DARKTRACE

多層的AI装備

サイバーセキュリティを形成する
ツールを理解する

概要

ダークトレースはルールベースのセキュリティツールをすり抜ける新しいサイバー攻撃と戦うことを目的として、コーポレートネットワーク内で脅威となるデジタルアクティビティを検知するという課題に対し、教師付きおよび教師なし機械学習を適用する、研究開発主導型の企業です。

本ホワイトペーパーでは、DarktraceのCyber AIを構成する多層的な機械学習テクノロジーと、これらがどのように組み合わされ、自律的な、自己を更新する、人間の入力に応え ーしかしそれを必要としないー システムがどのように構築されているかを説明します。

謝辞

Tim Bazalgette、
PhD、SVP AI / ML、R&D

Nicole Carignan、
SVP、Security & AI Strategy、Field CISO

Hanah-Marie Darley、
MSc、Director、Security & AI Strategy、Field CISO

Jack Stockdale、CTO

Brittany Woodsmall、
MBA、Manager、Product Marketing、AI

Lily Steinberg、
Manager、Content Marketing

上記の方々に対し本書に対する貢献に感謝します。

目次

サイバー脅威の新時代

過去 10 年間、サイバー戦争は間違いなくエスカレートし、犯罪者や国家、あるいは機に乗じた個人が、デジタル化やインターネットへの接続を利用して、金銭的利益、評判の毀損、あるいは戦略的優位性を得るためにシステムに侵入しようとしています。

また、サイバー脅威アクター達は絶えず進化を続けています。ファイアウォールやシグネチャベースのゲートウェイ等、従来型の防御をすり抜けるために新しい攻撃ツールやテクノロジーを開発し、最も脆弱なファイル共有やアカウントに侵入して来るのです。

最近では、生成AIの普及に伴い、攻撃者達も自動化を取り入れて攻撃の量、速度、多様性を増大させています。

78%

CISO の 78% が、AI を使ったサイバー脅威が既に組織に対して大きな影響を及ぼしていると認めています。

95%

CISO の 95% が、サイバー防御のスピードと効率を向上できると考えています。

攻撃にAIを使用することで、高度な戦術、技術、手順（TTP）をより迅速に更新して脆弱性をエクスプロイトし、検知をすり抜けています。また、ますます拡大しているサービスとしてのマルウェア（MaaS）やサービスとしてのランサムウェア（RaaS）など、サブスクリプションベースのツールは経験の少ない攻撃者の参入障壁を引き下げ、複雑な多段階の攻撃の実行がより簡単になっています。

また、今日の平均的組織では複数の拠点や複数のデバイスおよびテクノロジーサービス ー クラウドサービスやSaaS（Software-as-a-Service）ツール、信頼のおけない家庭用ネットワークから、公認されていないIoT（Internet of Things）デバイスまで ー に渡ってデータが共有されており、攻撃者たちはこうしたデジタル環境の複雑さを悪用しています。それと同時に、悪意のある内部関係者も常に脅威として存在します。

これらの新しい脅威の出現によって技術的な変化が必要となります：

それは、私達が予期できないものを機械学習を使ってどう検知するか？ということです。つまり、以前の攻撃についてのデータセットに依存することなく、その環境で起こっている他のすべてのこととその動作を比較することにより、脅威が何であるかを学習するシステムを構築できないか？ということです。

本ホワイトペーパーはAIおよび機械学習に対するダークトレースのアプローチを解説し、ダークトレースの多層的なAIアーキテクチャの中核となる、教師なし機械学習、教師付き機械学習のユニークな連携に光を当てます。

サイバーセキュリティへの従来のアプローチ

従来のセキュリティパラダイムにおいて、ファイアウォール、エンドポイントセキュリティソリューション、およびSIEM（Security Information and Event Management）やIDS（Intrusion Detection Systems）などその他のツールは、特定のポリシーを徹底し、既知の脅威からの保護を提供するために展開されています。

これらのツールが企業の全体的防御計画において果たす一定の役割はありますが、急激に変化を続ける新時代のサイバー脅威に対抗するには、特にエンタープライズインフラの多様化が進むなかで、十分な装備とは言えません。

現代のAIテクニックの強みと課題

境界制御は範囲が狭く、侵入点で攻撃を見逃してしまった場合には失敗を意味し、それ以上のアクションを取ることができません。

エッジセキュリティは主として既知のルールやシグネチャに依存しており、以前に識別されている攻撃の検知に特化しています。これらのツールが未知の攻撃を検知できることもありますが、これらのツールの得意とするものではなく、新手の脅威に対しては効果が低いと言えます。これらのツールに依存することへのもう1つの問題点は、2024年に見られたエッジセキュリティツールの脆弱性開示件数の増加で、この傾向は今後も続くと思われます。

ログツールおよびSIEMデータベースは、組織全体で一貫したデータ収集を行い、セキュリティチームの脅威予測と整合させるための作業を行うのに人手がかかります。これらのツールを運用するには、起こり得るあらゆる事態をセキュリティチームが想定しつつ、多数のアラームでアナリストを圧倒してしまわないようにしなければなりません。

いわゆる**行動アナリティクス**は、多くの場合特定の役職または機器がどのように行動「すべき」であるかを設定し、そこからの逸脱を探すという、ルールベースのパラダイムに依存しています。また、人間が主導する教師付き機械学習モデルを使用して、疑わしいあるいは異常な動作を学習させる製品は、検知が過去の静的な知識に限定されてしまいます。これらのアプローチは今日のビジネスの複雑性と規模に対応できません。

結局のところ、従来のシステムはこれらの根本的制約により、今日のビジネスの複雑性と攻撃者のイノベーションのペースに後れをとっています。

これらのアプローチでは次が必要になります：

- 過去の攻撃をすべて知っていること
- 業務内容および業務特有のルールを完全に理解していること
- 新しい攻撃に対する高品質な情報を共有する万全な方法があること
- 脅威ランドスケープを予測してゼロトラストを実現すること
- 上記すべての情報を有効なルールやワークフローに変換できること

機械学習はサイバーセキュリティ業界に大きなチャンスをもたらし、多くのベンダーが既に何年も利用しています。しかし、機械学習をサイバーセキュリティに適用することの潜在的メリットは大きいものの、すべてのAIツールや機械学習モデルが同じというわけではありません。

たとえば、AIを使ってプロセスを自動化しているツールであっても、侵害事例からのデータがなければ、ソリューションを提供できません。予測不可能な動きの速い攻撃の時代においては、このようなアプローチはまったく不十分なものとなりました。

AIのさまざまなタイプと、サイバーセキュリティにおけるそれらの役割を知ることは、堅牢な多層型のAIセキュリティソリューションに何が必要かを理解するのに役立ちます。以下のセクションでは、主なタイプのいくつかを説明します。

