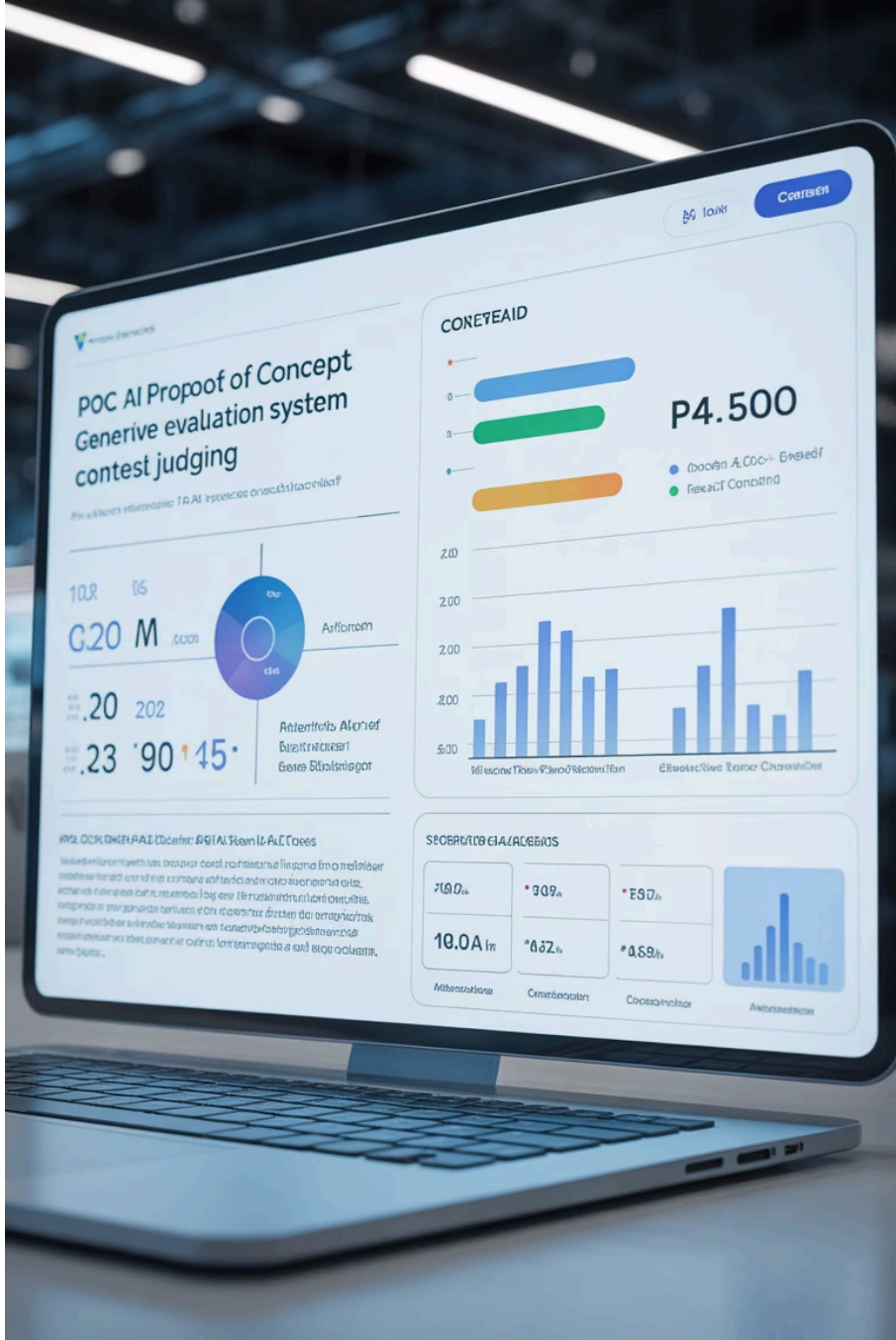
表彰コンテスト審査における生成AI評価システム PoC実施提案

ビジネスプラン、製品・サービス、学術論文、社会課題解決、また従業員表彰制度など様々な表彰事業の審査プロセスの課題に対して、生成AI評価システムを適用し、実用性を検証するPoCサービスをご提案いたします。

貴組織の実際の応募データを用いた精度検証を通じて、最適なシステム設計の方向性を確定します。

Award of



エグゼクティブサマリー

本提案は、生成AI×ルーブリック評価システムの本格導入に先立ち、**詳細ヒアリング(1-2週間)**と**PoC(概念実証)(4-6週間)**を実施するものです。

