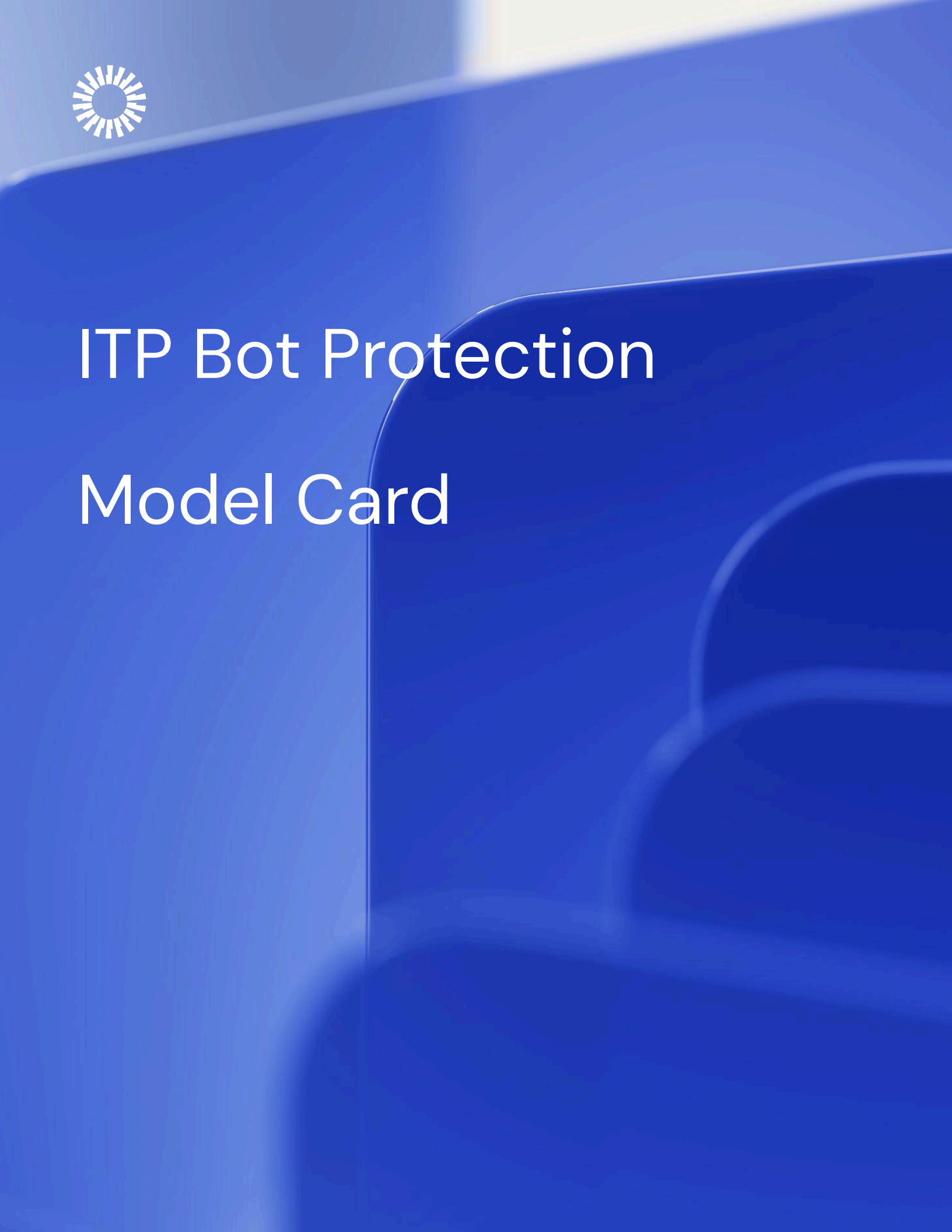

ITP Bot Protection

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: ITP Bot Protection

Overview

- **Product/Feature Name:** ITP Bot Protection
- **Description:** The model analyzes login and signup activity to detect automated "bots." It does this by checking for patterns associated with bots, such as anomalous request characteristics and behaviors. Users do not interact with the bot detection system directly. It runs in the background. If the model flags the request as "suspicious," the feature may challenge the user's browser as a proof-of-work mechanism or provide other configuration options to impede automated attacks.
- **Primary Function:** Analysis & Insights: Provides predictions, scores, or analytical insights from data.

