APAC全域で生成AIを安全かつ確実に管理

ベライゾン
サイバーセキュリティコンサルティング、
シニアディレクター
Chris Novak

はじめに

人工知能（AI）を使いこなすことは、競争上の優位性を獲得するために極めて重要です。企業は1年ほどで投資回収が可能になり、平均回収額は投資額1ドルあたり4ドルに近づいています[1]。

しかし、AIを上手く活用するには、安全対策と倫理的基礎となるガバナンス、つまり「責任あるAI」が不可欠です。また、技術的な側面、攻撃の手段、堅牢なセキュリティ対策など、生成AIに関連する特定の脅威と脆弱性に対応することも重要です。

最新の「2024年度データ漏洩/侵害調査報告書（DBIR）[2]」で扱う攻撃の中では、AI絡みの攻撃の割合はまだ少数ですが、今後は増加する可能性があるため重要な課題です。

新しいAIモデルを導入することで、世界中の医療、金融、気候変動、エネルギー、災害、インダストリー4.0、生産性、商取引などの最も重要な課題に対処することができます。ベライゾンは、信頼できるパートナーとして、企業のAIを活用したこれらの重要な課題への取り組みをサポートしています。

高速、低レイテンシー、大容量の5GをAI、クラウドおよびエッジコンピューティングと組み合わせることで、ビジネスネットワーク全体でデータが自由かつ容易に移動できるようになります[3]。ただし、このイノベーションのセキュリティを確保することは重要な課題であることを理解することが大切です。これは、最高情報セキュリティ責任者（CISO）と経営幹部の双方が今日取り組んでいる課題です。

私たちは、世界人口の65%[4]、世界のGDPの54%[5]以上を占めるアジア太平洋（APAC）地域における、この革新的テクノロジーの可能性と脅威について深掘りしてみました。APACのこうした規模を考慮すると、イノベーションを保護し、倫理的な展開を確実にするために、効果的なガバナンスとセキュリティ対策を実施するための早急な対応が必要となります。

1 https://news.microsoft.com/source/wp-content/uploads/2023/11/US51315823-IG-ADA.pdf
2 https://www.verizon.com/business/en-au/resources/reports/dbir/
3 https://www.verizon.com/business/resources/articles/s/5g-and-ai-creating-connected-global-business/
4 https://asiapacific.unfpa.org/en/populationtrends#:~:text=The%20Asia%20and%20the%20Pacific,populous%20countries%2C%20China%20and%20India
5 https://www.worldeconomics.com/Thoughts/The-Future-is-Asian.aspx#:~:text=Today%2C%20the%20Asian%20share%20of,for%20less%3A%20about%2033%25

生成AIがすべてを解決？

予測AIとは異なり、生成AI（GenAI）は、システムに明確にプログラムされていない新しいコンテンツ、アイデア、またはデータパターンを作成または生成することができます。

1. インフラの強化：生成AIは、より複雑なAIモデルのトレーニングに必要な大量のデータ処理と転送を可能にし、ネットワークのパフォーマンスと信頼性を向上させます。
2. 運用の変革：生成AIは、特に営業とエンジニアリングにおける社内業務に影響を及ぼします。チャット形式の生成AIツールは、過去の実績、設計の選択、顧客ソリューションを明示し、情報を横断的に精査できます。
3. 製品開発と顧客サービス：生成AIは、ビデオの文字起こしや即時の顧客サポートなど、ほぼリアルタイムでデータ分析や顧客とのやり取りなどを実現します。これにより、より動的で対応力の高いサービスを提供することができます。

最近、Verizon Connectは先進的なAI Dashcamソリューションを APAC市場に導入しました。このドライブレコーダーシステムは、車両のドライバーにとって信頼できる副操縦士として機能します[6]。混雑した道路では、他の車両に近づきすぎたときに安全な距離を保つように優しく注意を促すなど、リアルタイムで情報を提供します。

世界中で、ベライゾンの5Gプラットフォームを利用している企業は、分散型ネットワークからのデータを迅速にデジタル化するための新しい方法を導入しています。医療機関は、モニタリングデバイスからリアルタイムのインサイトを活用して、治療への意思決定を改善しています。

