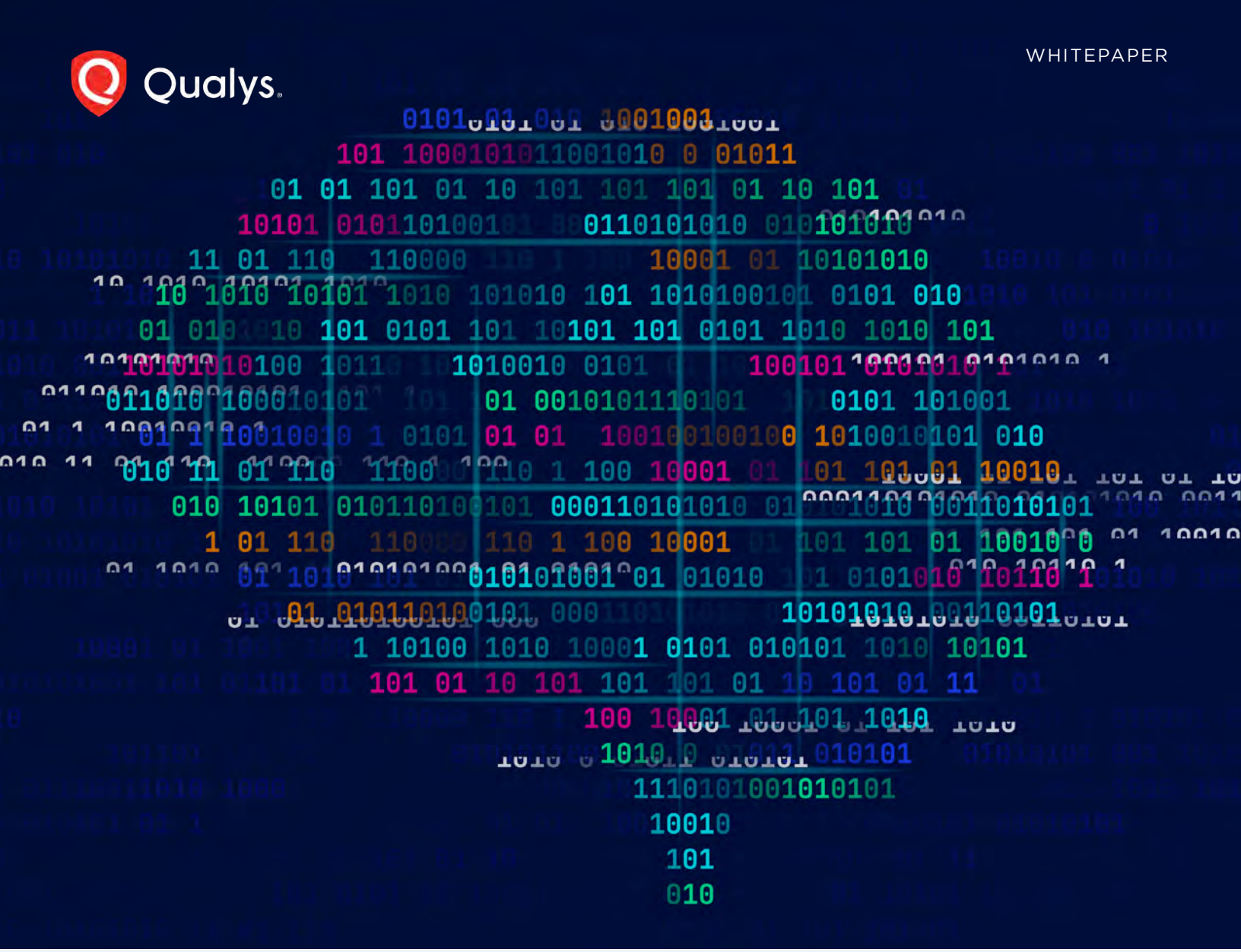
Qualys

WHITEPAPER



企業におけるLLMと生成AIの未来の確保

目次

エグゼクティブサマリー

組織による大規模言語モデル（LLM）や生成AIの急速な導入に伴い、堅牢なセキュリティおよびガバナンスの枠組みを構築することが最優先事項となっています。最近のガートナー社のレポートによると、CISOの71%がAI関連の脆弱性に懸念を抱いており、AIのプライバシーやセキュリティに関する責任は、IT部門（82%）や情報セキュリティチーム（66%）を中心に組織全体で共有されるようになっています。こうした状況下で、統一されたアプローチの必要性は明らかです。本資料では、AIイノベーション、サイバーセキュリティ、ガバナンスの相互関係を考察するとともに、新興技術特有の課題に対応する包括的ソリューション「Qualys TotalAI」を紹介します。
AIがもたらす変革の可能性や攻撃対象領域（アタックサーフェス）の拡大、そして企業環境を保護するために求められる高度なセキュリティ対策について検証します。主要なセキュリティ課題、規制への対応、AIガバナンスのベストプラクティスに関する知見を提供し、CISOが複雑なAIセキュリティおよびコンプライアンスの状況を効果的に管理・対応できるよう支援します。

企業環境におけるLLM（大規模言語モデル）と生成AIの台頭

変革をもたらす可能性と導入の動向

LLM（大規模言語モデル）や生成AIは産業に革命をもたらしており、近年の予測では、これらの技術が2030年までに世界経済に15兆7,000億ドルの経済効果をもたらす可能性があるとされています。特筆すべき点として、すでに54%以上の組織が4つ以上のAIシステムやアプリケーションを導入しています。こうしたAIの導入は、主要な分野全体でイノベーションを推進しています。

ヘルスケア: 診断精度の向上、創薬の加速、そして個別化医療の実現。

金融サービス: AIアルゴリズムの活用による、不正検知、リスク評価、および顧客サポートの改善。

製造業: サプライチェーンの最適化、予知保全の実現、そして製品設計プロセスの効率化。

小売業: 高度なチャットボットの活用や、一人ひとりに合わせたショッピング体験の提供。

こうしたAIの急速な導入は、単なるトレンドにとどまりません。それは、デジタル時代における企業の事業運営、イノベーション創出、そして競争のあり方を根本から変革しつつあります。

> **15兆7000億ドル**
> 2030年までにAIが世界経済にもたらし得る貢献
