Trellix

We bring security to life.

# Artificial Intelligence and Machine Learning Data Sheet

Trellix is a pioneering leader in cybersecurity powered by artificial intelligence (AI) and machine learning (ML), and is committed to deploying AI/ML in a safe, transparent and responsible way – consistent with our customers' high expectations and in compliance with all regulatory requirements, including the European Union's Artificial Intelligence Act (EU AI Act) and other evolving laws and regulations worldwide.

The purpose of this document is to provide our customers with information on how Trellix uses AI/ML in our products and services, and to help our customers understand and assess the impact of Trellix's data processing on their overall AI/ML compliance posture.

For more information about Trellix's AI vision and guiding principles, see the AI section and FAQs on our Trellix Trust Center, as well as our testimony to the U.S. House of Representatives.

## Overview of AI/ML in Trellix Solutions

In general, Trellix's advanced cybersecurity solutions rely on "human-machine teaming," which combines the speed, efficiency and scale of AI/ML with the deep expertise and judgment of human security analysts.

On the "machine" side, Trellix solutions use AI/ML to analyze telemetry data received from customer networks and devices worldwide to identify patterns and anomalies that may suggest malicious activity, and to proactively and adaptively investigate and respond to threats in real-time. Furthermore, Trellix Wise™, our integrated generative AI feature, automatically facilitates threat intelligence by generating and answering questions, exploring hypotheses, and summarizing evidence from diverse sources.

In turn, on the "human" side, our customers' security analysts use the actionable intelligence and insights provided by our solutions to optimize their organization's security posture and cyber readiness.

Below is a summary of how AI/ML capabilities are embedded in Trellix's core security capabilities.

1. **Advanced Threat Detection:** The following Trellix endpoint protection products use evolving security tools and techniques to proactively identify, analyze, and mitigate

threats that bypass traditional signature-based defenses.

- **Trellix Endpoint Security solutions** include ML Protect, which is a core classification module utilizing supervised learning to determine if a file is malicious or unknown. It employs an ML classifier based on feature extraction and score calculation against a predefined threshold.
- **Trellix Endpoint Security (ENS)** consists of four security modules that work independently and use AI/ML to provide several layers of security.
  - **Threat Prevention** uses XGBoost and RandomForest algorithms to automate malicious file detection within a customer's network by analyzing files against malware definition files (DAT) and Trellix Global Threat Intelligence (GTI).
  - **Firewall** integrates allowlisting ML modules in its "Adaptive" mode. ENS Firewall automatically permits all traffic that does not align with an existing Block rule. It then generates dynamic Allow rules for this non-matching traffic, enabling Trellix customers to store and transfer these administrative rules.
  - **Web Control** employs ML modules to determine the safety rating and content of websites and downloads, querying Trellix GTI for reputation information.
  - **Adaptive Threat Protection** includes ML Protect for automated behavior analysis both on the client system and in the cloud.
- **Trellix Endpoint Detection and Response (EDR)** includes several ML modules to enhance traditional signature-based threat identification by capturing and analyzing suspicious endpoint behavior. Trellix EDR incorporates Trellix Wise™ to offer guidance during investigations and provide sophisticated threat landscape analytics. The ML modules in EDR include:
  - **Behavior-Based Detection:** This module uses powerful cloud-based analytics, including multiple analytic engines and ML, to inspect endpoint activity and uncover suspicious and stealthy behavior.
  - **MITRE Mapping:** The results from this behavior-based detection are directly mapped to the MITRE ATT&CK framework. This mapping helps analysts understand the phase of a threat, determine its associated risk, and prioritize the appropriate response actions.
  - **AI-Guided Investigations:** This module uses AI/ML-guided investigations that provide analysts with machine-generated insights into attacks. This dynamic investigation process automatically asks and answers questions – and gathers, summarizes, and visualizes evidence to understand the root cause.
- **Trellix Endpoint Forensics (HX):** Trellix HX provides multi-layered endpoint protection spanning on-prem, cloud, and disconnected environments in a single agent and managed from a single source. HX features Trellix's Malware Guard, an ML model utilized to classify malware and benign objects. Malware Guard

automates the detection of malicious files using LightGBM Algorithm.