教師付き機械学習とサイバーセキュリティ

一部のサイバーセキュリティベンダーは教師付き機械学習をさまざまな方法で10年以上も試しています。そのほとんどは、ビッグデータサイエンス、サイバー脅威インテリジェンス、既知または報告済みの動作、および分類器などを使って脅威検知ベースの報告データを自動化しています。

教師付き機械学習はラベル付けされたデータでトレーニングされます。
サイバーセキュリティでは、これは多くの場合これまでに見られた動作のデータベースを意味します。そこに含まれる各動作は悪意のあるもの、あるいは良性のものとして知られているものであり、そのようにラベル付けされています。その上で新しい動作が分析され、それらが悪意あるクラスと良性のクラスのどちらに近いかを判断します。悪意のあるものである可能性が十分に高いと評価されたものには、脅威のフラグが立てられます。

教師付き機械学習システムは過去の知識に基づいて明確な答えを出すのに最適です。
たとえば、多数の既知のランサムウェアのサンプルをシステムに入力することにより、システムはマルウェア共通のインジケータを学習し、将来類似の攻撃が発生すると、それを検知することができます。これはシグネチャベースのアプローチよりも有利です。特定のシグネチャにそれほど限定されず、より複雑なプロパティの組み合わせに基づいて脅威を識別することを学習できるからです。

しかし、教師付き機械学習においては”オーバーフィッティング”がよく問題となります。これはモデルパラメータをトレーニングデータに細かく合わせすぎる問題です。そのカテゴリの本質を学習するのではなく、機械学習は特定の例を学習するのです。たとえば、教師付き機械学習の分類器は既知のランサムウェアの明示的なTTPを識別することを学習できますが、攻撃者がTTPを変化させ、環境寄生テクニックを使用し、あるいはシグネチャと一致しないツールを使用すると、そのグループでは検知されません。そのためこれらの教師付き機械学習モデルは、データセットから外れたパターンや特徴を見逃すことになります。

過去5年間において、私達はより多くのベンダーが行動アナリティクスや異常検知方面に拡張していく様子を目にしていますが、これらはベースラインの（静的な）正常性についての理解に基づいて、1、2種類の機械学習モデルを使って異常を検知することに限定されています。

この手法は、デジタルエステート内の各デバイスまたはユーザーの事前に定義された動作をベースとしたものです。この手法では、動作プロファイルが作成される際の学習と、その後の異常検知が分離されています。そのため、学習は継続的ではなく、ビジネスオペレーションの動的変化に対応するためには定期的な更新と再トレーニングが必要で、擬陽性と偽陰性が毎日大量に発生することとなります。

主要なAIサイバーセキュリティ手法の違いと、ダークトレース独自のアプローチについてより詳しく知るには、学術記事 “Anomaly-Based Threat Detection:Behavioral Fingerprinting Versus Self-Learning AI” をお読みください。

教師付き機械学習だけに依存するシステムには次のような根本的な制約があります。

- モデルが特定の種類のアクティビティを検知するようトレーニングされている場合、それとある程度異なる新しい種類のアクティビティは識別できません。
- トレーニングデータはラベル付けする必要があり、それには大量の人間による入力、あるいは機械ラベリングや、データセットの既存のプロパティに基づいたラベルの決定などその他の手法が必要となります。
- ラベル付けの誤ったデータ、あるいは人間のバイアスが入力されると、システムが新しい動作を正しく分類する能力が大幅に損なわれます。
- 教師付き機械学習は本質的に静的であり、定期的な再トレーニングと更新を必要とします。

教師なし機械学習とサイバーセキュリティ

サイバーセキュリティに応用されるもう1つのAI手法は、教師なし機械学習です。教師付きアプローチとは異なり、ラベル付けされたトレーニングデータを必要としません。データの重要なパターンや傾向を、人からの入力を必要とすることなく識別することができるのです。

教師なし機械学習は、過去の脅威の知識に依存するのではなく、単独でデータを分類し、説得力のあるパターンを検知する力を持っています。これにより人間による入力や、学習を導くための関与に頼ることがなくなります。精度と有効性は、個別のユースケースにどのようなデータと機械学習テクニックが用いられるかということに大きく依存しています。
使用されるテクニックによっては、教師なし機械学習を使って継続的に学習し、更新し、自己調停することが可能になります。

たとえば、Darktraceは教師なし機械学習を使用して、インフラ全体に渡ってデバイス、ユーザー、あるいはクラウドコンテナやセンサーなどの「正常な」動作の理解を構築し、この変化する「生活パターン」からの逸脱、すなわち脅威の出現の可能性を検知します。

しかしながら、他のすべての機械学習テクニック同様に、教師なし機械学習にもいくつかの弱点はあります。入力パラメーターに制約され、出力の精度を確保するには入念なセットアップを必要とします。さらに、教師なし機械学習はデータに含まれるパターンを見つけることができますが、それらのパターンの中には関係のないものや紛らわしいものが含まれていることがあります。また、その出力はどのような脅威が存在するかではなく、どのような異常があるかという形で提供されるため、それらをセキュリティチームへの情報に変換するには、人間による解釈や、さらなるモデリングが必要となります。

大規模言語モデル（LLM）とサイバーセキュリティ

市場における最近の LLM の爆発的普及に伴い、多くのベンダーはチャットボット、RAG（Retrieval-Augmented Generation）システム、エージェント、エンベディング等に使用するため、生成 AI を製品に取り入れることに躍起となっています。LLM は生成 AI の原動力となり、教師付き、教師なし、両方の手法で使うことができます。これらは膨大な量のデータであらかじめトレーニングされ、人間言語、機械言語その他に適用することができます。

セキュリティ分野では、生成 AI は自然言語のクエリーを変換してさまざまなツールや異なるデータセットに対するクエリーを生成することにより、データ取得を最適化するのに役立ちます。

また、レポート生成プロセスの一部としてサマリーを作成し、高度なフィッシング攻撃のエミュレートを可能にすることで予防的セキュリティにも貢献できます。しかし、これはセマンティック分析であるため、LLM ではセキュリティ分析と一貫性のある検知に必要な推論が困難な場合があります。

責任を持って適用しないと、生成 AI は架空のデータを参照したり矛盾した応答を返したりする「ハルシネーション」による混乱を招くことがあります。これは異なるセキュリティチームメンバーによって記述されるプロンプトに確証バイアスが含まれるためです。

Darktraceの機械学習：複数のテクニックを組み合わせる

機械学習テクニックの各タイプそれぞれに強みと弱みがあるため、機能を強化しつつ 1 つの手法の弱点を克服するには、多層的な、複数の手法を組み合わせたアプローチが理想的です。Darktrace の Cyber AI テクノロジーは複数の機械学習アプローチで構成され、これらを組み合わせてサイバー防御を実現しています。

