DARKTRACE

WHITE PAPER

Darktrace Antigena: The Future of AI-Powered Autonomous Response

Executive Overview

The current era in cyber security is characterized by three fundamental challenges: the complexity of the enterprise network and connected infrastructure, the speed of new-age attacks, and the strain of overwhelmed incident responders.

In many ways, the expansion of networks and continual adoption of new technologies - from cloud services to the Internet of Things - has expanded the attack surface and introduced new entry-points through which attackers can gain a foothold. This, combined with the ready availability of exploit kits on the Dark Web, has led to the 'vicious circle of the SOC', where incident responders are so busy fire-fighting that they rarely have time to implement the critical patches that would prevent the problem at source.

As security teams struggle to keep up with the rising volume of routine attacks, a new generation of cyber-threat has also emerged, characterized in large part by fast-moving threats that make an impact well before humans have time to respond. These threats span from indiscriminate ransomware campaigns that move at machine speed, through to insider data theft and polymorphic malware that can hide amid the noise of the network and evade traditional controls.

In general, traditional security tools work by pre-defining 'benign' or 'malicious' behavior to identify and block known threats. Yet this approach is severely limited, as it cannot detect novel threats, and the new devices and systems that make up the digital business are so complex and unfamiliar that it is often unclear what benign or malicious would look like in the first place.

When these tools were first introduced, the aim was to deliver some measure of automation to assist human security teams, but they have proven insufficient and sometimes counter-productive, especially in the context of disruptive blocking actions and false positives. Given these limitations, a more sophisticated approach is needed to help strained security teams cope with digital complexity and fast-moving threats like ransomware.

Thanks to the latest advances in machine learning and AI, this approach has been realized in autonomous response technology, which uses artificial intelligence to help augment the human responders. This AI-powered technology is defined by its proven ability to contain high-confidence threats within seconds, without causing disruption to the business. As cyber-threats gain in speed and severity, this approach is being harnessed to transform even the most complex and vulnerable organization into a resilient, self-defending digital business.

With over 7,000 deployments across 105 countries worldwide, Darktrace's cyber AI platform has delivered the first proven, enterprise-grade autonomous response technology on the market: Darktrace Antigena. Like a digital antibody, the technology works by learning the normal 'pattern of life' for every user and device in the business and taking surgical action to contain in-progress threats in real time, before they have time to escalate into a crisis.

This white paper will explore the critical challenges that security teams face in this new era of cyber-threat, and how Darktrace Antigena is leveraging AI to autonomously fight back against advanced attacks, giving human responders the critical time needed to catch up.

> "Darktrace Antigena acts faster than any security practitioner could to prevent damage from attacks such as ransomware."
>
> **451 Research**

Legacy Strategies in Incident Response

To keep pace with emerging threats and inevitable digital growth, organizations have historically turned to three basic strategies in incident response. While these strategies may have been viable in the past, they are no longer sufficient and often introduce new challenges of their own.

By considering their shortcomings, we can better understand why thousands of organizations are adopting AI-powered autonomous response technology to fight back against advanced attacks.

Hire More Incident Responders

Skilled cyber security professionals are at the heart of any mature cyber security strategy and at the center of the SOC. However, it has become clear in recent years that the complexity and speed of attacks will outpace even the best security team. By the time human responders have sifted through the daily torrent of alerts and determined that action is required, it is often too late to save critical data or sufficiently protect essential parts of the business.

While security leaders should not be replacing their incident responders with machines altogether, automation will be critical in enabling their skilled personnel to defend the network effectively, especially given the scale of our businesses and the volume and speed of incoming attacks.

Rather than bring in more humans to manage the increasing workload, the business's objective should be to minimize the urgent fire-fighting and help security teams prioritize the most strategically important activities - from modernizing IT systems and implementing critical patches, to supporting DevOps and other teams on new applications and business rollouts.

Pre-Programmed Response Tools

To supplement incident responders, security teams have almost always leveraged some measure of automation via a host of pre-programmed response tools - from 'next-gen' firewalls and anti-virus solutions, through to Intrusion Prevention Systems and secure email gateways.