> Source: PWC

拡大する攻撃対象領域と新たなセキュリティのパラダイム

AI導入には大きなメリットがある一方で、攻撃対象領域が拡大し、従来のサイバーセキュリティ対策では完全には対処できない新たなセキュリティリスクも生じています。IAPP-EYの「年次プライバシー・ガバナンス・レポート」によると、プライバシー関連業務における最優先事項として、AIガバナンスの重要度がわずか1年で9位から2位へと急上昇しました。拡大した攻撃対象領域には、以下のようなものが含まれます。

AIモデルの脆弱性: 敵対的攻撃やその他の手法によって悪用され得る弱点。

データパイプラインのリスク: AI開発におけるデータ収集、前処理、および学習の各段階に潜む脅威。

APIエンドポイント: 「AI-as-a-Service」の提供やサードパーティとの連携に伴う、外部への露出拡大。

クラウドインフラとワークロード: AIワークロードやデータストレージの分散的な性質に起因するリスク。

こうした新たな攻撃経路の出現により、セキュリティ戦略にはパラダイムシフトが求められています。すなわち、従来の境界型防御から、AIの特性を考慮した、より包括的なセキュリティ体制への転換が必要です。

AIを取り巻く環境における主要なセキュリティ課題

1. データのプライバシーと機密情報の漏洩

LLM（大規模言語モデル）や生成AIにおける大きなリスクの一つは、機密情報が意図せず外部に漏洩することです。
実際、データ関連のリーダーの55%が、これを最大のセキュリティ懸念事項として挙げています。パブリックなLLMに入力されたデータは、しばしば組織の管理下を離れてしまうため、このリスクはさらに増大します。主な課題は以下の通りです。

意図しない情報の記憶（記憶化）：LLMは、学習データに含まれる機密情報を記憶し、それを再現してしまう可能性があります。

プロンプトを介したデータ漏洩: ユーザーが意図せず機密情報を入力してしまい、その情報が出力結果に現れる恐れがあります。

推論攻撃: 巧妙に作成されたクエリ（質問）を用いることで、AIモデルから機密情報を抽出される可能性があります。

こうしたリスクを軽減するために、組織は強固なデータガバナンスを確立し、データの匿名化技術を活用するとともに、データへのアクセスや出力内容をきめ細かく制御できるセキュアなAIプラットフォームを導入する必要があります。

> **データ関連のリーダーの55%が、**
> 機密情報の意図しない漏洩を最大のセキュリティ懸念事項として挙げています。
> Source: Securitymagazine.com

