

CDATA

ホワイトペーパー

25% の 精度ギャップ

業務システムにおける MCP プロバイダーの性能比較

CRM、プロジェクト管理、データウェアハウス、ERP の 4 プラットフォームにわたる 378 件のクエリの結果です。

このベンチマークを実施した理由

MCP (Model Context Protocol) は、AI エージェントと業務システムをつなぐデフォルトのインターフェースになりつつあります。企業がチャットボット型のコパイロットから、リアルタイムの業務データを読み書きし操作する自律型エージェントへと移行するにつれて、MCP プロバイダーは変換レイヤーとしての役割を担います。自然言語のクエリを受け取り、適切な API 呼び出しに変換し、構造化された結果を返すのです。

その約束はシンプルです。営業担当者が AI エージェントにパイプラインデータの抽出を頼めば、エージェントは MCP プロバイダー経由で CRM に問い合わせ、回答を返します。プロジェクトマネージャーがスプリントバックログの内容を尋ねれば、エージェントはプロジェクト管理プラットフォームに問い合わせ、該当する issue を表示します。手動でのフィルタリングや、ツール間のコンテキストスイッチは不要です。

しかし、実際にはどの程度うまく機能しているのでしょうか。MCP プロバイダーの評価の多くは、接続の網羅性—どれだけのプラットフォームに対応しているか—に焦点を当てており、正解率、つまり「そのクエリに対して正しいデータを返しているか」は見過ごされがちです。これは看過できないギャップです。CRM に接続はできても「今四半期にクローズする商談」を誤って解釈する MCP プロバイダーは、接続していない場合よりも悪質です。間違った回答が、正しい回答のように見えてしまうからです。

本ベンチマークは、そのギャップを埋めることを目的としています。5 つの MCP プロバイダーを 4 つの業務プラットフォームにわたって 378 件の実際のクエリでテストし、シンプルな 1 点を測定しました。プロバイダーは正しい結果を返したか、返さなかったか、です。

検証内容

市場における主要な設計アプローチを代表する 5 つの MCP プロバイダーを評価しました。CData Connect AI（リレーショナル／セマンティックレイヤー）、統合 API プロバイダー、CRM ネイティブの MCP サーバー、MCP ゲートウェイ、iPaaS ベースのプロバイダーです。それぞれを、CRM、プロジェクト管理、データウェアハウス、ERP の 4 つの業務プラットフォームに対してテストしました。

各プロバイダーは、プラットフォームごとに 16 件の標準化されたプロンプトを実行しました。単純な検索から複数条件のフィルタリングクエリ、書き込み操作まで多岐にわたります。これにより合計 378 件のテスト実行が行われました。評価者は、事前に設定された正解データに基づいて各回答を採点しました。MCP プロバイダーが正しいデータを返したか、返さなかったかのいずれかであり、部分点はありません。

プロンプトは、実際の業務を反映するように設計されました。

- 「今四半期にクローズ予定の商談が何件あるか教えてください」 — 会計カレンダーの認識と相対的な日付計算が必要
- 「Alex Johnson にアサインされ、ステータスが『To Do』の issue をすべて一覧表示してください」 — 異なるフィールドをまたいだ複数条件のフィルタリングロジックが必要
- 「スプリント 112 の issue を、優先度が高い順にすべて抽出してください」 — スプリントのナビゲーションと、テキストベースの優先度の列挙値に対する意味的な理解が必要
- 「商談『Acme Corp Deal』のステージを『契約書送付済み』に変更してください」 — 書き込み操作におけるワークフローの検証が必要
- 「過去 30 日間に作成されたすべての注文を検索してください」 — 相対的な日付計算と正しいフィールドマッピングが必要

これらは特殊なケースではありません。営業チーム、プロジェクトマネージャー、財務アナリストが日常的に尋ねる、ごく普通の質問です。

全体の結果

ほとんどの MCP プロバイダーは、15-42% の頻度で誤った結果を返しました。これは、およそ 2〜3 件に 1 件の割合でエラーが発生することを意味します。提案エンジンであれば許容できる失敗率かもしれませんが、自律型エージェントにとっては受け入れられないレベルです。

プラットフォーム	CData の正解率	他社プロバイダー	CData との差
CRM	100%	75%-100%	差なし
プロジェクト管理	94%	45%-50%	+45-50 pt
データウェアハウス	100%	75%	+25 pt
ERP	100%	20%	+80 pt
全体	98.5%	65%-75%	+25 pt

CData Connect AI は 98.5% の正解率（68 件中 67 件が正解）を達成しました。他のプロバイダーは 65% から 75% の範囲でした。25 ポイント以上のギャップはプラットフォームを問わず一貫して見られましたが、その内訳はプラットフォームごとに異なりました。

