AIシステム
脆弱性診断サービス

セキュリティの穴を塞ぎ、
AIの品質を高めるソリューション

運用中のAIシステム、セキュリティは万全ですか？

「すでにAIシステムの運用を開始しているが、セキュリティ対策はまだ調べられていない…」
そんな漠然とした危機感を抱えていませんか？

よくある盲点と誤解

WAFやWebセキュリティを入れているから安心
既存の対策はWebサーバーを守るもので、
AIへの悪意ある命令(プロンプト等)は素通り

自動化スキャンツールでチェックしているから安心
ツールは既知のパターン照合に過ぎません
誤検知(重要でない指摘)も多く、逆に手動でしか発見できない
論理的な矛盾を見逃す可能性がある

開発ベンダーが安全に作ってくれているはず
AIの出力挙動は確率的なため、
従来のソースコードレビューだけでは脆弱性を防げない

1つでも当てはまる場合は診断を推奨

- ☑ 社内データを読み込ませたRAGシステムを活用している
- ☑ AIシステムに新しい機能追加やシステムアップデートを行った
- ☑ セキュリティの観点から、顧客や取引先より対策要求が出ている
- ☑ 業界規制やコンプライアンス要件(個人情報保護等)を満たしたい

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

脆弱性の放置による致命的なリスク

AIシステムの脆弱性を放置すると、不正アクセスによる個人情報漏洩で巨額の損害賠償、
監督官庁からの業務停止命令など、事業継続そのものに深刻な打撃を与える可能性が高まります

既存の脆弱性リスク(Web層)

Webアプリケーション共通の重大なセキュリティ欠陥
攻撃者に悪用されると、システム全体の根幹が揺らぐ可能性

SQLインジェクション

データベースを不正操作され、数万件規模の全顧客情報や
機密データが外部に完全流出する恐れ

OSコマンドインジェクション

サーバー権限を乗っ取られ、データを全削除または暗号化
身代金要求(ランサムウェア)やサービス停止の可能性

認可制御の不備

URLの数値を書き換えるだけで他人のマイページや
契約書が見えてしまう。個人へのプライバシー侵害で炎上

ファイルアップロードの不備

サーバー上で任意の不正スクリプトを実行され、
システム停止やランサムウェア被害へ

AI特有の脆弱性リスク

LLMやAIエージェントの仕組みそのものを狙う新しい攻撃
従来のセキュリティ対策では検知不可能な領域

プロンプトインジェクション

「今までの指示を無視して」といった指示で、
数千万円かけて構築したプロンプト(ノウハウ)が盗まれる

システムプロンプトの奪取

攻撃者の巧妙な指示により、企業が隠していた裏側の
制御命令や設計思想が盗まれる

RAG(知識ソース)の汚染・機密漏洩

参照データに悪意ある情報を混入されAIの回答を操作されたり、
社外秘のデータが一般ユーザーに漏洩する

不適切コンテンツ生成

AIに偽情報や差別発言をさせ、ブランドイメージを破壊
法的な責任(名誉棄損)を問われる可能性

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

脆弱性に関する国内の事故・防御事例

脆弱性によりサービス閉鎖に追い込まれた事例と、事前の防御により被害をゼロに抑え込んだ事例をご紹介します
脆弱性対策はサービス運営を続けるための必須の対策となります

脆弱性による事故事例

大手ハウスメーカー の事例

会員サイトにて個人情報漏洩（ID・パス・氏名等）が発覚

原因

使われていない古いWebページがサーバー内に残っており、古いページに残っていた脆弱性が攻撃を受けた

攻撃手口

SQLインジェクション攻撃

影響

約80万人分の個人情報が漏洩・または漏洩の懸念が発生し、後に会員サイトは閉鎖

脆弱性の防御事例

金融系システム開発企業 の事例

WAF導入により、Webアプリの脆弱性攻撃を可視化・遮断
不正アクセスの攻撃はあったが、事前の防御でブロック

対策

定期的な監視と防御壁（WAF）の二重構築

成果

監視を実施したところ、海外などからSQLインジェクションをはじめとする、Webサービスの脆弱性を狙った大量の不正アクセスが届いていたことが発覚
事前に防御層を構築していたため、すべての攻撃をブロック