高度なビデオ監視や機器追跡などのAI対応ソリューションにより[7]、医療現場では診断手順、手術の分析、患者の安全性を向上させるための方法が再考されています。

攻撃対象領域の拡大

しかし、AIは、クラウド移行や分散型5G機能の需要を促進していると同時に[8]、これまで考慮されていなかった攻撃対象領域を露呈しています。

攻撃対象領域の拡大は、生成AIの高度な機能と相まって、何が問題になるのか十分に認識・検討することなくAIソリューションを性急に導入する企業にとっては重大なリスクとなります。

ますます巧妙化する脅威と並んで、基本的な手法が多くの攻撃の原動力となっています。脆弱性の悪用は、攻撃者が企業にアクセスするために使用する上位3つの手法の1つです。ベライゾンの「2024年度データ漏洩/侵害調査報告書」では、ゼロデイ脆弱性が急増しており、パッチ管理の改善と即時適応が極めて重要であることが明かにされています[9]。

6 https://www.verizon.com/about/news/new-verizon-connect-ai-dashcam-delivers

7 https://www.verizon.com/business/resources/5g/5g-business-use-cases/workforce-productivity/patient-data-analytics/#solution

8 https://semiengineering.com/how-ai-in-edge-computing-drives-5g-and-the-iot/

9 https://www.verizon.com/business/resources/reports/2024-dbir-data-breach-investigations-report.pdf

新興のAIシステムを詳細に調査することで、企業のAI戦略の改善点が明らかになり、長期的なセキュリティ体制を強化できます。これはITリーダーにとって重要な事項であり、2025年半ばまでに世界中で共通脆弱性識別子（CVE）の数が25%増加すると予想されています[10]。

生成AIの闇

こうした生成AIの活用は興味深いものですが、プライバシー関連の課題も生じます。生成AIテクノロジーは潜在的に機密性の高い膨大な量の情報を処理および分析するため、AIの導き出す回答の正確性とデータの倫理的使用に重点を置くことが最も重要です。

大規模言語モデル（LLM）は、フェイクと呼ばれる誤った主張をすることがよくあります。その答えは説得力があるように見えますが[11]、正しい情報源から得られるのはごくまれです。このため、LLMにより、医療業界などでは誤った情報による深刻な事態が生じる可能性があります。

また、研究者らはAIモデルであるChatGPTに、単語を何度も繰り返すだけでトレーニングデータ[12]を公開させることが可能であることを発見しました。これにより個人情報が漏洩し、AIが記憶した機密情報が意図せず漏洩することが明らかになっています。

このため、対話型アシスタントに機密データを入力する従業員には、意図しない情報開示や侵害のリスクがあります。このリスクは、独自の情報をAIにトレーニングさせ、データ保護法に違反し、権限のないユーザやサードパーティのサーバーに機密情報が公開される可能性です。

企業は、イノベーションとユーザデータおよびプライバシーを保護する責任のバランスを取りながら、AIの実装に慎重なアプローチを取る必要があります。

新たなAIの脆弱性

AIの敵対的な脅威の状況は[13]、現実世界のサイバー攻撃とセキュリティの課題を分析することで把握でき、AIシステム特有の脆弱性が明らかになります。

これは検証中のプロセスですが、いくつかの危険な領域が明らかになっています。

- データストリームのポイズニング：攻撃者がAIトレーニングデータを操作して、エラーや悪意のある攻撃を仕込みます。この「ポイズニング」により、AIが巧妙に再プログラムされ、特定の条件下でアクティブになる脆弱性やバックドアが埋め込まれ、システムの整合性と信頼性が損なわれます。

10 https://www.securitymagazine.com/articles/100426-cves-expected-to-increase-25-in-2024

11 https://www.ox.ac.uk/news/2023-11-20-large-language-models-pose-risk-science-false-answers-says-oxford-study

12 1 Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A.F., Ippolito, D., Choquette-Choo, C.A., Wallace, E., Tramèr, F. and Lee, K., “Scalable Extraction of Training Data from (Production) Language Models,” arXiv preprint arXiv:2311.17035, Cornell University, 2023. https://arxiv.org/abs/2311.17035