これにより Darktrace はコーポレートネットワーク、クラウドコンピューティングサービスおよび SaaS、IoT、産業用制御システム（ICS）、ならびに E メールシステムを含む組織のデジタルエステート全体を保護することができます。

多層的AI

Darktraceはさまざまなタイプの機械学習を組み合わせて、Darktrace ActiveAI Security Platformの製品全体に適用されるAIを構築しています。

組織のインフラとサービスにプラグインされたDarktraceのAIは、環境内のデータとその相互動作を取り込んで分析し、個別のユーザーやデバイスの詳細情報に至るまで、その環境の通常の動作についての理解を形成します。システムは「何が正常か」についての理解を、変化する証拠に基づいて絶え間なく更新します。

正常についての動的な理解を持つことにより、AIエンジンは異常かつ良性と考えられないイベントや動作を、非常に高い精度で特定できます。

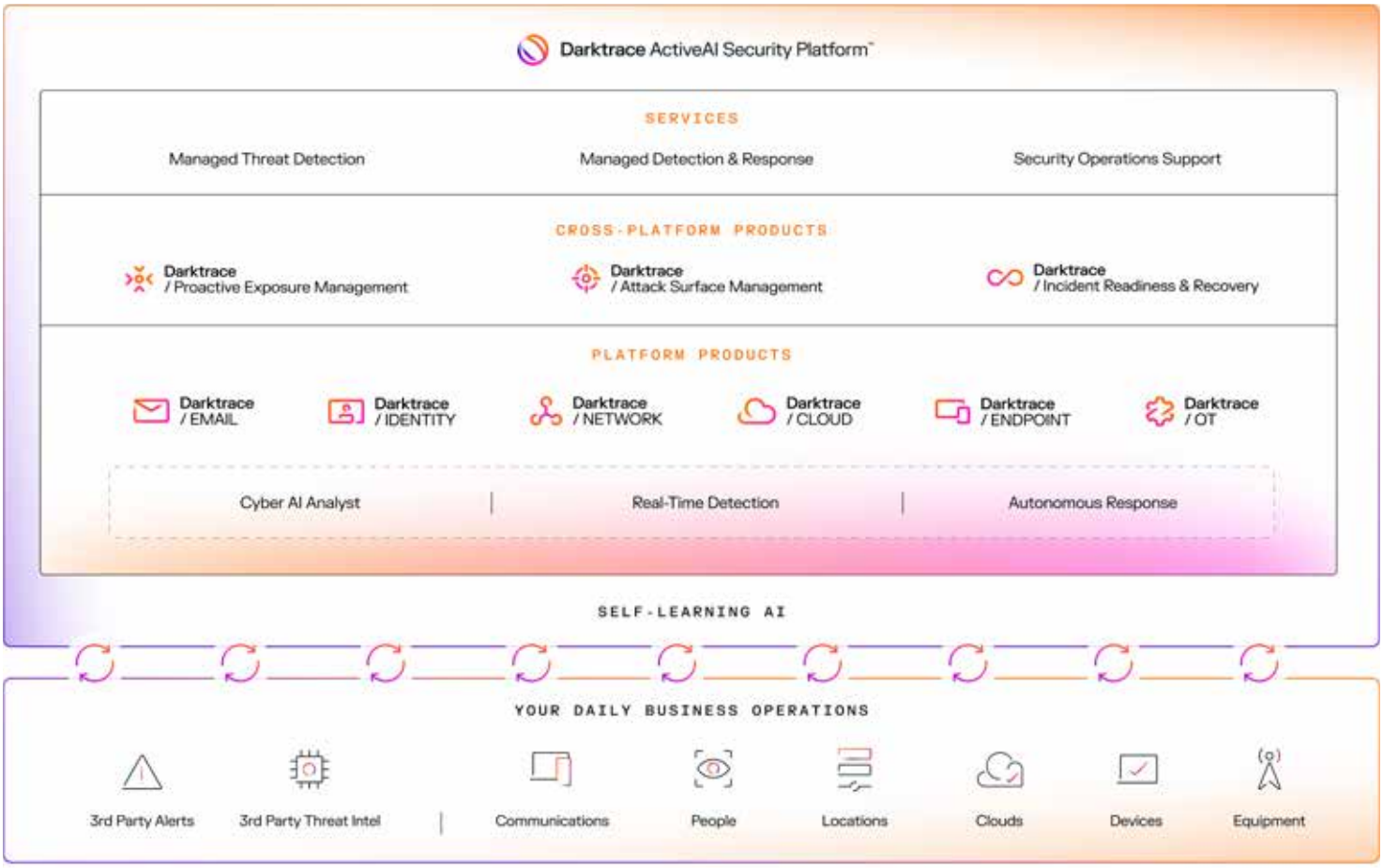


図 01: Darktrace ActiveAI Security Platformは、従来サイロ化していたデータセットすべてに渡って検知、遮断、およびプロアクティブなセキュリティを提供します。

攻撃者の最初の足跡を示すアクティビティを、事前の知識やインテリジェンスなしに識別できる能力は、今日の脅威アクターに対抗できるAIの有効性の根幹です。AIの支援により、人間のセキュリティチームは、日々の正常なデジタル処理の圧倒的なノイズの中から見つけ出すことが難しい、かすかな兆候を検知することが可能になります。

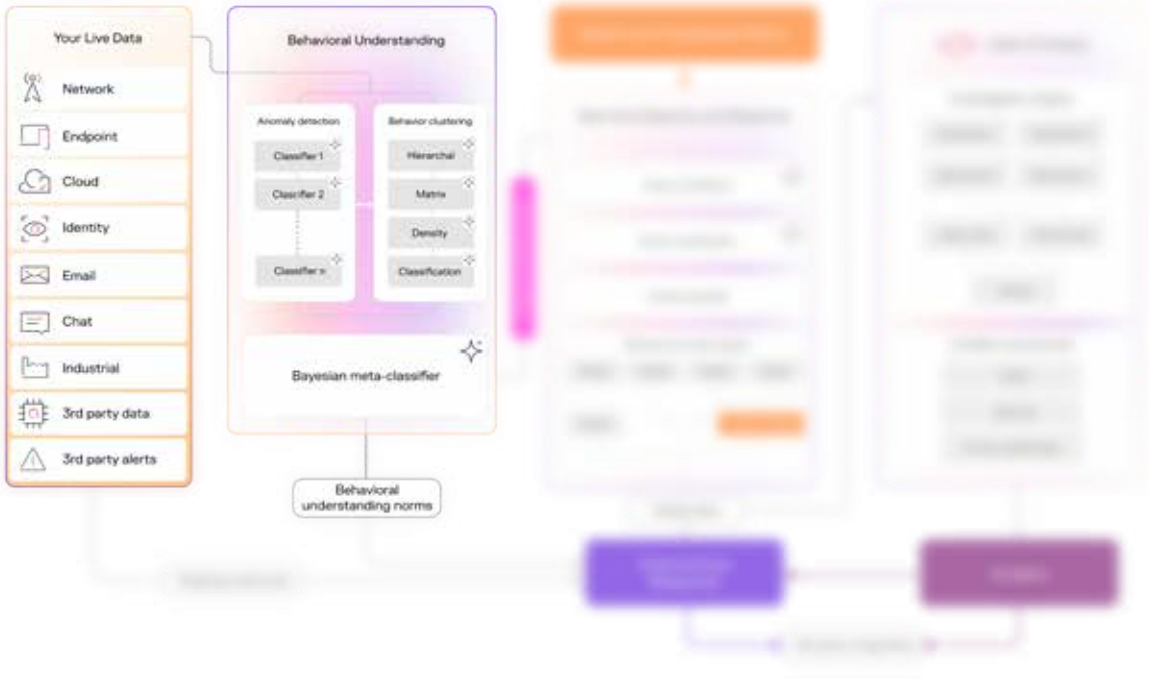
Darktraceにより検知される、しばしば見過ごされがちな脅威には次のようなものがあります：

