

EDUCATION GUIDE

Data in service of detection and response

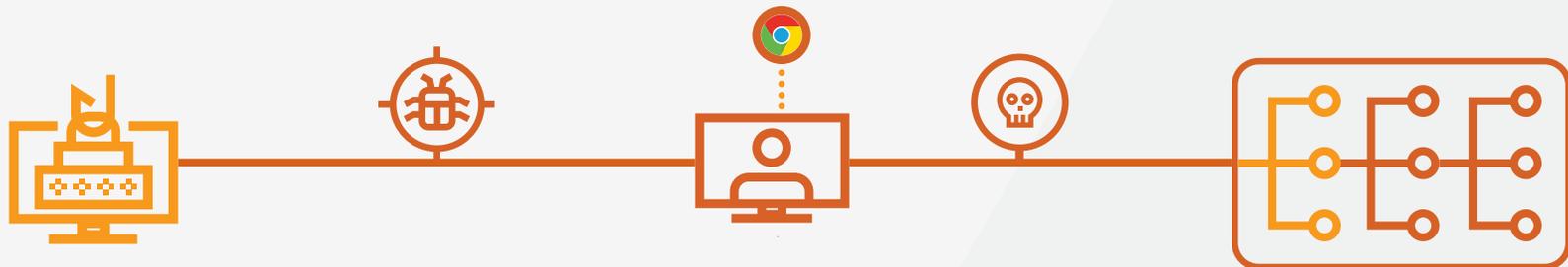
Why is the data so important?

Consider the following scenario.

Your team has learned about a banking trojan that uses a fake Google update to gain a command-and-control foothold on the target system.



After an initial compromise via spearphishing or drive-by download, the exploited host downloads the full payload in the guise of a Chrome update before establishing the command-and-control (C&C) channel and allowing further reconnaissance and lateral movement deeper into the network.



How do you discover and identify the potentially malicious communication? And if you find that communication is malicious, how do you respond?

In this case, the implant calls back periodically to the attacker's command-and-control infrastructure, which we would observe as beaconing behavior. Beaconing can be a weak indicator of potential malicious activity serving as the foundation for a command-and-control channel, or the call-back to fetch malware.

However, most commonly, beaconing is part and parcel of innocuous behaviors, such as your Smart TV or teleconferencing device reaching back to its home hub. Stock tickers and sports score updates are also notorious for beaconing.

In this threat hunting/investigation example, you want to be able to answer questions like:

1 Are there instances of beaconing observed in my network?

2 Which hosts are potentially infected, not just the IP address?

3 Is the payload size something I would normally see?

4 Is the traffic going to an unusual external destination?

5 Are beaconing sessions obfuscated within a single, long connection?

6 What external destinations are being beaconed to?

7 Does the beaconing cadence demonstrate unusual request/reponse frequency?

8 Does the beacon have a rare or unusual JA3 hash?

9 What is the privilege level of the hosts that are beaconing?

10 Does the connection use unusual services and protocols?

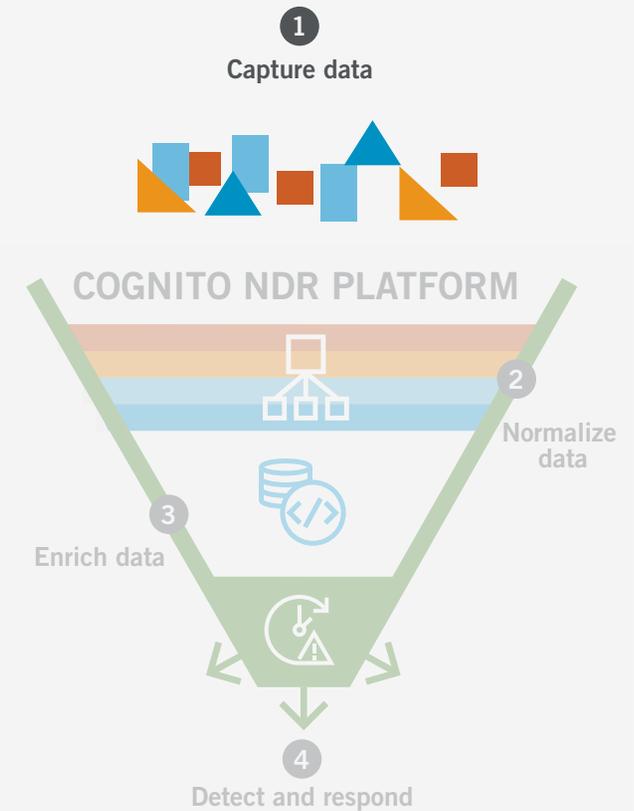
The first step is to make sure that the attributes necessary to answer these questions are readily available to the security analyst.