- **Trellix Email Security - Cloud:** Trellix Email Security - Cloud utilizes FAUDE to automate URL-based threat detection, identify phishing, and filter spam. Further, it automatically flags malicious emails, helping to ensure that email administrators are promptly notified when a suspicious email enters the email stream.

2. **Behavioral Analytics:** Trellix uses standard user and system behavior to detect deviations that may signify insider threats or sophisticated cyberattacks. Behavioral analytics ML is integrated into Trellix Network Detection and Response (NDR), which is designed for continuous monitoring and real-time detection, investigation, and containment of emerging threats across the customer's network infrastructure. Trellix NDR automates anomaly detection by utilizing unsupervised learning and clustering methodologies through its integrated Trellix Network Security solution.

   Trellix NDR integrates SmartVision advanced correlation and analytics to detect suspicious lateral (east-west) network traffic. SmartVision includes over 180 rules specifically for lateral movement, provides full kill-chain detection, targeting server-facing deployments. SmartVision also incorporates an ML framework for:

   - Data-exfiltration detection;
   - JA3 detection for identifying encrypted communication;
   - Web shell detection (providing visibility into web server attacks); and
   - Detection of malware lateral movement.

3. **Vulnerability Management:** Trellix uses predictive analytics and real-world threat intelligence to prioritize vulnerabilities to address the most sophisticated attacks across multiple vectors. Trellix's predictive intelligence ML modules are integrated within Trellix Insights – which is our cross-product telemetry analytic engine utilizing local and global threat intelligence to offer comprehensive visibility into emerging threats and risks. Trellix Insights delivers actionable analytical observations, enabling proactive defense against global threats, significantly reducing detection and resolution time. Key components that use ML include:
   - **Processing Engine**: Monitors streamed telemetry across products via batch processing for global metrics intelligence.
   - **Retrospective Engine**: Identifies retrospective events for customers, alerts them to campaign events, and modifies backend data.
   - **Dynamic Application Containment:** Runs suspicious or unknown files with restrictions in a sandbox-like environment to limit their actions and prevent them from causing damage.
   - **MITRE Explorer**: Analyzes correlations between campaigns, threat actors, and relevant MITRE tools/techniques.
   - **IVX Integration**: Enables file scanning for malware via Trellix IVX - Cloud

(described in detail below).
        - **Suspicious Correlations**: Highlights artifacts not explicitly clean or malicious but potentially linked to known threats/campaigns.

4. **Automated Incident Response:** Trellix products orchestrate rapid containment and remediation actions in response to detected threats. Automated incident response ML modules are available in Trellix Managed Detection and Response (MDR). By combining AI and ML-led advanced threat intelligence, 24/7 monitoring, and expert response capabilities, Trellix MDR helps businesses detect and mitigate cyber threats with increased precision and speed. Trellix MDR delivers continuous threat monitoring, incident response, and remediation of threats and integrates with Trellix Endpoint Security solutions.

5. **Threat Intelligence Automation:** Trellix products process and correlate global threat data to enhance the capabilities of human analysts. The following Trellix products use Threat Intelligence Automation:
    - **Trellix Global Threat Intelligence (GTI)** is our global threat correlation engine and intelligence server. It uses communication behavior, volume, and network traffic patterns to predictively adjust reputations across all threat areas. GTI provides real-time, cloud-based threat intelligence services, including web reputation/categorization, network connection (IP) reputation, and file reputation.
    - **Trellix Threat Intelligence Exchange (TIE):** Shares file and threat information instantly across Trellix customer networks using the Trellix Data Exchange Layer (DXL), allowing local control over file reputation.
    - **Trellix Insights:** Trellix's cross-product telemetry analytic engine utilizing local and global threat intelligence to offer comprehensive visibility into threats and risks (described in detail above).
    - **Trellix Intelligent Virtual Execution - Cloud (IVX - Cloud):** An API-driven threat detection cloud service that scans content on demand to identify resident malware by rendering a verdict and providing supporting details.
    - **Advanced Research Center (ARC):** The Trellix ARC brings together an elite team of security professionals and researchers to produce insightful and actionable real-time intelligence to advance customer outcomes and the industry at large. The Trellix ARC utilizes local and global threat data to identify AI/ML technologies and/or techniques to bolster threat detection across Trellix's suite of products.

