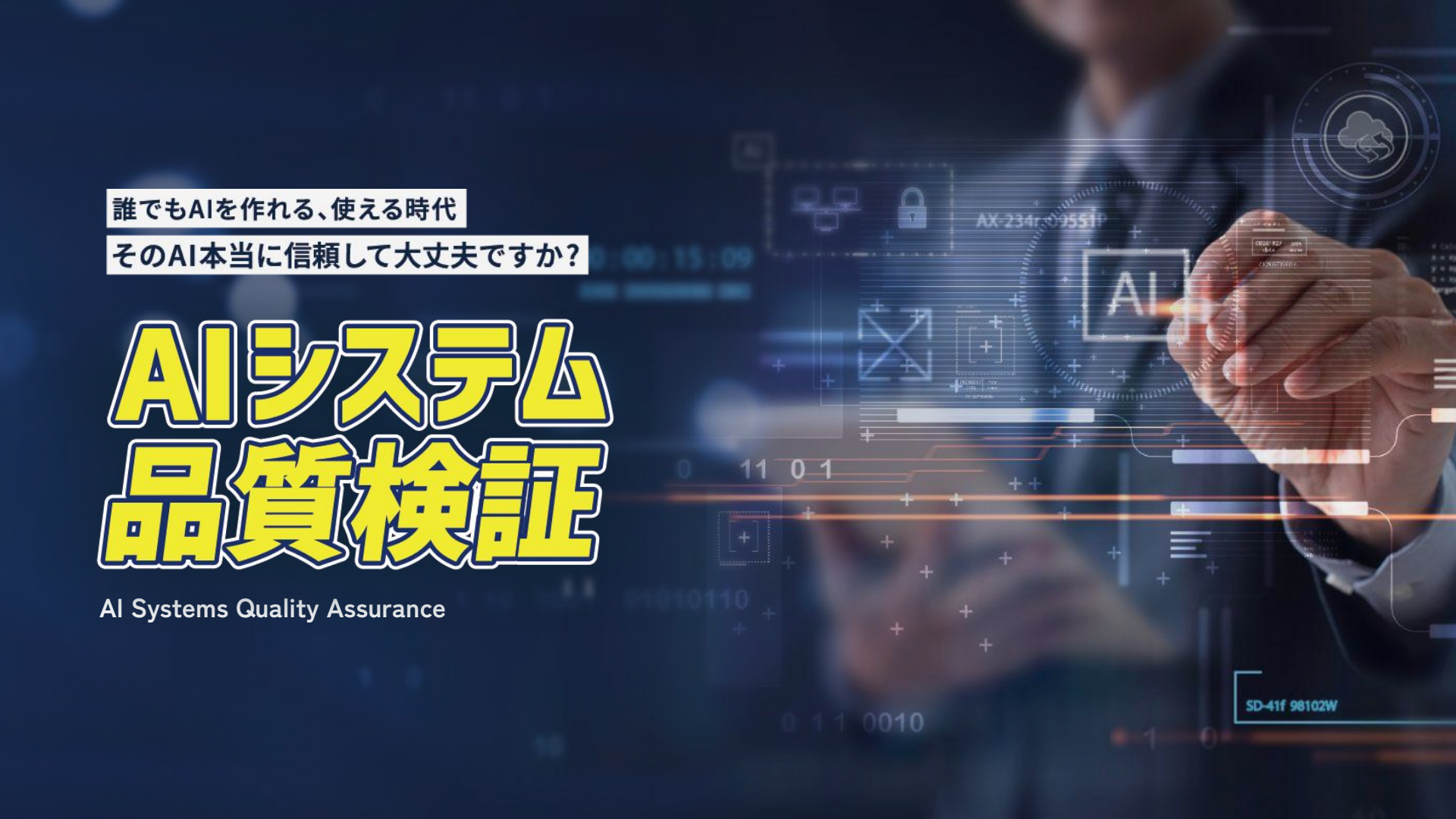
誰でもAIを作れる、使える時代

そのAI本当に信頼して大丈夫ですか？

AIシステム品質検証

AI Systems Quality Assurance



AIが当たり前になった現代、
AI時代ならではの課題が広がっています

課題 1 顧客向けAIサービスの品質保証

例：カスタマーサポートチャットボット・注文受付AI・AI審査システム

- AIが誤回答しても、リリース前に発見できる仕組みがない
- 「ここまで確認すればOK」という判断基準がない
- 本番稼働後、品質が劣化していることに気づけない