全体のスコアと同じくらい目立ったのは、そのばらつきでした。CData はどのプラットフォームでもほぼ同じ正解率を維持しました。一方、他のプロバイダーは大きく変動しました。MCP ゲートウェイは CRM の 95% からプロジェクト管理の 50% まで落ち込みました。統合 API プロバイダーは 100% から 50% まで低下しました。iPaaS プロバイダーは 75% から 45% まで幅がありました。複数のシステムにわたって AI を展開する企業にとって、この不安定さは、単純な正解率の数字と同じくらい大きな問題です。

プロバイダーが失敗するポイント

失敗はランダムに発生したわけではありません。4つの繰り返しパターンに集約されました。

相対的な日付ロジック

15件の失敗が、相対的な日付を含むクエリで発生しました。「今四半期にクローズする商談」や「過去30日間に作成された注文」といったクエリでは、プロバイダーが文脈から日付の範囲を計算する必要があります。iPaaS プロバイダーは、2025年第4四半期という明確な期間であっても、「今四半期」が何を指すのか毎回確認を求めてきました。他のプロバイダーは、間違った日付フィールドを使用しました。クエリで `CREATED_DATE` を指定していても `ORDER_DATE` を取得するなどあるいは日付フィルタを完全に省略し、複数年にわたるデータセットを返しました。

パイプラインを確認する営業責任者にとって、これは結果が何も得られないか、手動でのフィルタリングが必要なデータの塊を渡されるかのどちらかを意味します。まさに、AIが排除するはずだった作業そのものです。

複数条件のフィルタリングクエリ

12件の失敗が、2つ以上の条件を同時に満たす必要のあるクエリで発生しました。「プロジェクト Acme Project のバックログにある issue」というクエリでは、プロジェクト ID とステータスの両方が一致する必要があります。MCP ゲートウェイと統合 API プロバイダーは、バックログの6件ではなく、そのプロジェクトの50件すべての issue を返しました。プロジェクトのフィルタは適用されたものの、ステータスの条件が完全に欠落していたのです。

実務上の結果として、プロジェクトマネージャーは無関係なチケットを手動でふるい分けることになり、ツールが処理するはずだったまさにその種の雑務に時間を費やすことになりました。

意味的な解釈

8件の失敗は、優先度とソートのロジックに関するものでした。プロジェクト管理プラットフォームでは、通常、数値ではなくテキストベースの優先度の値—Highest、High、Medium、Low、Lowest—が使われます。「Highest priority」の issue を求められた際、プロバイダーは「High」の優先度の項目を返す（近いが誤り）、すべての優先度レベルを混在させる、あるいはクエリそのものに完全に失敗するという結果になりました。CData も一度だけこれを誤り、「Highest」ではなく「High」を返しました—これが全プラットフォームを通じた唯一のエラーです。

その結果、エンジニアリングチームは間違っただけの項目に取り組むことになります。本来すぐに対応すべき issue が、優先度が混在したリストに埋もれてしまうのです。

書き込み操作

7 件の書き込み操作が失敗しました。商談ステージの変更、注文ステータスの更新、issue のワークフロー状態間の移動は、いずれもプラットフォーム固有の検証ルールの理解を必要とします。どのデータを送信するかだけでなく、どの遷移が許可されているかも理解する必要があります。CRM ネイティブの MCP サーバーは、プラットフォームへの直接アクセスを持っていたにもかかわらず、すべての商談ステージの更新に失敗しました。iPaaS プロバイダーは不安定な挙動を示しました。最初の試行では失敗し、リトライで成功するというもので、検証処理の扱いが不安定であることを示唆しています。

書き込み操作が失敗すると、業務プロセスが停滞します。商談は前進しません。ステータスは更新されません。自律型エージェントは、手動でのフォローアップの待ち行列を作り出してしまいます。

設計が重要な理由

これらの失敗パターンは、パッチで対処できる些細な癖ではありません。MCP プロバイダーの動作の仕方における、根本的な設計上の分かれ道を反映しています。

ほとんどのプロバイダーは、自然言語のクエリを直接 REST API の呼び出しに変換します。これは「ID で連絡先を取得する」といった単純な検索であれば問題なく機能しますが、クエリが何らかの解釈を必要とする場合—相対的な日付の計算、複数条件のフィルタリング、「Highest」が「High」の同義語ではなく特定の列挙値であるという理解、あるいは特定の CRM のワークフローでどの商談ステージの遷移が有効かの把握—になると破綻します。

CDATA のアプローチは、エンティティ間の関係性、ビジネスロジック、プラットフォームの規約を理解するセマンティックモデルを備えた、リレーショナルな抽象化レイヤーを使用します。これによって、「今四半期」を確認なしに処理し、複数のフィルタを正しく適用し、ワークフローのルールに沿って書き込み操作を検証できます。