These tools are 'pre-programmed' in two crucial respects. First, they rely on pre-defining what 'benign' or 'malicious' look like based on a knowledge of past attacks, typically using rules, signatures, or training data. While this approach may block many routine attacks, it will fail to detect constantly evolving threats and subtle insiders who can lurk beneath the surface and exfiltrate data over weeks and months. It is also notorious for flooding security teams with false positives, adding to their workload and sometimes even distracting them from responding to the most serious threats.

In another respect, these tools are 'pre-programmed' to the extent that their actions are rarely more sophisticated than a simple block or quarantine. This essentially applies a one-size-fits-all approach and inevitably leads to actions that are either too disruptive - as even a minor signature match could get a client kicked off the network - or too brittle, as pre-programmed responses would be unable to adapt to the dynamic behavior of a polymorphic attack or resourceful insider.

Finally, these tools tend to have a fairly limited scope, as their purview is typically confined to the perimeter, endpoint, or email gateway. This limits their ability to make real-time decisions based on insights correlated across the digital business, and to continuously respond to threats as they develop.

Orchestration Solutions

In recent years, security teams have sought to streamline integrations and automate workflows by deploying a range of orchestration solutions, which are designed to correlate insights from different tools and facilitate the creation of playbooks that the machine can execute on your behalf. While these tools may increase flexibility, and even offer the ability to perform automated clean-ups – from wiping laptops to replacing software and Operating Systems – they suffer from significant limitations and complexity.

In a SOC environment plagued by limited resources and overwhelmed responders, security leaders have often grown frustrated with orchestration tools that require a great deal of effort before their teams can start to gain value out of the approach. In particular, security teams would first need to develop their playbooks and understand how these and their security processes work together with their business processes before they can even begin to write down the rules for the machine to implement.

Additionally, these solutions and the many hooks involved naturally require a great deal of ongoing maintenance and upkeep to remain viable. In most cases, it is more efficient to simply patch a vulnerability than wait for something to go wrong and orchestrate around it.

More generally, even if your security team has gone through the trouble of configuring an orchestration solution properly, the technology is only as good as the data it takes in. This means that these tools not only require more manual labor and expertise, but also fail to solve the urgent problem of evolving threats that evade traditional controls, and the need to take targeted actions rather than broad-brushed quarantines.

Autonomous Response: The Machine Fights Back

Having experienced these limitations firsthand, forward-looking security teams are now opting for a more innovative approach - one which leverages AI-based technology to contain in-progress threats at machine-speed, without causing needless disruption to the business.

This proven approach has been realized in autonomous response technology, an instructive title that helps us distinguish between early attempts at automation and some of today's most sophisticated technology in cyber AI. Broadly speaking, a solution that falls into this category must meet the following four conditions:

- Takes surgical action to contain high-confidence cyber-threats
- Doesn't cause disruption to business functions
- Responds in real time
- Fights back against varying threat types, including insider threat

Takes Surgical Action to Contain High-Confidence Cyber-Threats

Autonomous response technology leverages machine learning and AI to cut through the noise and take surgical action based on high-confidence calculations of a given alert's level of threat. While these alerts may not always point to the most serious attacks, it has become abundantly clear that automation in this area can only be effective if it relies on a sophisticated approach to real-time threat detection; and indeed, the application of artificial intelligence in cyber security is nowhere more auspicious than in its ability to tell friend from foe at a granular level, and only surface up alerts which pose a significant risk. By leveraging cyber AI and only neutralizing high-confidence cyber-threats, autonomous response technology reduces analyst workload and avoids needless disruption to the business.

Doesn't Cause Disruption to Business Functions

Even if certain response tools were to succeed in avoiding false positives on the detection side, the risk of taking overly disruptive actions could still pose a real challenge. Indeed, autonomous response technology is not only distinctive in its ability to take action on high-confidence threats, but also – and perhaps more importantly – in its ability to take very targeted actions that are in each case proportionate to the threat detected. In many ways, surgical automation in incident response amounts to the ability to respond to the unique situation in a way that only interrupts potentially threatening activity, and leaves everything else alone.