Model Details

- **Model Type**: ML + Rule Engine
- **Model Origin**: In-house
- **Model Provider**: Okta
- **Model Version**: Bot Detection v1
- **Model Architecture**: Predictive Model

Intended Use & Limitations

- **Intended Use Cases:** The model analyzes login and signup activity to detect automated "bots." It does this by checking for patterns associated with bots. Its goal is to prevent automated bot access while letting real users access the service normally.
- **Out-of-Scope Use Cases: T**he Bot Detection model relies on the ability to correctly identify the originating client IP address to evaluate reputation and behavior. If requests are passed to Okta through trusted proxy IP addresses (such as CDNs or load balancers) without proper configuration, the model cannot identify the originating client IP address. If administrators do not configure trusted proxies properly to preserve and forward the client IP, the model is unable to accurately assess the source and is less effective at detecting threats. The model is designed to detect anomalous behavior in external authentication traffic. It is generally not intended for use in detecting threats within trusted internal network zones where automated traffic (e.g., from service accounts) may be expected and legitimate. The model provides a probabilistic

risk score and should not be used as the sole determinant for critical access decisions in high-stakes or safety-critical systems without accompanying authentication factors (like MFA).

- **Known Limitations:** (1) False Positives/Negatives: Like all predictive models, the model cannot be expected to provide 100% accuracy. It effectively provides a "risk determination" (High/Low probability) rather than a definitive judgment, meaning it can produce false positives (blocking real users) or false negatives (missing bots). (2) New Threats: The model may be less effective against brand-new, unseen attack patterns that were not present in its training data (until it is retrained). (3) Adversarial attacks: Highly advanced bots that mimic human behavior perfectly or extensively disguise their origins can still evade detection.
- **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:** Any predictive model trained on behavioral data carries a standard risk of bias if the training data is unrepresentative of current threats or legitimate intended use patterns (e.g., flagging traffic from certain regions as "suspicious" disproportionately).
  - ☐ **Harmful or Inappropriate Content:** The model could generate offensive, unsafe, or otherwise inappropriate content.
  - ☑ **Other:** False Positives / False Negatives: The model may incorrectly classify legitimate users as bots (false positives), leading to friction or blocked access, or fail to detect actual bots (false negatives), allowing malicious activity to proceed. Model Staleness: If the model is not retrained regularly, it risks becoming "stale" and failing to detect new, evolving threat patterns that were not present in its original training data

---

Data and Privacy

- **Model Inputs:** The primary inputs for the Bot Detection model are technical data points and behavioral signals gathered during authentication or request events.
- **Model Outputs**: The Bot Detection model analyzes request telemetry and outputs a risk score or classification (e.g., High Bot Probability vs. Legitimate User). This internal score is used to determine the immediate handling of the request.
- **Data Minimization**: The model processes only essential security telemetry required for bot classification (e.g., IP, device data).
- **Training Data**: The model is trained on anonymized and aggregated security telemetry generated by activity across the entire network The data is collected from the authentication traffic of the Okta service itself.
- **Is the model trained on Customer Data** (as defined in Okta's Master Subscription Agreement at https://www.okta.com/legal)**?** The model is not trained on Customer Data.

Evaluation and Security

- **Methodology:** The model's performance is evaluated using a combination of "offline" historical analysis and "online" real-time validation, testing the model's ability to generalize to new, unseen traffic rather than just memorizing the examples it was trained on.
- **Performance Metrics:** We prioritize metrics that balance security effectiveness with user experience friction: (1) Precision (Positive Predictive Value): The percentage of flagged requests that were actually bots. (2) Recall (Sensitivity): The percentage of actual bots that were correctly detected. (3) Challenge Rate & Block Rate: Operational metrics monitored to confirm the model isn't aggressively over-blocking traffic during normal baseline activity.

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 Feb 25, 2026