VECTRA®

Building security that thinks

Machine learning fundamentals for cybersecurity professionals

Chris Morales, Head of security analytics

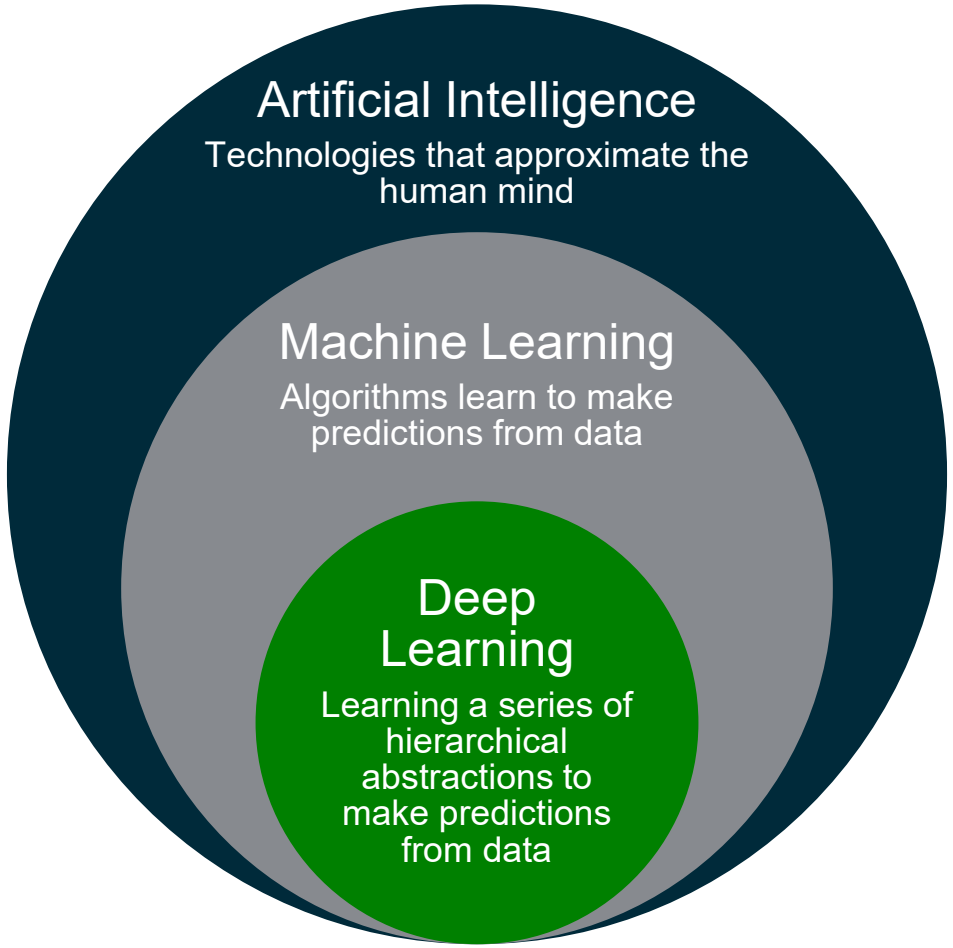
chris@Vectra.ai

The five questions that data science answers

- Is this A or B (or C or D)?
  - Classification
- How much / How many?
  - Regression
- How is this data organized?
  - Clustering
- Is this weird?
  - Anomaly
- What action should be taken?
  - Reinforcement