Responds in Real Time

Time is of the essence when faced with machine-speed attacks, and minutes or hours – let alone days – can quickly lead to loss of valuable data and monetary or reputational costs to the business. With self-spreading ransomware campaigns, patches might become available at the beginning of the campaign, but incident responders tasked with defending large and complex infrastructures will struggle to implement those patches overnight. Accordingly, the ability of autonomous response technology to take action within seconds is critical and explains its increasing appeal in this new age of cyber-threat.

Fights Back Against Varying Threat Types

Whereas legacy response tools typically encourage a stilted 'stove-pipe' approach to cyber defense, autonomous response technology is designed to power the self-defending digital business in its entirety. As the attack surface continues to expand, the scope of our cyber defense technologies must expand along with it. This is why autonomous response technology is invariably agnostic to diverse digital environments, and to the type of threat detected. From insiders gone rogue and hacked connected devices, through to slow and stealthy cyber campaigns that go undetected for months, autonomous response technology provides resilient coverage across the digital business and, derivatively, across the entire 'lifecycle' of both external attacks and insider threats.

Fight Back: Darktrace Antigena

Powered by Darktrace's multi-award-winning AI, Darktrace Antigena represents the first proven application of autonomous response technology in the enterprise. Having pioneered a unique self-learning approach for real-time threat detection, Darktrace's cyber AI platform has evolved to deliver surgical automation that fights back at machine speed, taking proportionate action to contain in-progress threats before they have time to escalate into a crisis.

Rather than pre-defining the threat in advance, Darktrace's AI works by analyzing rich data sources across the digital business to learn the 'pattern of life' for every user, device, and all the relationships between them, using its evolving understanding of 'normal' to identify subtle deviations indicative of an in-progress attack.

When the system detects a high-severity threat, Darktrace Antigena responds within seconds - taking proportionate action to neutralize the threat and give the security team time to catch up. For example, ransomware - which can infect dozens of computers in just under a few minutes - can be detected and contained within approximately 2 seconds, avoiding spread beyond the initial point of compromise.

While human responders continue to strain under the weight of digital expansion, Darktrace's AI flourishes amid the increasing complexity, as every new data point contributes further detail to the system's contextual awareness of what 'normal' looks like, and whether disparate indicators of abnormality should be correlated and contained before a threat has time to develop. This stands in sharp contrast to the vast majority of legacy response tools, which only make decisions based on isolated data points within stunningly brief timeframes.

Darktrace Antigena's probabilistic approach lies at the heart of its claim to surgical automation, combining as it does the context-based confidence of anomaly detection with a measured response that need only interrupt the unusual. The range of actions that Darktrace Antigena can take spans from reconfiguring networks or perimeters and interrupting connections via TCP resets, through to changing permissions, freezing accounts, and even excluding devices when necessary. In every case, the system's intimate understanding of 'normal' allows it to deliver responses that are proportionate to the threat, and which are neither too brittle, nor too disruptive.

Autonomous Response: Dynamic & Surgical

In practice, this amounts to an autonomous response that adapts to the shape of the threat as it unfolds and can become more suspicious over time. Suppose, for example, that Darktrace's AI surfaces a subtle pattern of anomalous activity emanating from a device on the network. Once the device's threat level scores above a minimal threshold, Darktrace Antigena responds by taking granular action and blocking a single abnormal connection – shutting down a command and control channel beaconing out to Portugal.

Yet over time, the malicious implant subtly switches to fallback mechanisms. Adapting to the new behavior, Darktrace's AI factors the previous event into its calculation and takes further action in response, escalating from the interruption of a single connection to enforcing the device's group 'pattern of life', which only allows it to make connections and data transfers that it or any of its peer group typically make. As and if the threat escalates further, Darktrace Antigena escalates its response in kind, only permitting the device to operate within the confines of its individual 'pattern of life', and containing the behavior long enough for the security team to investigate and ultimately remediate the threat.

