

Aixis Audit Methodology

# 監査方法論書

独立系AI監査機関 Aixis が用いる

評価フレームワーク、採点ロジック、品質保証プロセスの全容。

METHODOLOGY

v1.0.0

PUBLISHED

2026年4月11日

PUBLISHER

株式会社Aixis

## CONTENTS

# 目次

1. はじめに — なぜ独立監査か	02
2. 5軸評価フレームワーク	02
3. スコア算出ロジックと採点式	03
4. グレードスケールと判定基準	04
5. 品質保証プロセス	05
6. 監査信頼度メタ評価（BenchRisk準拠）	06
7. 監査ライフサイクル	07
8. 再監査サイクルと臨時トリガー	08
9. 方法論バージョンニング	08
10. 独立性4原則	09
11. 参考リンク	09

## SECTION 01

## はじめに — なぜ独立監査か

国内のAIツール市場では、ベンダー自身による発信、有償レビューサイト、個人ブログが入り混じり、購買担当者が中立的な比較情報に到達することが難しくなっている。多くのレビューサイトは紹介手数料や広告費を主要収益としており、「高評価を付けるほど収益が増える」構造的な利益相反を抱えている。

Aixisはこの構造的課題を解決するために設計された独立系AI監査機関である。評価対象ベンダーから紹介手数料、広告費、掲載料、スポンサー費用を含むいかなる名目の報酬も一切受け取らないことを公開ポリシーとして宣言し、収益源を利用者組織からのサブスクリプション契約および個別監査レポートの販売のみに限定している。

本書は、platform.aixis.jp で公開している監査プロトコル、監査プロセス、透明性ポリシー、独立性宣言、スコア改訂履歴の各ページに記載された内容を、単一の参照文書として再構成したものである。購買担当者・情報システム部門・法務部門のそれぞれが社内稟議やベンダー選定資料の一次資料として引用できることを意図している。

本書に記載されている数値・手順・ポリシーはすべて platform.aixis.jp 上で公開されている内容と一致する。相違がある場合は、常にウェブ上の最新ページを正として扱う。

## SECTION 02

## 5軸評価フレームワーク

Aixisはすべての監査対象ツールを、実務適性・費用対効果・日本語能力・信頼性/安全性・革新性の5つの独立した評価軸で採点する。各軸は0.0から5.0までの連続スコアを持ち、標準の総合スコアでは5軸を均等に扱う。

軸	評価観点	重み
実務適性 Practicality	業務フローにおける到達点、UXと操作性、 既存環境への統合容易性、出力品質の一貫性	均等
費用対効果 Cost Performance	料金体系の透明性、無償枠の実用性、 有料プランの妥当性、乗り換えコスト、応答速度	均等
日本語能力 Japanese Readiness	UI日本語化、ビジネス日本語の適切性、 日本語ドキュメント、レイアウト・フォント対応	均等
信頼性・安全性 Safety	データ保護とアクセス制御、監査ログ、 保存場所と越境移転、インシデント対応体制	均等
革新性 Uniqueness	独自技術・差別化機能、エコシステム、 ロードマップ透明性、APIラッパー依存度	均等

業界別ランキングでは、ユースケースに応じた重み付けが適用される場合がある。重み付けが適用される場合は、該当ランキングページ上でその旨を明示する。

### SECTION 03

## スコア算出ロジックと採点式

各評価軸のスコアは、実環境での自動テスト結果と、AI定性評価による加重平均で算出する。自動テストと定性評価の比率は固定で、それぞれ60%と40%である。両者はいずれも0.0から5.0の範囲に正規化されたうえで合成される。

### 3.1 軸スコア

$$\text{軸スコア} = \text{自動テスト結果} \times 0.6 + \text{AI定性評価} \times 0.4$$

自動テストは実環境でのプロトコル実行により、応答速度・成功率・UI操作性・日本語処理精度などの定量データを取得する。定性評価は主としてLLMによる自動評価を基盤とし、必要に応じて手動チェックリストで補完する。

