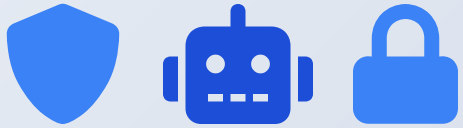
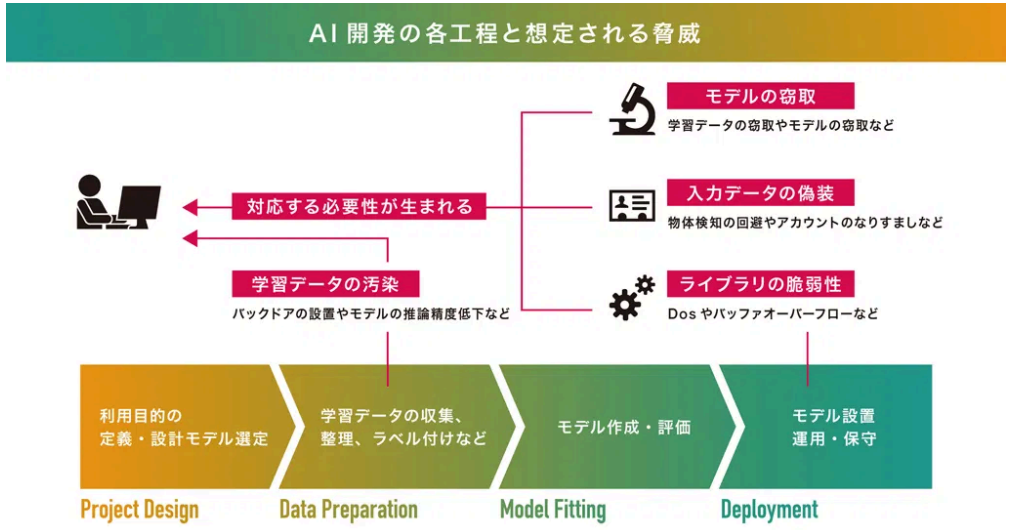


生成AIのセキュリティ

10の鉄壁セキュリティ対策

はじめに

- AIのセキュリティ対策は、**個人情報や企業秘密の漏洩リスクを大幅に減らし**、AIを安全に活用するために不可欠な新しい常識となっています。
- 生成AIの普及により、従来のセキュリティ対策だけでは**不十分**になっています。
- AIツールの利用には、**特有のリスク**が存在します
  - 入力データの学習利用による情報漏洩
  - 誤った情報（ハルシネーション）の生成
  - 意図しない機密情報の混入
- 本プレゼンテーションでは、**10の鉄壁セキュリティ対策**を紹介し、AIを安全に活用するための具体的な方法を解説します。



対策1-2: 信頼性確認と情報保護

対策1: AIツール提供元の信頼性を確認する

目的: 怪しいツールに大切な情報を渡さないための基本

- 運営会社名とその実績を確認する
- プライバシーポリシーをチェックする
- 利用規約の禁止事項、免責事項を確認する
- ユーザーレビューや評判を参考にする

⚠
リスク: 情報抜き取りや情報漏洩の可能性

対策2: 個人情報や機密情報を伏せて指示する

目的: AIに渡す情報を最小限に抑える

- 個人情報や企業名は仮名や記号に置き換える
- 数値データはダミーの値や概算値を使用する
- 具体的な地名や製品名は一般的な名称に置き換える
- 必要に応じてローカル環境で動作するAIツールを検討する

⚠
リスク: 入力情報が学習データとして利用され、外部に漏洩する可能性

対策3-4: 学習設定と生成物確認

対策3: モデルのトレーニング設定をオフにする

目的: 入力データがAIの学習に勝手に使われるのを防ぐ

- 多くのAIサービスでは、ユーザー入力データをモデル改善に利用する設定がデフォルトでオン
- ChatGPTの場合: アカウント設定から「すべての人のためにモデルを改善する」をオフに
- 業務利用の場合は、データが学習に利用されない組織向けプランの検討

