Okta ThreatInsight

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: Okta ThreatInsight

Overview

- **Product/Feature Name**: Okta ThreatInsight
- **Description**: ThreatInsight analyzes sign-in attempts for potentially suspicious activity. If suspicious events are found, it records the events and can deny access to the requests if configured to do so. Two separate machine learning (ML) models are used: one to detect when an org is under attack ("Model One"), and another to detect suspicious requests to rate limit them ("Model Two").
- **Primary Function**: Analysis & Insights: The model provides predictions, scores, or analytical insights from data.

Model Details

- **Model One - Org Attack Detector Model**
    - **Model Version:** v1
    - **Model Type**: ML
    - **Model Origin**: In-house developed
    - **Model Provider**: Okta
    - **Model Architecture**: Streaming Anomaly Detector
    - **How is the model accessed?**: The model is accessed offline to detect when tenants are under large credential based attacks.
- **Model Two - Suspicious Request Detector Model**
    - **Model Version:** v1
    - **Model Type**: ML
    - **Model Origin**: In-house developed
    - **Model Provider**: Okta
    - **Model Architecture**: Supervised Ensemble Learning Model
    - **How is the model accessed?**: The model is accessed online to detect suspicious requests and isolate suspicious traffic from legitimate traffic for rate limiting purposes.

Intended Use & Limitations

- **Intended Use Cases**:
    - **Model One** - Org Attack Detector Model: The model detects when an org is under attack by establishing a baseline of normal login failures and identifying and flagging anomalies that may indicate an attack. When the model detects such an attack, we alert the customer administrator with a SystemLog event and apply aggressive heuristics to identify and block potentially malicious IP addresses.
    - **Model Two** - Suspicious Request Detector: The model identifies and isolates suspicious requests to prevent them from overwhelming legitimate traffic. By rate-limiting these requests instead of blocking them, the model helps ensure that genuine users aren't affected during a large-scale credential-based attack.
- **Out-of-Scope Use Cases**: When requests come through proxies, for Okta to correctly identify the originating client IP address, trusted proxies should be configured in Network zones.
- **Known Limitations**: ThreatInsight blocks some malicious traffic, but it can't guarantee it will detect and block every malicious IP address or threat.
- **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 model can produce false negatives (i.e. malicious requests are not blocked) or false positives (i.e. valid requests are blocked).

Data

- **Model Inputs**: The primary inputs are data points collected during user authentication attempts across the Okta network. Inputs may include IP address, geolocation data, user agent and device information, and authentication telemetry (data about the success, failure, and velocity of login attempts).
- **Model Outputs**:
    - **Model One** - Org Attack Detector Model: The model detects when an org is under attack by establishing a baseline of normal login failures and identifying and flagging anomalies that may indicate an attack.
    - **Model Two** - Suspicious Request Detector: The model identifies and isolates suspicious requests.
- **Data Minimization**: The models process telemetry data needed for threat detection.
- **Training Data**: The training data comes from the Okta authentication service itself. The models learn from the continuous stream of successful and failed login events generated by all customers using the Okta platform. The models are trained on non-personal telemetry.
- **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 and Security

- **Methodology**: The performance of the models is continuously evaluated using a combination of historical data analysis and live, real-world monitoring. The models are subject to continuous monitoring and tuning by Okta.
- **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 monitoring of the models is continuous. Okta does not publish specific performance metrics, as this information could be exploited by attackers to understand the model's strengths and weaknesses.

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