実際の過去応募データを用いてAI評価の精度・実用性を検証し、貴表彰事業にとって最適なシステム設計の方向性を確定します。投資判断に必要な具体的データと実装計画を提供いたします。

早期導入により、審査プロセスの効率化と公平性向上を実現し、応募者へのより充実したフィードバック提供が可能になります。

提案概要

PoC特別割引価格: ¥550,000

実施期間: 6-8週間

成果物:

- **AI評価精度レポート**
- **デモシステム**
- **実装推奨プラン**
- **ROI詳細試算**

詳細ヒアリング実施計画

01

初回ヒアリング(Week 1)

事務局責任者、審査委員長、Award Force運用担当者（導入済みの場合）との対面/オンラインミーティング(2-3時間)を実施します。現状の審査プロセス、応募データの詳細、評価基準の運用状況、Award Forceの利用状況について包括的にヒアリングいたします。

02

データ収集(Week 1-2)

過去3年分の応募要項、応募書類サンプル(10-20件)、審査員向けガイドライン、評価シート、過去の審査結果データ、Award Force設定情報などをご提供いただきます。機密情報はNDA締結後に安全に取り扱います。

03

中間報告(Week 2)

ヒアリング結果を分析し、現状分析サマリー、課題整理、AI評価システム要件定義(初版)、PoC実施計画書を作成します。PoC計画承認会議(1-2時間)で合意形成を図ります。

ヒアリングの3つの重要項目

審査プロセス詳細

応募受付から結果発表までの全体フロー、審査段階、各段階での審査方法と評価基準、審査委員の構成、所要時間などを詳細に把握します。

応募データ分析

年間応募件数の推移、カテゴリー別内訳、応募書類の構成、添付資料の種類、書類のボリュームなどを確認します。

評価基準運用

現行ルーブリック等の評価基準の使われ方、審査員による解釈のばらつき、過去の評価データ、改善要望について理解を深めます。

PoC(概念実証)実施計画

実際の過去応募データを用いて、生成AI評価システムの評価精度、実用性、技術的実現可能性、費用対効果を4-6週間かけて徹底的に検証します。

1

Phase 1: 環境構築

Week 1-2

Dify開発環境セットアップ、GPT-4 Turbo API接続、ルーブリック評価ワークフロー構築、ナレッジベース構築、評価ロジック実装を行います。

2

Phase 2: AI評価実施

Week 3-4

過去応募データ10-20件を選定し、AI評価を実施します。複数のプロンプトパターンとモデルで比較検証を行い、処理時間とコストを計測します。

3

Phase 3: 精度検証

Week 4-5

人間評価との相関係数算出、項目別一致度分析、審査員2-3名による妥当性評価を実施、定量・定性の両面から精度を検証します。

4

Phase 4: 連携検証

Week 5

Award Force APIアクセス確認、Zapier連携テスト、データフォーマット、セキュリティ検証を行い、実装可能性を確認します。

5

Phase 5: 最終報告

Week 6

デモシステム構築、最終報告書作成、報告会実施(1-2時間)を行い、実装推奨プランとROI詳細試算を提示します。

検証する4つの重要指標

評価精度の検証

AI評価と人間評価の一致度を測定します。総合得点の相関係数(目標r>0.7)、項目別得点の相関、受賞/落選判定の一致率(目標>80%)を算出し、評価の信頼性を定量的に検証します。

実用性の検証

審査業務への適用可能性を評価します。審査委員による評価妥当性スコア、コメント品質評価、処理時間とコストの測定を通じて、実務での活用可能性を判断します。

技術的実現可能性

Award Force連携の確認を行います。Zapier連携テスト、データフォーマット検証、エラーハンドリング確認、セキュリティ検証を通じて、システム統合の技術的課題を特定します。

費用対効果の算出

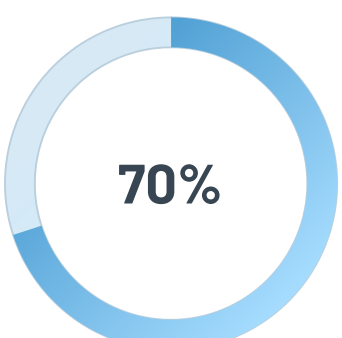
ROIの定量的試算を実施します。審査時間の短縮効果、人件費削減額、3年間の累計効果を算出し、投資判断に必要な具体的データを提供します。