6. **Management and Orchestration:** Trellix incorporates AI/ML modules in the management and orchestration functionality across our suite of products, including:
    - **Trellix ePolicy Orchestrator - SaaS (ePO - SaaS)** is a cloud-based management service that reduces incident response times and simplifies risk management. It integrates ML models like Adaptive Threat Protection for

real-time file/process reputation analysis, and provides access to Trellix Insights.

- **Trellix Insights** uses local and global threat intelligence, including ML modules that act upon such intelligence to identify and alert customers on malware events.
- **Trellix Wise™** uses generative AI to accelerate investigation and root cause analysis. Features of Trellix Wise™ include:
  - **Automated Investigation:** Trellix Wise™ automatically asks and answers questions, explores multiple hypotheses, and gathers, summarizes, and visualizes evidence from multiple sources.
  - **"Ask Wise" Feature:** Security Analysts can use the 'Ask Wise' button to conduct a detailed analysis of any threat and receive recommended next steps.
  - **"Search" Feature:** Trellix Wise™ builds a search query using the "Search" feature, leveraging AI/ML capabilities (such as deep learning for network anomaly detection) to enhance threat detection, with natural language questions in up to 50 different languages.
- **Trellix Hyperautomation:** Trellix Hyperautomation enables SecOps teams to automate no-code capabilities, integrations, and a drag and drop workflow builder to enable analysts of any level to automate critical processes. Customers may integrate nearly any tool that has an API with Trellix Hyperautomation to accelerate responses to threats like phishing emails, credential theft, lateral movement attacks and more. In addition, Trellix Hyperautomation incorporates feeds from third party sources, including the Cybersecurity and Infrastructure Security Agency (CISA), to trigger automated detection of zero-day vulnerabilities or other critical scenarios in Trellix customers' network environments.

## Regulatory Compliance

Regulatory requirements related to AI/ML continue to evolve worldwide. Trellix complies with applicable laws in every jurisdiction where we do business, and is committed to supporting our customers' compliance journeys as well.

In general, Trellix does not engage in activities that are deemed "high risk" from an AI/ML perspective. For example, Trellix AI/ML engines are not used in any way that could fundamentally violate human rights or manipulate human behavior, allow social scoring, or make automated decisions that could result in discrimination. Rather, our Trellix AI/ML is used solely by our customers' security analysts to detect, manage and respond to potential cybersecurity threats, and for no other purposes. Automated decision rules made by our AI/ML solutions are configured by our customer security analysts who understand and fully expect that such decisions are implemented via AI/ML. Further, the results of automated decisions typically result only in the quarantine of suspicious data and/or turning off access or administrative privileges to prevent proliferation of potentially malicious activity – which in no way affect fundamental human

rights. In all events, our customers' human security analysts can transparently review and understand such automated decisions, and make corrections where appropriate. (For more information about cybersecurity in the context of high risk systems, see Recital 15 of the EU AI Act.)

But even though the use of AI/ML in Trellix solutions is itself not high risk, we recognize that many Trellix customers do directly engage in high risk activities, including the management or support of critical infrastructure, and the processing of highly sensitive data. We also recognize that Trellix solutions are integral to our customers' overall security posture. As a leader in responsible AI, Trellix has implemented various controls and processes that align with the requirements for high-risk AI systems. This commitment is designed to support our customers in meeting their security and compliance requirements, as further detailed below.