顧客への誤情報提供による
法的責任・ブランド毀損

課題 2 社内向けAIシステムの品質保証

例：社内文書検索AI・業務申請自動化AI・社内FAQボット

- AIシステムの品質を検証する仕組みがない
- モデル更新・データ変更時の品質変化を確認できない
- 社内AIが扱う顧客データの漏洩リスクを把握できない
- 社内向けAIの出力がCSなど顧客対応の現場で使われ、サービス品質に直結している

機密情報・個人情報の流出による
コンプライアンス違反・損害賠償

この状態が続くと、何が起きるか

AIインシデント件数増加

+55%

2024年→2025年（Workato 2026）

データ侵害の平均対応コスト

5.5億円

IBM 2025年 日本企業

企業への信頼低下

71.4%

ギャプライズ 2026年

AIの回答が購買行動に影響

80%以上

ギャプライズ 2026年

AIシステムに潜む、4つのリスク

従来型ソフトウェアとの構造の違いと、AI固有のリスク

出力品質リスク

テストで合格した品質が本番環境で維持される保証がない

テスト時と同じ結果が常に得られるとは限らず、
動作の再現性を担保しにくい性質

セキュリティリスク

悪意ある入力により、AIへの指示が乗っ取られることがある

本来出力すべきでない情報を開示したり、
システムの動作指示を上書きされたりする脆弱性

AIエージェントの誤動作リスク

人間が意図しない操作を行うことがある

自律的に複数の処理を実行するAIエージェントは、
意図しないファイル削除・誤送信・連鎖処理などを
「良かれと思って」行うことがある
処理の内容を事後に追跡することも困難

バイアス・倫理リスク

**AIは学習データに含まれるバイアスを
そのまま出力に反映させることがある**

特定の条件で不公平な判断をしたり（例：順序バイアス）、
悪意ある入力によって有害なコンテンツを出力するリスク

これらのリスクは、従来のソフトウェアテストの手法だけでは検出・防止できません

国内AIインシデント事例と規制の最新動向

AI品質リスクの放置が企業にもたらす経営上の懸念

財務的損害

損害賠償・緊急対応コストが企業を直撃

- AI誤作動・情報漏洩に起因する損害賠償・対応費用
- 日本企業の平均コスト：5.5億円（IBM調査 2025年）
- 業務停止・顧客補償も追加発生

法的責任

「AIだから免責」は法廷で通用しない

- AI出力を根拠にした意思決定の損害責任は利用側へ
- AI提供者でなく利用事業者が問われる判例が確立しつつある

ブランド毀損

炎上は数時間でグローバルに拡散する

- AIトラブルはSNSで制御不能な速さで広がる
- 失った顧客信頼の回復には数年単位の投資を要する

規制違反

行政指導・調達基準除外のリスクが高まっている

- 個人情報保護委員会による行政指導・社名公表の対象に
- 大手取引先の調達基準除外、IPO審査での問題化も現実的

AIのリスクは、現行のシステム内でも潜在的に進行している可能性があります

AIの品質問題を放置すると、何が起きるか

他社の話ではありません。同様の状況が貴社にも起きうる可能性があります

SaaS・プロダクト

カスタマーサポートAI 炎上・損害賠償請求(国内)

AIチャットボットが差別的な回答を出力し大規模炎上。
弁護士経由で3,000万円の損害賠償請求に発展し、
「AIの発言は会社の公式見解」として使用者責任が問われた事例。

SaaS・情報漏洩

対話型AIサービス プロンプト・個人情報漏洩(国内)