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

『通常(Web層)の診断』×『AI特有の診断』によるAIリスクへの対応

AIシステムの安全性を確保するには、システム基盤の「Web層」と「AI層」の両方への対策が不可欠です
両方の診断により、経済的・社会的損失リスクを包括的に軽減することが可能となります

土台を守る 通常の脆弱性診断

サーバー、インフラ、Web画面、
クラウドの穴を塞ぐ

AIが稼働するシステム基盤そのものを防御します。
AIのプロンプト対策をしても、土台のサーバーや
Webサイトの防御が突破されると、データベースから
すべての情報が盗まれる可能性があります

×

AIを守る AI特有の脆弱性診断

プロンプトインジェクション、
RAG汚染を防御する

AIへの悪意ある指示や、学習データの汚染といった
「AIシステム特有の未知の攻撃」を遮断します。
これにより、AIの暴走、機密情報の不正な引き出し、
開発ノウハウ(技術的な機密情報)の流出を防ぎます

ProVisionが提供する「AIシステム脆弱性診断」は『通常の診断』に『AI特有の診断』を加えた

AIシステムに特化した脆弱性診断サービス

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

ProVisionが提供する

AIシステム脆弱性診断とは？

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

「AIシステム脆弱性診断」6つの特徴

国際基準に適合した高度な診断体制と、ProVisionならではの「品質検証(QA)」へのシームレスな連携体制により、
セキュリティの穴を塞ぐだけでなく、AI全体の価値と精度向上までをトータルでサポートします

01 国際基準に適合

経済産業省策定の**情報セキュリティサービス基準に適合した専門ベンダー**が対応。
セキュリティの国際基準「OWASP TOP 10」「OWASP Top 10 for LLM」および国内IPAの基準に準拠した、**極めて高い品質の診断**

02 有資格者によるダブルチェック

すべての案件を必ず**2名体制**で実施
そのうち1名は国家資格である「情報処理安全確保支援士」を保有。
徹底したダブルチェックにより、**高品質の維持とツール特有の「誤検知」を排除**

03 手動×ツールのハイブリッド

高度な自動スキャンツール診断に加え、セキュリティエンジニアによる手動診断を実施。
機械的なツールでは検知できない、**複雑な処理ロジックの隙を突く脆弱性も検出**

04 高いコストパフォーマンス

専門ベンダーへ直接依頼するよりも安価な提供を実現。
リクエスト・API単価の低減により、
約1割程度のコスト削減が可能

05 品質検証への移行も可能

脆弱性診断を通じてシステムの挙動や品質状況を把握できるため、**診断完了後、早急かつ適切な品質検証の体制を構築可能。**
継続的なAIの精度向上まで、一気通貫で伴走できるのはProVisionの強みです

06 Slack等による円滑な連携

SlackやChatworkを常時活用。
診断状況の共有から、緊急性の高い脆弱性が発見された場合の「リアルタイム速報」まで、**円滑でスピーディなコミュニケーションを実現**

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

診断領域
土台を守る

通常の脆弱性診断(Webアプリの一例)