How to get the right data

1. Capture data

Sensors extract relevant metadata, logs and telemetry from all network traffic in the cloud, SaaS, data center, IoT, and enterprise. Our unique software architecture developed from Day 1, along with customized flow engines, enable data capture and processing with massive scale.

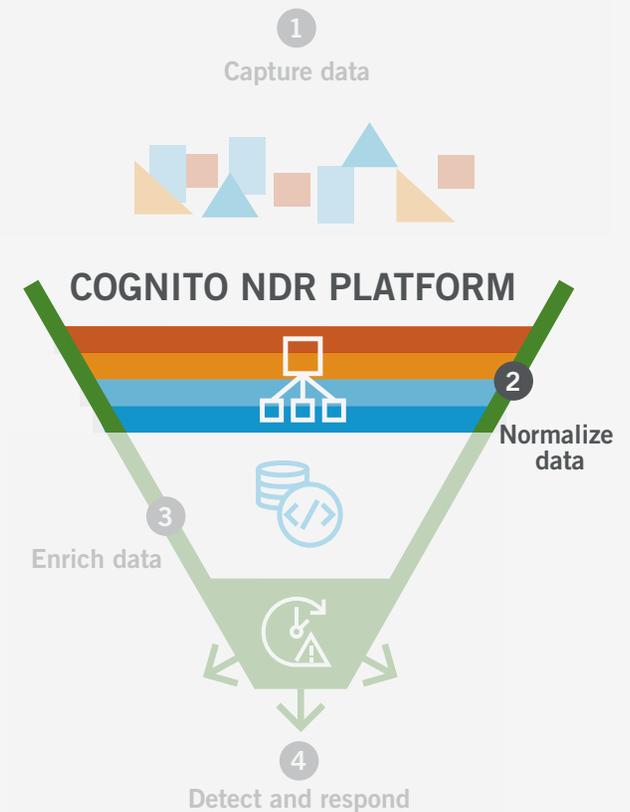
- NETWORK TRAFFIC
- THREAT INTELLIGENCE
- ACTIVE DIRECTORY LOGS
- OFFICE 365
- DHCP LOGS



2. Normalize data

Traffic flows are deduplicated and custom flow engines extract metadata to detect attacker behaviors. Characteristics of every flow are recorded, including ebb and flow, timing, direction, and packet size. Each flow is then attributed to a host instead of an IP address.