ERP の結果はこれを明確に示しています。ERP ベンダー自身のエンジニアが構築したネイティブの ERP MCP サーバーは完全に失敗し、単純な検索に対して「Record 'account' was not found」というエラーを返しました。iPaaS は 40% の正解率にとどまり、構成の甘いクエリが原因で結果が大きく不安定でした。CDATA は、自社のコネクタ基盤を通じて動作し、100% を達成しました。プラットフォームへのアクセスがあるだけでは十分ではありません。クエリと API 呼び出しの間にある抽象化レイヤーこそが、正確性を左右する場所なのです。

AI の導入についての意味

必要な正解率のしきい値は、ユースケースによって異なります。人間が実行前にすべての提案を確認するコパイロット型のツールであれば、75% の正解率でもおそらく運用可能です。—煩わしいものの、機能はします。人間がエラーを捕捉できるからです。

しかし、業界は自律型エージェント—CRM のレコードを更新し、ワークフローを起動し、財務データを変更し、人間を介さずに行動する AI—へと向かっています。正解率 75% では、そのエージェントは 4 回に 1 回の割合で行動に失敗します。日付ロジックのエラーは時間を浪費させます。複数条件フィルタリングの失敗は、重要な項目を埋もれさせます。意味的な誤りは、チームを間違った優先事項に取り組ませます。書き込み操作の失敗は、業務プロセスを止めてしまいます。そして重要なのは、この不正確さが積み重なっていくことです。75% の正解率を 5 段階のプロセスにわたって適用すると、プロセス全体が成功する割合は 24% 未満になります。**75% の正解率が、そのまま 75% の失敗率に変わってしまうのです。**

自律型エージェントが実用に耐えるためには、MCP プロバイダーはおおよそ 98% の正解率を超える必要があります。それを下回ると、エージェントは節約する以上の後始末の作業を生み出してしまいます。

そこに到達するための設計パターンは、今では比較的明確になっています。構文を変換するだけでなくビジネスロジックを解釈するセマンティックレイヤー、洗練されたクエリを構築できる抽象化レイヤー、そしてプラットフォーム固有の規約を理解する成熟したコネクタです。MCP プロバイダーを評価する企業は、自社のデータとクエリで検証すべきですが、本ベンチマークの結果は、設計上のアプローチこそが正解率を最も強く予測する要因であることを示唆しています。

CData について

CData は、企業内での AI とデータのやり取りに対して、接続性・コンテキスト・ガバナンスを提供します。同プラットフォームは、350 以上のデータソースにわたるリアルタイムアクセスとデータのレプリケーションを提供し、AI ワークロードの正確性を高めるために設計されたセマンティックインテリジェンスを備えています。CData は、Anthropic、Databricks、Microsoft、Google、Palantir をはじめとする 10,000 社以上の顧客の AI 基盤を支えています。

免責事項

本ベンチマークは、MCP エコシステムに関する正解率データを提供するために、CData Software, Inc. (以下「CData」) が社内ですべてのテスト、評価、分析は CData の従業員によって行われました。結果について第三者による検証は行われていません。

主な制約事項

- テストは 2025 年第 4 四半期に、Methodology セクションに記載された特定の構成を用いて実施されました
- サンプルサイズは、各 MCP サーバーのプラットフォーム対応状況に応じて異なります
- 採点は二値（オール・オア・ナッシング）方式で行われ、部分的に正しい回答も失敗として扱われました
- 「Efficiency Score」は CData が独自に開発した指標であり、業界標準ではありません
- テストの複雑さのプロファイルは、プラットフォームによって異なります
- 結果は管理されたテスト条件下のものであり、すべての本番環境でのパフォーマンスを表すものではない可能性があります

パフォーマンスの変動について：実際のパフォーマンスは、デプロイ構成、データの特性、ネットワーク条件、クエリのパターン、LLM のモデルバージョン、API の変更、プラットフォーム固有の要因によって大きく異なる可能性があります。テスト対象のプラットフォームおよび MCP サーバーは、テスト実施後に更新されている可能性があります。

専門的な助言ではないことについて：本レポートは情報提供のみを目的としており、専門的、法的、または技術的な助言を構成するものではありません。企業は、購入または導入の意思決定を行う前に、自社の具体的なユースケースを代表するデータとクエリを用いて、独自の検証を実施する必要があります。

AI のためのデータレイヤー。エージェント、アプリ、チームへ、信頼できるリアルタイムの企業データアクセスを。

CData は AI のためのデータレイヤーとして、AI の正確性・効率性・柔軟性を高めます。単一のプラットフォームが数百の企業システムにわたるライブデータを接続し、AI が正確に応答するために必要なセマンティックコンテキストを付与し、AI とデータのあらゆるやり取りを統制します。エージェントの構築、アプリケーションへのデータアクセスの組み込み、あるいは基幹システムへのセルフサービスアクセスの提供、いずれの場面でも、CData は企業データを AI で活用できる状態にします。CData は、Salesforce、Databricks、Microsoft、Google、Palantir をはじめ、10,000 社以上の顧客の AI ワークロードを支えています。