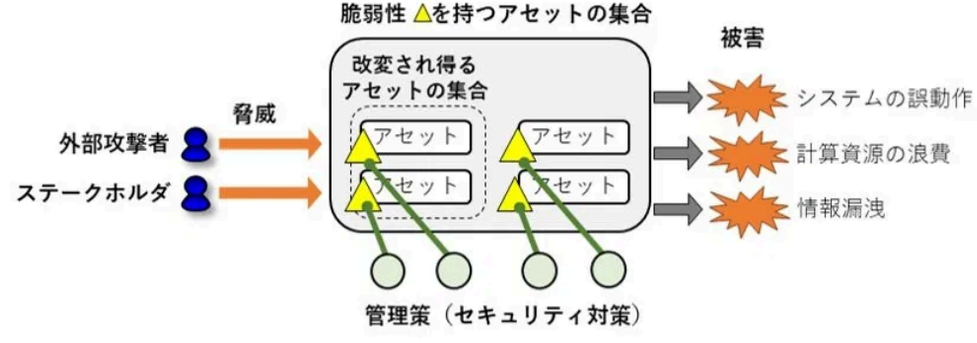
⚠
リスク: 入力データが学習に使用され、他のユーザーへの回答生成に利用される可能性

対策4: 生成物の内容をよく確認する

目的: AIが生成した情報を鵜呑みにしない

- 事実に基づいた情報や数値データは必ず信頼できる情報源でファクトチェック
- 文章の論理的な矛盾や不自然な点がないか確認
- 著作物との類似性や引用の適切さをチェック
- 機密情報や個人情報の混入がないか確認

⚠
リスク: 誤った情報（ハルシネーション）、偏った意見、著作権侵害



対策5-6: 認証とアップデート

対策5: 2段階・多要素認証を設定する

目的: アカウント乗っ取りを防止する

- AIツールのアカウントや連携アカウントに2段階認証を設定
- 認証アプリ（Google Authenticatorなど）を使用する方法
- SMSで認証コードを受け取る方法
- セキュリティキーを使用する方法

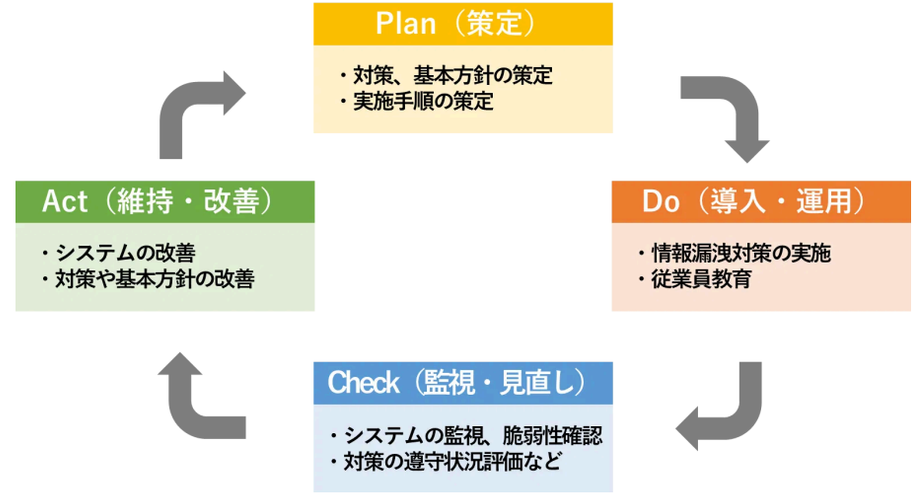
⚠
リスク: ID・パスワード漏洩時の不正アクセス

対策6: 最新のアップデートを使用する

目的: ソフトウェアの弱点を放置しない

- AIツール、OS、Webブラウザ、セキュリティソフトを常に最新版に
- 自動更新機能を有効にする
- 定期的に手動でアップデートがないか確認
- 公式サイトや正規のアプリストアからアップデート

⚠
リスク: セキュリティ脆弱性、マルウェア感染、不正アクセス



対策7-8: 権限とWi-Fi

対策7: 不要な権限を与えない

目的: 必要以上の情報アクセスを許可しない

- アプリやツールが求める権限が本当に必要か判断
- 最小権限の原則：機能に必要な権限だけを許可
- 常時許可ではなく、利用中のみ権限を許可する設定を選択
- インストール済みアプリの権限設定を定期的に確認