一般的な脆弱性項目を網羅的に診断します
診断内容は日々アップデートを行い、最新事例にも対応した品質の高い診断サービスを提供しています

診断メニュー	具体的な診断内容(例)	手法
SQL インジェクション	データベースを利用する箇所において、SQL文の生成不備を悪用し、 会員情報やクレカ情報など全顧客データを外部に流出させるバグがないか検証	手動＋自動
クロスサイト スクリプティング	画面に入力された値を出力する箇所において、悪意あるスクリプトを埋め込み、 アクセスした一般ユーザーのパスワードを盗み出す罠がないか検証	手動＋自動
OSコマンド インジェクション	シェルコマンドを利用する箇所において、意図しないコマンドを実行可能にし、 外部からサーバーを完全に乗っ取られる危険性がないか調査	手動＋自動
ディレクトリ トラバーサル	ファイル名やディレクトリ名を送信するパラメータにおいて、URLを細工し、 本来非公開のシステム設定ファイルやソースコードを覗き見られるリスクを検証	手動＋自動
ファイルアップロード の悪用	ファイルをアップロードする箇所において、画像等に見せかけた不正プログラム (ウイルス)を植え付けられ、サーバー内で勝手に実行される不備がないか診断	手動＋自動
アカウントロック機能	連続してログインに失敗した際に試行回数制限(ロック)があるか診断 制限がない場合、パスワードの自動総当たり攻撃で突破される可能性がある	手動＋自動
権限昇格	一般会員アカウントなどの下位権限において、通信データを書き換えることで、 「他のユーザーの削除や強制退会」など上位権限(管理者)の操作を 強制実行できるバグがないか診断	手動＋自動
権限外情報への アクセス	ユーザーIDなどを格納するパラメータ(URL等)を操作することで、「?user_id=1002」 に変えるだけで他人の住所や注文履歴が見えてしまう重大なバグがないか診断	手動＋自動

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

診断領域
AIを守る

AIシステム特有の脆弱性診断(一例)

OWASP Top 10 for LLM 、MITRE ATLAS、AI事業者ガイドライン、
NIST AI RMF、WSTGを考慮した最先端の診断メニューを標準提供しております

診断メニュー	具体的な診断内容(例)	手法	提供区分
プロンプトインジェクション	悪意あるプロンプト（指示）を入力し、 システムプロンプトの奪取や制御の回避が可能か確認	手動＋自動	標準メニュー
間接的インジェクション	外部のWebサイトやRAGの参照ファイルを介して、 AIに意図しない不正な動作を誘発できるか検証	手動＋自動	標準メニュー
機密情報漏洩診断	LLMの学習データやRAG内部に含まれる個人情報・社外秘情報が、 回答を通じて流出するリスクを確認	手動＋自動	標準メニュー
RAG汚染診断	データソースに悪意ある情報を混入させ、 AI全体の回答を特定の意図通りに操作・歪曲できるか耐性を評価	手動＋自動	標準メニュー
不適切コンテンツ生成	差別、暴力、公序良俗に反する表現や、著しく不正確な情報をわざと 生成させ、フィルター強度を評価	手動＋自動	標準メニュー
セキュアな出力処理	AIが生成した回答文字列に悪意あるスクリプトが混入した場合、 ブラウザ側で実行（XSS等）されないか防止策を確認	手動＋自動	標準メニュー
AIエージェント権限逸脱	AIにDB操作やAPI実行などの自律ツール操作権限を与えている場合、 それが本来のユーザー権限を超えないか確認	手動＋自動	標準メニュー
レッドチーミング	実際の高度な攻撃シナリオに基づき、多角的な疑似攻撃を行うことで、 AIシステムの限界値を測定	手動メイン	要相談

※システムの負荷耐性などのパフォーマンスをチェックする「性能診断」は、
脆弱性診断とは別枠のオプションにて対応可能です。合わせてご検討ください。

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

診断領域 全方位をカバーする脆弱性診断メニュー

インフラ、アプリ、ネットワークまで、あらゆる攻撃経路をカバーする脆弱性診断メニューを提供しています

Webアプリケーション診断

ブラウザ経由の攻撃を診断
SQLi、XSS、認証不備等、
Webアプリ特有の診断項目を網羅
AIのチャット画面等のUI層を守る

プラットフォーム診断

サーバー、OS、ミドルウェアの
設定不備を診断
IPアドレスを指定するだけで、
土台となるシステムの穴を
ネットワーク内外から調査

ペネトレーションテスト

「侵入」を目的とした高度な疑似攻撃
脆弱性を組み合わせて、実際に
機密情報まで辿り着けるかをプロが
検証する実戦形式のテスト

スマホアプリ診断

iOS/Androidアプリを対象に、
端末内のデータ保存状況、
通信の暗号化、リバースエンジニア
リング耐性等を診断し、アプリ経由の
流出を防ぐ

クラウド診断

AWS/Azure/GCPの設定不備等を診断
S3の公開設定やIAM権限ミス等、
クラウド特有の設定ミスによる
大規模な漏洩リスクを未然に防止

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

脆弱性診断はいつ実施すべきか？(推奨タイミング)