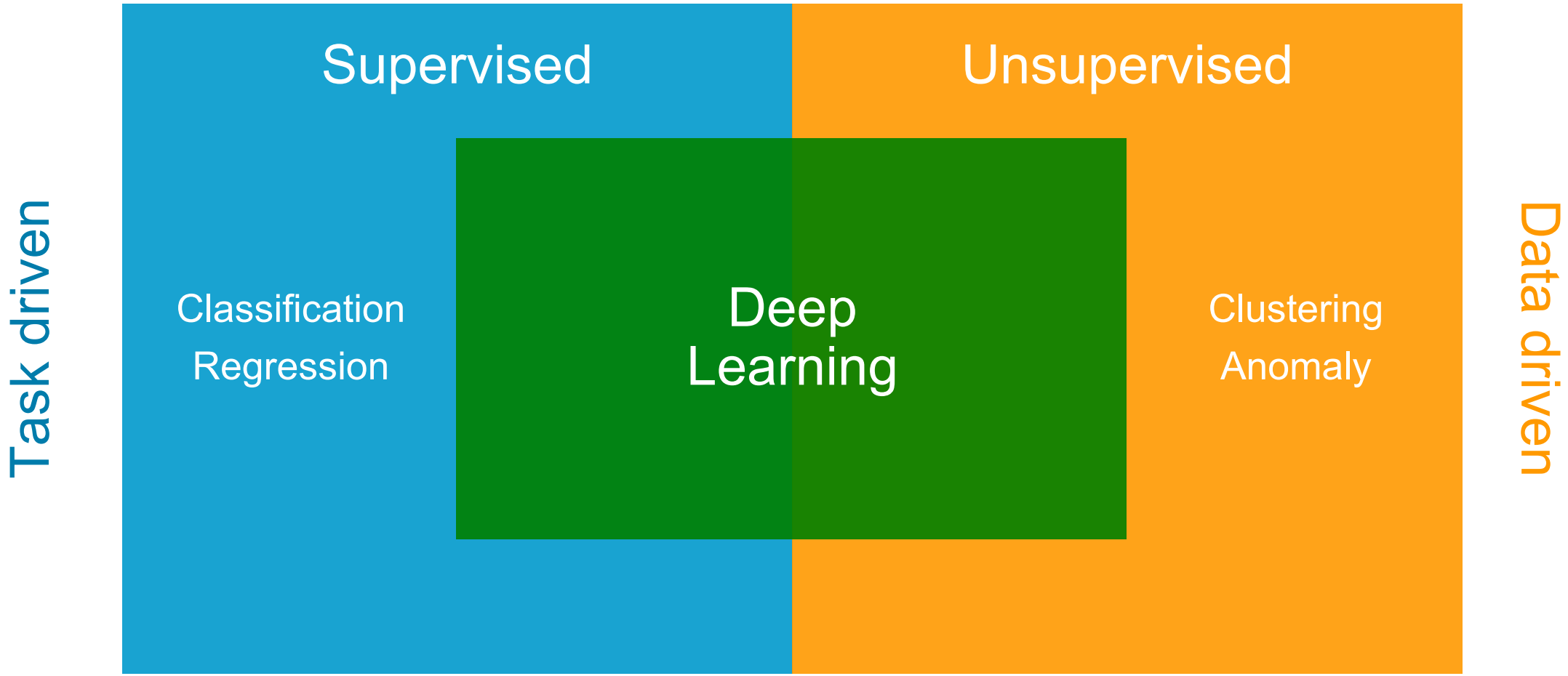
VECTRA

How does data science create intelligent machines?



VECTRA

Supervised and unsupervised machine learning

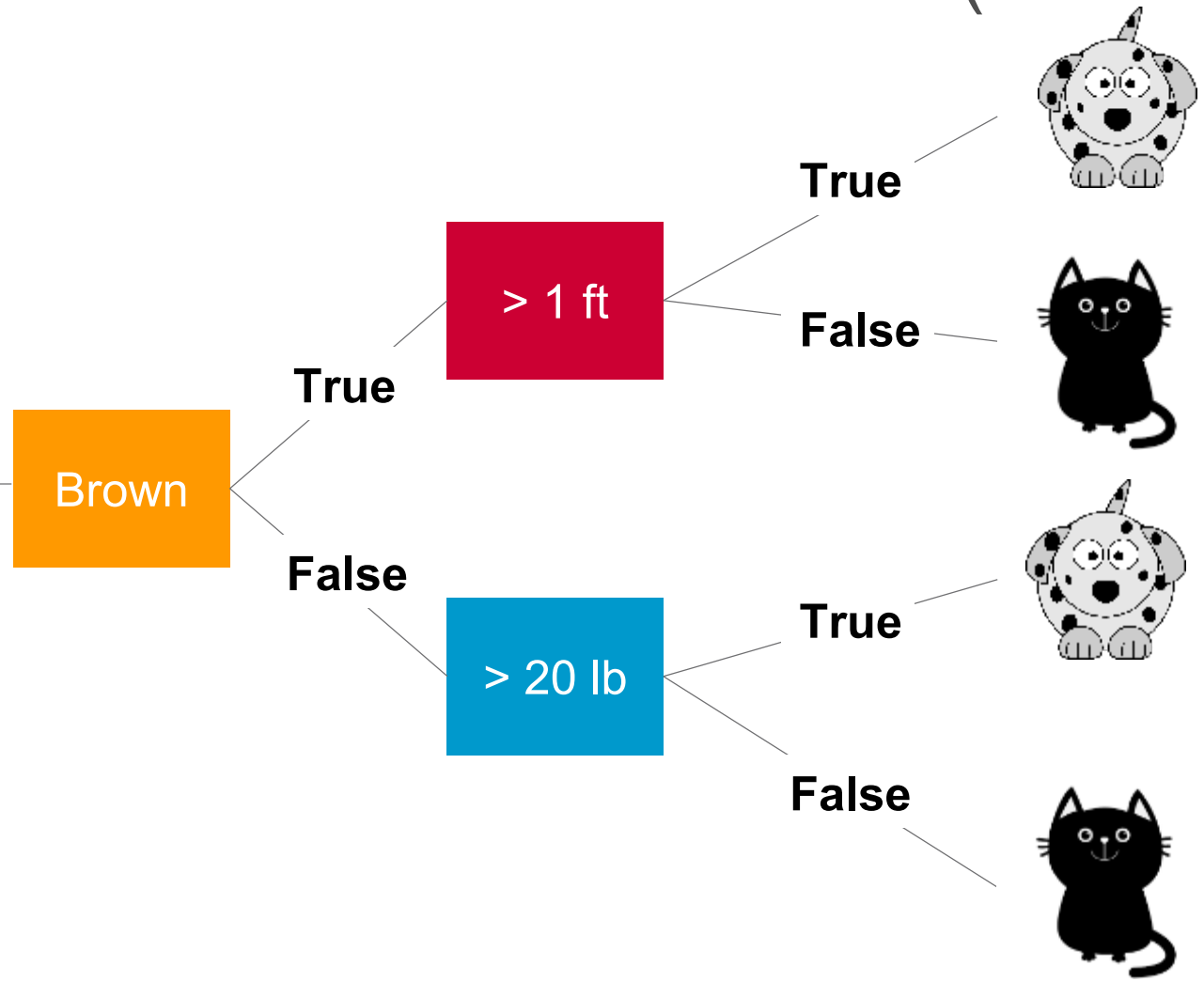


VECTRA

Supervised Classification - Is this A or B (or C or D)?