- 内部関係者からの脅威 ー 悪意のあるもの、偶発的なもの
- ゼロデイ攻撃 ー これまでに見たことのない、新たなエクスプロイト
- 潜在的脆弱性、内部、外部を含む
- マシンスピード攻撃 ー 急速に伝播および/または変異するランサムウェアおよびその他の自動化された攻撃
- クラウドおよびSaaSベースの攻撃
- サイレントなステルス性の攻撃
- 高度なスピアフィッシング

こうした幅広い脅威に対し、Darktraceがどのように AIを使って検知および遮断を行っているか詳しく見ていきましょう。以下のセクションでは、DarktraceのAI、および教師付きと教師なし学習テクニックをどのように組み合わせているか、そしてダークトレースが展開するグローバルな環境からの情報を統合して脅威検知を強化しているかを具体的な例を使って説明します。

Darktrace / EMAIL、 / Attack Surface Management、 / Proactive Exposure Management 等、プラットフォーム製品の一部は若干異なるAIアーキテクチャを使っているものもあります。しかし、システム全体としては同じAIアプローチにより、複数の種類のAIを使ってそれぞれの環境を理解しています。

1.ビヘイビア予測



トレーニングデータ

Darktrace は組織固有のデジタル環境全体からライブデータを取り込んで継続的にトレーニングを行い、サードパーティデータおよびアラートと統合することにより、より大きな全体像を構築します。注目すべき点は、こうしたタイプのデータを使うことにより、セキュリティがそれぞれ固有の環境に合わせたものになることです。他の AI ベースのツールは大規模なデータレークから学習を行って汎用的なモデルをトレーニングするため、過剰な単純化や仮定がおこりがちです。それぞれの組織独自のデータでトレーニングされる Darktrace AI は、多数の企業のデータを比較してそれらとの動作の違いを推定しているのではありません。そうではなく、それぞれの企業専用の AI システムを構築することにより、高精度な検知を実現します。

ベイズ確率手法

デジタルアクティビティに対する多数の分析を組み合わせるため、Darktrace は新しい情報や情報の変化に応じて更新されていく、ベイズ確率モデルを使用しています。ベイズ確率手法には、過去の情報と現在の情報両方に対して確率的モデリングを適用する、教師なし機械学習テクニックが含まれています。確率的モデリングを絶えず調整することにより、アセット、ピアグループ、そして組織全体に対する「生活パターン」が作成されます。

Darktrace はベイズ確率論によりそれまで未知であった関係を見つけ出し、独立してデータを分類することができます。ラベル付けのされていない未構成のデータからも、何が正常で何が異常かを示す説得力のあるパターンを検知することが可能です。このテクニックはその組織固有の動作についての継続的な学習、および仮説の修正を可能にします。新しいデータに照らして脅威レベルを継続的に再計算することで、Darktrace は従来のシグネチャベースの手法ではカオスにしか見えなかったデータフローの中から、攻撃を示す明確なパターンを識別することができます。

クラスタリングアルゴリズム

エンティティまたはアセットにとって何が正常と見なされるべきかをモデル化するため、Darktrace はその動作をネットワーク上の類似のエンティティとの脈絡で分析します。Darktrace は教師なし機械学習を使って意味のあるグループをアルゴリズム的に識別します。これは人手では困難かつコストのかかる作業です。

環境内の関係の全体像を作成するため、Darktrace は行列ベース、密度ベース、および階層的なクラスタリングを含む、様々なクラスタリング手法を適用します。その結果作成されるクラスタを使って、意味のあるグループ分けをアルゴリズム的に識別し、正常な動作のモデルに情報を与えます。

継続的なクラスタリングは疑わしいアクティビティが明らかに悪意のあるものとなる前にそれを識別するだけでなく、既に存在していた侵害も識別することができます。限定されたトレーニング期間を必要とする他のソリューションは、悪意ある動作であってもそれをベースラインとして学習してしまう可能性があります。一方、クラスタリングはアクセス権、ロール、ID、機能をピアグループと比較するため、他のデバイスが同じタイプの動作を行っていないときにそれを認識することができます。

"インストールして1週間たたないうちに、Darktraceは私達がまったく知らなかった脅威や脆弱性について通知してくれました。"

■ **前CIO**

Bunim/Murray Productions

異常検知モデルとベイズメタ分類器

Darktraceの多層的AIは、デジタルエステート全体のデータを特性づけるわずかに異なるレベルの証拠を区別することにより、あいまいさを把握します。単純に二極的な‘悪性’または‘良性’の出力だけでなく、これらの数学的アルゴリズムはリスクスコア、レアリティスコア、特徴量重要度を含む、脅威の可能性の度合いをマークした出力が可能です。これによりシステムのユーザーはアラートを厳密にランク分けし、最も緊急に対処する必要があるものを優先することが可能となります。

Darktraceはその中核においてデバイスの動作の次のような様々な指標の分析に基づき、「正常」と見なされる動作を数学的に特性付けます：

- プロトコルの使用
- 接続の頻度
- ピアツーピア通信
- アプリケーション使用パターン
- 外部通信パターン
- ポートアクティビティ
- トラフィック暗号化
- セッション継続期間
- 周辺デバイスの使用
- DNS要求パターン
- Eメールアクティビティ
- 疑わしいエンティティとのやりとり
- 多要素認証（MFA）の使用

ほとんどの異常検知モデルは、適切なデータタイプに適用された場合にのみ正確な結果を出します。このことは、1つの異常検知機械学習モデルをすべてのデバイスおよびすべてのデータタイプに適用することは正確性の問題が生じるためできない、ということを意味します。