2. 敵対的攻撃とモデルの操作

敵対的攻撃は脅威として増大しており、リーダーの52%がその影響について懸念を表明しています。
こうした攻撃はモデルの完全性を損ない、誤った出力や悪意のある出力を生成させる可能性があります。
主な敵対的攻撃の種類は以下の通りです。

回避攻撃: 入力を操作し、誤分類やエラーを誘発させる攻撃。

ポイズニング攻撃: 学習データセットに悪意のあるデータを混入させ、モデルの性能を低下させる攻撃。

モデル反転: モデルのパラメータから、機密性の高い学習データを復元する攻撃。

これらのリスクを軽減するには、堅牢な学習手法、入出力の継続的な監視、敵対的攻撃検知システムの導入など、多層的な防御策が必要です。

> **リーダーの52%が、**
> 敵対的攻撃に対する懸念を表明しています。
> Source: Immuta.com

3. モデルの盗用および知的財産に関するリスク

アルゴリズムや学習データといったAI資産の保護は、知的財産を守り、競争上の不利益やデータ侵害を防止する上で極めて重要です。主なリスクには以下のようなものがあります。

モデル抽出: 繰り返しクエリを送信することで、独自のモデルを再構築すること。

学習データの盗用: 価値ある学習データセットへの不正アクセスを行うこと。

アルゴリズムのリバースエンジニアリング: モデルの挙動から独自のアルゴリズムを推測すること。

これらのリスクを軽減するために、組織は厳格なアクセス制御を実施し、保存中および転送中のモデルやデータを暗号化するとともに、不審なクエリパターンを検知する監視システムを導入する必要があります。

4. プロンプトの挿入と入力操作

悪意のある攻撃者は、巧妙に作成されたプロンプトを用いてLLMの脆弱性を悪用し、不正な操作やデータ漏洩を引き起こす可能性があります。この新たな脅威は、AIセキュリティ戦略において特に注意を払う必要があります。主なリスクは以下のとおりです。

コマンドインジェクション: 悪意のあるコマンドを埋め込んでモデルの動作を操作する。

コンテキスト操作: 会話のコンテキストを操作して意図しない応答を引き起こす。

ジェイルブレイク: モデル内の倫理的制約やコンテンツフィルタを無効化する。

これらのリスクを軽減するために、組織は堅牢な入力検証を実施し、プロンプト設計のベストプラクティスを適用し、AI対応のセキュリティツールを活用して悪意のあるプロンプトを検出・ブロックする必要があります。

5. コンプライアンスおよび規制上の課題

AI技術の進展に伴い、組織はますます複雑化する規制環境に直面しています。EUの「AI法（AI　Act）」のような新たな枠組みでは、コンプライアンスを確保するために、先を見据えたガバナンス体制の構築が求められています。主な検討事項は以下の通りです。

