carahsoft

crest(data)

Ɩ ƛƎKπvdzI·fƛűè [I·ōŜfŜŘ 5I·úI·ăŜűă ŦƧℕJ ! L9yƔI·ōfŜŘ ¢KℕJŜI·ű 5ŜűŜŎűƛƨƴ ¢KℕJƨdzƎK 9ÈLJŜℕJűπ[ŜŘ {ŜŎdzℕJƛűè ! yƴƨűI·úƛƨƴ

!L ƻ a[ï 5I·úI· [I·ōŜfƛyƎ ŦƧℕ {ŜŎdzℕJƛűè

Thank you for your interest in exploring this content.

Carahsoft is the **Trusted Government IT Solutions Provider**® supporting a broad portfolio of industry-leading technologies and a wide range of contract vehicles.

As the **Master Government Aggregator**®, Carahsoft connects government agencies, industry partners, and technology providers to deliver innovative, mission-focused solutions.

In partnership with Crest Data, we provide technology solutions that ŘŜfƛØŜℕ ŜÈLJŜℕűπfŜŘ ŘI·úI· fI·ōŜfƛyƎ I·yŘ I·yƴƨúI·úƛƨƴăŜℕØƛŎŜă I·ú ăŎI·fŜΣ Ŝƴ'I·ōfƛyƎ ƧℕƎI·yƛÎI·úƛƨyã ú2 ōdzƛfŘ KƛƎKπIjdzI·fƛűè űℕI·λyƛyƎ ŘI·úI·ăŜűă úKI·ú ƛƳLJℕƧØŜƴ ! LπŘℕƛØŜƴ'úKℕŜI·ú ŘŜúŜŎúƛƨƴ I·ŎŎdzℕI·Ŏè I·yŘ I·ŎŎŜfŜℕI·ûŜ úKŜ ŘλăŎƧØŜℕè ƧŦ ŎƧƳLJfŜÈ I·úúI·Ŏ'ę LJI·úúŜℕyã®

To learn more about how Carahsoft can support your technology needs, please visit carahsoft.com

Explore More Resources:
carah.io/CrestDataResources

Join Events & Webinars:
carah.io/CrestDataEvents

Discover Technology Solutions:
carah.io/CrestData

Learn About Procurement:
carah.io/CrestDataContracts

Connect With Our Team:
crestdata@carahsoft.com
703-581-6680

crest(data)

CASE STUDY

AI & ML · DATA LABELING FOR SECURITY

High-Quality Labeled Datasets for AI-Enabled Threat Detection Through Expert-Led Security Annotation

500K+
High-quality labeled security datasets

Millions/day
Security events processed & labeled

Petabyte-scale
Raw security logs handled

Fortune 500
Enterprises & agencies protected

Executive Summary

THE OPPORTUNITY

A global cybersecurity company struggled to build its next-generation **AI-based threat detection system** because of the sheer volume of raw security data to be processed. With petabyte-scale logs and subtle signs of compromise hidden inside normal system activity, the client grappled with data quality problems and an acute need for **accurately labeled datasets** to train its machine learning models. Existing labeling services lacked the security domain expertise to identify threats, and the client's own security team didn't have the capacity to build these training sets at scale.

Crest Data tackled this with a holistic data labeling strategy for high-consequence security use cases — assembling a **labeling team with deep cybersecurity expertise** and a scalable infrastructure that processes millions of events per day. The result: more than **500,000 high-quality labeled datasets** used to train models across a range of detection use cases. This expert-driven process improved threat detection rates, reduced false positives to eliminate alert fatigue, and enabled the client to discover previously unknown, complex attack patterns.

About the Customer

A **leading global cybersecurity platform provider** serving Fortune 500 companies and government agencies, looking to enhance their threat detection capabilities through machine learning.

Their security products **protect millions of endpoints worldwide** and rely on increasingly sophisticated detection algorithms to identify emerging threats.

CUSTOMER PROFILE

INDUSTRY
Cybersecurity — Security Platform Provider

CUSTOMERS
Fortune 500 companies & government agencies

FOOTPRINT
Protects millions of endpoints worldwide

CRITICAL NEED
ML-ready training data for next-gen detection

The Customer Challenge

SCALE, EXPERTISE & DATA QUALITY

As a world-leading cybersecurity company, the client faced a number of **operational and technical challenges** in developing its next-generation AI-based threat detection system.

Large-scale, complex data. Petabytes of raw security logs and threat data had to be ingested — and subtle signs of compromise were rare and easily hidden in legitimate activity.