**Risk Management System.** Trellix has established a risk management system that spans the lifecycle of our AI/ML systems, from initial design through deployment and post-market monitoring. This system includes (1) proactive identification of risks to safety and fundamental rights, (2) assessment of the severity and probability of identified risks, and (3) technical and organizational measures to reduce risks to an acceptable level.

**Data Governance and Management.** Trellix has appropriate data governance and management practices that cover how we train, validate and test our AI/ML models.

- **Data Quality:** We implement comprehensive data quality checks to help ensure the accuracy, completeness, and consistency of datasets.
- **Data Relevance and Representativeness:** We meticulously curate datasets to be relevant to the intended purpose of the AI system and sufficiently representative to mitigate bias.
- **Data Collection Practices:** We collect data lawfully and ethically, adhering to privacy and data protection regulations worldwide. We use anonymization and pseudonymization techniques where appropriate.
- **Use Limitation: We do NOT use customer data (e.g., systems information, IP addresses, email addresses, host names, file paths, log information, etc.) to train our AI models.** Our AI models are computation models that do NOT self-mutate or train based on customer data. Though we use customer detections and events for inference to provide security-related insights, such data does not update our AI models. All such insights are based on information about **attacks** – NOT on information about our **customers**.
- **Data Labeling:** We establish clear protocols for data labeling to help ensure consistency and accuracy.

**Privacy and Data Protection**

- **Compliance**: Our AI/ML data handling practices comply with all applicable privacy and data protection requirements worldwide, including the EU General Data Protection Regulation (GDPR) and other EU and Member State legislation. To the extent our products and services process personal data (e.g., usernames and file paths) in the course of detecting potentially malicious events, we adhere to strict privacy and security policies and procedures to protect all such information. We collect and process only the data necessary for the intended purpose of our AI systems, and have implemented robust technical and organizational security measures to protect all data used in our AI/ML processes from unauthorized access and manipulation. For data residency compliance, all AI/ML inference requests originating in the EU are processed using regional cloud services, ensuring the data remains within the European Union.
- **Customer Control:** Customers using Trellix solutions maintain significant control over their data usage. This includes, where appropriate, options to customize data flows for specific environments and the ability to choose when to enable or disable AI/ML functionality. For detailed information about customization and opt-in/out choice for AI inference and data usage, see our relevant product documentation, as features vary by product. Or, contact your Trellix support team directly with any questions.

For more information on how our various products and services collect, store, use, secure and delete customer data, see our Privacy Data Sheets on the Trellix Trust Center.

### Technical Documentation

We maintain comprehensive technical documentation for all Trellix AI/ML systems, including:

- A description of the AI system's general characteristics, capabilities, and limitations;
- Information pertaining to the data utilized for training, validation, and testing;
- A description of the design specifications, development process, and algorithms employed;
- Information regarding the risk management system and its implementation;
- Results of conformity assessments and testing; and
- Instructions for use, including information on human oversight.

This confidential and proprietary documentation is regularly updated, and may be made available to relevant regulatory authorities upon request.

### Record-Keeping

Trellix AI/ML systems are engineered to facilitate our customers' automatic logging of

events to help ensure the traceability of operations. Such logs may enable:

- Recording of decisions rendered by the AI system;
- Monitoring of key performance indicators; and
- Tracking of human interventions and overrides.

Customers may maintain logs, with Trellix support as needed, for durations meeting regulatory and audit requirements.

### Transparency and Provision of Information to Customers

Trellix is committed to providing our customers with clear and comprehensible information regarding our AI/ML systems. In general, the intended purpose of our AI/ML systems is to help our customers detect, manage and respond to potential cybersecurity threats. As with any cybersecurity solution, there are inherent limitations to the ability of our AI/ML systems to detect and respond to all possible threats, as malicious actors are constantly evolving their attack vectors.

