Okta Identity Threat Protection

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 Identity Threat Protection

Overview

- **Product/Feature Name:** Okta Identity Threat Protection
- **Description:** Okta Identity Threat Protection continuously assesses users and their sessions to identify and enable responses to potential threats in real-time. It identifies potential threats based on changes in user behavior, device health, and other contextual signals during active sessions.
- **Primary Function:** Analysis & Insights

Model Details

- **Model Version:** v1
- **Model Type:** Machine Learning (ML)
- **Model Origin:** In-house developed model
- **Model Provider:** Okta
- **Model Architecture:** The model's architecture is Gradient Boosted Trees. It is a type of Predictive Model that provides predictions from data.

Intended Use & Limitations

- **Intended Use Cases:** This model is designed to protect users' sessions from potential threats, such as hijacking, by identifying changes to users and their sessions.
- **Out-of-Scope Use Cases:** Any use other than the intended use case is out of scope and not recommended.
- **Known Limitations:** The model's output may be less effective if the user's request or data is inaccurate for any reason, such as a false positive from a device.
- **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 behavior is not detected) or false positives (i.e. valid behavior is incorrectly identified as high risk).

Data

- **Model Inputs:** The inputs are system-generated telemetry from Okta's authentication pipeline. These inputs may include data attributes such as: IP, device type, browser, operating system, device identification code, device name.
- **Model Outputs:** The model's output is a risk level. This output determines if a user or a user's session has changed and generates a Syslog event for changes with the corresponding risk level.
- **Data Minimization:** The model processes telemetry needed for threat detection.
- **Training Data:** The model uses historical usage data, telemetry data, and synthetic data for training.
- **Is the model trained on Customer Data** (as defined in Okta's Master Subscription Agreement at https://www.okta.com/legal/)**?** The model does not train on Customer Data.

Evaluation and Security

- **Methodology:** We run the model in production and continuously monitor the performance of the model.
- **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. 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