Predict Cat or Dog given

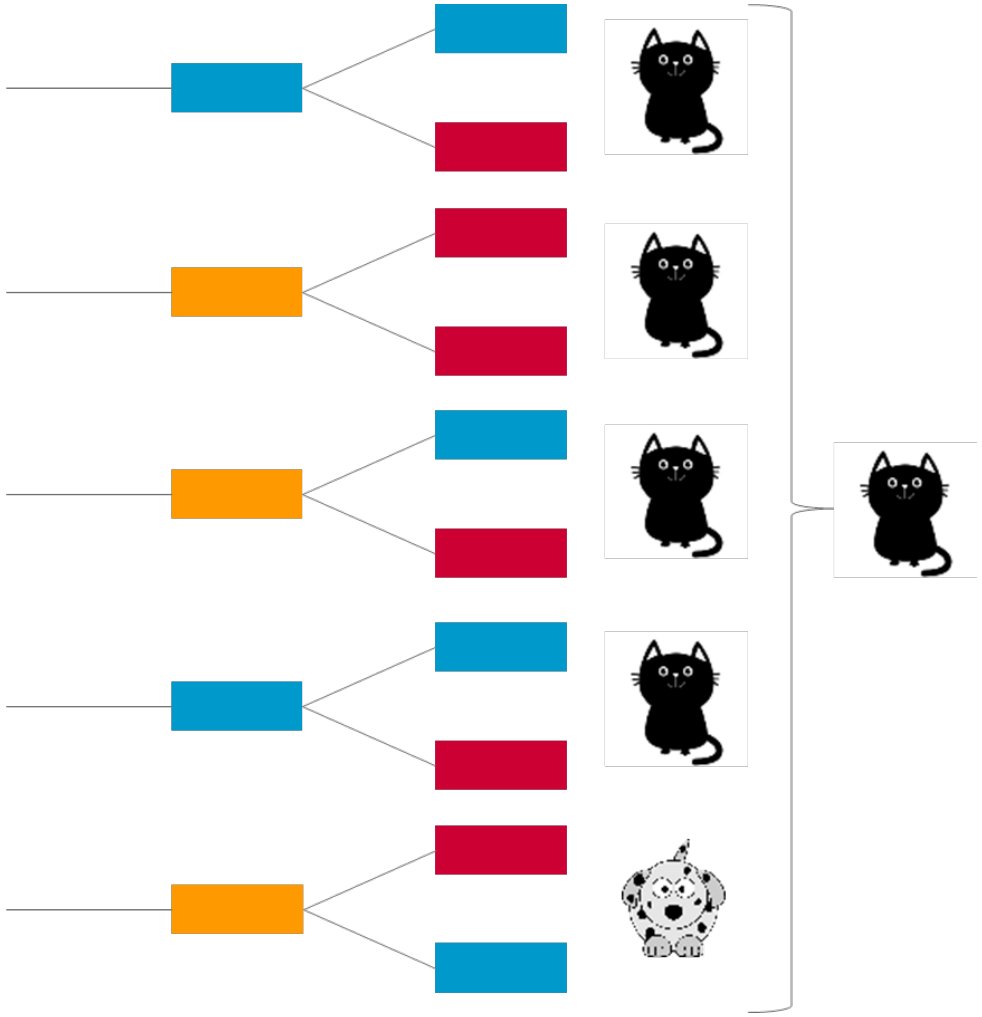
[Brown (T/ F) ,
Weight (0 -200 lb),
Height (0 - 3 ft)]



VECTRA

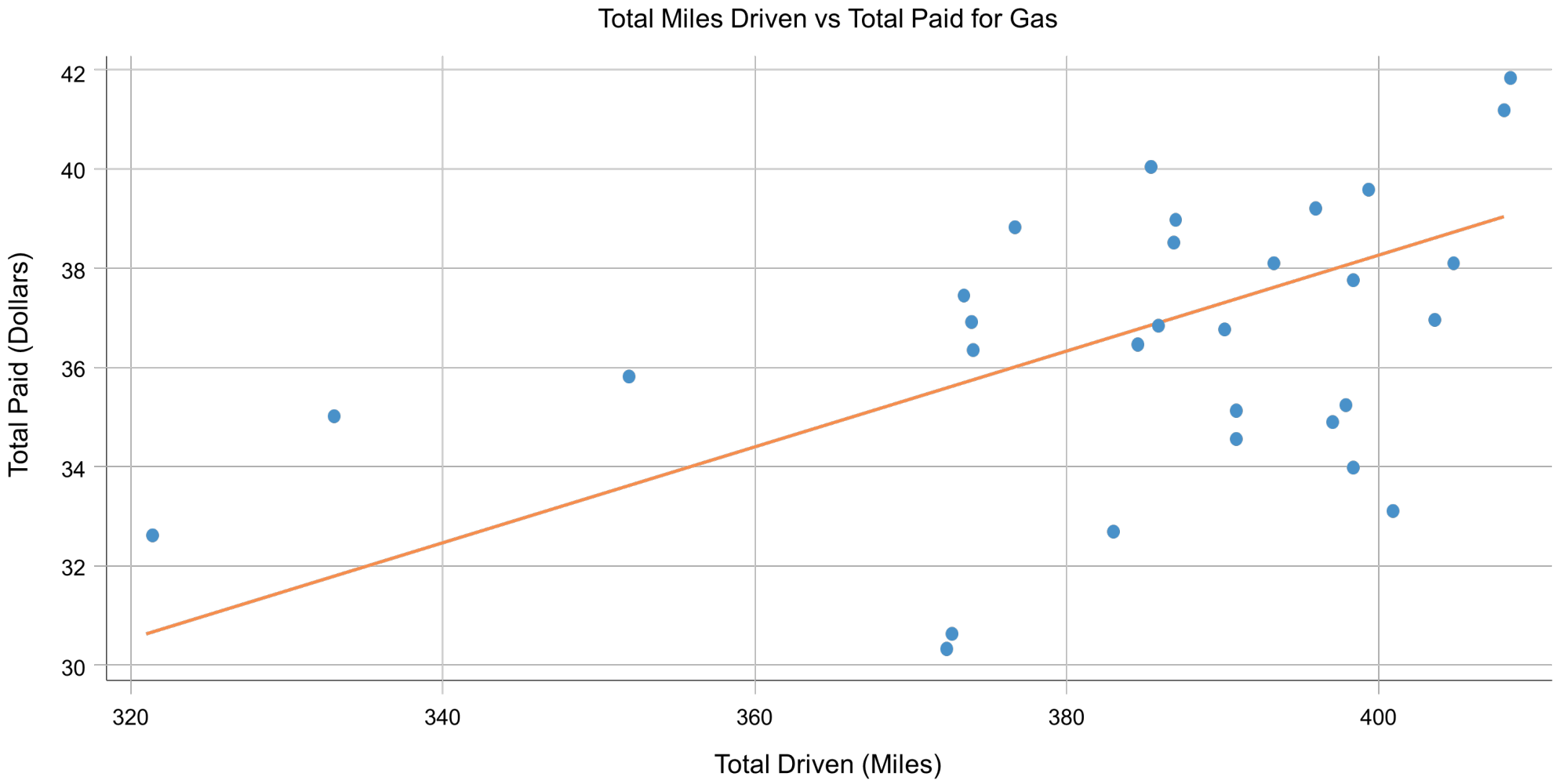
Supervised Classification - Random Forest