Here as elsewhere, Darktrace Antigena's surgical actions were anchored in an understanding of the device's normal behavior in relation to its past, to its peer group, and to the wider organization. While Darktrace's AI escalated as the threat developed, the response was proportionate in every case, and was able to maintain the device's pattern of normality by design. Throughout the incident, the user of the device would have been free to use normal applications and access the usual file shares, without even noticing that Darktrace's AI was working behind the scenes to proactively protect the business.

> "Darktrace Antigena enables organizations to fight back against cyber-threats without disrupting daily activities."
>
> **IDC**

Self-Defending Digital Business

Finally, Darktrace Antigena responds to cyber-threats across the digital business, from enterprise and industrial networks, through to cloud containers, SaaS applications, and even email communications. This breadth not only refines the quality of Darktrace Antigena's decision-making - as it enriches the system's holistic understanding of 'normal' and brings even more context to bear on its evolving measure of threat probability – but also extends its scope of action well beyond the limited reach of legacy response tools.

Indeed, whereas most pre-programmed response tools are isolated and therefore only have one shot to stop an emerging attack, say, at the perimeter, Darktrace Antigena's breadth cuts across the entire kill chain - whether the attacker is attempting to move from the corporate network to the cloud, from the industrial network to the enterprise, or even from an IoT device in one area of the business to a critical server in another.

By providing this broad protection, Darktrace Antigena represents autonomous response at its finest – taking proportionate action on high-confidence cyber-threats at computer-speed, wherever the threat resides.

Darktrace Threat Visualizer & Mobile App

Darktrace's graphical Threat Visualizer interface provides a single view from which anomalous activity and Antigena actions can be visualized and investigated in real time. The Threat Visualizer is designed for users of all maturity levels, from forensic experts, to business executives and less experienced members of the IT team.

A wealth of information can be variously queried and exposed using the interactive features within the Threat Visualizer, including a dynamic dashboard where users can filter incidents based on their level of severity, and an interactive Play-Back tool which lets users replay incidents and assess the real-time context around each event.

The Darktrace Mobile App also lets users access these rich insights and confirm Antigena actions when on the move. Designed to offer maximum flexibility and increase the speed of mitigation, the app offers push notifications of in-progress attacks and one-click confirmations of Antigena's suggested responses. When an attack transpires, security teams can view and contain threats within seconds, even when between shifts or out of office.