For more information about our AI system characteristics and limitations, computational and hardware resources needed for our AI systems, and mechanisms that allow customers to properly collect, store and interpret logs, see our Trellix Trust Center.

### Human Oversight

It is critical that our customers' human security analysts can understand, interpret, and intervene (when necessary) in the automated decisions and actions of our Trellix products and services.

As discussed above, automated decision rules made by our AI/ML solutions are configured by our customer administrators and security analysts who understand and expect that such decisions are implemented via AI/ML. In addition, human security analysts (Trellix or customer) can transparently review, understand, and correct these automated decisions and actions.

For instance, Trellix solutions enable human oversight through Human-in-the-Loop (HITL) mechanisms. We provide customer security analysts with intuitive control panels, dashboards and alerts that enable them to review, validate and override AI-generated recommendations or automated actions.

In addition, Trellix AI/ML solutions use Retrieval Augmented Generation (RAG) for step-by-step processing, citing sources for human security analysts to ensure accuracy, reliability, and to prevent AI "hallucinations."

Customers always have access to standard documentation and training focused on effectively interacting with and overseeing AI/ML features across all Trellix offerings.

### Accuracy, Robustness, and Cybersecurity

Trellix has implemented appropriate measures to ensure the high performance and security of our AI/ML systems.

- **Accuracy:** We continuously maintain and enhance accuracy in threat detection and response through rigorous testing and validation using diverse datasets.
- **Robustness:** We have designed our AI models to exhibit resilience to errors, inconsistencies, and adversarial attacks, including comprehensive testing for data poisoning, model evasion, and other adversarial techniques.
- **Cybersecurity:** We have implemented robust cybersecurity measures to safeguard our AI systems (as well as related training data, models, and underlying infrastructure) from unauthorized access, manipulation, or attacks.

Importantly, our AI/ML products have customer-visible telemetry that provide details related to the operations of the models themselves. For example, our ML-based anomaly detection provides a variety of metrics around the scoring and confidence, which help users understand how anomalous a given security event is. Similarly, our LLM-based detection products provide a confidence score; and the LLM provides detailed reasoning on exactly why it is sure or unsure of its decision.

### Quality Management System

Trellix integrates the development and deployment of its AI/ML systems within its existing comprehensive Quality Management System (QMS), which includes:

- Organizational structure and defined responsibilities;
- Established processes for design, development, testing, and deployment;
- Rigorous risk management procedures;
- Comprehensive documentation and record-keeping protocols; and
- Processes for continuous improvement.

### Ongoing Monitoring

To help identify any emerging risks and to ensure continued compliance, Trellix engages in ongoing monitoring of the performance and behavior of our AI/ML systems. This includes:

- Performance monitoring (e.g., analysis of false positive/negative rates),
- Incident reporting and analysis,
- Collection of real-world performance data, and
- Regular review and updates of AI models based on new data and evolving threat landscapes.

## Supporting the AI Journey of Our Customers

Among the most concerning developments for security teams is the rise of agentic cyberattacks that use AI to operate independently, make decisions, and adapt in real time with unprecedented speed and scale. Customers need next-generation, AI-driven cybersecurity solutions to combat these threats within a complex operational and regulatory environment. At Trellix, we are committed to supporting our customers as they navigate the complex requirements of an AI landscape that is evolving faster than ever. Guided by our foundational Trellix AI principles, we are dedicated to protecting fundamental human rights – and building a safer, more secure, and more trustworthy digital future powered by responsible AI.

## Contact Information

For more information about Trellix's AI/ML practices or our controls and measures to comply with applicable laws, please contact your designated Trellix representative. Or reach out to our privacy and AI compliance team at privacy@trellix.com.

### About This Data Sheet

Please note that the information provided with this document concerning technical or professional subject matter is for general awareness only, may be subject to change, and does not constitute legal or professional advice, warranty of fitness for a particular purpose, or compliance with applicable laws. This Data Sheet is reviewed and updated on an annual, or as needed, basis.