Predict Cat or Dog given



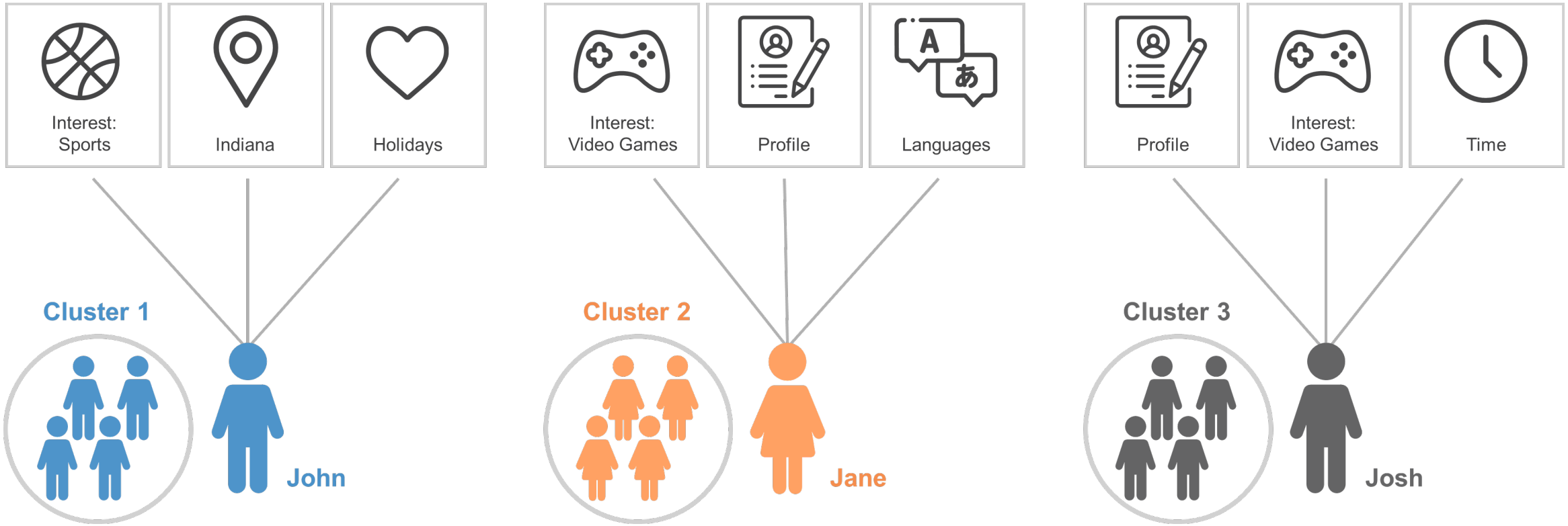
VECTRA

Supervised Regression - How much / How many?