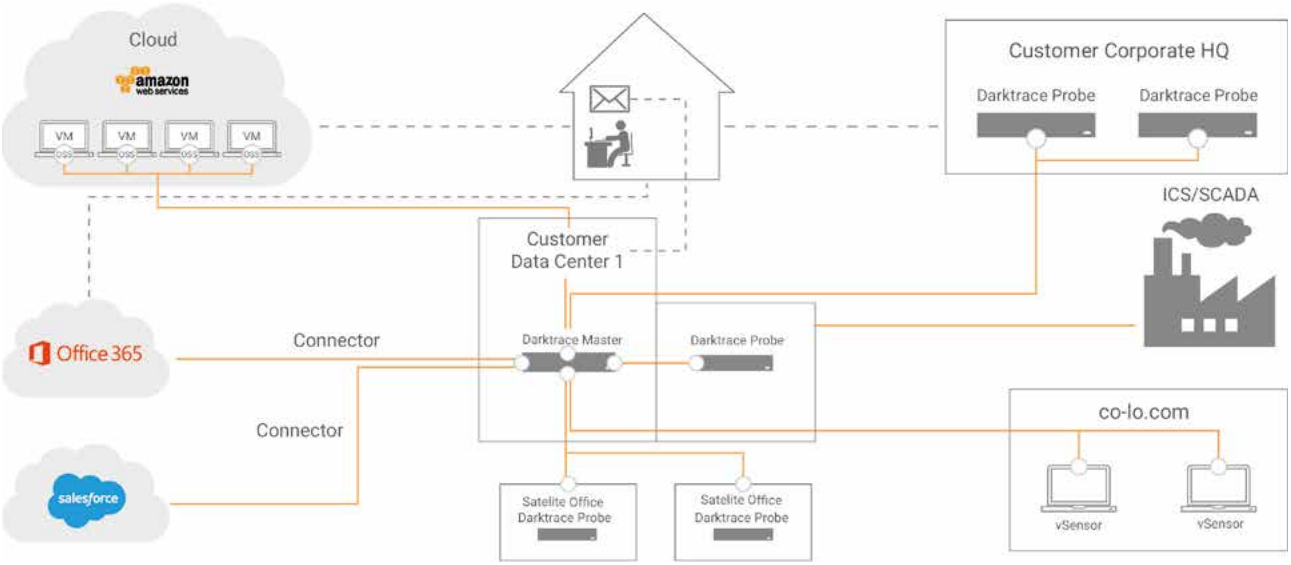

Figure 1: Darktrace Antigena across the digital business

Rules of Engagement: Building Initial Trust

Darktrace Antigena is fully customizable and controllable, allowing customers to stay in the driving seat and transition to a fuller use of AI in their enterprise.

To build initial trust, organizations invariably benefit from the ability to configure Darktrace Antigena's scope as it relates to the system's level of automation and focused use cases, as well as the environments in which it can be deployed.

Configurable Automation

Darktrace Antigena can be configured in one of two modes to allow for varying degrees of automation. Depending on your organization's risk appetite, either mode can be applied globally across the entire organization, or locally for a given device, system, or user group.

Human Confirmation Mode
In this mode, Darktrace Antigena generates responses which must be validated by the security team before action is taken. Customers can approve Antigena's proposed actions easily via the Darktrace Mobile App or Threat Visualizer. This allows users to build confidence in Antigena's decision-making before switching to Active Mode.

Active Mode
In Active Mode, Darktrace Antigena is fully autonomous within its defined operating parameters. This means that a threat may be instantly contained without a security analyst needing to log on. Actions initiated in Active Mode can also be monitored through Darktrace's Threat Visualizer or the Darktrace Mobile App.

> "With Darktrace, talk about AI in cyber security has turned into action."
>
> **Ovum**

Focused Use Cases

Organizations can also build trust by starting their deployments with a use-case oriented approach. If desired, Darktrace Antigena can be configured to trigger on specific risk categories based on the type of threat detected. These categories include external attacks, insider threats, and compliance risks, though the system can also be configured to trigger on any significant anomaly regardless of threat type.

External Attacks
Darktrace Antigena can stop even the most automated variant of ransomware in its tracks, given the system's ability to fight back at computer-speed. In this context, Darktrace Antigena would typically interrupt concerted attempts to encrypt internal network shares, or neutralize emails carrying ransomware payloads before they reach the user. Darktrace Antigena can also take targeted action on other forms of malware, by blocking the download of malicious files from rare external sources or attempts to beacon out to command and control centers.

Insider Threats
Darktrace's AI is also capable of correlating a rich constellation of weak indicators to identify the early signs of an insider threat. If configured for this use case, Darktrace Antigena can take action on a range of anomalies, from unusual privileged user activities and large outbound data volumes, through to unexpected large downloads from internal servers to client devices. Darktrace Antigena's actions in this context might include blocking unusual SSH and RDP connections from non-administrative systems, or stopping devices from sending unusually large volumes of data to external devices they don't normally communicate with.

Compliance
Darktrace Antigena can also be leveraged to promote cyber hygiene and compliance, and prevent attacks that may appear in future. For instance, Darktrace Antigena can be configured to selectively stop devices from communicating with file share services like Google Drive, block users from connecting to the TOR anonymizing network, or stop the use of FTP either internally, externally, or both.

Autonomous Response Across Your Digital Infrastructure

As Darktrace's AI is fundamentally agnostic to diverse digital environments, organizations can also build trust by starting their deployment in environments where they think it will add the most value, and then extending its reach from there.