脆弱性診断は「リリース前の最終確認に一度だけ」ではなく、システム変更後も継続して実施することを推奨しています
手戻りを防ぎ、トータルコストを抑えるためには、システムの変更点や定期的なサイクルに合わせて
継続実施することが重要です

01

機能の追加・変更時

AIモデルそのものの差し替え、
コア機能のアップデート、
外部API連携の追加、およびAPI側の
仕様変更があったタイミング

**データが通るルートが変わるため、
新しい攻撃経路が
生まれやすい瞬間に行う**

02

RAG・参照データの更新時

AIが社内知識として参照する
データベース(RAG)や参照ファイル、
学習データの構造を大きく更新・
変更したタイミング

**意図しないデータ漏洩や、
回答操作(汚染)のリスクを
再検証する必要**

03

定期的な実施(半年に1回)

システムを全く変更していなくても、
世の中には毎日新しい未知の
攻撃手法やソフトウェアの脆弱性が
発見されている

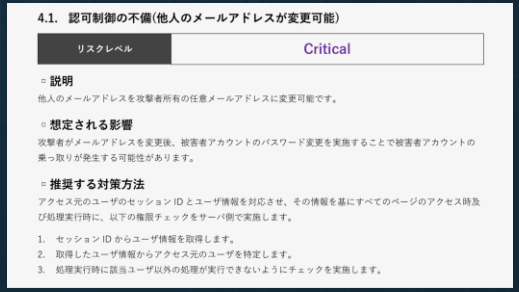
**半年に1回などの定期診断を
行うことで、常に安全であることを
顧客や株主に証明可能**

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

成果物　診断結果の可視化と具体的な報告書の提出

CVSS(※脆弱性の深刻度を国際標準でスコアリングする評価指標)に基づく国際基準のスコア化、再現キャプチャ、
具体的な修正コードの提示を行うことで、「修正に向けてすぐに動ける」レポートを提供します

評価/スコア	リスクレベル	概要・お客様が取るべきアクション
D 9.0～10.0	Critical	非常に容易に悪用され、深刻な影響を招く **速やかな対策実行を強く推奨**
D 7.0～8.9	High	重要な情報漏洩やシステムダウンの危険性 **速やかな対策が必要**
C 4.0～6.9	Medium	複数の要因が重なることで実際の被害につながる可能性。対策を推奨。
B 0.1～3.9	Low	被害を拡大させる潜在的リスク 順次対応を推奨
A 0	Information	脆弱性ではないが、セキュリティ向上のための開発改善



4.1. 認可制御の不備(他人のメールアドレスが変更可能)

リスクレベル	Critical

▫ 説明

他人のメールアドレスを攻撃者所有の任意メールアドレスに変更可能です。

▫ 想定される影響

攻撃者がメールアドレスを変更後、被害者アカウントのパスワード変更を実施することで被害者アカウントの乗っ取りが発生する可能性があります。

▫ 推奨する対策方法

アクセス元のユーザのセッション ID とユーザ情報を対応させ、その情報を基にすべてのページのアクセス時及び処理実行時に、以下の権限チェックをサーバ側で実施します。

1. セッション ID からユーザ情報を取得します。
2. 取得したユーザ情報からアクセス元のユーザを特定します。
3. 処理実行時に該当ユーザ以外の処理が実行できないようにチェックを実施します。

← レポートサンプル抜粋

報告書に記述される要素

検出された脆弱性がもたらす	開発者が自社でバグを完全に再現・確認できる	サーバー側で具体的にどうプログラムを直すべきかを示す
具体的な想定される脅威・ビジネス影響	**詳細な確認手順・キャプチャ**	**推奨する対策方法・記述例**

リアルタイム共有とスピード納品

診断期間中、緊急性の高い脆弱性が発見された場合、最終報告書を待つことなく**発見された当日中に速報レポートとして リアルタイムで共有**。全体の診断終了後は、詳細な正式報告書一式をスピーディに納品