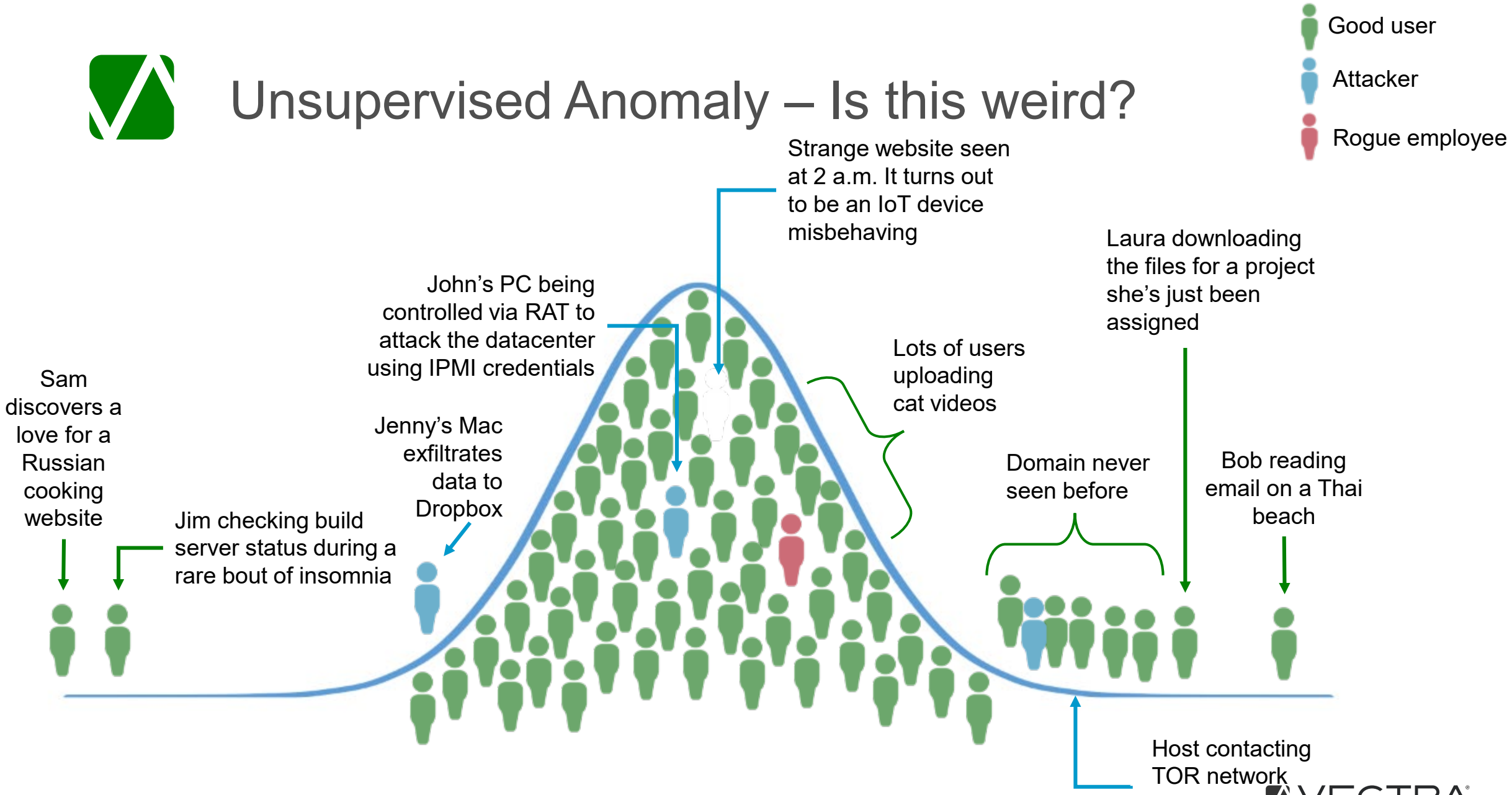
VECTRA

Unsupervised Clustering - How is this data organized?



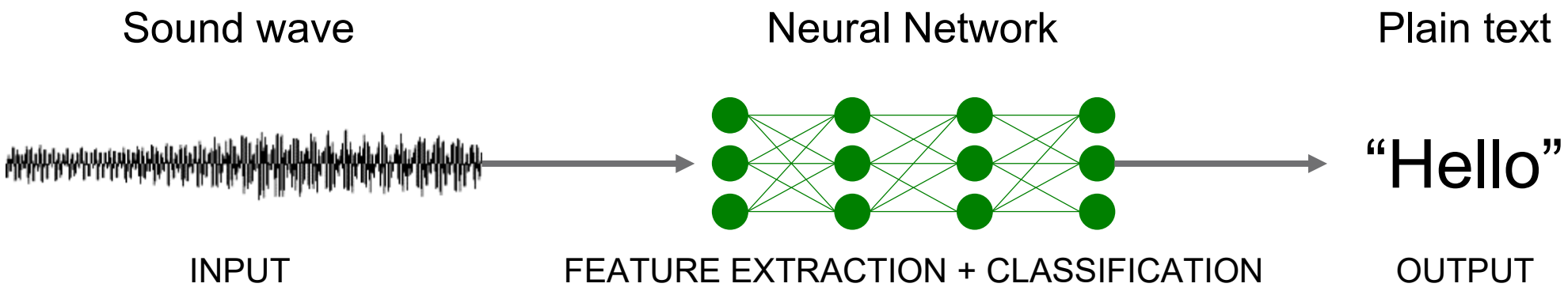
VECTRA

Unsupervised Anomaly – Is this weird?