Lack of specialized expertise. Conventional data labeling services lacked the expert security knowledge to identify threats, making it difficult to build ML datasets.

Lack of time and resources. The client's security teams didn't have the capacity to manually label training data at the scale their global operations demanded.

Data quality issues. Security datasets often had inconsistent structure and incomplete data, making it hard to train suitable models.

Critical training needs. The client needed highly accurate datasets to train models for detecting complex attacks and preventing false alarms — but had no efficient method of generating them.

> The client needed a partner who could combine **genuine security domain expertise with industrial-scale labeling** — turning massive, messy log data into trustworthy training sets for next-generation detection models.

The Customer Solution

EXPERT-LED, AT SCALE

Crest Data delivered an **end-to-end data labeling solution** for critical security use cases — pairing expert analysis with the unprecedented scale required to train the client's next-generation AI threat detection systems.

Expert Security Annotation

A team of **cybersecurity experts** refined the annotation process for security-specific data, applying precise labels to intricate threat and anomaly patterns under robust quality control.

Industry-Leading Scale

A scalable platform processed and labeled **millions of security events per day**, producing over **500,000 quality-labeled datasets** for targeted ML training.

Domain-Specific Training & QA

Tailored curricula for attack-pattern recognition and alert prioritization, plus a **multi-level verification process** and dedicated QA teams with deep security capability.

The solution focused on four advanced technical areas and a structured, phased implementation — detailed overleaf.

Advanced Technical Focus

FOUR CORE DETECTION AREAS

The labeling effort concentrated on the security problems that matter most for high-fidelity detection — each demanding nuanced, context-aware judgement that only domain experts can provide.

Threat Classification

Accurate classification of events by threat type and severity, including the stages of attack campaigns.

Anomaly Identification

Contextual classification of statistical anomalies to discriminate normal system behaviour from genuine security threats.

Attack Pattern Recognition

Detection of multi-stage attack campaigns by inter-relating disparate, seemingly unconnected security events.

Alert Prioritization

Risk classification and business-impact analysis to rank potential threats by what matters to the business.

Structured Implementation Approach

PHASED & VALIDATED

1

Security Assessment & Workflow Design

Began with a thorough security assessment and design of the labeling workflow.

2

Pilot, Testing & Validation

A focused pilot project with rigorous testing and validation of approach and quality.

3

Large-Scale Production

Scaled into full production labeling across millions of daily security events.

4

Continuous Improvement

Ongoing refinement driven by structured feedback loops and QA insights.

500K+
labeled datasets

EXPERTISE MEETS SCALE

Half a million expert-labeled datasets, built for security

Millions of events processed every day and labeled by cybersecurity specialists — multi-level QA ensured the consistency and accuracy that next-generation detection models depend on.

Outcomes

BUSINESS IMPACT

Crest Data's custom data labeling solution had a profound impact on the client's threat detection system, translating directly into key business outcomes.

Higher Model Accuracy

The client achieved a **higher level of accuracy** in detecting threats across all of their security solutions.

Lowered Alert Fatigue

A significant reduction in **false positives** helped alleviate fatigue across the security team.

Enhanced Detection Scope

Detection expanded to cover **complex, previously unknown attack patterns**, improving overall security coverage.

Sustainable Competitive Advantage

Superior detection of **zero-day and advanced malware** threats let the client maintain a competitive edge and better serve customers.

About Crest Data

AI-FIRST · SECURITY-LED

Crest Data is a data and AI-driven technology solutions provider for enterprises and technology innovators across **cybersecurity and cloud security**, helping them move faster and more securely. We offer specialized services in **AI & ML and MLOps** to enhance next-generation threat detection capabilities. Our specialization includes **Data Labeling & Annotation for Security**, efficiently bridging the gap between massive raw data volumes and robust machine learning through expert-led annotation for complex security logs.

TRAIN DETECTION MODELS ON DATA YOU CAN TRUST

Turn raw security logs into AI-ready, expert-labeled datasets.

Talk to Crest Data's AI & ML experts about security-specific data labeling, scalable annotation pipelines, and multi-level QA built for next-generation threat detection.

Talk to Our Experts →

www.crestdata.ai/contact

EXPLORE AI & ML

crest(data)

USA · CA
3031 Tisch Way #602
San Jose, CA 95128

INDIA · AHMEDABAD
CDS House, SG Road
Makarba, Ahmedabad 382210

INDIA · PUNE
Amar Madhuban Tech Park
Baner, Pune, MH 411045

+1 (408) 551-0051 info@crestdata.ai www.crestdata.ai