AIサービスの周辺DBの設定不備により、ユーザーのプロンプトや
LINE ID・メールアドレスが第三者から閲覧可能な状態に。
外部ユーザーの指摘で発覚し、全面システム改修を余儀なくされた。

AI事業者ガイドラインにおける規制化の方向性

2024年4月
v1.0 策定

全AI事業者にガバナンス体制の構築を推奨

2026年3月
v1.2 改訂（最新）

利用者・提供者・開発者への具体的義務要件を明文化

今後の見通し

欧米の義務化を先例に法的強制力への進行が見込まれる

このまま義務化が進むと…

- 大手取引先の調達条件から除外
- IPO審査・内部統制の確認項目化
- 個情委：行政指導・社名公表
  （現行法でも適用）

AI固有のリスクに対応する専門的なテストの必要性

従来のQAをそのままAIに適用することはできません
そのため、AI固有のリスクに対応した専門的なテストが追加で求められます

従来のシステムテスト

- **機能テスト**（仕様通りに動くか）
- **表示テスト**（画面表示は正しいか）
- **性能テスト**（応答速度・負荷）
- **セキュリティテスト**（基本的な脆弱性）
- **回帰テスト**（変更による影響確認）

＋

AI固有のテスト

- **ハルシネーション検知**
- **回答の一貫性・ゆらぎ確認**
- **プロンプトインジェクション耐性**
- **RAG参照精度の確認**
- **バイアス・公平性テスト**
- **ロバストネス**（異常入力耐性）

一般的な方法論が標準化されていない今、
ProVisionはこれまでの実案件を通じて、AI特有のリスクを包括した検証観点を体系化

検証観点の詳細 —— 4つの層

各レイヤーでの検証内容、および担当範囲の補足

出力品質・正確性

ハルシネーション検知

- AI誤作動・情報漏洩に起因する損害賠償・対応費用

RAGテスト

- 読み込み資料が正確に参照されているか1資料3観点で検証

回答の関連性テスト

- 入力に対し適切な回答が返るか確認(入力ゆらぎ耐性含む)

安全性・倫理・ガバナンス

バイアス・公平性テスト

- AIが特定条件下で偏った判断・回答をしていないかを確認

有害性テスト

- 有害コンテンツの誘発がないか一般倫理観のもと確認

セキュリティ

プロンプトインジェクション

- 指示乗っ取り・情報開示誘発など約10パターンで検証

ロバストネス・非機能

入力ゆらぎ耐性テスト

- 別称・言い換え・誤字など異なる表現での挙動を検証

応答速度テスト

- 通常負荷時の応答速度・タイムアウト条件を確認します

※ 上記以外の観点はご要望・案件内容に応じてご相談ください。

サービスの進め方

ヒアリングから継続的なモニタリングまで、ProVisionは4つのステップでAIシステムの品質に向き合います

step 1 ヒアリング・現状把握

AIシステムの概要・利用状況・品質への懸念をヒアリングし「何をどう確認するか」の方針を共同設計

- AIシステムの現状整理
- 品質懸念の言語化
- 検証の方向性

step 2 検証観点の設計

ヒアリング内容をもとに4層の中から御社のAIに適した観点を優先度づけして設計

- 検証方針の合意
- 検証スコープの確定

step 3 検証の実施

設計した観点に基づきテストを実施。確認した内容や気になる挙動を丁寧に整理

- 品質の現状把握
- 課題・懸念点の可視化
- 改善ポイントの整理

step 4 報告・継続対応

確認内容を整理し、改善の方向性を一緒に検討。定期的なQBRで継続的な品質維持を支援

- 改善への着手点
- 改善施策の方向性
- 定期的な品質レビュー

なぜ、継続的に検証し続けることが必要なのか

AIの品質は、一度の検証で永続的に保たれるものではありません。
主要モデルのバージョンアップ、RAGが参照するデータの更新、AI環境の切り替え、AIガイドライン等の動向による評価基準の変化―リリース後もこうした変化が常に発生します。
これらに対応するため、継続的なモニタリングと定期的な再検証が不可欠です。