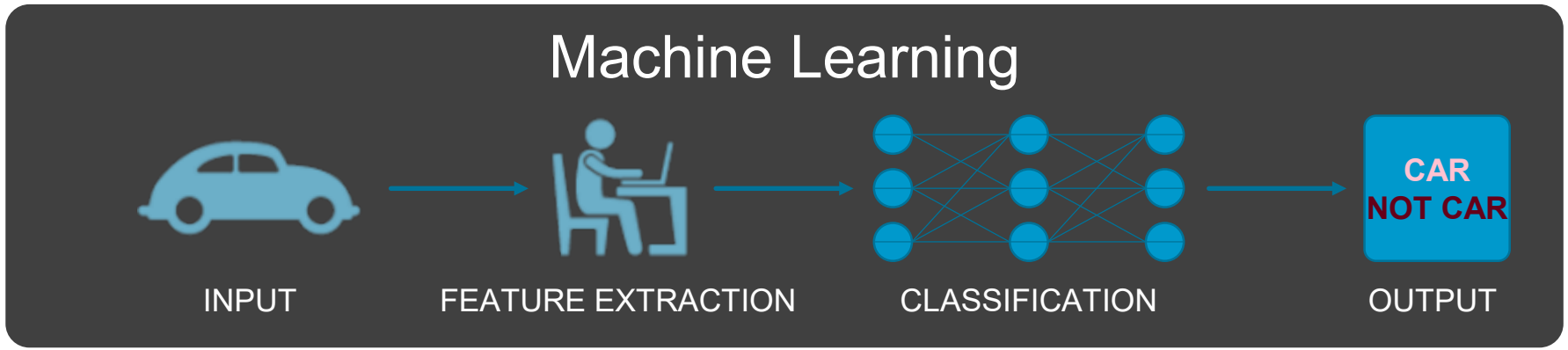


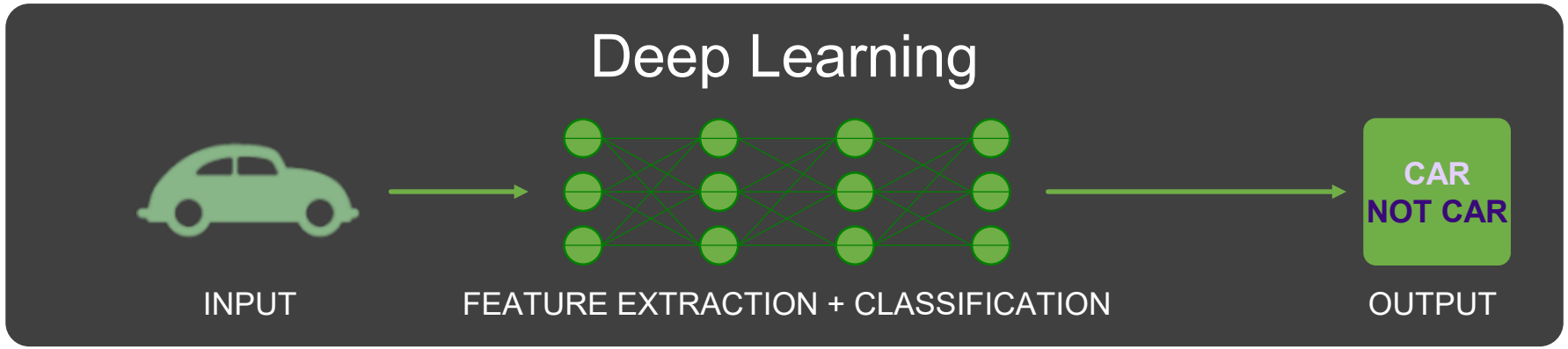
VECTRA

Deep learning - Using artificial neural networks



Choosing traditional machine learning or deep learning





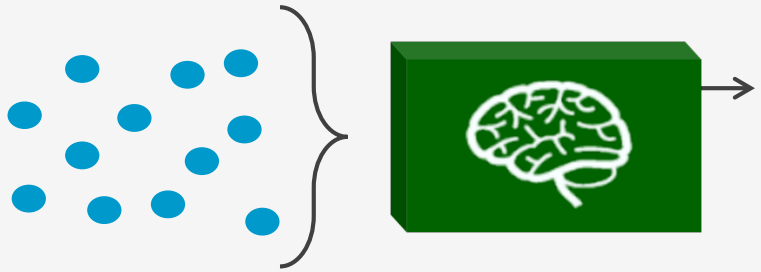
Applying data science to threat detection

Signature

Short-lived reactive intelligence

- How the threat looks
- Find threats that you've seen before
- Snapshot in time
- No local context