データ保護: GDPR（EU一般データ保護規則）やCCPA（カリフォルニア州消費者プライバシー法）などのプライバシー規制の遵守。

アルゴリズムの透明性: AIによる意思決定における説明可能性と公平性の確保。

業界固有のコンプライアンス: 医療や金融など、各業界におけるAI関連規制への対応。

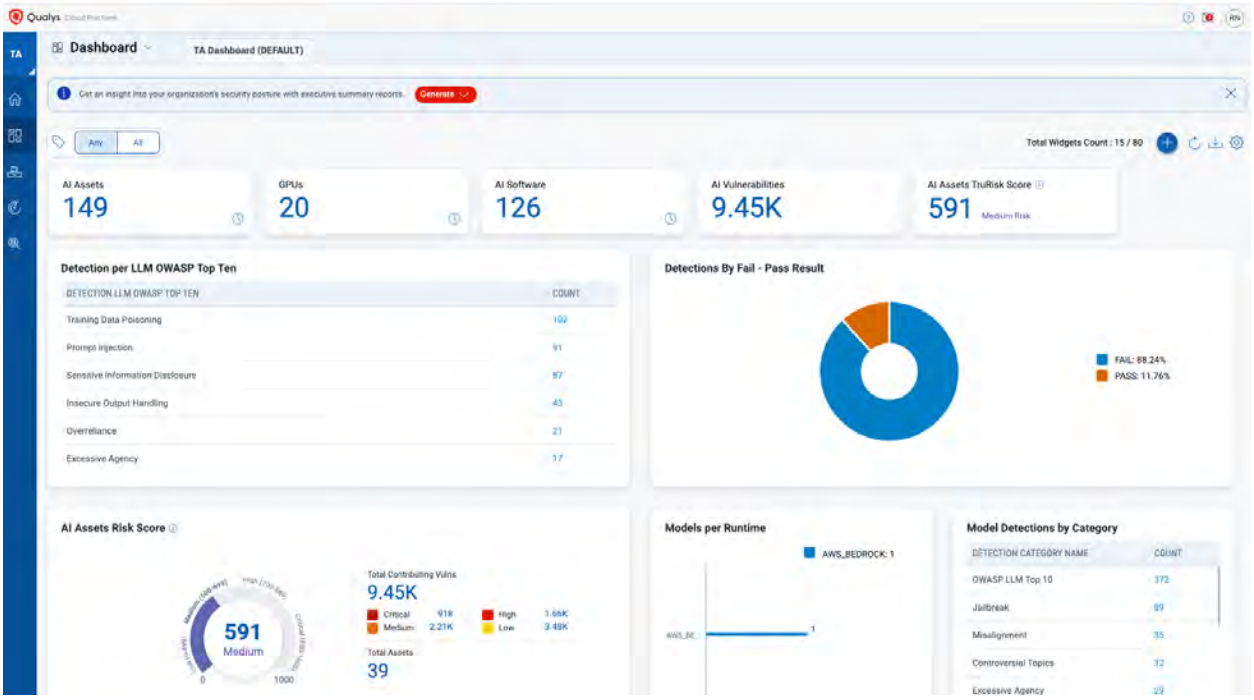
規制を遵守し続けるためには、組織は現行の規制に適合しつつ、将来的な変化にも柔軟に対応できる、包括的なAIガバナンスの枠組みを構築する必要があります。

Qualys TotalAI: 包括的なAIセキュリティとガバナンス

包括的な検出と脆弱性評価

Qualys TotalAIは、業界をリードする当社の資産可視化および脆弱性検出技術を基盤とし、AIおよびLLM（大規模言語モデル）資産に対する包括的な検出とセキュリティ機能を提供します。AI資産の検出を自動化することで、環境全体にわたるモデル、データセット、およびそれらを支えるインフラストラクチャを継続的に特定・スキャンします。これにより、組織は自社のAIエコシステム全体を完全に可視化できるようになります。

AIに特化した最適化された脆弱性スキャンは、従来の評価手法の枠を超え、モデルの脆弱性やインフラストラクチャの不備など、AIコンポーネント特有の設定ミスやリスクを明らかにします。
さらに、高度な分析機能を組み合わせることで、AIシステムへの潜在的な影響度に基づいて脆弱性の優先順位を付け、セキュリティチームが最も重大なリスクへの対応に注力できるよう支援します。