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

実績 幅広い業界での豊富な診断実績

官公庁、金融、大手インフラ企業からWebベンチャーまで、極めて高いセキュリティ基準を求められる
数多くのプロジェクトを成功に導いています （※ベンダーの実績一例）

業界・業種区分	プロジェクト名	概要と診断のポイント
某官公庁・自治体 様	感染症ワクチン予約アプリ診断	数百万人の住民が利用する社会インフラ 個人情報の完全な保護と、大量アクセス時の一切の脆弱性を 排除するための診断
新聞社 様	社内用情報共有Webアプリ診断	機密情報の流出を防ぐため、不正アクセスを防ぐ認可制御の 不備を中心に診断
産業IoT企業 様	IoT機器 Webコンソール画面診断	物理的なスマート機器を制御するWeb管理画面の脆弱性を診断 乗っ取りによる機器の不正操作リスクを未然に防止
SIer企業 様	社内Webシステム脆弱性診断 (オンサイト)	開発ベンダーとしての高度なセキュリティ要件を満たすため、 細やかな診断・連携を実行
サイト運営企業 様	Webクチコミ情報サイト診断	悪意あるスクリプトの埋め込みによる閲覧者への攻撃(XSS)や、 不正なリダイレクト処理がないかを網羅的に検証

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

価格 サービスメニューと費用感

お客様のご予算やシステム構成に合わせて対象を絞り込み、費用を最適化するご提案も可能です (見積もり無料)

簡易診断 クイック診断

AIへの入力形式で最大リスクを発見・可視化

30万円 税抜

入力リスクに特化した「安全度」「対策」の把握

基本3項目	全AI共通の基本チェック (悪質な指示・機密情報漏洩の防止など)
入力1形式	4つの形式(テキスト・チャット、文書・PDF、画像、音声)から1つの形式を選択し重点テスト

総合診断 本診断

AIシステム全体への網羅的な脆弱性の把握

100万円 税抜〜

AI特有リスク項目網羅チェック
社内データ・RAGの汚染、権限の暴走、不適切発言など

＋

システム土台全方位
Web画面、サーバー、クラウド、ネットワーク、スマホアプリ等

以下の診断メニューをバランス良く実施

- AIセキュリティ診断
- Webアプリケーション診断
- プラットフォーム診断
- スマホアプリ診断
- クラウド診断

本診断の料金算出基準 (参考)

診断メニュー	基本診断価格	「単位」の数え方と仕組み
Webアプリケーション診断	診断単価 × リクエスト数	リクエスト数：画面のボタン等を押した際にサーバーへ送られる通信の数
プラットフォーム診断 (サーバー・インフラ)	診断単価 × IPアドレス数	サーバーやネットワーク機器を外部から調査するための、グローバルIPアドレスの総数に基づき一括で確定

・基本単価は診断内容に応じて変動いたします。詳細についてはお問い合わせください　・ペネトレーションテスト、スマホアプリ診断、クラウド診断は別途お見積もりとなります

