The Bot Detection feature performs pre-login assessments across critical user flows, including login, signup, and password reset to determine if the login, signup, and password reset requests are from a bot. Two separate machine learning (ML) models are used: one for login and password reset attempts, and another for signup attempts. Any attempt to log in, create an account, or reset a password from signals indicating elevated risk, such as a suspicious IP address or unusual traffic pattern, may be required to complete an additional verification step.
- **Primary Function:** The models mitigate scripted attacks by detecting when a request is likely coming from a bot. These types of attacks include credential stuffing attacks or list validation attacks. Bot Detection provides support against certain attacks and adds protection with minimal friction for legitimate users.

Model Details

- **Model Version:**
    - **Model Version for login model:** v5
    - **Model Version for signup Model:** v1
- **Model Type:** Two separate ML models are used
- **Model Origin:** In-house developed models
- **Model Provider:** Okta
- **Model Architecture:** The feature utilizes two models with similar architecture: an ensemble-based supervised learning model for classification using threshold-based decision logic. One model is dedicated to login and password reset flows, while the other is used for signup flows.

Intended Use & Limitations

- **Intended Use Cases:** The models are intended to protect login, signup, and password reset flows from automated attacks. The login model is intended to protect the login and password reset flows, while the signup model is intended to protect the signup flow.
- **Out-of-Scope Use Cases:** Any use other than the intended use case is out of scope and not recommended.
- **Known Limitations:** Bot Detection does not guarantee 100% detection of bots.

- **Potential Risks:** What are the potential risks or ways the model could fail or produce problematic outputs? Check all that apply and briefly explain:
    - ☐ **Factual Incorrectness (Hallucinations):** The model may generate information that is not factually correct.
    - ☐ **Bias:** The model may produce outputs that are biased against certain demographic groups or reflect societal stereotypes.
    - ☐ **Harmful or Inappropriate Content:** The model could generate offensive, unsafe, or otherwise inappropriate content.
    - ☑ **Other**: The models can produce false negatives or false positives.

Data

- **Model Inputs:** The models use a wide range of signals from user login, signup or password reset attempts which may include IP address, ASN, User Agents, and third-party bot scores. Additional inputs may be used to provide and improve the models.
- **Model Outputs:** A risk determination of whether the request is likely coming from a bot. The output is used by an internal service to determine if a challenge is required based on the tenant's settings.
- **Data Minimization**: The models process the telemetry required for threat detection.
- **Training Data**: Both models are trained on a continuous stream of structured, system-generated telemetry from the Okta authentication service. The training process leverages a vast and diverse set of non-personal network and behavioral signals.
- **Is the model trained on Customer Data** (as defined in Okta's Master Subscription Agreement at https://www.okta.com/legal/)**?** The models do not train on Customer Data.

Evaluation

- **Methodology:** The performance of both models is monitored across Okta's production regions to help ensure adequate performance across regions and groups.
- **Performance Metrics:** Internal dashboards and tools are used to monitor the efficacy and performance degradation of the models. Alerts are in place for performance degradations, and model monitoring is continuous. While Okta does not publish detailed performance metrics, the models are benchmarked internally using measures such as precision and recall, with continuous monitoring for drift.

Artificial Intelligence (AI) Principles

Okta strives to safely use and develop AI to strengthen the connections between people, technology, and our community. When it comes to AI innovation, we aim to live our core values and harness the power of AI in a way that reflects said values. This kind of thinking is part of our DNA. That's why we take a values-based approach to AI. Okta's Responsible AI Principles underscore (i) transparency; (ii) building customer trust through security, privacy, and safety; (iii) accountability; and (iv) innovating responsibly regarding inclusivity, fairness, and ethics.

These principles are aligned with Okta's values: "Love our customers." "Always secure. Always on." "Build and own it." "Drive what's next."

Our developers adhere to responsible AI principles regarding privacy, security, responsible innovation, and more general principles and obligations regarding Customer Data. For more information, please see the published full version of Okta's Responsible AI Principles on Okta.com.

Security and Privacy

- Okta adheres to its existing commitments regarding security, privacy, and confidentiality in connection with Okta products and features that leverage AI that are offered as part of the Okta services.
- Okta follows industry standard processes for testing, developing, and making available products and features that leverage AI for customers.
- Okta has policies and programs in place regarding the use of and governance over AI.
- The data validation measures Okta takes for products and features that leverage AI may vary by product and feature and may include measures like input sanitization, having an allow list of characters that can be passed in the input, having a block list of terms that will be rejected, and having a custom post processing step that validates the output depending on the use case.
- The measures Okta has in place to help ensure that the models leveraged by Okta in Okta's product offerings are accurate and unbiased may vary by product and feature and may include monitoring the performance of models, auditing data to identify inaccuracies or missing information, having a diverse team of developers and data scientists that develop, maintain and improve Okta's products that leverage AI, and having a human in the loop when necessary.

Last Updated: September 30, 2025