"リモートワークにより新しい業務モードにシフトしていく中でも、Darktraceは全員がキャンパス内で仕事をしていた時と同じ機能をすばやく提供してくれました。"

■ **前CIO**

Salve Regina University

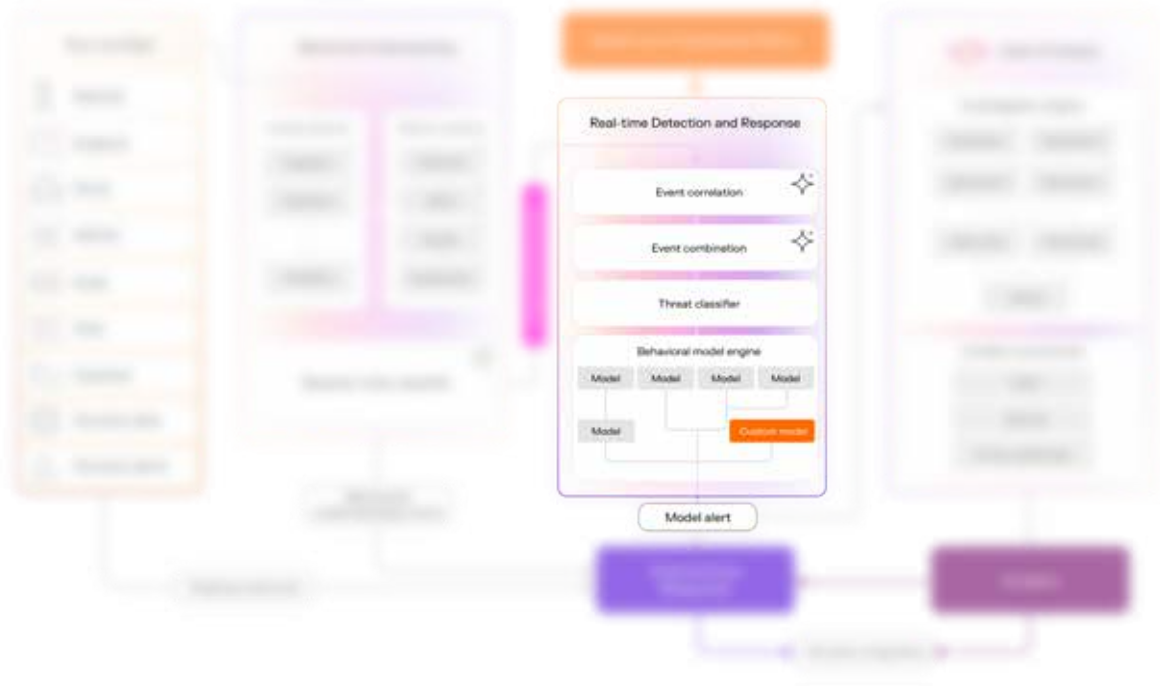
それを踏まえて、Darktraceはアンサンブル的なアプローチを取り、異常検知を何百ものメトリックに渡ってスケール化します。その上でベイズメタ分類器を適用し、基盤となる教師なし機械学習モデルをファインチューニングし、継続的な再調整を行うことで継続的な学習を可能にしています。

ベイズメタ分類器は、複数の分類器からの出録を組み合わせて全体の出力を生成するとともに結果に磨きをかけます。たとえば、ある分類器が過剰に発動している場合、その環境または取り込んでいるデータに対して正確でないことを意味します。それに対応してベイズメタ分類器はそれを使わないようにし、その特定の環境についてのアラートに対する重要性の重みを下げます。

その結果、それぞれの運用環境において特定の分類器がより強調され、その組み合わせは環境に合わせた独自のものとなります。

これは、ジャーナリズムにたとえて考えることができます。ソースが多様であるほど、すぐれた確証と正確性が得られるということです。異常検知モデルをベイズメタ分類器と組み合わせることにより、AIの出力はより協調的かつ適応型となり、より精度の高い結果を得ることができます。

2.リアルタイムの脅威検知



確率モデルと決定木モデル

イベントの相関づけと組み合わせが実行され、関連するアクティビティのより広い範囲を確認します。正常についての真の理解に基づき、DarktraceのAIエンジンは異常なイベントとリスクの高い振る舞いを結びつけます。異常な振る舞いはMITRE ATT&CKフレームワークに照らして評価され、異常に対して脅威のコンテキストが与えられます。

3.カスタマイズ可能なモデルエディター

処理をコントロールするためのロジック

検知のための機械学習モデルの上に、脅威分類とコンテキスト化を行うためのカスタマイズ可能な論理ステートメント、条件、モデルのレイヤーが存在します。この“Model Engine”レイヤーは、セキュリティチームに対して AI エンジンの処理に対する可視性と、出力をカスタマイズする機会を提供するものであり、説明可能性、解釈可能性、コントロール性、および監査により AI 出力の運用を変更する能力を高めることができます。モデルエディターには標準モデル（広範に更新があります）が付属していますが、セキュリティチームはこの論理レイヤーを編集することにより、処理を行う AI エンジンをコントロールすることができます。モデルエディターを使って、セキュリティチームは値、優先度、しきい値、アクション等を指定することができます。これにより、ユースケースやビジネスの必要性に基づいたカスタム検知を作成できます。また、新しい検知モデルを作成したり、既存の検知モデルの優先度を高めるなどして、AI の挙動を変更し、コントロールすることが可能です。

たとえば、セキュリティチームによっては、モデルエディターを使用して既存の機能をカスタマイズし、デバイスのプロファイリング、タグの割り当て、設定ミスの特定や自律遮断のトリガーなどを行わせているケースもあります。これらのカスタムモデルを使って、Darktrace を既存のセキュリティプロセスに統合する、あるいはSOCプレイブックを再現することができます。