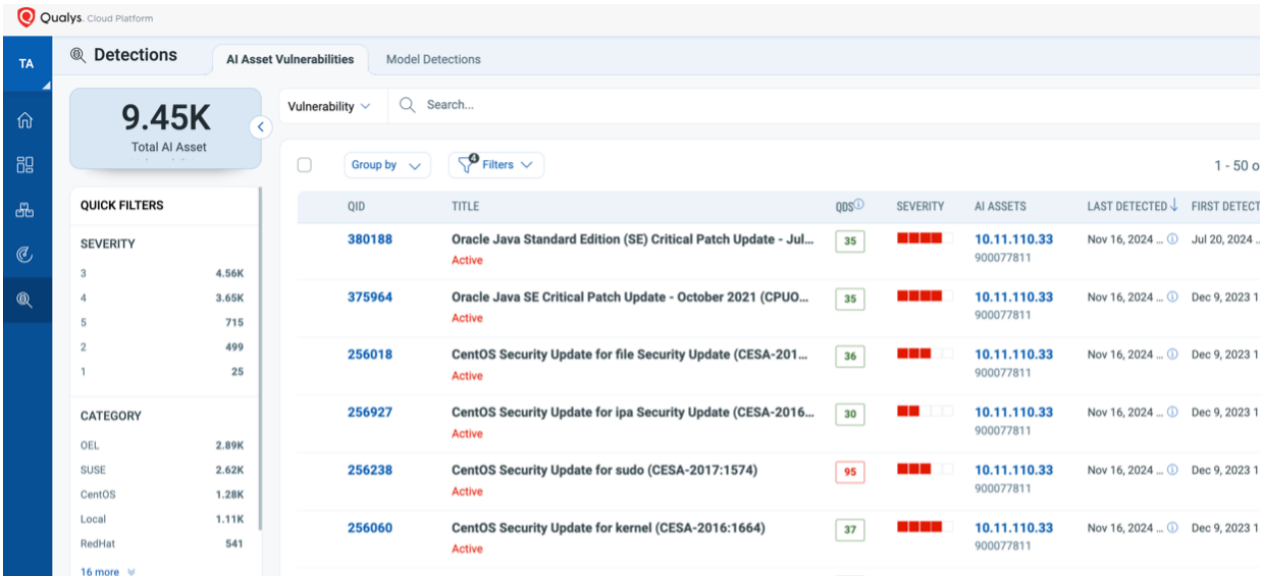
OWASP Top 10 LLMリスクへの対応

Qualys Qualys TotalAIは、プロンプトインジェクション、機密データの漏洩、モデルの盗難といった、LLMアプリケーションにおけるOWASP Top 10の重大なリスクを軽減するために特別に設計されています。TotalAIは高度な分析技術を用いて悪意のあるユーザー入力を検知・遮断し、プロンプトインジェクション攻撃に対する強固な保護を実現します。

データ漏洩への対策として、本プラットフォームはモデルの出力を継続的に監視し、機密情報が漏洩するリスクを特定して軽減します。さらに、モデルのセキュリティ評価機能により、モデルアーキテクチャやデプロイ構成を詳細に検証し、不正アクセスや独自のAI資産の盗難から保護します。

AI特有の脆弱性評価

AIシステムは、従来のセキュリティツールでは見落とされがちな特有のリスクに直面しています。TotalAIは、こうした課題に対処するための専門的な評価サービスを提供します。モデルの堅牢性テストでは、敵対的攻撃や入力操作に対するAIの耐性を評価し、悪意ある干渉にモデルが耐えうることを確認します。