ProVisionが選ばれる3つの理由

どのAI環境にも対応できる柔軟性と対応力、深い専門知見で、お客様のAIシステムの品質を一貫して守り続けます

01 どのAI環境にも対応可能

GPT・Gemini・Claude・社内LLM—
どのAI製品を使っていても対応可能です
マルチベンダー対応だからこそ、
お客様のAI環境に最適化した
柔軟な検証方法を提案できます。

02 現場に入り込み、品質を守り続ける

提案やレポートに留まらず、エンジニアが現場に深く入り込み、お客様のチームと一体となって品質を守り続けます。
深く関与するほど、**AIの変化に的確に対応できるパートナー**になります。

03 QA専業20年の信頼基盤

創業2005年・QA一筋20年。
取引社数約700社・
リピート率90%以上。
既存QA顧客との信頼関係を起点に、
AIフェーズへスムーズに移行できます。

事例のご紹介 1

1 背景・課題

AIエージェントを顧客向けサービスで本番稼働していたが、品質基準が未定義のまま運用。
「意図しない回答をしていないか」という不安を抱えながら、確認方法がわからない状態だった。

2 ProVisionが実施したこと

- 4層（出力品質・安全性・セキュリティ・ロバストネス）で体系的に検証
- 全網羅でなくリスク重点のセッションベースドテストを採用
- 問題指摘から修正後の確認テストまで一貫担当

結果 1	結果 2	結果 3
品質基準が言語化 **リリース判断の軸が確立**	別称入力の誤認識・ 順序バイアス等の問題を複数発見 **修正まで支援**	品質保証のノウハウが内製化 **リリースを自走できる状態に**

事例のご紹介 2

1 背景・課題

内製AIテスト設計ツールを導入していたが、「本当に効果が出ているのか」を評価する手段がなかった。
ツールの社内浸透・定着も思うように進んでいない状態だった。

2 ProVisionが実施したこと

- AIツール生成テストケース(200〜500件/案件)を精査・評価し課題を整理
- モデル更新後の品質変化を継続監視・報告
- 各チームへの導入推進・定着まで継続的に伴走

結果 1	結果 2	結果 3
AIツールの効果測定基準が確立 **活用判断の根拠が明確に**	AI活用チーム立ち上げを支援 **テスト設計工数削減を実現**	継続的なパートナーシップへ発展 **長期的な連携を確立**

プラン・体制案のご紹介

ご状況に合わせて最適なプランをご提案します

まずはセキュリティのチェックから

AIシステム脆弱性診断

診断内容

01 AIへの不正操作（プロンプト操作）
02 機密情報・社内データの漏洩
03 AI出力の悪用・権限の逸脱

診断レポート

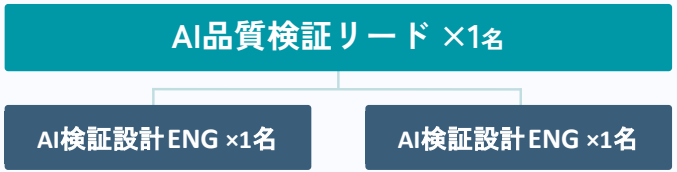
・CVSSスコア × OWASP LLM分類によるリスク評価
・RAG設定の推奨事項

50万円～（リクエスト数により変動）

※ 専門セキュリティベンダー（情報セキュリティサービス基準適合）が実施。
ProVisionが窓口となりワンストップで対応します。

より本格的な品質保証を進めたい場合はこちら

AIシステム品質検証プラン



・初回導入に最適なシンプル体制で費用を抑えてスタート
・リード主導で品質方針設計から検証・報告までワンストップ対応
・月次レポートと継続モニタリングで品質を長期維持

250万円～/月

※実際の費用はご相談のうえ、案件内容・規模に応じてご提案します

お問い合わせ

お見積りやサービス内容に関する質問など、お気軽にお問い合わせください。

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/