PoC成功基準と期待される成果

定性的成功基準

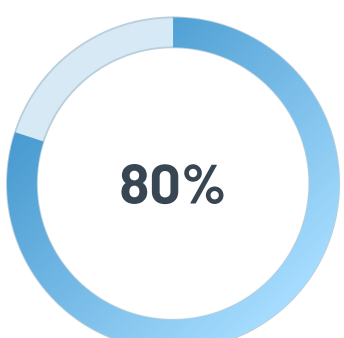
- **審査委員の受容度**
  審査委員の過半数がAI評価を「有用」と評価し、審査業務への統合に前向きな姿勢を示すこと。
- **技術的実現可能性**
  Award Forceとの連携に重大な技術的懸念がなく、本番環境での運用が現実的に可能であること。
- **投資対効果**
  3年間のROIがプラスになる見込みがあり、長期的な費用削減効果が期待できること。
- **セキュリティ適合**
  セキュリティ・コンプライアンス要件をクリアし、データ保護が十分に確保されること。

定量的成功基準



相関係数

総合得点の相関係数r>0.7を達成(目標r>0.8)



判定一致率

受賞/落選判定の一致率>80%(目標>90%)

評価妥当性

審査員による評価妥当性スコア平均4.0以上

PoC費用詳細と含まれるサービス

¥550,000（税抜）

費用に含まれるもの

ヒアリング・分析
¥150,000

詳細ヒアリング実施、現状分析、要件定義、課題整理を含みます。

開発・実装
¥150,000

Difyプロトタイプ開発、ワークフロー設計、RAGナレッジベース構築を含みます。

評価・検証
¥120,000

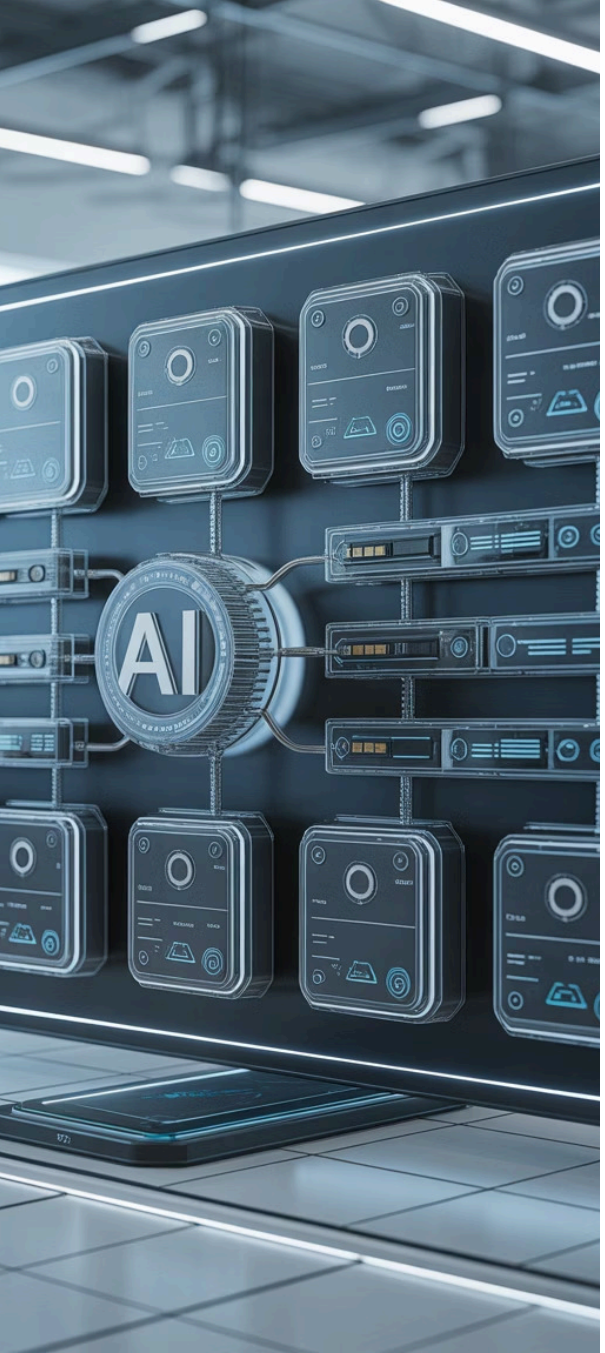
10-20件のAI評価実施、複数パターン検証、精度分析レポート作成を含みます。

連携・管理
¥130,000

Award Force連携検証、プロジェクト管理、週次進捗報告、最終報告会を含みます。

追加で含まれるサービス

- OpenAI API使用料(PoC期間中)
- Dify Cloud利用料(PoC期間中)
- Zapier Professional利用料(1ヶ月)
- デモシステム(1ヶ月間利用可能)
- 最終報告書(詳細分析付き)
- 実装推奨プラン
- ROI詳細試算
- 技術ドキュメント一式
- 週次進捗レポート