13 https://atlas.mitre.org/

- 偽装による防御回避：LLMプロンプトインジェクションと呼ばれる手法により[14]、攻撃者はAIモデルを欺くデータを作成し、誤解や偽情報を生成せさます。この手法はAIによる防御を回避し、脆弱性を悪用して、警備の目をかいくぐるように意図しないアクションを実行します。
- アーキテクチャの事前調査：攻撃者は、Discover Machine Learning（ML）Model Ontology[15]と呼ばれる手法を使用してAIシステムのアーキテクチャを調査し、弱点を特定します。AIのフレームワークを理解することで、攻撃者は悪用可能な脆弱性を特定し、システムの防御を確実に弱体化させる正確な攻撃を実施します。

現実的な脅威に確実に防御を集中させる

ベライゾンの2024年度DBIRによると、サイバー侵害の大部分（68%）は、権限の悪用を除き、ソーシャルエンジニアリング攻撃やエラーなど、依然として人的要素が関与しています[16]。

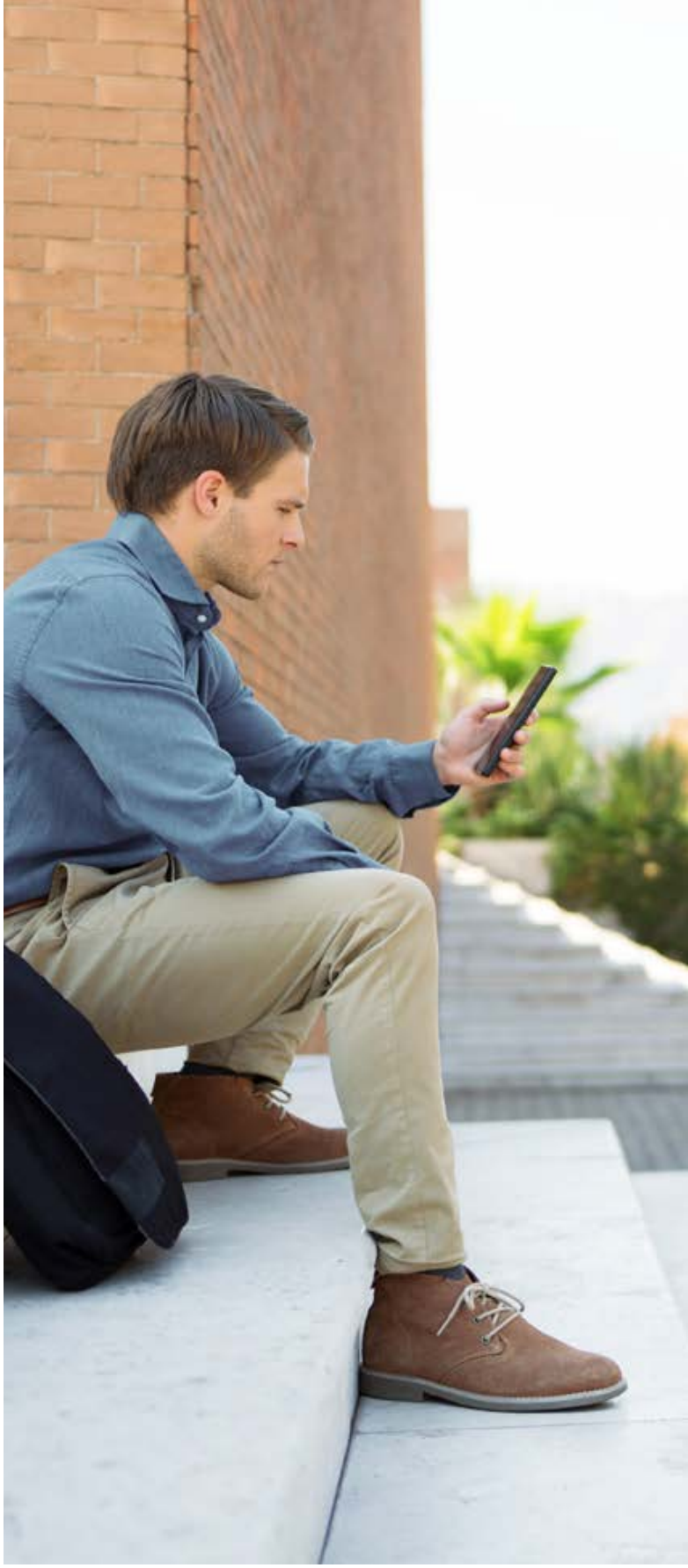
多くのサイバーセキュリティカンファレンスでは、AIを利用した異常な脅威の例が話題になっており、それが過度の警戒を招くことになっています。

