AI SAFETY & MODEL GOVERNANCE

Version 1.1

Last updated: May 28, 2026

1. Purpose

This Statement summarizes Blueshift's approach to AI safety, governance, and responsible development for the machine learning, agentic automation, and autonomous action capabilities across the Covered AI Systems, including the IntelliThreat AI Platform and the IntelliThreat Autonomous tier. Contractual commitments regarding AI and data use are set forth in the Platform MSA, the INTELLITHREAT AI PLATFORM ADDENDUM, the INTELLITHREAT AUTONOMOUS ADDENDUM (where applicable), and the DATA PROCESSING ADDENDUM — INTELLITHREAT AI PLATFORM. This Statement is informational only and does not create independent contractual obligations.

2. Governance Framework

Blueshift maintains an internal AI Governance Program that may include, as applicable:

- Executive oversight and crossfunctional review
- Model lifecycle management
- Risk assessments for new AI features
- Documentation, as appropriate, of training data sources and evaluation methods

3. Safety Principles

Blueshift's AI systems are designed to support the following principles:

- **Security First:** AI features prioritize security outcomes and minimize operational risk.
- **Human Oversight:** Customers can configure automation levels and approval workflows based on available Platform functionality. For the IntelliThreat Autonomous tier, Blueshift implements a two-tier consent model: (1) Advisory Mode (AI recommendations, no automated execution — default); and (2) Autonomous Mode (full autonomous execution, enterprise-level explicit written consent required). Customers may activate a Kill-Switch to suspend Autonomous Mode; Blueshift targets cessation of future or pending Autonomous Mode actions within 60 seconds of activation. All Autonomous Actions are logged in tamper-evident audit logs retained for a minimum of 12 months.
- **Transparency:** AIgenerated insights are intended to be explainable where operationally and technically feasible.
- **Privacy:** Blueshift does not use identifiable Customer Data to train public models and does not provide such data to third parties for training their public models.

4. Model Training & Data Use

- Blueshift may use **deidentified, anonymized, or aggregated** telemetry, as permitted by applicable agreements and law, to improve detection logic, service performance, and related models.
- Blueshift does **not** sell Customer Data and does not use identifiable Customer Data to train public models, except as otherwise expressly permitted by applicable agreements.
- Model updates are subject to internal validation and safety review, as appropriate to the nature of the update.

5. Autonomous Actions

The IntelliThreat Autonomous tier and applicable autonomous features of the IntelliThreat AI Platform may execute real-time actions within Customer-authorized environments only upon explicit Customer enablement. Blueshift describes below certain controls applicable to autonomous execution within the Covered AI Systems.

- Two-Tier Consent Model: Autonomous execution is gated by two escalating consent tiers — Advisory (default, no execution), and Autonomous (full autonomous execution, enterprise explicit written consent). Customers may not access Autonomous Mode without completing the required consent and configuration steps.
- Kill-Switch: Customers may activate a Kill-Switch at any time to suspend Autonomous Mode actions. Blueshift targets cessation of future or pending Autonomous Mode actions within 60 seconds of activation. Reactivation requires deliberate Customer action.
- Audit Logging: All Autonomous Actions are recorded in tamper-evident audit logs retained for a minimum of 12 months and made available to Customer via the Platform interface and API export, with full exports available upon request.
- Incident Notification: Blueshift provides notification to Customer for Autonomous Action incidents within 2 hours for Severity-1, 4 hours for Severity-2, and 24 hours for Severity-3, in each case subject to the INTELLITHREAT AUTONOMOUS ADDENDUM.
- Customer Responsibilities: Each Customer is solely responsible for validating its automation configuration, maintaining human oversight where available and desired, monitoring operational impact, and determining that Autonomous Mode is appropriate for its environment and risk tolerance.
- Liability: Liability for Autonomous Actions is governed by the Platform MSA and, where applicable, the INTELLITHREAT AUTONOMOUS ADDENDUM.

6. AI Regulatory Alignment & Risk Management Framework

Blueshift aligns its AI governance and risk management practices with applicable regulatory frameworks and industry standards. Key alignments include:

- NIST AI Risk Management Framework (AI RMF 1.0): Blueshift’s AI Governance Program is designed with reference to the GOVERN, MAP, MEASURE, and MANAGE

functions of the NIST AI RMF. An AI RISK MANAGEMENT FRAMEWORK ALIGNMENT STATEMENT is made available to enterprise customers upon request.

- EU AI Act: Blueshift monitors EU AI Act implementation as it applies to the Covered AI Systems. To the extent the EU AI Act applies to the IntelliThreat Autonomous tier or other Covered AI Systems based on the manner of deployment, use case, functionality enabled, and applicable guidance or legal interpretation in the relevant jurisdiction, Blueshift expects to implement applicable provider-side conformity, transparency, logging, and human oversight measures and to update this Statement and applicable addenda accordingly.
- ISO/IEC 42001 (AI Management Systems): Blueshift is evaluating alignment with ISO/IEC 42001 as part of its ongoing AI governance maturity roadmap.
- Responsible Disclosure: Security researchers who identify AI-specific vulnerabilities or safety issues may report them via Blueshift's Responsible Disclosure & Vulnerability Reporting Policy.

This Statement is informational only. It describes Blueshift's practices as of the effective date and does not constitute a warranty, guarantee, or contractual representation. Blueshift's actual contractual obligations regarding AI capabilities, data use, autonomous actions, liability, and incident response are exclusively governed by the PLATFORM MASTER SUBSCRIPTION AGREEMENT, the INTELLITHREAT AI PLATFORM ADDENDUM, the INTELLITHREAT AUTONOMOUS ADDENDUM (where applicable), and the DATA PROCESSING ADDENDUM — INTELLITHREAT AI PLATFORM. In the event of any conflict between this Statement and those agreements, the applicable agreement controls.