Corporate Estate & Email

In most piecemeal deployments of this kind, organizations often start by deploying Darktrace Antigena to cover their corporate estate and email communications, as the system will likely add the most value in areas of the business that have the highest volume of work.

In particular, Darktrace Antigena's email capability is an especially valuable application of AI for organizations that are just starting to ramp up their autonomous response deployments. Notably, Darktrace Antigena is able to bridge the traditional security knowledge gap between what occurs at the external email layer, and what occurs inside the network. It recognizes that a malicious email is designed to provoke actions and activities in the network which Darktrace AI can see, and which can inform Antigena's autonomous actions against similar emails which are targeting the business.

Crucially, this means that Darktrace Antigena is able to contain email-borne attack campaigns against multiple users after a single user has already been infected in the network. For example, Darktrace can recognize that malicious activity in the network had an email as its source, and allow Darktrace Antigena to hold back similar emails targeting the business by either removing them from corporate inboxes, or preventing them from reaching the user.

This preemptive protection would stop the spread of an emerging attack campaign, and give human responders time to catch up.

> "Darktrace's AI detects threats and stops them immediately."
>
> **Penn Highlands Healthcare**

Cloud Containers & SaaS Applications

Organizations will typically extend their deployment beyond the corporate estate and email to cover cyber-threats in the cloud.

The types of actions Darktrace can take in this context vary depending on the specific cloud environment or SaaS application being used, as illustrated in the lists below which are not exhaustive nor definitive across all SaaS or cloud platforms.

To neutralize in-progress attacks in cloud environments like AWS and Azure, Darktrace Antigena can:

- Terminate a virtual machine or edit its properties
- Edit S3 bucket permissions in AWS
- Temporarily disable a user's programmatic access
- Reset user passwords to disable management access
- Edit user permissions
- Temporarily stop sharing a document

In SaaS applications like Office 365, Salesforce, G-Suite, and Box, Darktrace Antigena can:

- Kill a user's active sessions
- Temporarily disable users
- Restrict or delete file sharing settings from certain files and folders
- Restrict a user from accessing certain parts of the cloud environment
- Suspend members from teams and hence their access to certain shared files (in Dropbox, for example)

Industrial Control Systems

Given the safety-critical nature of industrial environments, Darktrace Antigena is typically deployed on the border of OT networks, or between enterprise and industrial networks.

For instance, users might configure the system to confine sub-contractors to their individual 'patterns of life' when interacting with industrial equipment, or stipulate that corporate systems can only access the stock levels in the power station according to routine ways.

Darktrace Antigena
Real-World Threat Discoveries

Stopping Sophisticated Ransomware

At a global financial services firm, an employee circumvented corporate policy to check her personal webmail on a company laptop. She opened what she believed to be a Word document, but was actually a malicious ZIP file containing a ransomware payload. The device contacted a rare external domain and began downloading a suspicious EXE file.

Darktrace's AI recognized this activity as highly anomalous. When the executable began to encrypt SMB file shares, this represented a deviation from the device's normal 'pattern of life'. At this stage, the system determined the threat was serious enough to require an immediate response.

The security team was not on site to take action to remediate the situation. Darktrace Antigena took autonomous response and interrupted all attempts to write encrypted files to network shares. In so doing, Antigena neutralized the threat just seconds after the malicious activity began.

Ransomware attacks like these are increasingly common, and as new and more insidious strains emerge on the dark web every day, ransomware will inevitably bypass even the most sophisticated legacy tools. Moreover, ransomware is capable of encrypting an entire network in a matter of minutes. Human security teams cannot keep up with such fast-moving attacks, and autonomous response has become vital in today's threat landscape.

> "Darktrace Antigena is a force multiplier."
>
> **City of Las Vegas**

Disrupting an Insider Threat

At a major hotel chain in Asia, Darktrace's AI detected a sudden spike in anomalous activity. External servers were attempting thousands of remote-desktop connections by guessing default usernames and passwords.