また、ディープフェイクのロボコールや、選挙をめぐる新しく高度なソーシャルエンジニアリング攻撃にAIが利用されることについても懸念されています。

では、このような攻撃の実際の頻度に注目してみましょう。

- 現時点では、高度なAIテクノロジーを駆使した広範囲にわたる攻撃が発生する可能性は非常に低いです。
- AIがより高度な脅威に利用される例もありますが、こうした事例はまれであり、通常は一般人ではなく著名人をターゲットにしています。
- ほとんどの人は、AIを利用したフィッシング攻撃よりも、電子メールやテキストメッセージなどの従来のフィッシング攻撃の影響を受けやすい傾向にあります。

オーストラリアの銀行に対する5,000万件のサイバー攻撃に関するニュースが報じられましたが[17]、本格的なAI生成ソースが絡んだ攻撃はごくわずかであることがわかっています。こうした視点は、実際のリスクの状況を理解し、最も必要な場所に防御を集中させるのに役立ちます。



14 https://atlas.mitre.org/techniques/AML.T0051

15 https://atlas.mitre.org/techniques/AML.T0013

16 verizon.com/dbir

17 https://www.mpamag.com/nz/news/general/bank-chief-reveals-the-volume-of-cyberattacks-banks-are-dealing-with/424684

AIガバナンスの役割

企業は現在、テクノロジーを進化させ、より巧妙化する攻撃者といたちごっこを繰り広げています。

さらに懸念されるリスクは、イノベーションの推進や企業や政府を保護するために新たに現れるAIモデルに対して、脆弱または不完全なガイドラインが講じられる際に生じます。

オーストラリアのClare O'Neil内務大臣が[18]、サイバー攻撃を受けた際に連絡すべき政府内の30～40人のリストがあると委員会から伝えられたと認めたことは、AIがいかにして組織のセキュリティを強化し、対応を効率化しなければならないかを示しています。

組織は、内部の生成AIソリューションの構築に熱を上げています。しかし、最初に尋ねる必要がある質問の1つは、テストのために何をするのかということです。AIに詳しくない人がランダムにテストしても、生成AIソリューションが本当に安全かどうかはわかりません。自宅とペンタゴンのセキュリティが違うことと同じで、アプローチはそれぞれにカスタマイズして定量化する必要があります。

複雑なAI環境に通常の侵入テストを実施した場合、脆弱性が生じます。特に攻撃対象領域がIoT環境やインダストリー4.0において典型的な自己最適化システムに拡大するとそれは顕著です。攻撃するとき、攻撃者はルールに従って行動しているわけではありません。彼らは何らかの被害を引き起こすことを目的しているため、防御側は最善策を施し、攻撃者の一歩先を考えていなければなりません。

サプライチェーンが脆弱な理由

AIをテストしている企業は、より高度なリソースを持つ国家の重要なインフラと見なされる可能性があります。今後10年間でAPACが世界の成長の70%を牽引すると予想されており[19]、サプライチェーンのセキュリティ確保は非常に重要です。

日本の大手航空宇宙メーカー[20]と防衛関連企業に対する最近の攻撃は、サプライチェーンの脆弱性の重大度が直接のパートナーをはるかに超える「第四者、第五者リスク」を実証しています。年間収益が10億ドルを超え、従業員が約1万人のこの企業は、国家の防衛に大きく貢献しています。サイバー攻撃により業務は混乱し、さらなる不正アクセスを阻止するためにWebサイトを停止しました。