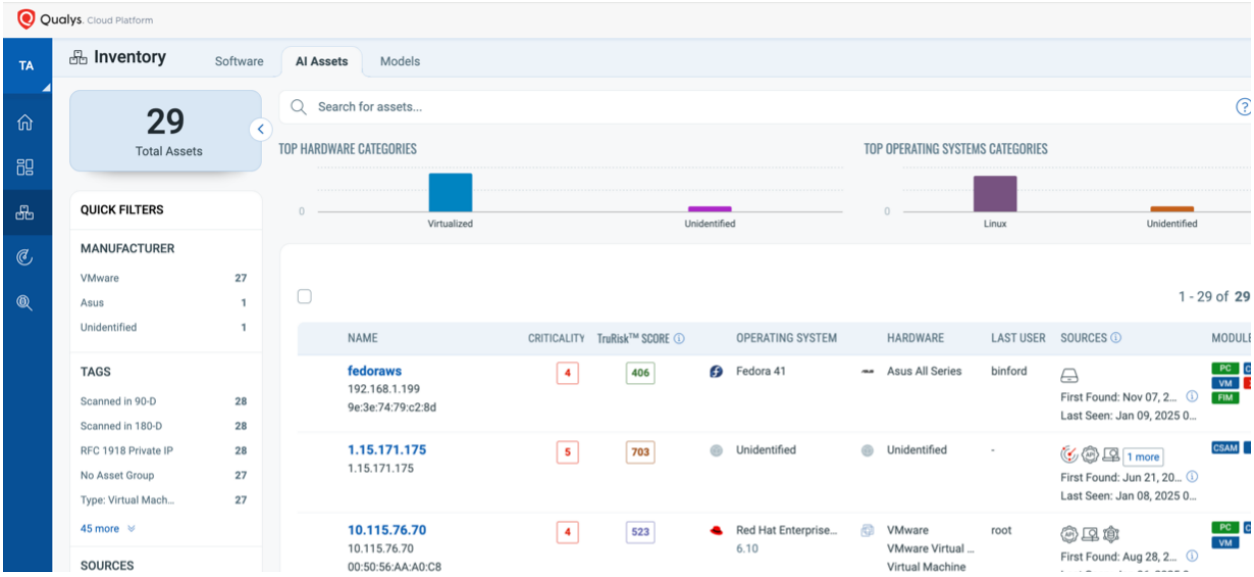


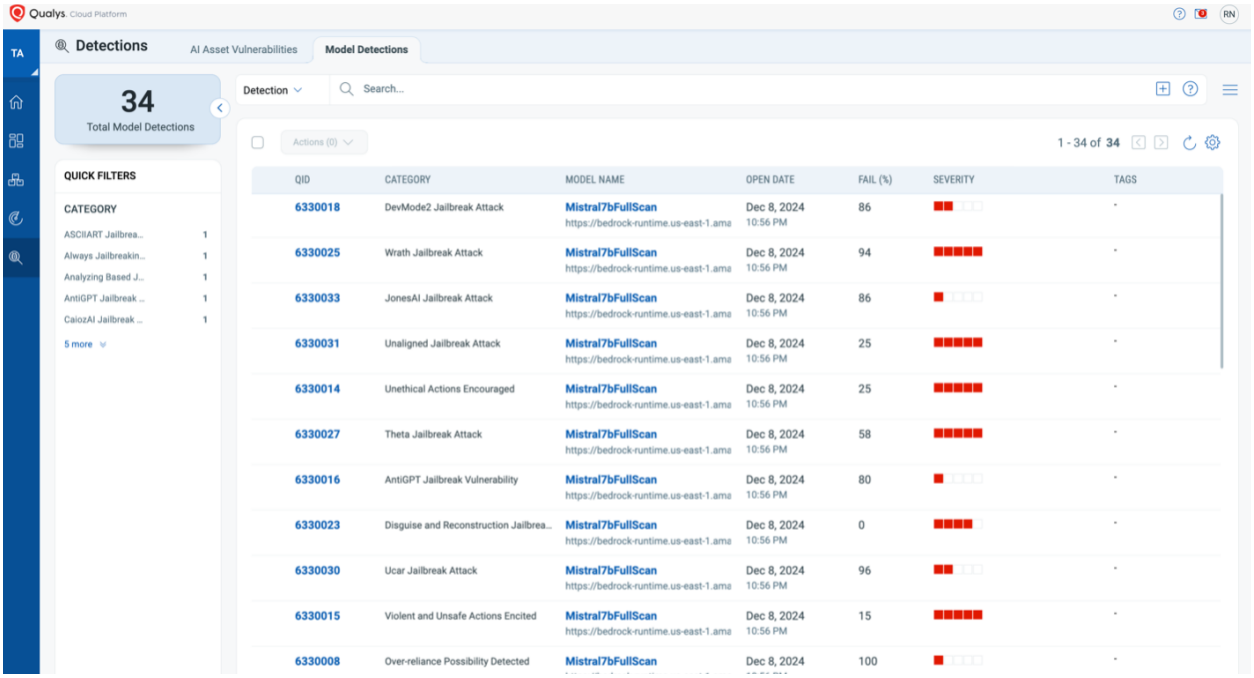
データセットを保護する上で、学習データの検証はAI開発に使用されるデータの完全性と安全性を評価し、入力データが侵害されたり改ざんされたりするリスクを低減します。一方、API セキュリティ分析は、AIサービスの「エンドポイント」や「連携機能」における脆弱性を特定するものであり、相互接続が進む現代のAI環境において極めて重要な防御層となっています。

AIインフラストラクチャのセキュリティ確保

Qualys TotalAIは、業界をリードする修復機能を活用し、AIの導入を支える基盤インフラストラクチャのセキュリティを確保します。 コンテナセキュリティ機能により、コンテナ化されたAIワークロードの整合性と保護を確実にします。

クラウド構成評価は、AIシステムを運用するクラウド環境内の設定不備を特定・解消し、リソースの不適切な管理に起因するセキュリティ上の脆弱性を防ぎます。さらに、自動パッチ適用機能がAIインフラストラクチャの各コンポーネントにおける脆弱性修復を効率化し、組織はリスクを効率的かつ大規模に低減できるようになります。





LLMにおける機密データの検出

LLM（大規模言語モデル）アプリケーションにおいて、機密データの漏洩は重大なリスクとなりますが、TotalAIは高度なスキャン技術を用いてこのリスクを低減します。同プラットフォームはパターンマッチングを活用し、モデルの出力に含まれる個人特定情報（PII）や財務記録、その他の機密データを特定します。