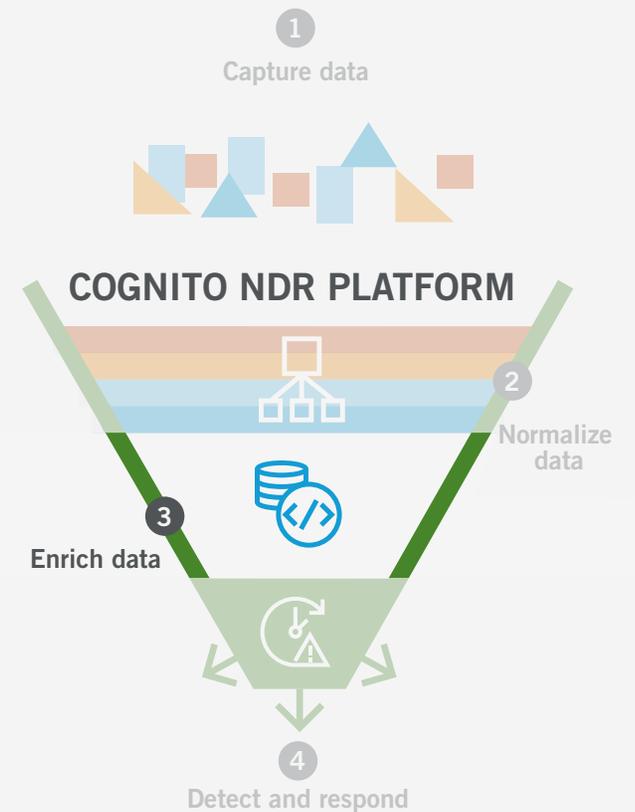
IP-TO-HOST NAME ASSOCIATION
TRAFFIC DIRECTIONALITY
DEDUPLICATION
HOST ID



3. Enrich data

Security researchers and data scientists build and continuously tune scores of self-learning behavioral models that enrich metadata with machine learning-derived security attributes, including attack behaviors, account scores, host scores, and correlated attack campaigns.

- SECURITY PATTERNS (E.G. BEACONS)
- NORMAL PATTERNS (LEARNING)
- PRECURSORS (WEAK SIGNALS)
- ATTACKER BEHAVIORS
- ACCOUNT SCORES
- SAVED SEARCH
- HOST SCORES
- CAMPAIGNS



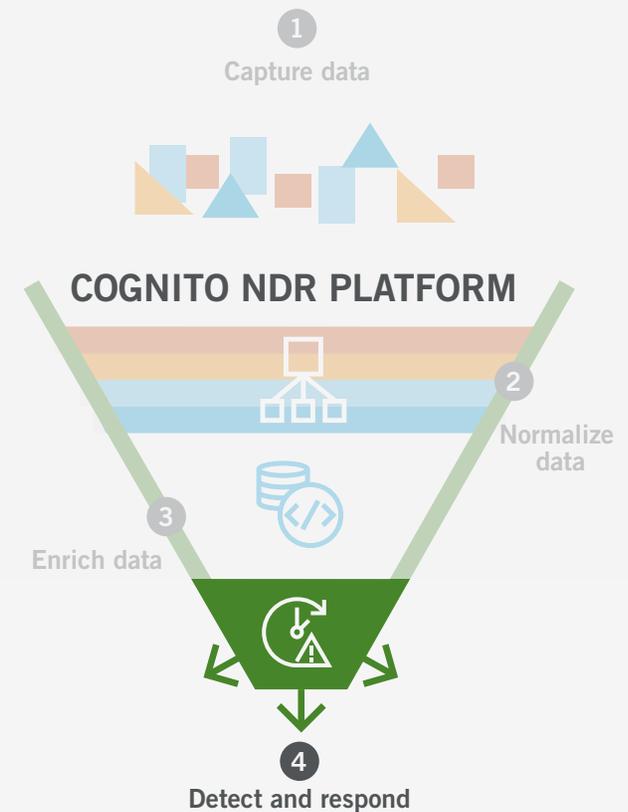
4. Detect and Respond

Detect

- Scores of always-learning attacker behavior algorithms detect threats automatically and in real time, well before a data breach occurs.
- Detected threats are automatically triaged, prioritized based on risk level, and correlated with compromised workloads and host devices.
- Automating Tier-1 tasks condenses months of work into minutes and enables security analysts to focus on threat hunting and investigations.