一方、最近の調査によると、昨年、インド企業の83%がサプライチェーンの侵害を含むサイバー攻撃を受け[21]、多大な経済的損害を被ったとされています。これらの企業のうち、サイバーセキュリティの課題に備えていると考えているのはわずか52%であり、サプライチェーンの防御を強化する必要があることが浮き彫りになっています。

サプライチェーンのリスクは世界的な懸念事項

オーストラリアの資産は、重要なインフラを含め、およそ6分ごとにサイバー攻撃を受けていることはよく知られています。オーストラリア証券投資委員会（ASIC）は最近、調査対象となった企業の44%が、サプライチェーンパートナーからのデータ漏洩を阻止する対策案をまだ策定していないことを明らかにしました[22]。

また、2032年のブリスベンオリンピックなどの大規模な世界的スポーツイベントへの影響も考慮しなければなりません。これらのイベントでは、AIがサイバー侵害の主な原動力となり、多大な財務リスクとインフラリスクをもたらす可能性があります。

2024年第58回スーパーボウル開催中、Verizon Frontline公共安全対策チームは数十の連邦機関と協力し、化学、生物、放射線、核の脅威からサイバーセキュリティ攻撃まで、あらゆる脅威に対抗する防衛チームの態勢を整えました。チームは、潜在的な攻撃に先手を打つために、インフラと会場のセキュリティ評価を継続的に実施しました。

生成AIの保護

O'Neil大臣は、オーストラリアだけでなくAPAC全体におけるAIの活用と保護に関して懸念を述べています。「私たちは、第二次世界大戦以来最も困難な地政学的状況[23]に直面しています。戦略的競争の地域に住んでおり、サイバーの領域は今後10年間の出来事の展開に不可欠な要素となるでしょう」

各国は、2030年までにセキュリティを強化し、サイバー領域における安全の確保に注力する必要があります。そのためには、民間部門を含むAPACの各国が「責任あるAI」に基づくアプリケーションの構築を開始する必要があります。

強力なガバナンスがなければ、AIは良いことよりも悪事の原因となり、非倫理的な結果、偏ったモデル、誤った情報をもたらす可能性があります。

ベライゾンの2024年度DBIRによると、公務におけるインシデントが最も多く（12,217件）、データ漏洩/侵害が確認されたのは1,085件でした[24]。

18 https://minister.homeaffairs.gov.au/ClareONeil/Pages/afr-cyber-summit-18092023.aspx

19 https://www.pwc.com/gx/en/about/pwc-asia-pacific/global-supply-chains-the-race-to-rebalance.html

20 https://westoahu.hawaii.edu/cyber/global-weekly-exec-summary/defense-contractor-japan-aviation-electronics-falls-victim-to-a-cyber-attack/

21 https://timesofindia.indiatimes.com/gadgets-news/over-80-indian-companies-hit-with-cyber-attacks-last-year-report/articleshow/103394017.cms

22 https://asic.gov.au/about-asic/news-centre/find-a-media-release/2023-releases/23-300mr-asic-calls-for-greater-organisational-vigilance-to-combat-cyber-threats/

23 https://minister.homeaffairs.gov.au/ClareONeil/Pages/afr-cyber-summit-18092023.aspx

24 https://www.verizon.com/business/resources/reports/dbir/2024/industries-intro/public-administration-data-breaches/

生成AIを中心に据えた動き

AIの原則は、APACにおけるテクノロジーを社会の利益のために導くことを目的としています。しかし、真の試練は、この多様性に富んだ地域全体での実施にあります。

倫理ガイドラインは、厳格な規制なしにAIの誤用を抑制することを目指していますが、理念と実行可能な実装の間には依然として大きなギャップがあります[25]。

- AIの原則を運用化するためのガイダンスとツールの採用はAPACの各国間で大きく異なり、一貫性についての疑問が生じています。
- 立法化の取り組みは有望ではあるものの、AIの急速な進化に追いつくかという課題に直面しています。
- 日本、マレーシア、オーストラリアなどの国々では、意欲的な国家戦略により、開発におけるAIの役割に大きな期待が寄せられています。