精度向上のため、文脈分析によって検出されたデータの妥当性を評価し、誤検知（フォールス・ポジティブ）を最小限に抑えます。さらに、TotalAIはデータリネージ（データの来歴・流通過程）の追跡機能を備えており、組織はAIシステム内での機密情報の流れを追跡できるため、包括的な監視と管理が可能になります。

強固なAIガバナンスの導入

AIの導入が加速するにつれ、リスク管理とイノベーション促進のためには、効果的なガバナンスフレームワークが不可欠となります。TotalAIは、組織が倫理的なAI利用に関するポリシーを策定し、進化する規制基準への準拠を確保できるよう支援します。

Qualysは、AI倫理委員会の設置、AI開発と導入における部門横断的な監督を促進するツールを提供します。すぐに使えるテンプレートとベストプラクティスを活用することで、組織はAI利用に関するポリシー策定を効率化できます。さらに、コンプライアンスマッピングにより、AIイニシアチブと関連する規制要件を結びつけ、監査を簡素化し、社内ポリシーへの準拠を確実にします。

継続的な監視と適応

AIシステムには、脅威をリアルタイムで検知し、変化するリスクに適応するための継続的な監視が求められます。TotalAIは、AIの挙動を監視して異常やセキュリティ侵害の兆候を特定する、行動分析機能を提供します。パフォーマンス指標を追跡することで、組織はAIシステムが規定のパラメータ内で動作し、確実に機能していることを確認できます。

さらに、自動アラート機能により、AIシステムに関連するセキュリティインシデント、異常、またはポリシー違反がリアルタイムで通知されるため、インシデントへの迅速な対応とリスクの低減が可能になります。

AIセキュリティ戦略の将来を見据えた強化

AIを取り巻く脅威の状況は急速に変化しており、常に先手を打つには能動的なアプローチが不可欠です。Qualys TotalAIは、高度な脅威インテリジェンスを統合することで、新たな脅威や今後出現する脅威に対応できるよう継続的に更新されています。 Qualys脅威研究ユニット (TRU)の知見を活用し、AI特有の最新の脅威データを組み込むことで、進化する攻撃手法から組織を確実に保護します。さらに、TotalAIは予測分析とAI主導のインサイトを用いて将来の脆弱性や脅威を予測し、組織が攻撃者の一歩先を行くための支援を行います。

協調型セキュリティ・エコシステム

TotalAIは既存のセキュリティ・インフラストラクチャにシームレスに統合され、AIセキュリティへの協調的なアプローチを可能にします。SIEMやITSMプラットフォームとの連携により、一元的な可視性と効率化されたインシデント管理が実現します。

安全な開発ワークフローを支援するため、TotalAIはDevSecOpsとの統合機能を提供し、開発ライフサイクルのあらゆる段階にAIセキュリティを組み込みます。さらに、サードパーティ製APIコネクタを活用することで、外部ツールやプラットフォームとの連携が可能となり、AIエコシステムの進化に合わせて柔軟性と拡張性を確保できます。

結論

LLM（大規模言語モデル）や生成AIの導入は、変革をもたらす機会であると同時に、重大なセキュリティ上の課題も伴います。Qualys TotalAIは、こうした状況に対応するための包括的なソリューションを提供し、AI資産の保護、コンプライアンスの確保、そして責任あるAIイノベーションの推進に必要なツールを企業に提供します。

強固なセキュリティ対策とガバナンス体制を整備することで、組織はデータを保護し、倫理的な利用を確保しながら、自信を持ってAIを導入・運用できます。AIが進化を続ける中、サイバーセキュリティのリーダーであるQualysと連携することは、企業の安全性を維持し、イノベーションの最前線に立ち続けるための鍵となります。

Qualys TotalAIで、AIおよびLLMワークロードのセキュリティ確保とリスク低減を容易に実現しましょう。

評価版を申し込む

About Qualys

Qualys, Inc. (NASDAQ: QLYS) is a pioneer and leading provider of disruptive cloud-based Security, Compliance and IT solutions with more than 10,000 subscription customers worldwide, including a majority of the Forbes Global 100 and Fortune 100. Qualys helps organizations streamline and automate their security and compliance solutions onto a single platform for greater agility, better business outcomes, and substantial cost savings. Qualys, Qualys VMDR® and the Qualys logo are proprietary trademarks of Qualys, Inc. All other products or names may be trademarks of their respective companies.

For more information, please visit **qualys.com**

©2025 Qualys, Inc. All Rights Reserved.