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

価格

サービス価格体系と費用感

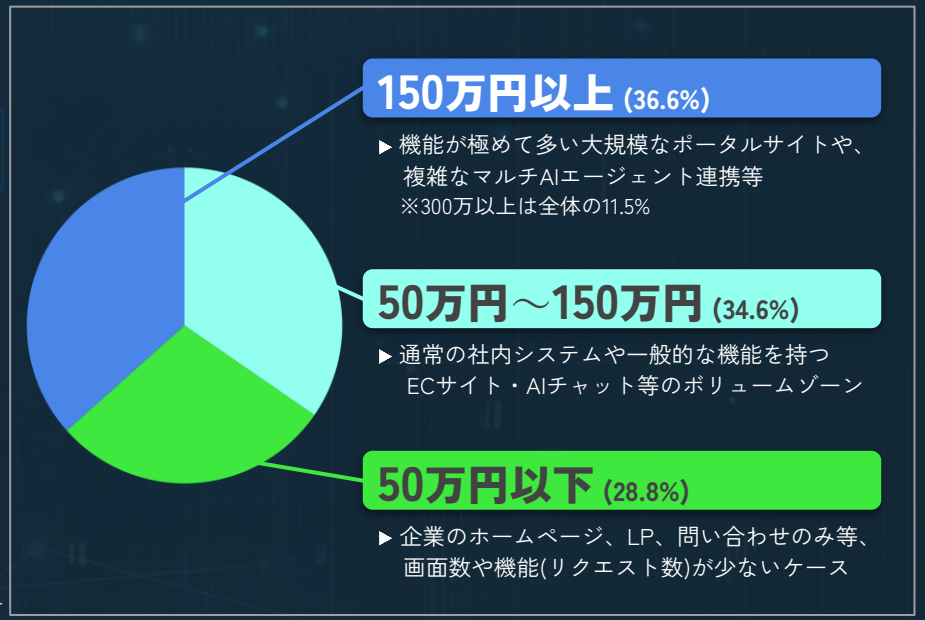
AIシステム脆弱性診断の価格体系は診断メニューごとに異なります
お客様のご予算やシステム構成に合わせて対象を絞り込み、費用を最適化するご提案も可能です (見積もりも無料で実施)

料金の算出基準 (参考)

診断メニュー	基本診断価格	「単位」の数え方と仕組み
Webアプリケーション診断	診断単価（※1）× リクエスト数	**リクエスト数：** 画面のボタン等を押した際にサーバーへ送られる通信の数 ※全通信を洗い出した後、診断対象を選定
プラットフォーム診断 (サーバー・インフラ)	診断単価（※1）× IPアドレス数	サーバーやネットワーク機器を外部から調査するための、グローバルIPアドレスの総数に基づき一括で確定

※1：診断内容に応じて基本単価は変動いたします。詳細についてはお問い合わせください
※2：ペネトレーションテスト、スマホアプリ診断、クラウド診断は別途お見積もりとなります

実際の総額費用の割合 (参考)



Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

作業フロー

お問い合わせから完了までの詳細フロー

ProVisionとセキュリティ専門ベンダーが強固に連携し、お客様の手間を最小限に抑えながら、
スムーズかつ迅速に診断を実施します

01 見積・ヒアリング

対象アプリの画面や仕様を基に、全リクエストを無料で洗い出し
精査の上、最も効率的な診断範囲とプランをご提案

02 環境準備・確認

お客様にて検証用テストアカウント等をご用意いただき、弊社側でAPIエンドポイント等の事前疎通・接続確認を実施

03 セキュリティ診断

手動×ツールのハイブリッドで疑似攻撃を実施

※緊急性の高い脆弱性は当日中にリアルタイム共有

04 報告書納品

診断終了後、5日目安で詳細な報告書を納品
国際基準CVSSスコアに基づく品質評価や、プログラム修正案を提示

05 修正 & 再診断

報告書の再現手順に則りお客様側でバグを修正
完了後、修正が適切かを確かめる「再診断」を無料で実施し、最終報告

追加オプション

報告会の実施：「報告書を渡されるだけでなく、エンジニアから直接役員や開発チームに解説してほしい」という要望に対し、
追加オプションにて対応も可能

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

FAQ よくあるご質問への回答

ご質問	回答
通常の脆弱性診断（Web層）を除いて、AI特有の脆弱性診断のみを実施することは可能ですか？	原則として**「通常の診断」と「AI特有の診断」を組み合わせたパッケージ**としてサポートしております。 AI層の対策を徹底しても、土台となるWebサーバーやインフラの防御が突破されてしまうと、 データベースからすべての情報が盗まれるリスクが残るため、両層の包括的な診断を推奨しております。 ご予算やシステム構成に合わせて、診断対象（リクエスト数やIPアドレス数）を絞り込み、 費用を最適化するご提案は可能ですので、お気軽にご相談ください。
見積もりを依頼するにあたり、事前にどのような情報を用意すればよいですか？	**診断対象となるAIシステムの画面遷移図、仕様書、またはAPIの仕様がわかるもの**をご用意いただけますとスムーズです。 詳細な情報がない段階でも、対象アプリの画面や挙動を基に、全リクエストを無料で洗い出し、 最も効率的な診断範囲とプランをご提案いたしますので、まずはお気軽にお問い合わせください。
本番運用中のシステムしかありませんが、診断は可能ですか？	安全性を最優先とするため、**原則として「検証用（テスト）環境」のご用意**をお願いしております。 診断では実際に擬似攻撃を行うため、データが書き換わったり、システムに負荷がかかったりするリスクがあるためです。 どうしても検証環境の準備が難しい場合は、影響の出ない時間帯での実施や、診断手法の調整などを個別にご相談させていただきます。