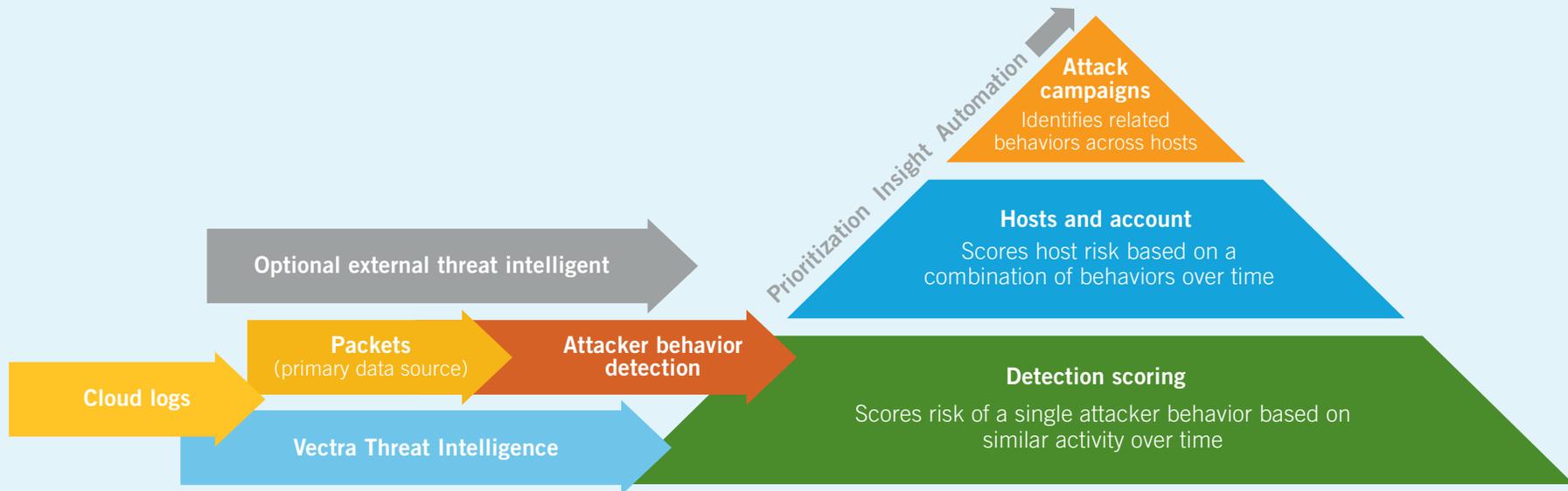
Respond

- Machine learning-derived attributes like host identity and beaconing provide vital context that reveals the broader scope of an attack.
- Custom-engineered investigative workbench is optimized for security-enriched metadata and enables sub-second searches at scale.
- Actionable data is always at your fingertips by augmenting threat detections with security insights and context that eliminate guesswork.



The power of having the right data to identify attack campaigns

In the domain of cybersecurity, there are myriad sources of data that may indicate the presence or absence of advanced persistent threats, the stage those threats are at, and the infected machines and accounts involved in the threat. These data sources range from network traffic – unencrypted and encrypted – to metadata from endpoints to behavioral profiles of users to logs from cloud services.



By uniquely combining the domain expertise of security researchers with the data know-how of data scientists, Vectra AI has been able to build nearly a decades worth of the right datasets, all of which can be used to inform design choices and train novel algorithms in manner that would have been otherwise nearly impossible.

Applications and constraints of AI and machine learning

The successful application of AI and machine learning requires the appropriate selection of relevant data, the determination of the methods by which to preprocess it, and an understanding of which models can answer which types of questions about that data.

In addition to being designed to achieve a specific goal, they are still subject to constraints:

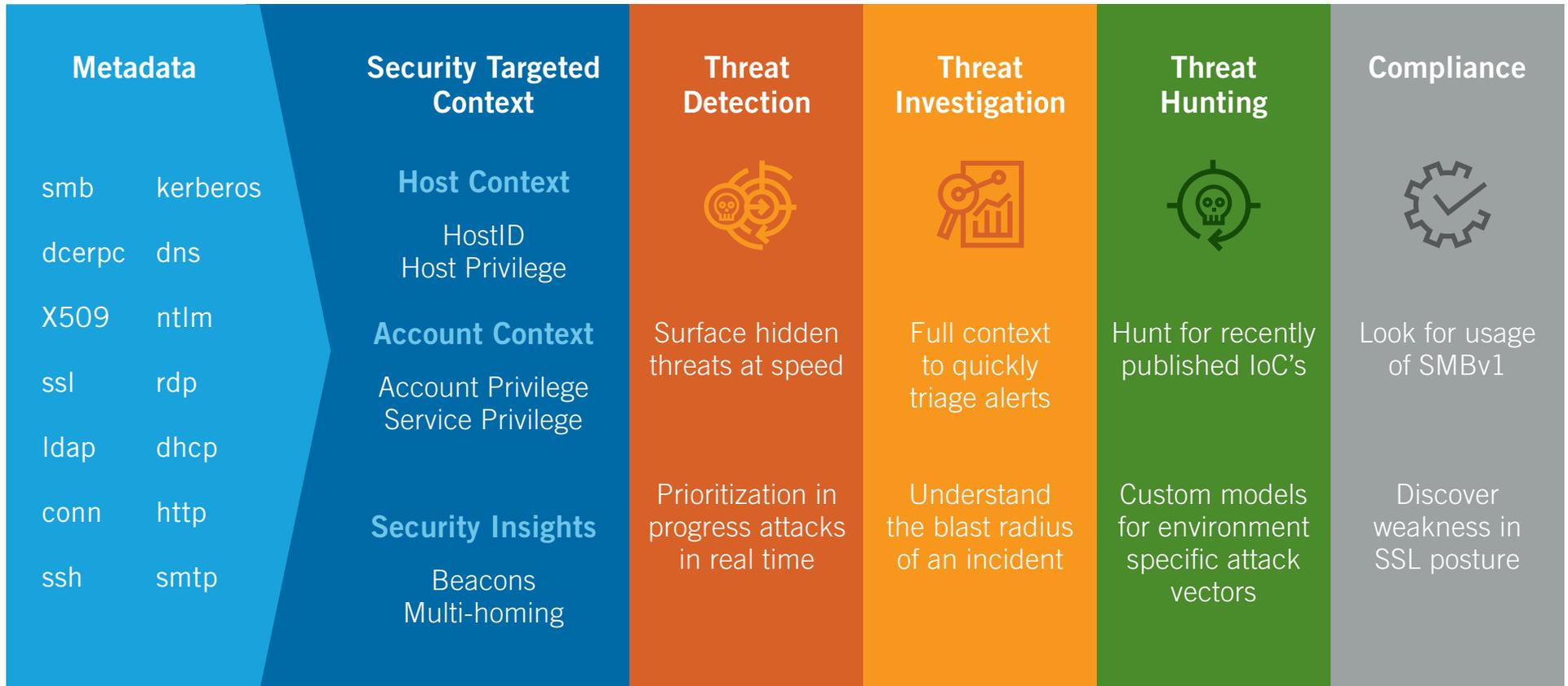
- 1 There are costs associated with data storage, so not all data can be kept indefinitely.
- 2 Labels for data in many cases may be impossible obtain, either due to lack of domain expertise (as is often the case in cybersecurity) or simply due to epistemic uncertainty (a hand-written digit may never be distinguished as being a 7 or a 1, with certainty).
- 3 Other types of data may be forever impenetrable and only have useful and informative properties (e.g., the temporal profile of bytes and packets sent between a client and server).



Driven by AI, the Cognito NDR platform automates manual and mundane security tasks.