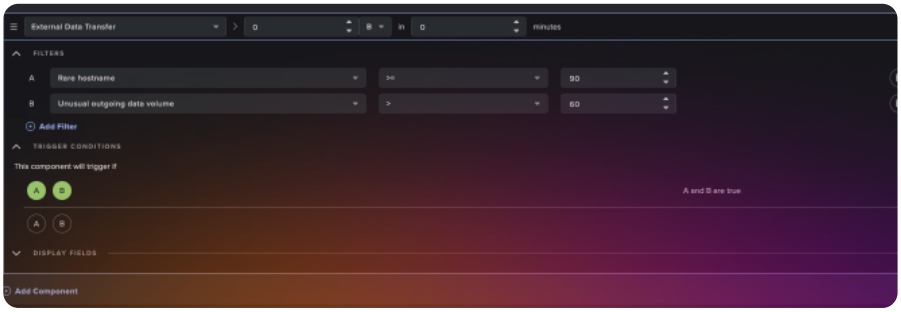


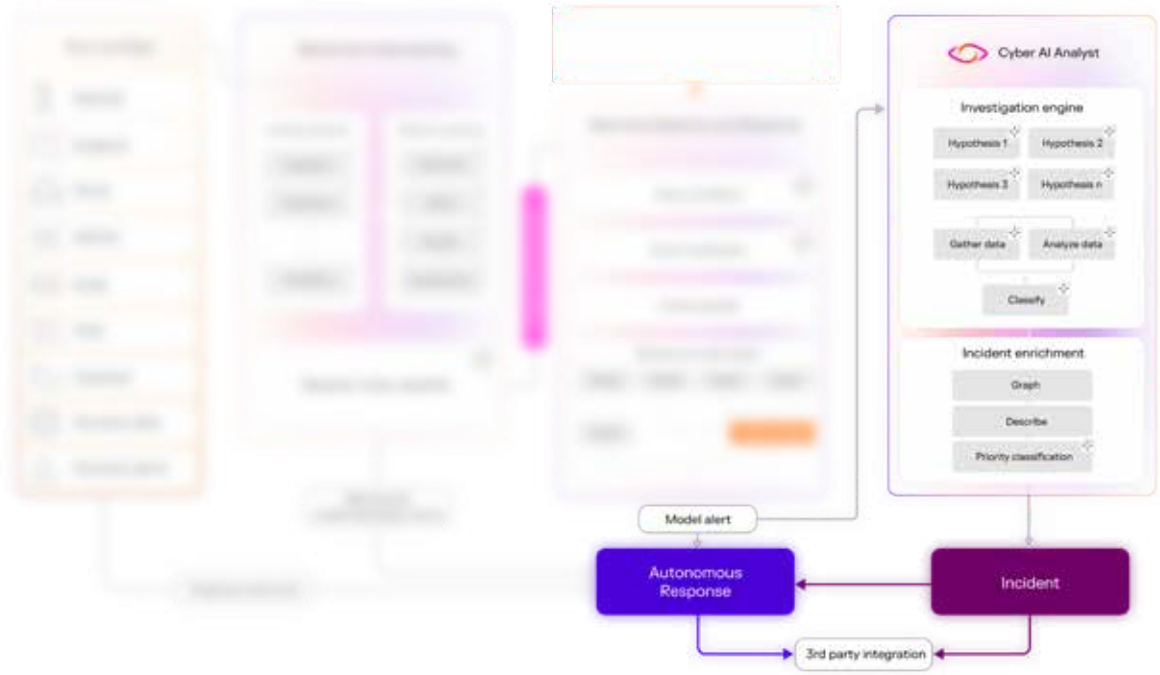
図 02: アラートは、モデルエディターを使って、レアリティや特異性のスコアに対するしきい値の編集を含め、さまざまな方法でカスタマイズすることができます。

また、モデルエディターでは他のAIベンダーが使用しているような一般公開されているモデルガーデンの持つリスクである、モデルの改ざんリスクを回避できます。これは、モデルが公開の場所でホスティングされていないためであり、また属性に基づいた変更管理機能があること、そしてモデルロジックに透明性と説明可能性があることにより、ユーザーは変更が行われたかどうかを確認できます。

モデルのカスタマイズを行わない場合も、モデルエディターを通じてユーザーが AI の意思決定に対する可視性と理解を持つことが可能です。AI を使った各機能は、明確に示されたスコアや説明とともにその結果を提示します。たとえば、Darktrace は特定のアラートをトリガーしたしきい値や、AI が行った調査の詳細を示し、どのようにその結論に至ったかを確認することができます。

AIサイバーセキュリティツールの解釈可能性の高さは、調査や修正のアクションを進めるのに役立つ多くの情報を提供し、人間のセキュリティチームのレベルおよびスキルの引き上げに貢献します。また、ROIを最大化する上で不可欠である、人間のセキュリティチームとAIツールの間の信頼関係を構築します。

4.調査と対処



Cyber AI Analyst

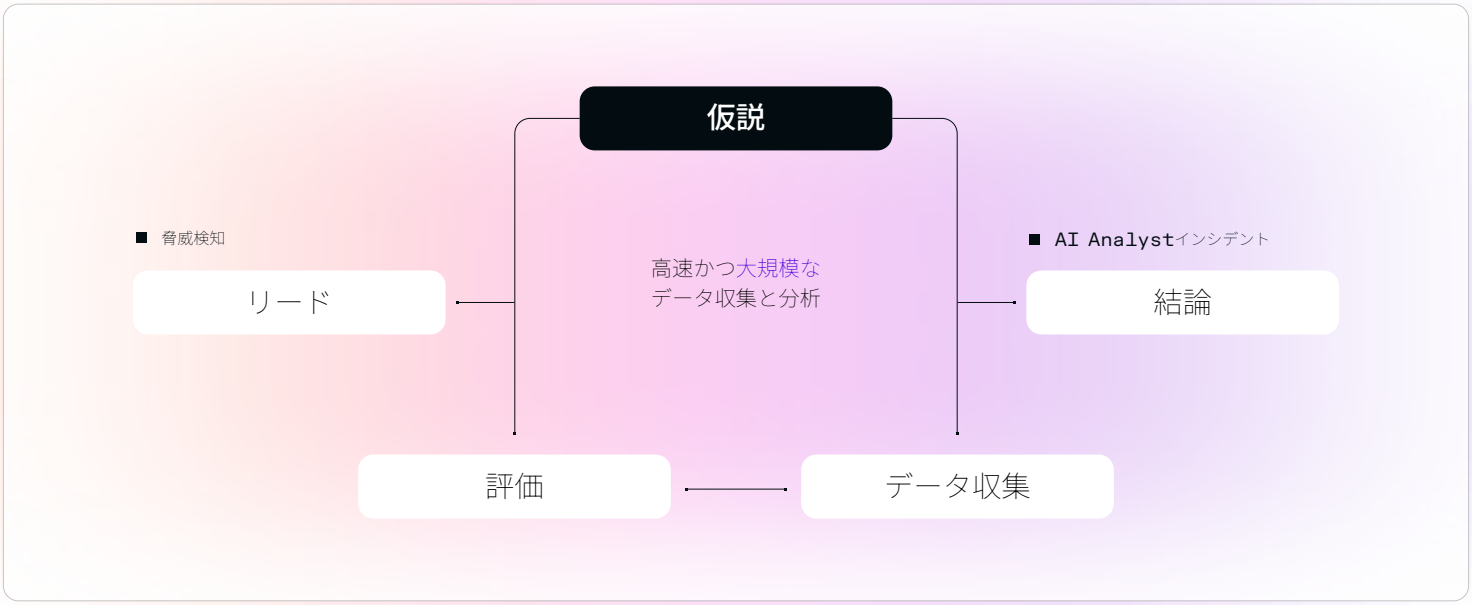
Darktrace は多様な機械学習テクニックを使って、モデルにより見つかった異常の詳細な分析と調査を行い、とりわけ SOC チームのレベル 1 およびレベル 2 の作業を自動化します。これは調査ワークフローの中で行われる繰り返しの、時間のかかるタスクを自動化するのに役立ち、セキュリティチームは時間とリソースを節約することができます。この機能を私達は Cyber AI Analyst と呼んでいます。