AI評価システムの技術アーキテクチャ

Award Forceと連携した生成AI評価システムは、最新のAI技術とクラウドサービスを組み合わせた堅牢なアーキテクチャで構築されます。

Award Force

応募データの一元管理プラットフォーム。応募書類、評価シート、審査結果を統合管理します。

Zapier連携

Award ForceとDifyを接続する自動化プラットフォーム。データの取得・送信を自動処理します。

Dify + GPT-4

AI評価エンジン。ルーブリック評価、RAG検索、スコア算出、コメント生成を実行します。

評価結果

詳細な評価レポート。項目別スコア、総合評価、改善提案コメントを審査員に提供します。

生成AI評価の急速な普及

人事、教育、パフォーマンス評価など、様々な分野でAI評価システムの導入が急速に進んでいます。これらの成功事例は、顕彰制度への適用においても大きな可能性を示しています。

81%

人事・採用分野

2027年までに81%の組織がAI評価を導入するとGartnerが予測。現在、88%の企業がHR機能でAIを活用し、審査ツールの精度は89-94%を達成しています。

92%

教育評価

2025年までに大学生の92%がAIツールを使用すると見込まれ、高等教育のテスト採点においてAI導入が50-100%に達すると予想されています。

70%

パフォーマンス評価

2025年までに70%の組織がAIを活用すると予測されており、リーダーシップ評価では80%の予測精度を実現しています。

顕彰制度への適用

これらの先行事例を踏まえ、ビジネスコンテストや社内審査といった顕彰制度においても、今後3〜5年でAI評価の導入が大幅に拡大すると予想されます。早期導入により、審査の公平性向上、効率化、応募者へのより充実したフィードバック提供が実現します。

よくあるご質問(FAQ)

Q1. PoCで使用したシステムは本実装でそのまま使えますか?

PoCは検証目的のため、本番環境の20-30%程度の完成度です。本実装では、パフォーマンス最適化、エラーハンドリング強化、UI改善などが必要になります。ただし、PoCで開発したコアロジックやプロンプトは最大限活用します。

Q2. PoC期間中に審査シーズンと重なった場合は?

過去年度のデータで実施するため、現在進行中の審査には影響しません。むしろ、審査シーズンの課題を観察する良い機会になります。リアルタイムでの課題把握により、より実践的なシステム設計が可能になります。

Q3. 過去データが十分にない場合は?

最低10件(受賞7件、落選3件)あれば実施可能です。ただし、精度検証の統計的信頼性は件数に依存するため、20件以上を推奨します。データが不足する場合は、複数年度のデータを組み合わせることも検討します。

Q4. AI評価が人間より厳しい/甘い場合は?

プロンプト調整により評価傾向をコントロール可能です。PoC中に「厳格モード」「標準モード」「寛容モード」を比較検証し、最適なバランスを見つけます。審査委員の評価傾向に合わせたカスタマイズも可能です。

Q5. 特定カテゴリーだけ精度が低い場合は?

カテゴリーごとに専用プロンプトを作成することで対応可能です。拡張プランではカテゴリー別の詳細分析を行い、各カテゴリーの特性に最適化された評価ロジックを開発します。

Q6. セキュリティ監査は可能ですか?

可能です。貴社のセキュリティポリシーに応じて、第三者監査の受け入れやセキュリティチェックリストの提出に対応します。追加費用が発生する場合は事前にお見積もりいたします。

審査プロセスDXの最初の一歩！

生成AI技術は審査プロセスを大きく変革しますが、その導入には慎重な検証が不可欠です。PoCは大きな投資前の「試着」のようなもの。

貴組織のコンテストが、より公平で効率的、応募者にとって価値ある顕彰制度へ進化するための第一歩として、ぜひPoC実施をご検討ください。Award ofが貴組織の成功を全力でサポートいたします。

無料相談を予約

株式会社テンプロクシー
mo4ma プロジェクト Award ofサクセスチーム
Email: info@award-of.net
Tel: 03-3413-2267
東京都世田谷区下馬2丁目11-3-104