# Telephony Toll Fraud Mitigation

## 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:** Telephony Toll Fraud Mitigation

---

## Overview

- **Product/Feature Name:** Telephony Toll Fraud Mitigation
- **Description:** Okta's machine learning model detects and mitigates telephony toll fraud (SMS pumping) attacks by analyzing behavioral patterns and request metadata in real-time. The model operates as an invisible security layer that identifies suspicious authentication requests and applies targeted rate-limiting to fraudulent pathways without disrupting legitimate user access.
- **Primary Function:** The model provides real-time fraud detection and risk scoring for SMS and voice OTP (One-Time Password) delivery requests within Okta's Multi-Factor Authentication (MFA) system. It analyzes request velocity, behavioral signals, and telemetry patterns to classify transactions as legitimate or fraudulent, enabling Okta to prevent financial losses from SMS pumping attacks while maintaining seamless user experience for 99% of legitimate users.

---

## Model Details

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

---

## Intended Use & Limitations

- **Intended Use Cases:** Real-time detection of telephony toll fraud (SMS pumping) attacks during OTP delivery.  Risk scoring of authentication requests to identify suspicious patterns. Targeted rate-limiting of

fraudulent request pathways to disrupt attacker profitability. Protection of Okta customers from financial losses due to premium-rate SMS/voice call exploitation

- **Out-of-Scope Use Cases:** The model is not designed for (i) use in high-risk or safety-critical systems beyond fraud detection; (ii) providing financial or legal advice; (iii) demographic profiling or identity verification beyond fraud risk assessment; and (iv) applications to use cases outside the telephony authentication domain
- **Known Limitations:** Limitations include (i) Zero-day fraud pattern failure; and (ii) Extreme Class Imbalance where the model trained on heavily imbalanced data can naturally develop a bias toward predicting the "majority class" (legitimate traffic), causing them to miss subtle or low-volume fraud campaigns to avoid ruining accuracy metrics.
- **Potential Risks:**
  - ☑ **Factual Incorrectness (Hallucinations):** False negatives can occur, specifically with novel or sophisticated fraud patterns which differ from historical training data.
  - ☑ **Bias:** The model may exhibit bias toward the majority class (legitimate traffic), reducing sensitivity to low-volume fraud campaigns.
  - ☐ **Harmful or Inappropriate Content:** The model could generate offensive, unsafe, or otherwise inappropriate content.
  - ☑ **Other:** Legitimate users may experience temporary rate-limiting or delays in receiving OTP codes if flagged as suspicious.

---

## Data and Privacy

- **Model Inputs:** Structured telemetry and event metadata extracted from live authentication requests.
- **Model Outputs**: Risk score classification.
- **Data Minimization**: All input data is anonymized prior to storage and processing. Training data consists of aggregated telemetry records rather than individual transaction details. Tenant identifiers are anonymized to preserve customer isolation. No raw personally identifiable information (PII) or identifiable Customer Data is retained in the model or training pipeline
- **Training Data**: Historical Okta syslog event data derived from authentication activity across the platform.
- **Is the model trained on Customer Data** (as defined in Okta's Master Subscription Agreement at https://www.okta.com/legal)**?** No

---

## Evaluation and Security

- **Methodology:** A held-out test set that is temporally separated from the training data, containing authentication events from a later time window than those used for training. This prevents data leakage and ensures evaluation reflects real-world generalization. Cross-validation is used during hyperparameter tuning on the training set only.
- **Performance Metrics:** The following metrics are used to assess and evaluate the on-going performance of the model:
  (i) Precision and Recall to balance the cost of false positives against false negatives;
  (ii) F1 score as the primary performance indicator;

(iii) AUC-ROC to measure overall discriminative ability,
(iv) False Positive Rate to enable minimal disruption to legitimate users.

---

### 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 Aug 10, 2026