この調査AIは関連のあるすべてのアラートを詳しく調べます。教師なしおよび教師付き機械学習を使用し、そのトレーニングデータのうち1つのソースは、ダークトレースのエキスパートアナリストの考察および行動から構成されています。

こうして、Cyber AI Analystは人間が脅威調査プロセスを実施するやり方を再現しています。Cyber AI Analystは次を行います：

- 関連するアラートに対して1つまたは複数の仮説を調査する
- 関係したアセットのログを分析する
- グラフ理論分析を使って異なるドメイン間でインシデントの相関づけを行う
- 仮説に対する分析を評価し、仮説を確認、反論、変更（および必要に応じて分析を反復）する
- 侵害のライフサイクル全体を追跡したよりハイレベルのインシデントの一部として、詳細な技術的情報を含むアラートを生成する
- 優先度スコアおよびAIがなぜその結論に至ったかについての説明を含む調査サマリーを作成する

Cyber AI Analystのプロセス：



調査プロセスの初期段階において、Cyber AI Analystは何が起こっているかについての幅広い仮説を少なくとも1個立て、その後カスタムアルゴリズムおよびその他の機械学習テクニックを使って、この情報に対して人間が行うように問い合わせと分析を行います。

たとえば、LLMを使用してホスト名の目的を分析し、教師付き機械学習を使用してホスト名がドメイン生成アルゴリズム（DGA）を使って生成されたものかどうかを識別する、あるいは教師なし機械学習を使って通常の通信のパターンとの比較を行うなど、さまざまな処理が行われます。

また、Cyber AI Analystはグラフ理論を使って組織のデジタルエステート内のアセット間の関係をマッピングします。これにより数学的計算が行われ、アセットの重要性、頻度、類似性などの重み付けが行われます。

セマンティック分析をすべてのタイプのデータに適用する生成AIベースのチャットボットソリューションとは異なり、Cyber AI Analystは特定のタイプのデータを評価するための特定のモデルを使用します。その結果、何千ものデータポイントに基づいてインテリジェントに判断を行い、自動的にアラートを調査することで時間を節約することができます。

Cyber AI Analystは通常、年間にして最大30名の追加のフルタイム従業員がレベル2分析を行い手動でレポートを作成するのに相当する能力を提供します。詳細な情報を含むハイレベルのインシデントのアラートを作成することで、人間のアナリストがLevel 3のタスクに集中できるようにし、セキュリティオペレーションを増強します。

調査が完了しCyber AI Analystが脅威を理解すると、これらの結果を教師付き機械学習によって分類し、AIによってのみ可能なスピードおよび規模で目的のインシデントを見極めます。

これにより、あらゆるレベルのアナリストの調査作業を、より効率的でシンプルなものにすることができます。また、セキュリティチームがより高価値な戦略的作業、たとえばリスクの管理や業務全般に渡る改善への取り組みなどに集中するために必要な、貴重な時間を確保することができます。

Cyber AI Analystは人間が見逃してしまう、または識別するための時間がないような詳細な事象をしばしば検知し、最初の仮説が成り立つかどうかを数分の間に判断することができます。さらに重要な点として、このテクノロジーはこれらの調査の結果を分類および保存し、任意の時点において優先度の高い少数のインシデントのみを表示することにより、アラート疲れを軽減することができます。

調査結果と推奨される対処をユーザーインターフェイスで伝えるとともに、任意の時点においての優先度の高いいくつかのインシデントのみを自然言語で説明した、詳細なPDF形式のレポートを出力することもできます。また、他の幅広いシステムやサービスと統合することのできるアラートを生成します。これらのアラートおよび調査結果に対してはその後、コンテキスト情報やセキュリティ上の考察が追加され、エグゼクティブもエンドユーザーも同様にレビューし、アクションを取ることが可能になります。

重要な点は、Cyber AI Analystは新たな前例のない状況にその場で適応できるため、ユーザーは多数のアラートの処理に時間をかけるのではなく、意味のある戦略的な仕事に時間をかけ、優先して処理することができるようになることです。

自律遮断

ダークトレースの多層的な AI は、既知か新手かに関わらずサイバー攻撃を検知できるだけではありません。それらを遮断することもできます。Darktrace ActiveAI Security Platform の中核的機能である自律遮断機能、Autonomous Response は、進行中の攻撃をマシンスピードで自律的に無害化するためにとるべきベストなアクションを計算します。

その結果、セキュリティチームが調査と修正を行うまでの間サイバー攻撃を封じ込めることができます。24 時間、週 7 日、人間のセキュリティチームがオフィスにいない間も、組織を確実に保護することができます。

Autonomous Response は前述の検知機能と直接連携しています。デバイスおよびそのグループに対して導き出された「正常」からの著しい逸脱、特定の悪意ある兆候または望ましくないアクティビティ、あるいは小さいが意味のあるインジケータおよび期待された動作からのかすかな逸脱の組み合わせによって、モデルアラートや Cyber AI Analyst からトリガーすることができます。

このテクノロジーは、教師なしモデリングおよびクラスタリングが判断する「生活パターン」による検知、およびユーザーの関係、アクティビティのパターンおよびユーザーの意図測定するさまざまな最先端の教師なし分類器によって支えられています。

Autonomous Response は程度に見合ったピンポイントの対処を生成し、日々の業務を中断することなく正確なアクションを実行することができます。アクションは脅威の発生源に的を絞ったものであり、必要な場合にのみエスカレーションされます。たとえば、Autonomous Response は普段とは異なる FTP 接続を切断する、あるいは異常な IP 範囲からの Office 365 へのアクセスをブロックする、などの対処を行うことができます。さまざまな AI テクニックとの組み合わせにより、このソリューションは、観測したデータおよびそれ自身から受動的に学習します。たとえば、Autonomous Response がアクションを生成することにより、フィードバック強化ループがトリガーされます。こうした使い方により、Cyber AI テクノロジーは人間の役割を置き換えるものではなく、強化するのに役立ちます。Autonomous Response が人間よりも速くアクションを起こすことにより、セキュリティチームが事態に追いつくための貴重な時間が生まれます。

