News release

2024 年 9 月 19 日
PwC コンサルティング合同会社

PwC コンサルティング、AI サービスのビジネスリスクを特定・改善する「AI レッドチーム」によるサービスを開始

疑似的なサイバー攻撃で脆弱性を予見、インシデントを未然に防ぐ

PwC コンサルティング合同会社（東京都千代田区、代表執行役 CEO：安井 正樹、以下「PwC コンサルティング」）は本日から、当社の AI セキュリティに精通したエンジニアがクライアント企業の AI サービスに対して疑似攻撃を行うことで、その脆弱性や想定される脅威、生じ得るビジネスリスクを特定し、改善に向けて支援する「AI レッドチーム」によるサービス（以下、本サービス）を開始します。これにより、企業は AI を利用したサービスを正式に始める前に、想定されるリスクを予見することができるとともに十分な対策を講じることで、インシデントを未然に防ぐことが可能となります。

生成 AI の普及に伴い、ビジネスにおける AI 活用の機運が高まっています。企業は、AI を利用した新たなサービスを開発したり提供したりする一方で、AI 特有のリスクに直面しています。AI は従来の IT サービスとは異なり確率的プロセスに基づいて動作するため、誤った情報の提示や不適切なコンテンツの生成など、想定外の結果を招く恐れがあります。また、そのリスク領域は「技術リスク」のみならず、「倫理リスク」、「法律リスク」と広範に及びます。
経済開発協力機構（OECD）が公開している「OECD AI Incidents Monitor」によると、2023 年 2 月頃から AI サービスに関連するインシデントが急増しており、それまで 1 カ月あたり 100 件に満たなかったのが 2024 年 2 月には 800 件を超えたことが報告されています。こうした状況を受けて、国際的に AI の信頼に関する議論が活発に行われ、さまざまな規制やガイドラインなどの整備が進んでいます。
2023 年 12 月に公開された広島 AI プロセスの成果文書「全ての AI 関係者向けの広島プロセス国際指針」および「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」では、「AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる」という原則が示されており、その行動を遵守する具体的な取り組みの一つとして、レッドチームが提唱されています。
レッドチームとは一般に、サイバー攻撃者の立場で実際に想定される攻撃を疑似的に行うことで、セキュリティ対策の有効性を確認する組織を指し、リスクベースのアプローチによるセキュリティ対策の手法の一つとして注目されています。

■サービス概要
本サービスは、AI を利用したサービスに対してリスク起因者（サイバー攻撃者、犯罪者、愉快犯など）の観点から疑似攻撃を行うことで、脆弱性とそれに伴うビジネスリスクを特定し、その改善のためのアドバイスを提供します。企業が、自社の AI サービスの正式リリース前や運用中に本サービスを利用することで、実際にインシデントが発生する前にビジネスリスクを把握し、能動的に対策を講じることが可能となります。

PwC コンサルティングは、AI 脅威や AI のユースケースに関する豊富な知見を強みとしています。加えて、AI セキュリティに精通したエンジニアとコンサルタントが一体となって対応するチーム体制により、疑似攻撃の実施にとどまらず、現状把握、脅威分析から対策定義までを支援します。