注目すべきは、APACの3か国がAI規制に先駆しており、他の国々が追随する道すじを切り開いていることです。

- シンガポールは、AIの安全性のための実用的なツールの開発をリードしています。最近、情報通信メディア開発庁（IMDA）は、生成AI向けのAIガバナンスフレームワークの提案を協議用に公開しました[26]。これは、2024年半ばのAIガバナンスフレームワークの最終決定をサポートすることになります。
- 韓国の大胆なAI法は[27]、施行されれば、一般利用から高リスクのアプリケーションまで幅広く適用され、立法上初の事例となります。
- 中国はルールベースのアプローチを採用しており、特定のAI規制[28]によってAI管理への包括的なアプローチが策定されています。

企業はAIの力と可能性を最大限に活用するために、AI協議会を設立する必要があります。そうすることで、ビジネス モデル、マーケティング、ナレッジ管理、ソフトウェアエンジニアリングの改革を責任を持って安全に実行できるようになります。

成功するには、リスク管理は後追いで設定するのではなく、完全に組み込まれ、統合される必要があります。リスク定量化サービスは[29]、潜在的なプラットフォームの弱点やAIコンプライアンスのギャップを特定するのに役立ちます。

ベライゾンのCybersecurity Assessmentは、シミュレーション攻撃を駆使してAIなどの脅威を評価するRed Team Penetration Testingが含まれます。侵入テストでは、システムを調査して攻撃手段と脆弱性を検知し、攻撃対象の特定をサポートする自動テストを実行できます。

さらに、ベライゾンはAIガバナンスを導入し、データサイエンティストにAIモデルのレビュー登録を義務付け、大規模言語モデル（LLM）を精査して、潜在的なバイアスや有害な言語に対処し

25 https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/financial-services/deloitte-cn-fsi-acrs-gai-application-and-regulation-in-apac-en-231204.pdf

26 https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/public-consult-model-ai-governance-framework-genai#:~:text=SINGAPORE%20%E2%80%93%2016%20JAN%202024&text=The%20AI%20Verify%20Foundation%20(AIVF,last%20updated%20in%2020201

27 https://carnegieendowment.org/research/2024/02/koreas-path-to-digital-leadership-how-seoul-can-lead-on-standards-and-standardization?lang=en

28 https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

29 https://www.verizon.com/business/products/security/cyber-risk-management/governance-risk-compliance/

ています。これらの取り組みは、責任あるAIを推進する幅広い取り組みと一致しており、ベライゾンのGovernance, risk & compliance (GRC) servicesに統合されています[30]。

これらの戦術的アプローチは、ゼロトラストセキュリティの推進における重要な柱であるサイバーセキュリティのガバナンス、リスク管理、コンプライアンス（GRC）に準拠する必要があります。

APACのすべての組織は、2030年までに生成AIのギャップを克服するために協力する必要があります。倫理ガイドラインから法的枠組みまで、AIガバナンスの不均衡は、進化の格差と脆弱性につながり、経済成長からサイバーセキュリティまですべてに影響を及ぼす可能性があります。

シンガポールと米国は最近、AIガバナンスの枠組みを連携させました[31]。これは画期的な動きであり、AI政策の国際的整合に向けた大きな一歩です。

また、両国は日本、オーストラリアとも連携し、AIシステムを安全に活用する方法について提言しました[32]。さらに、AIシステムのセキュリティを強化するために、世界的に合意された初の「安全なAIシステム開発ガイドライン」を作成するためにも協力してきました[33]。

AIとテクノロジー向けのセキュアバイデザインフレームワークにより[34]、小規模な企業でも大規模なITチームを持たずに安全に取り組むことできます。

合計で世界21の機関がこの枠組みの下で活動し、AIシステム開発者が開発ライフサイクル全体を通じてサイバーセキュリティに重点を置いた意思決定を行えるよう支援しています。

ベライゾンと共に安全に生成AIを未来へ導く

多くの人は生成AIが「平等のための偉大な発明[35]」となることを期待していますが、現実は異なります。ほとんどの政府機関や組織は、地域の人々の安全と安心を脅かしかねない知識と人材の格差に直面しています。

