AI のリスク軽減と AI による
サイバーセキュリティの変革

AI の脅威の状況


データポイゾニング・
トレーニング攻撃
インファレンス
攻撃
回避攻撃
モデル 誠実さ・
バックドア
システミック
インフラ
ストラクチャー
コンプライアンス
抽出攻撃
抽出・統合問題

AI のリスク軽減 – 主要なリスクと軽減策

コンプライアンスのリスク

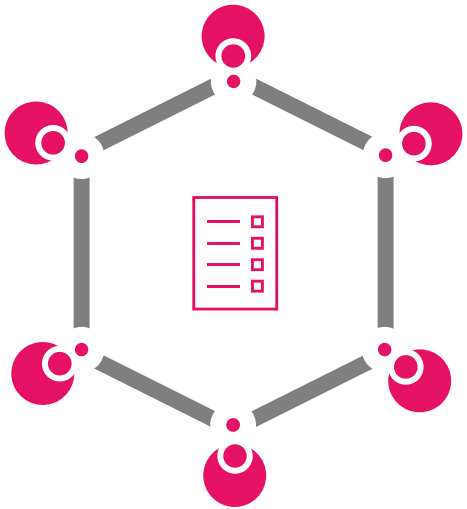
信頼できないソースから収集されたデータには、コンプライアンス違反につながる PII や機密データが含まれる可能性

データの整合性

データセキュリティと環境制御が弱いため、AI ライフサイクル中にデータの整合性が損なわれる可能性

モデルリスク

脆弱なモデル コード、敵対的な条件に耐えるために不適切にトレーニングされたモデル、トレーニングと幻覚、バイアスに対する過剰適合モデル



AI/ML セキュリティ ポリシー

AI/ML のセキュリティ制御を管理する文書化されたポリシー (例 – モデルのトレーニング、モデルのデプロイメント、データ収集)

音声データ管理 練習

信頼できるソースからのデータの利用、データの分類とラベル付け、適切な保護とデータ プライバシーの確保

信頼できる AI モデル

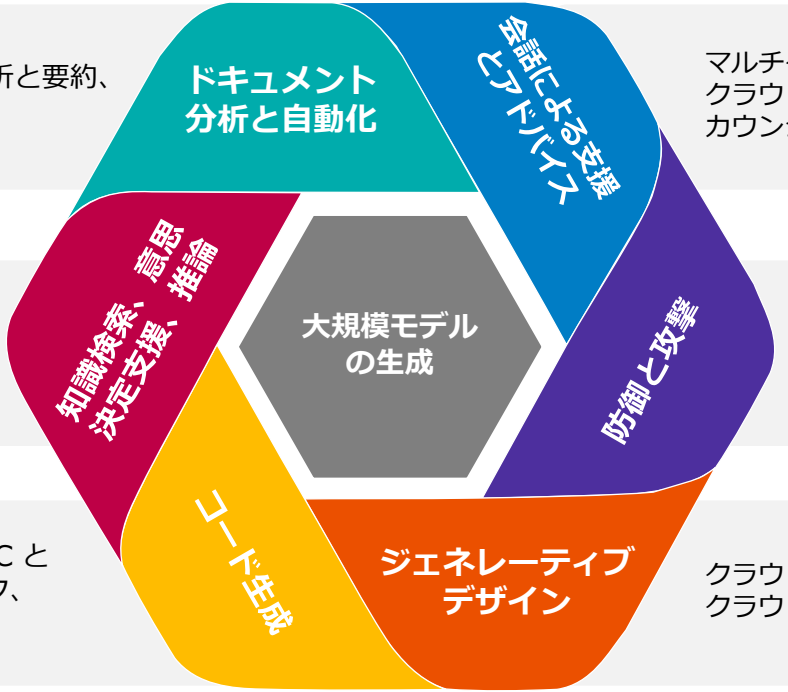
起こり得る状況（最悪の場合、破損したデータセット）に基づいてトレーニングされたモデル、モデルコード用の安全な SDLC、モデルのバージョン管理

ジェネレーティブ AI：機会と主なユースケース

クラウド セキュリティ レポートの分析と要約、コンプライアンス レポート

クラウド セキュリティ知識管理と知識検索 (Q&A)、トリアージと分析サポート、クラウドの復元力のための What-If シナリオ計画

コードのトラブルシューティング、IaC と PaC の生成、プロンプト プレイブック、コードのレビュー



マルチクラウド セキュリティ アドバイザー、クラウド セキュリティ コンプライアンス カウンシル

攻撃対象領域グラフ、脆弱性の優先順位付け、脅威ハンティングガイダンス、クラウド調査

クラウド セキュリティ テスト用の合成データ、クラウド セキュリティ設計ブループリント

Thank You