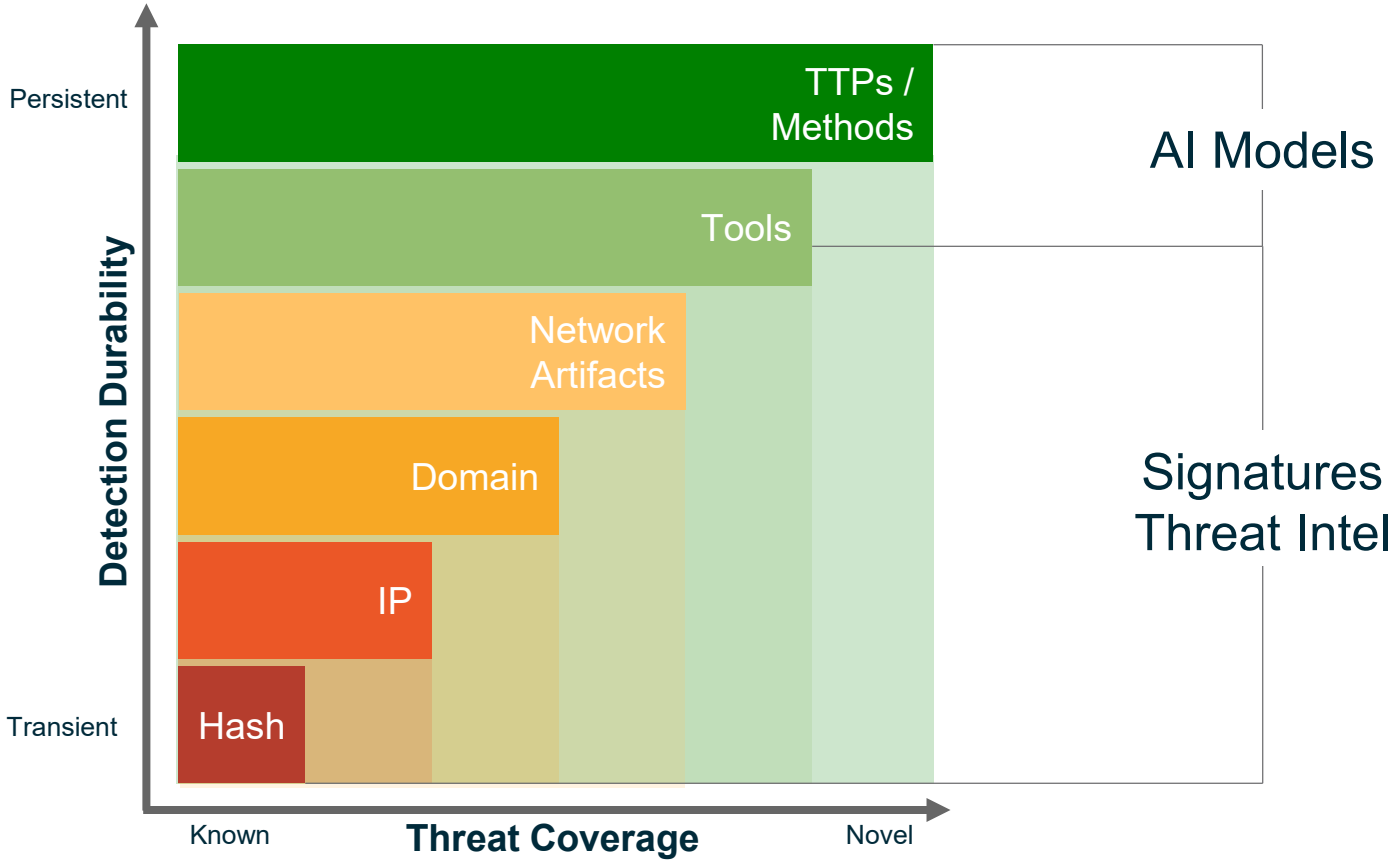
Data science



Long-lived predictive intelligence

- What the threat does
- Find what all threats have in common
- Learning over time
- Local learning and context

Looking for what the threat does



Durable coverage

- Both novel and known attacks
- Difficult and expensive to evade

Fast, labeled coverage of known threats

- Tools
- Exploits
- Known attacker infrastructure
- Environment-specific indicators

VECTRA

Combine data science with security research

Attacker Behavior models

- High-fidelity detection of things attackers must do
- No signatures: find known and unknown

Security Research

- Identify, prioritize, and characterize fundamental attacker behaviors
- Validate models

Data Science

- Determine best approach to identify behavior
- Develop and tune models

It’s all about detecting attacker behaviors

Command and Control	Reconnaissance	Lateral Movement	Exfiltration
Advanced C2: human control	Network sweeps and scans	Stolen accounts	Data movement
Botnet C2	Advanced: AD, RPC, shares	Exploits	Methods, e.g. tunnels
		Backdoors	

Security Research

- Identify, prioritize, and characterize fundamental attacker behaviors
- Validate models

Data Science

- Determine best approach to identify behavior
- Develop and tune models

VECTRA

Supervised deep learning to detect remote access

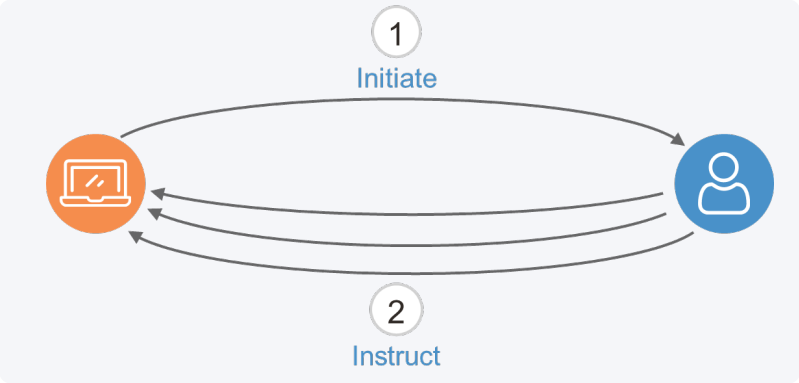
External Remote Access

- Traffic to a multi-dimensional time series
- Deep learning model to featurize the data flow
- Discovers the human on the outside taking control

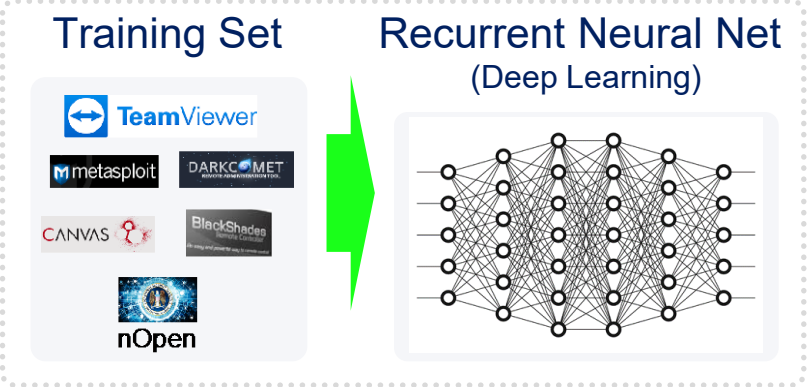
JQSnicker

LogMeIn®

Security Research



Data Science

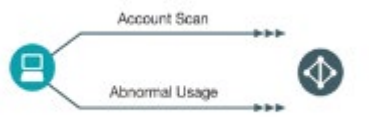


VECTRA®

Unsupervised learning to detect admin misuse

Suspicious Kerberos Client
Suspicious Admin
Suspicious Remote Exec

Security Research



Authenticate using a stolen credential

Administer a host using the stolen credential

Move laterally using credential for remote execution (RPC)