企業は、AIリスクを定量化するために[36]、将来ではなく今すぐ行動する必要があります。この行動の方向性とエネルギーは、経営幹部から発信される必要があります。セキュリティは、セキュリティチームだけの仕事ではなく、上層部から推進されるものです。

ベライゾンは、サイバーチームが組織横断的なAIステアリングチームを結成するのをサポートします。これは、組織が最初の生成AIアプリケーションを構築する前の重要なステップです。AIで先頭に立ち、サイバーセキュリティを確保するには、協力して作業することが不可欠です。

30 https://venturebeat.com/ai/verizon-exec-reveals-responsible-ai-strategy-amid-wild-west-landscape/

31 https://www.mci.gov.sg/media-centre/press-releases/singapore-and-the-us-to-deepen-cooperation-in-ai/

32 https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/governance/engaging-with-artificial-intelligence

33 https://www.cyber.gov.au/sites/default/files/2024-04/Guidelines%20for%20Secure%20AI%20System%20Development%20%28November%202023%29.pdf

34 https://www.cisa.gov/resources-tools/resources/secure-by-design

35 https://www.weforum.org/agenda/2024/02/generative-ai-society-equalizer/

36 https://www.verizon.com/business/products/security/cyber-risk-management/governance-risk-compliance/cybersecurity-assessments/

過去数年間、ベライゾンのチームは、ネットワークパフォーマンスの最適化、トレンドの特定、需要の創出、顧客サービスの向上に関連する日常的な業務を解決するための専用AI、またはAIを応用した技術の開発に多くのリソースを投入してきました。

ベライゾンは、大規模なデータセットでトレーニングを実施し、限定され明確に定義されたタスクを実行させます。つまりAIを、特定の運用またはビジネスニーズに対応するようにカスタマイズするのです。ベライゾンは、そのメリットとリスクを理解しています。

ベライゾンのネットワークは毎日700億のデータポイントを処理し、高度なAIシステムに入力しています。このデータは29,000の多様なソースから取得されており、デジタルエコシステムの規模と複雑さを示しています[37]。

ベライゾンのサポートはビジネスニーズに合わせてカスタマイズされ、詳細なセキュリティレポートや業界ベンチマークとの比較を含む確かなデータと標準に基づいた強力な防御計画を作成します。

AIによる安全な脅威検出と分析を実現

このフレームワークでは、AIを利用して攻撃者への対応を改善することができます。

- 継続的な監視：AIシステムは、ネットワークアクティビティを24時間365日監視し、人間が簡単に見落としてしまう脅威を検知します。
- 自動侵入テスト：このテストは、コンピュータシステム、ネットワーク、またはWebアプリケーションに対するサイバー攻撃をシミュレーションして、悪用される可能性のある脆弱性を特定します。
- トラフィック分析：AIが通常のトラフィックと疑わしいトラフィックを区別し、高度なサイバー脅威の検知を強化します。
- フィッシングの検知：フィッシングとスパムの特性を学習することで、AIは悪意ある電子メールを事前にブロックするのに役立ちます。
- マルウェアの識別：AIツールは既知のマルウェアサンプルを分析して、新しい亜種やゼロデイ（これまで知られていなかった）の脅威を認識します。
- パスワードのセキュリティ：AIは、複雑で解読が困難なパスワードを生成して推奨できます。
- タスクの自動化：日常的なサイバーセキュリティタスクはAIによって自動化され、専門家はより戦略的な問題に取り組むことができます。

生成AIは、企業、その従業員、顧客を支援するための強力なソリューションです。これまで述べてきた方向に進むことで、成長著しいAPAC市場で競争上の優位性が得られるだけでなく、AIはこの地域を人々にとってより安全で安心できる場所にすることができます。

ベライゾンのAIセキュリティについての詳細は、下記へお問い合わせください。

オーストラリア：+61 2 9434 5000
シンガポール：+65 6248 6600
日本：+81 3 5293 9000

37 https://www.sdxcentral.com/articles/interview/verizons-70-billion-network-data-points-highlight-genai-potential-and-challenges/2024/02/

本文書は、ベライゾンの情報と意見を取り入れて、InnovationAus.comが作成したものです。