> “Darktrace の提供する保護により、私達にとって可能な限りの防御を実現することができています。将来、サイバーインシデント発生の有無で組織が判断されることはなくなるでしょう。なぜならそれはほぼ不可避だからです。その代わりにインシデントからどのように復旧したかで判断されるようになるでしょう。その点で当社はダークトレースとそこで働く人たちが私達についていてくれるのだという自信があります。”

■ **ITマネージャー**
Bristows

> “サイバー攻撃の対象がデスクトップPCやサーバーに限られている時代は過ぎ去りました。Darktraceの機械学習は、戦いが始まる前に相手を撃退することができます。"

■ **CIO**
ラスベガス市

高速かつ的を絞ったアクション

既知および未知の攻撃を数秒で遮断

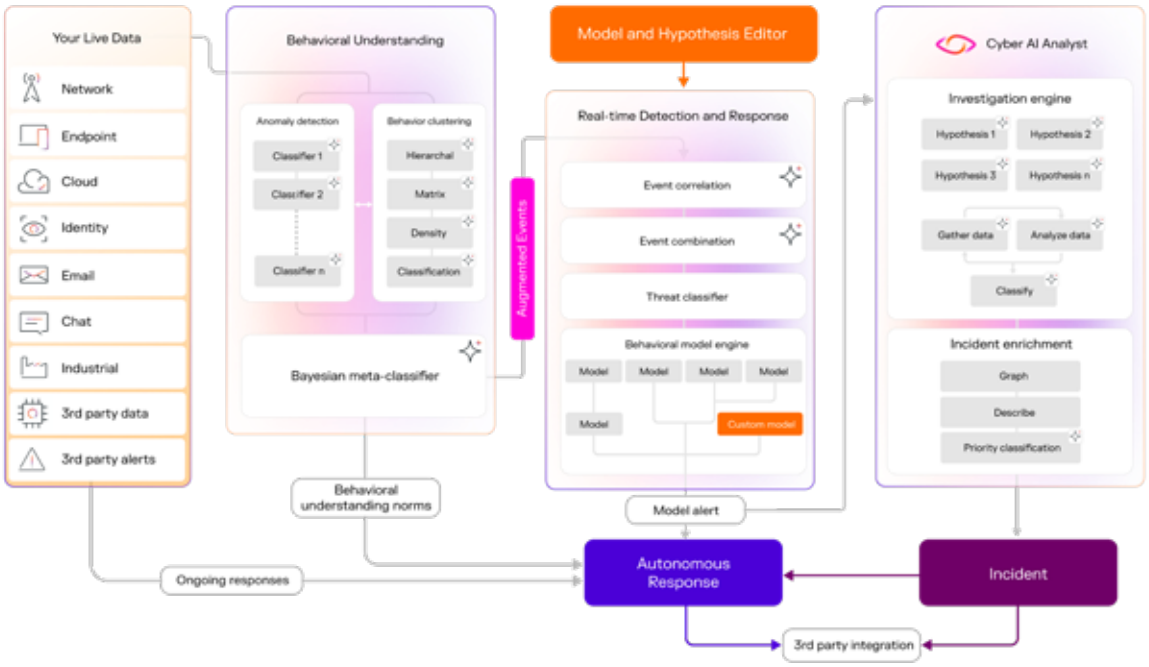
完全にカスタマイズ可能

- 完全に自律的に機能、または人間の確認がある場合にのみアクションを実行、あるいはその組み合わせが可能
- APIを介してファイアウォールや他のセキュリティツールと統合し、カスタマイズしたアクションを実行可能

あらゆるデジタル環境を保護

Darktrace ActiveAI Security Platform の中核的構成要素としてネットワーク、E メール、クラウド、アイデンティティ、エンドポイント、OT（Operational Technology）に渡り適用可能

各種AIレイヤーの統合



Darktrace の多層的な AI は一体となってビヘイビア予測、リアルタイム脅威検知および遮断、インシデント調査を実現すると同時に、可視性とコントロールでセキュリティチームを補強します。

AI の時代を迎え、私達の仕事のやり方にも大きなシフトが起こっています。ビッグデータを扱い膨大な計算を行う自動化および自律型ソリューションにより低価値な繰り返しのタスクを引き受けることのできる、新しいテクノロジーが広範に普及しています。AI が攻撃者の手に渡り、また企業のデジタルインフラの範囲と複雑性が増大する中で、セキュリティチームは急激に変貌するサイバー脅威ランドスケープへの対応という問題に直面しています。

従来のセキュリティ手法では、簡単な機械学習を一部使用するものを含めても、もはや不十分です。これらのツールはあらゆる攻撃ベクトルについていくことも、多種多様なマシンスピード攻撃にすばやく対応することもできません。これらは既知のパターンや予測されるパターンと比べて複雑すぎるからです。セキュリティチームは検知能力を一段高める必要があります。機械学習を使って環境を理解し、雑音をフィルタリングし、見つかった脅威にアクションを取らなければなりません。

Darktrace の多層的な AI テクノロジーは、ネットワーク全体を理解し、様々なレベルの動作を観測し、弱点となる可能性のある部分を検知しようとするセキュリティチームにとって重要なツールとなっています。機械学習テクノロジーは技術革新を続けるハッカーや内部関係者の脅威からシステムを守り、未知のサイバー攻撃手法への対応を形成する強い味方なのです。これはセキュリティチームに力を与えるよう設計された、サイバーセキュリティにおいて非常に大きな意味のある変化です。

“ダークトレースのテクノロジー、経験、および専門知識は、当社がサイバー攻撃の一歩先を行き、リスクを最小化し、セキュリティチームの生産性を増強するのに役立っています。”

■ **CISO**
Severfield

Darktraceが個別の環境においてどのようにAIを使用しているかについてのより詳しい情報は、以下のソリューション概要をご覧ください。

Darktrace / NETWORK

Darktrace / CLOUD

Darktrace / EMAIL

ダークトレースにお問い合わせの上、是非パーソナライズされたデモンストレーションをお申込みください

■ **ダークトレースについて**

ダークトレースはAIサイバーセキュリティのグローバルリーダーであり、日々変化する脅威ランドスケープに立ち向かう組織を支援しています。2013年に英国ケンブリッジで設立されたダークトレースは、それぞれのビジネスからリアルタイムに学習するAIを使用して未知の脅威から組織を保護する、必要不可欠なサイバーセキュリティプラットフォームを提供しています。ダークトレースのプラットフォームおよびサービスは2,400名を超える従業員により支えられ、世界でおよそ10,000社の組織を保護しています。より詳しい情報については、http://www.darktrace.com/jaをご覧ください。

北米：+1 (415) 229 9100 **ヨーロッパ：** +44 (0) 1223 394 100 **日本：** (03) 5456 5537 **ラテンアメリカ：**+55 11 4949 7696

darktrace.com/ja | japan@darktrace.com