## 3.2 総合スコア

$$\text{総合スコア} = (\text{軸1} + \text{軸2} + \text{軸3} + \text{軸4} + \text{軸5}) \div 5$$

5軸の算術平均が総合スコアとなる。総合スコアは第4節で定めるグレードスケールに従い、S～Dの5段階にマッピングされる。

### SECTION 04

## グレードスケールと判定基準

総合スコアは次の閾値でS～Dの5段階に区分される。いずれの閾値および判定基準も、[platform.aixis.jp/audit-protocol](https://platform.aixis.jp/audit-protocol) に公開されている確定値を転載したものである。

グレード	総合スコア	評価	判定基準
S	4.5 – 5.0	最高評価	全軸3.0以上 かつ 総合4.5以上
A	3.8 – 4.4	高品質・推奨	全軸2.0以上 かつ 総合3.8以上
B	3.0 – 3.7	標準的	致命的欠陥なし かつ 総合3.0以上
C	2.0 – 2.9	改善の余地あり	複数軸で基準未達
D	0.0 – 1.9	要注意	重大な品質問題あり

グレードは総合スコアだけでなく、各軸の最低スコアも考慮して判定する。特定の軸が極端に低い場合、総合スコアが高くても上位グレードに判定されないことがある。

## SECTION 05

## 品質保証プロセス

監査結果の品質は次の5段階の工程によって担保される。各工程は監査プロトコル (/audit-protocol) に「品質保証プロセス」として公開されている。

### 5.1 独立評価者の選定

評価対象ベンダーとの利害関係がない評価者を選定する。利益相反チェックを事前に実施し、評価を開始する前段階で構造的な中立性を確認する。

### 5.2 ダブルチェック評価

手動評価結果はLLMによる自動評価との整合性を確認する。大きな乖離が検出された場合はレビューを実施する。将来的に複数評価者体制への移行を予定している。

### 5.3 自動テスト検証

LLMによる評価基準（ループリック）は定期的に検証・改善し、評価の一貫性を確保する。

### 5.4 最終レビュー

シニアアナリストがすべてのスコアを最終レビューし、異常値の検出とデータ整合性の確認を行う。

### 5.5 公開前チェック

公開直前に、スコア・判定・コメントが監査プロトコルに準拠しているかを最終確認したうえでデータベースに反映する。

## SECTION 06

## 監査信頼度メタ評価

Aixisは評価結果そのものの信頼性についても定量的に検証する。

AVERIが提唱するBenchRiskフレームワークに着想を得た4次元の信頼度指標を各監査セッションに対して自動算出し、監査プロトコル上で公開している。

### 6.1 再現性 (Consistency)

同一条件での再実行時にスコアが安定するか。応答時間の変動係数とエラー率で計測する。手動評価比率が高い軸は構造的に低くなる傾向があるため、手動評価の必要性の根拠にもなる。

### 6.2 正確性 (Correctness)

評価エンジンの確信度分布と、有効なエビデンス（非エラー応答）の割合で計測する。自動スコアの信頼区間を定量化する役割を持つ。

### 6.3 網羅性 (Comprehensiveness)

テスト計画の完遂率とカテゴリカバー率で計測する。基本作成・構成力・日本語品質・正確性・応用機能の各カテゴリを網羅的に実行したかを評価する。

### 6.4 解釈性 (Intelligibility)

結果の解釈しやすさを評価する。応答データの充実度、軸スコアの詳細・強み・リスク情報の付与率で計測する。

## SECTION 07

## 監査ライフサイクル

Aixisの監査プロトコルは次の5段階で構成される。各段階の成果物は監査担当者間のレビューを経て次段階に進む。

### 01 ツール登録・選定

市場のAIツールを網羅的にリストアップし、監査対象として登録する。カテゴリ分類・基本情報の収集を実施し、評価準備を整える。

### 02 実環境テスト実行

実際の利用環境でテストプロトコルに基づく操作を実行し、客観的な定量データを収集する。応答速度・成功率・UI操作性・日本語処理精度などを網羅する。

### 03 AI品質評価

定量テストで捕捉できない品質を、AI解析による自動評価を基盤とし、必要に応じて手動チェックリストでの補完評価を実施する。UX品質・ドキュメント・サポート・セキュリティ体制を精査する。