Darktrace identified the activity as an anomalous deviation from the network's 'pattern of life' and further investigation revealed that these connection attempts used a specific pattern, indicative of an automated attack. Darktrace identified that some of these remote-desktop connections were using a known set of credentials.

The external server was being accessed from outside the network with an internal user account. The server then made remote-desktop connections between other company computers before arriving at the hotel property management system, from where a large volume of data was downloaded.

A comparable volume of data attempted to leave the network, going to the external device that initiated the original remote-desktop connection. These connections were deemed highly suspicious since they represented an extreme deviation from the devices' normal 'pattern of life'.

The company management reported that the user account in question belonged to a former employee who had only recently left the company. It is possible that he had sold his access credentials before they could be disabled, or he could have been attempting to retrieve the data himself before selling to a competitor.

Darktrace's AI was able to detect and take autonomous action in real time, preventing the attempted data exfiltration before the information left the network, and buying the security team precious time to intervene.

Responds within
2 seconds

7 threats blocked
every minute

10 hours a week saved
per security analyst

Conclusion: AI & the Next Generation of Cyber-Threat

The increasing sophistication of today's threat landscape together with the growing complexity of digital businesses has seen organizations of all sizes struggle to keep pace. Incident response teams find themselves overwhelmed and under-resourced, even as their legacy response tools continue to fall short in the face of emerging threats. And yet - while these threats already pose a nearly impossible challenge, they are rapidly taking on a new dimension.

In particular, cyber-criminals have begun to leverage advances in AI to introduce troubling new economies of scale to the fight. Traditionally, cyber-attacks could be located somewhere along a familiar spectrum – on one end, there were massive if unsophisticated attacks that affected the global landscape, and on the other, there were advanced, subtle attacks directed at only a few, high-value targets, as they were invariably constrained by the considerable manpower and time required to mount a bespoke attack. With AI, however, attackers are able to develop sophisticated tools that launch customized attacks at scale, such that machines can quickly analyze, say, email styles and LinkedIn profiles to send out targeted phishing campaigns to millions of victims at machine-speed.

As advances in AI continue to provide attackers with the opportunity to enhance the speed, scale, and sophistication of cyber-attacks, the war of algorithm against algorithm is being fought within corporate networks around the world. Organizations leveraging AI-based autonomous response technology are therefore clear-eyed about the shape of things to come, as today's businesses will stand or fall on their ability to rapidly contain even the most advanced attacks before they have time to escalate into a crisis.

Darktrace Antigena's cyber AI technology represents the first proven autonomous response technology in the enterprise, powering self-defending digital businesses to fight back against this new age of cyber-threat. By leveraging its evolving understanding of 'normal', Darktrace Antigena takes dynamic and surgical action to neutralize emerging threats – stopping their spread in real time, and giving security teams time to catch up.

Darktrace Antigena Key Benefits

- **Fights Back:**
  Takes surgical action to stop high-severity threats
- **Buys You Time to Catch Up:**
  Helps teams prioritize the most strategically important activities
- **Real-Time and Around-the-Clock:**
  Works in real time, 24/7
- **Customizable:**
  Configured according to your risk appetite
- **Flexible:**
  Controlled on the move via the Darktrace Mobile App

> "Darktrace's AI is unique and actually does what it says it will."
>
> **Forrester**

About Darktrace

Darktrace is the world's leading AI company for cyber security. Created by mathematicians, the Enterprise Immune System uses machine learning and AI algorithms to detect and respond to cyber-threats across diverse digital environments, including cloud and virtualized networks, IoT and industrial control systems. The technology is self-learning and requires no set-up, identifying threats in real time, including zero-days, insiders and stealthy, silent attackers. Darktrace is headquartered in San Francisco and Cambridge, UK, and has over 30 offices worldwide.

Contact Us

North America: +1 415 229 9100

Latin America: +55 11 97242 2011

Europe: +44 (0) 1223 394 100

Asia-Pacific: +65 6804 5010

info@darktrace.com

darktrace.com