⚠
リスク: 個人情報窃盗、盗聴・盗撮

対策8: 公共のWi-Fiでは慎重に利用する

目的: 安全でないネットワーク環境での情報漏洩リスク回避

- 公共Wi-FiでのAIツール利用は極力避ける
- VPN（仮想プライベートネットワーク）を利用して通信を暗号化
- URLが「https://」で始まっているか確認
- 重要なアカウントへのログインや秘密情報の入力は避ける

⚠
リスク: 通信内容の盗聴、情報漏洩

対策9-10: パスワードとデータ削除

対策9: アカウントに強力なパスワードを使用する

目的: 不正アクセスの最初の関門を強固にする

- 長さは12文字以上、できれば15文字以上を推奨
- 大文字、小文字、数字、記号を組み合わせる
- サービスごとに異なるパスワードを設定する
- パスワード管理ツールを利用して安全に管理・生成

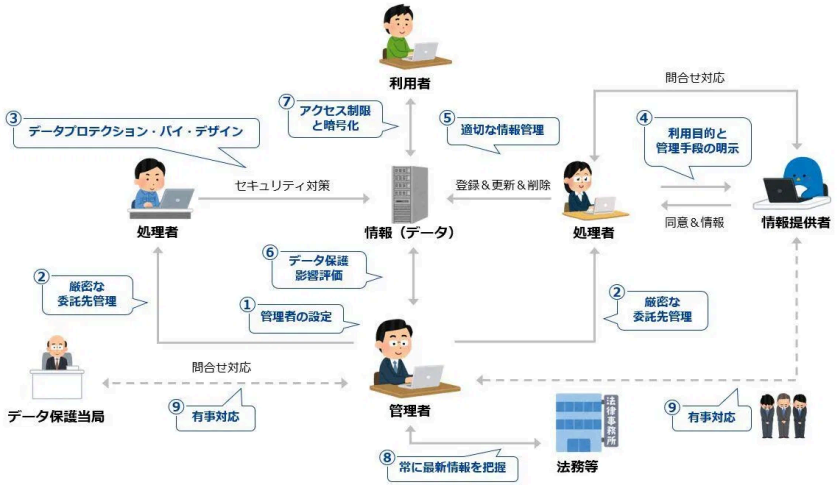
リスク: パスワード推測、芋づる式被害

対策10: 利用履歴・データ削除の方法を把握する

目的: 不要なデータをサービス提供元に残さない

- 利用規約やプライバシーポリシーでデータの保存期間を確認
- データ削除機能があれば、その手順を把握
- 機密情報を扱った後には定期的に削除を実行
- サービス退会時にデータも完全に削除されるか確認

リスク: サイバー攻撃、内部不正によるデータ漏洩



まとめとQ&A

AIセキュリティ対策の重要性

- 生成AIの普及により、**新たなセキュリティリスク**への対応が必須となっています
- 10の対策を**継続的に実践**することで、AIを安全に活用できます
- セキュリティは**最も弱い部分**が全体の強度を決めるため、総合的な対策が重要です

AIツールを利用する際に、まず何を一番に確認すべきですか？

AIツール提供元の信頼性を確認することが最も重要です。運営会社の実績、プライバシーポリシー、データ収集の目的などを確認しましょう。

AIが生成した情報の信頼性をどのように確認すればよいですか？

AIが生成した情報は鵜呑みにせず、事実に基づいた情報や数値データは必ず信頼できる情報源でファクトチェックを行いましょう。

AIに個人情報や機密情報を入力する際の注意点は？

AIに個人情報や企業秘密などの機密情報を直接入力することは極力避け、仮名や記号に置き換えるなどの工夫をしましょう。

「モデルのトレーニング設定をオフにする」とは具体的にどのようなことですか？

ユーザーがAIに入力したデータがAIモデルの学習に利用されるのを防ぐための設定です。ChatGPTの場合、「データコントロール」の項目にある「すべての人のためにモデルを改善する」をオフにします。