### 04 スコア算出・グレーディング

自動テストとAI定性評価の結果を統合し、第3節のロジックで5軸スコアを算出する。0.0～5.0の精密スコアに基づき第4節のグレードを付与し、詳細分析レポートを作成する。

### 05 データベース公開

スコア・グレード・詳細分析を監査データベース(platform.aixis.jp/tools)に反映し、利用者が閲覧・比較できる状態で公開する。

## SECTION 08

## 再監査サイクルと臨時トリガー

すべての監査対象ツールは原則として90日サイクルで定期再監査の対象となる。再監査では前回と同一のテストケースに加え、新たに追加されたテストケースも適用される。

定期サイクルを待たずに臨時再監査を実施するトリガーとして、AIモデルの大幅な変更（基盤モデルの切り替え等）、重大なセキュリティインシデント、料金体系・ポリシーの大幅な変更が挙げられる。

再監査の対象となる方法論の変更は、スコア改訂履歴ページ(platform.aixis.jp/score-change-log) に時系列で公開される。各リビジョンで何が変わり、どの範囲が再監査対象となったかを、改訂と同時に明示する。

## SECTION 09

## 方法論バージョンニング

監査方法論はセマンティックバージョンニングに準拠する3区分で管理される。公開されるすべてのスコアには、算出に用いた方法論バージョンが明示される。

- ・ Major — 評価軸の追加・削除、配分比率の大幅変更。全ツールが再監査対象となる。
- ・ Minor —  
計測項目の追加・テストプロトコルの改善。該当カテゴリのみが再監査対象となる。
- ・ Patch — 表記揺れ・文言の整理・誤記訂正。スコアへの影響はない。

スコア改訂履歴ページは追記専用（append-only）で運用される。過去エントリの内容は、訂正注記を伴う形でのみ更新される。

## SECTION 10

## 独立性4原則

Aixisは監査の独立性を制度的に担保するために、以下の4原則を透明性ポリシー(platform.aixis.jp/transparency)として公開している。

### 10.1 ベンダーからの報酬受領の完全禁止

紹介手数料、広告費、掲載料、スポンサーシップその他いかなる名目においても、評価対象ベンダーからの報酬を受け取らない。

### 10.2 収益源の限定と公開

収益は利用者組織からのサブスクリプション契約および個別監査レポートの販売に限定される。収益モデルは常に公開され、変更がある場合は事前に告知する。

### 10.3 掲載順序の中立性

ツールの掲載順序はスコアに基づいてのみ決定される。有料枠、優先掲載、スポンサード表示は一切存在しない。

### 10.4 評価者の利益相反チェック

手動評価を行う評価者は、評価対象ベンダーとの利害関係がないことを事前に申告・確認する。

## SECTION 11

## 参考リンク

本書の各章は、以下のページの公開内容を出典としている。

- ・ 監査プロセス — [platform.aixis.jp/audit-process](https://platform.aixis.jp/audit-process)
- ・ 監査プロトコル詳細 — [platform.aixis.jp/audit-protocol](https://platform.aixis.jp/audit-protocol)

- ・ 透明性ポリシー — [platform.aixis.jp/transparency](https://platform.aixis.jp/transparency)
- ・ 独立性宣言 — [platform.aixis.jp/independence](https://platform.aixis.jp/independence)
- ・ スコア改訂履歴 — [platform.aixis.jp/score-changelog](https://platform.aixis.jp/score-changelog)
- ・ よくある質問 — [platform.aixis.jp/faq](https://platform.aixis.jp/faq)
- ・ 監査データベース — [platform.aixis.jp/tools](https://platform.aixis.jp/tools)
- ・ コーポレートサイト — [aixis.jp](https://aixis.jp)

本書の内容についてのお問い合わせ先：

株式会社Aixis [info@aixis.jp](mailto:info@aixis.jp) ・ [platform.aixis.jp/contact](https://platform.aixis.jp/contact)