Data Science

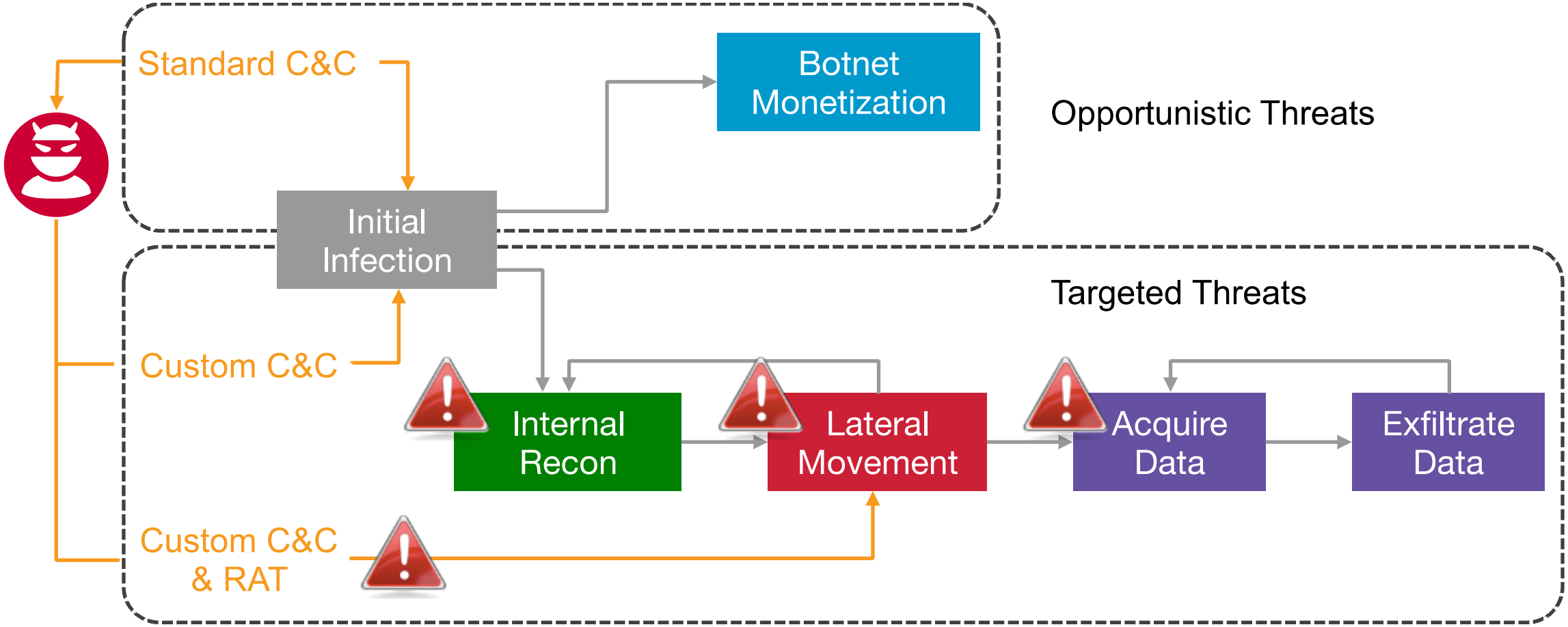
Controller for each host and identify mismatches

Learn which systems each host administers, via which protocols, and identify abnormal administration

Learn normal RPC usage (target, UUID, named pipe, account tuples) for each host and identify abnormal usage

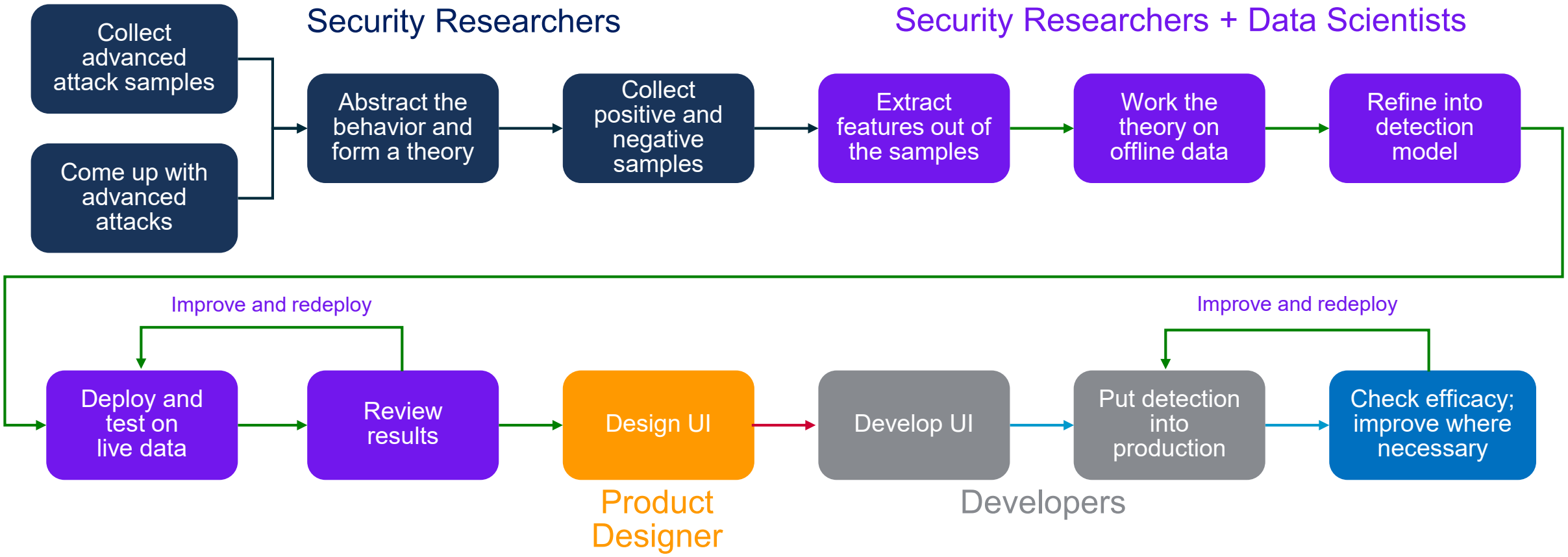
VECTRA

Prioritizing incidents using a time series of events



VECTRA

What it takes to a build an algorithm



VECTRA

What to ask when applying AI to threat detection

- What type of machine learning algorithms are used?
- How many machine learning algorithms are applied, and how are they categorized?
- How frequently are algorithms updated and new algorithms released?
- How many algorithms require a learning period, and how long does that take?
- How are critical and high-risk threats prioritized?
- What kind of efficacy can be expected?

VECTRA®

Thank you

Chris Morales

chris@Vectra.ai