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

FAQ よくあるご質問への回答

ご質問	回答
すでに社内で脆弱性診断を行っていますが、外部の第三者に依頼する意味や違いはありますか？	**専門企業による診断の最大の違いは「客観性」と「見落としのなさ（網羅性）」にあります。** システムを熟知している社内のメンバーほど、「こういう仕様で作ったから大丈夫だろう」という無意識の思い込み（バイアス）が働きやすく、診断が毎回決まったパターンになりがちです。 上場企業や医療機関など、極めて高いセキュリティを求められるお客様の案件でも、「社内テストはすべて合格していたが、弊社の第三者診断を通したところ重大な欠陥が見つかった」というケースを確認しています。 医療の「セカンドオピニオン」と同じように、**定期的に外部の客観的な視点を入れることで、安全な運用を進めることができます。**
自社で「AIによる自動脆弱性チェックツール」を入れているのですが、それでは不十分でしょうか？	最新の自動ツールであっても、ツール単体でセキュリティを万全にするのは困難です。理由は大きく2つあります。 1. **手動でしか見抜けない脅威がある** ツールは「過去の既知のパターン」を探すのは得意ですが、システムの「処理ロジック（矛盾）の隙」を突く攻撃や、人間の心理を突くような巧妙な攻撃は、手動診断でしか発見できません。 2. **「誤検知」の処理に手間がかかる** ツールは問題がない安全な箇所を「危険」と判定してしまう「誤検知」が一定数発生するため、本当に直すべき重大なリスクを見極めるには、結局人間の専門知識が必要になります。 「ツールの誤検知の仕分け」や「対策の検討」に自社のリソースを奪われるよりも、最初から「ツール×手動」のハイブリッドで対応し、修正案の提示まで行う専門家にお任せいただく方が、結果としてコストも手間も抑えられることがあります。
近年、AIシステムも含めてセキュリティ事故のニュースが増えているのはなぜですか？	原因の一つは**「攻撃者側もAIを悪用しており、攻撃のレベルとスピードが進化している」ことにあります。** 現在、悪意あるハッカーの間では、「WormGPT（ワームGPT）」や「FraudGPT（フラウドGPT）」といった、犯罪対策の制限（ガードレール）を排除した『攻撃専用の悪用AI』が実際に使われています。AIはプログラミングやセキュリティの分野が得意なため、これまでは人間が時間をかけて作っていた高度で多様な攻撃パターンを、AIがより速く、より多く生み出せるようになってきています。これにより、新しい攻撃手法が生まれるサイクルは以前より早くなっており、「一度診断したから当面は大丈夫」とは言い切れない状況に変化しています。 「システムを変更していないから大丈夫」が通用しにくい時代だからこそ、最新の脅威に対応した診断を行うことが重要になります。

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.

お気軽にご相談ください

貴社のAIシステムの価値と安全性を、確かな技術で守り抜きます

MAIL

prv-infoshare-ml@pro-vision.jp

TEL

045-872-4000

【営業時間】9:00–18:00
【定休日】土曜日・日曜日・祝日

ProVision

〒220-8140　神奈川県横浜市西区みなとみらい2丁目2番1号
横浜ランドマークタワー40階

株式会社ProVision

https://www.pro-vision.jp/

Confidential – Do Not Distribute Without Permission © ProVision Inc. All Rights Reserved.