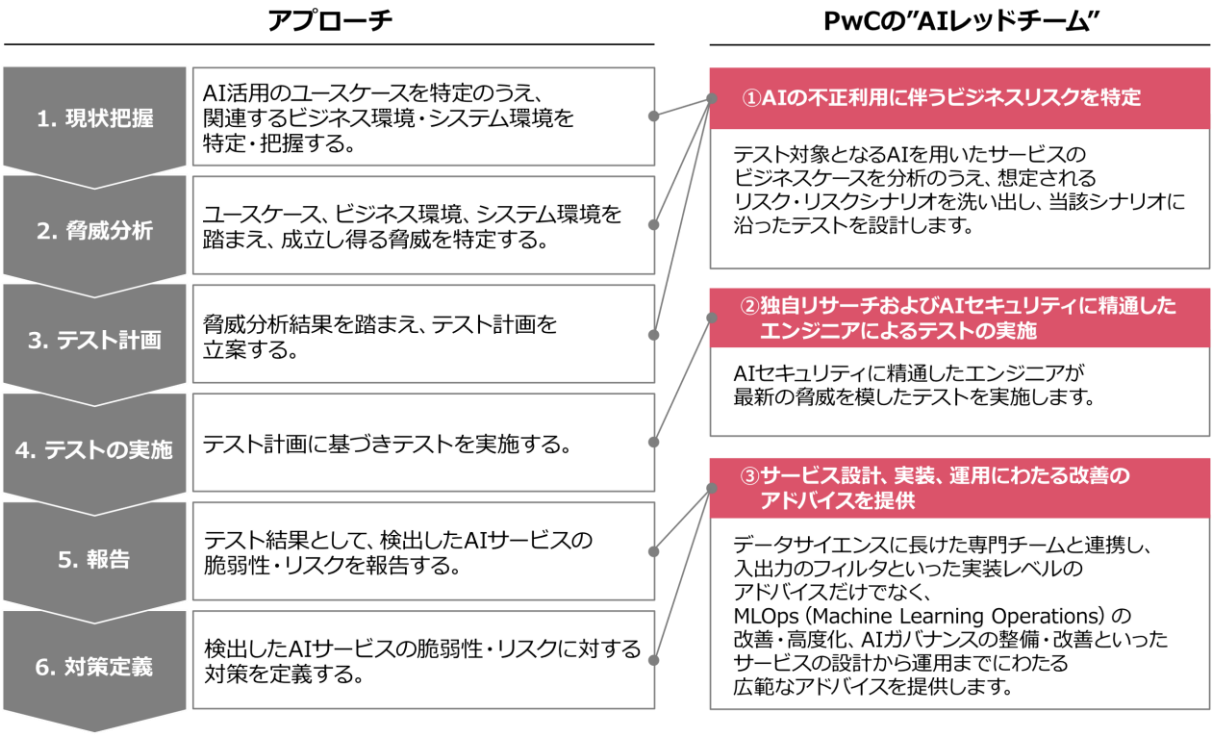
【提供開始日】2024 年 9 月 19 日
【対象者】AI を利用したサービスを開発または提供する事業者
【特長】

①AI の不正利用に伴うビジネスリスクを特定
攻撃や悪用といった AI の不正利用に伴って考慮すべきリスクの種類とインパクトは、そのビジネスユースケースに大きく依存します。本サービスでは、テスト対象となる AI を用いたサービスのビジネスケースを分析したうえで、想定されるリスクおよびリスクシナリオを洗い出し、当該シナリオに沿ったテストを設計します。これにより発見された課題がどのようなビジネスリスクに影響するかを特定します。

②独自リサーチおよび AI セキュリティに精通したエンジニアによるテストの実施
AI セキュリティに精通したエンジニアが、サイバーセキュリティ分野において権威ある研究機関や団体が公開するフレームワークやレポートなどのベストプラクティス（業界標準）や、日々発見・報告される新たな攻撃手法のリサーチ結果に基づいて、最新の脅威を模したテストを実施します。

③サービス設計、実装、運用にわたる改善のアドバイスを提供
検出された課題に対して、各種ベストプラクティスとの紐づけを行い、AI モデル構築や精度評価といったデータサイエンスに長けた専門チームと連携し、改善に向けたアドバイスを提供します。入出力のフィルタといった実装レベルのアドバイスだけでなく、MLOps（Machine Learning Operations）の改善・高度化、AI ガバナンスの整備・改善といったサービス設計・運用まで広範なアドバイスを提供します。

図表1：AIレッドチームのアプローチ（詳細）



詳細は、こちら
（https://www.pwc.com/jp/ja/services/consulting/analytics/responsible-ai/ai-red-team.html）を参照ください。

【関連情報】
AI サービスのリスクと AI レッドチーム
https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/ai-red-team.html

以上

PwC コンサルティング合同会社について www.pwc.com/jp/consulting

PwC コンサルティング合同会社は、経営戦略の策定から実行まで総合的なコンサルティングサービスを提供しています。PwC グローバルネットワークと連携しながら、クライアントが直面する複雑で困難な経営課題の解決に取り組み、グローバル市場で競争力を高めることを支援します。

PwC Japan グループについて www.pwc.com/jp

PwC Japan グループは、日本における PwC グローバルネットワークのメンバーファームおよびそれらの関連会社の総称です。各法人は独立した別法人として事業を行っています。
複雑化・多様化する企業の経営課題に対し、PwC Japan グループでは、監査およびブローダーアシュアランスサービス、コンサルティング、ディールアドバイザリー、税務、そして法務における卓越した専門性を結集し、それらを有機的に協働させる体制を整えています。また、公認会計士、税理士、弁護士、その他専門スタッフ約 12,700 人を擁するプロフェッショナル・サービス・ネットワークとして、クライアントニーズにより的確に対応したサービスの提供に努めています。