[Read the case study](#)

A platform powered by security-optimized network data



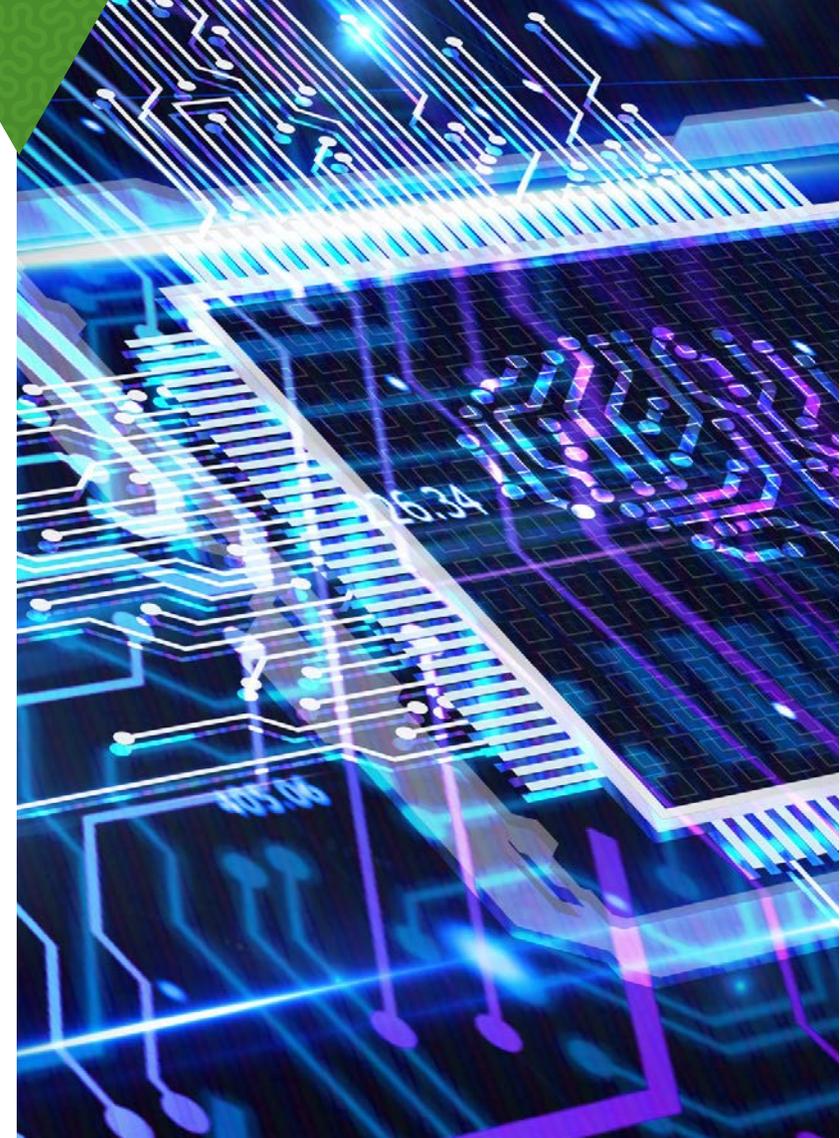
Why Vectra data is the gold standard

- 1 Vectra collects, stores and analyzes unlimited metadata at scale from all network traffic – in cloud, data center, IoT, and enterprise networks. This metadata is critical to the real-time detection and response to attacks.
- 2 Vectra metadata is enriched with security insights about attack behaviors – compromised workloads, host ID, location, accounts and privileges. Threat behaviors in a larger attack campaign appear as a single chain of related events.
- 3 These actions are automatic and require zero human intervention. With the ability to stop attacks fast, security teams have more time to hunt for threats and investigate incidents with unprecedented efficiency and precision.
- 4 Vectra is built by security researchers who deconstruct the latest attack techniques and data scientists who create self-learning machine learning algorithms that identify malicious behaviors.

For more information please contact a service representative at info@vectra.ai.

Email info@vectra.ai | [vectra.ai](https://www.vectra.ai)

© 2020 Vectra AI, Inc. All rights reserved. Vectra, the Vectra AI logo, Cognito and Security that thinks are registered trademarks and Cognito Detect, Cognito Recall, Cognito Stream, the Vectra Threat Labs and the Threat Certainty Index are trademarks of Vectra AI. Other brand, product and service names are trademarks, registered trademarks or service marks of their respective holders. Version 120220



Learn about the data science behind Vectra's AI threat detection models in this white paper

[Read the white paper](#)