

なぜRAGの精度が上がらないのか？

点検
チェック
リスト
付き

「データ品質」から見直す 生成AI活用



はじめに

AIから「期待通りの回答」が出ず、精度も安定しない

生成AIの業務活用が進む中、多くの企業が同じ壁に直面しています。特に、社内データを活用するRAG（Retrieval-Augmented Generation）を導入したにもかかわらず、業務で使えるレベルの精度に到達しないケースは少なくありません。

その原因は、AIモデルの性能ではありません。**本質的な問題は「データ」にあります。**

本資料では、生成AI（特にRAG）の精度を左右する「データ整備」の重要性を解説し、その考え方と具体的な進め方について紹介します。



生成AI活用における点検チェックリスト

- AIの回答内容にばらつきがあり、期待した精度に安定しない
- 必要な社内情報をAIが検索できず、欲しい回答が返ってこない
- 社内データがSaaS、基幹システム、ファイルサーバー、個人PCなど複数箇所に分散している
- 部門ごとにデータが管理され、情報がサイロ化している
- 同じ内容の資料が複数存在し、どれが最新・正確かわかりにくい
- PDF、Excel、テキスト、データベースなどデータ形式が混在している
- 不要情報や重複データの整理、構造化が十分にできていない
- データをAIが検索しやすい形（チャンク化・メタデータ付与）に整備できていない
- 古いマニュアルや更新前の情報をAIが参照してしまうことがある
- データ更新のルールや運用フローが整備されていない



なぜ、データがAI精度を左右する？

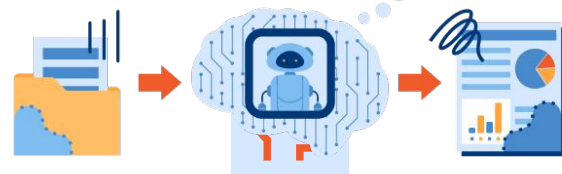
AI精度を左右する「GIGO」の法則

前頁のチェックリストで複数の項目に当てはまった方は、その原因を理解しておくことが重要です。

AIのアウトプットは、入力データの質に強く依存します。これは「GIGO (Garbage In, Garbage Out)」と呼ばれる基本原則です。RAGにおいては、AIは社内データを検索し、その結果をもとに回答を生成します。つまり、**検索対象となるデータの質が低ければ、どれほど高性能なAIであっても正確な回答は得られない**のです。

実際、RAGでは「**データの前処理や整備の質**」が精度に直結するため、適切に処理されていない場合、十分な性能が出ないケースが多く報告されています。

品質の低いデータからは
品質の低い結果しか生まれない！
Garbage in, Garbage out



RAG精度を阻害する「データ品質」における3つの課題

RAGの精度が思うように向上しない企業には、共通した構造的な課題が存在します。多くのケースでは、PoC（概念実証）の段階では一定の成果が見られるものの、実際の業務活用フェーズに移行した途端、精度に関する問題が顕在化します。具体的には、「回答内容にばらつきがある」「必要な情報が検索されない」といった現象が発生し、業務での活用に耐えない状態に陥ってしまうのです。

こうした状況に対し、AIモデルの変更やプロンプトの改善といった対策が取られることが多いものの、十分な効果が得られないケースも少なくありません。その理由は、問題が単発的な不具合や、個々のデータにおける質の問題ではなく、データの持ち方や管理方法そのもの＝「構造的な課題」に起因しているからです。

RAGの精度を阻害する要因

1

データの散在

2

フォーマットの不整合

3

情報の鮮度

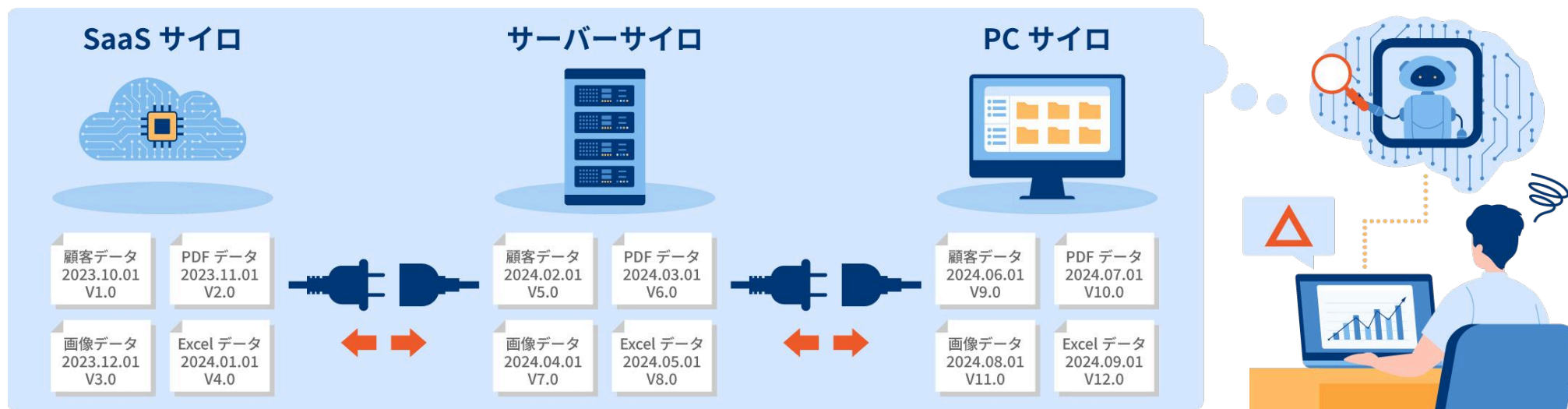
という3つの観点に集約されます。これらを構造的に捉え、**全体最適の視点で解決することが、AI活用を前進させるためには不可欠**です。

データの散在

多くの企業では、業務に必要なデータが一箇所にまとまっておらず、社内外のさまざまな場所に分散している状況がみられます。例えば、SaaS、オンプレミスの基幹システム、ファイルサーバー、さらには個人のPC内のファイル……など、データの所在は多岐にわたるでしょう。

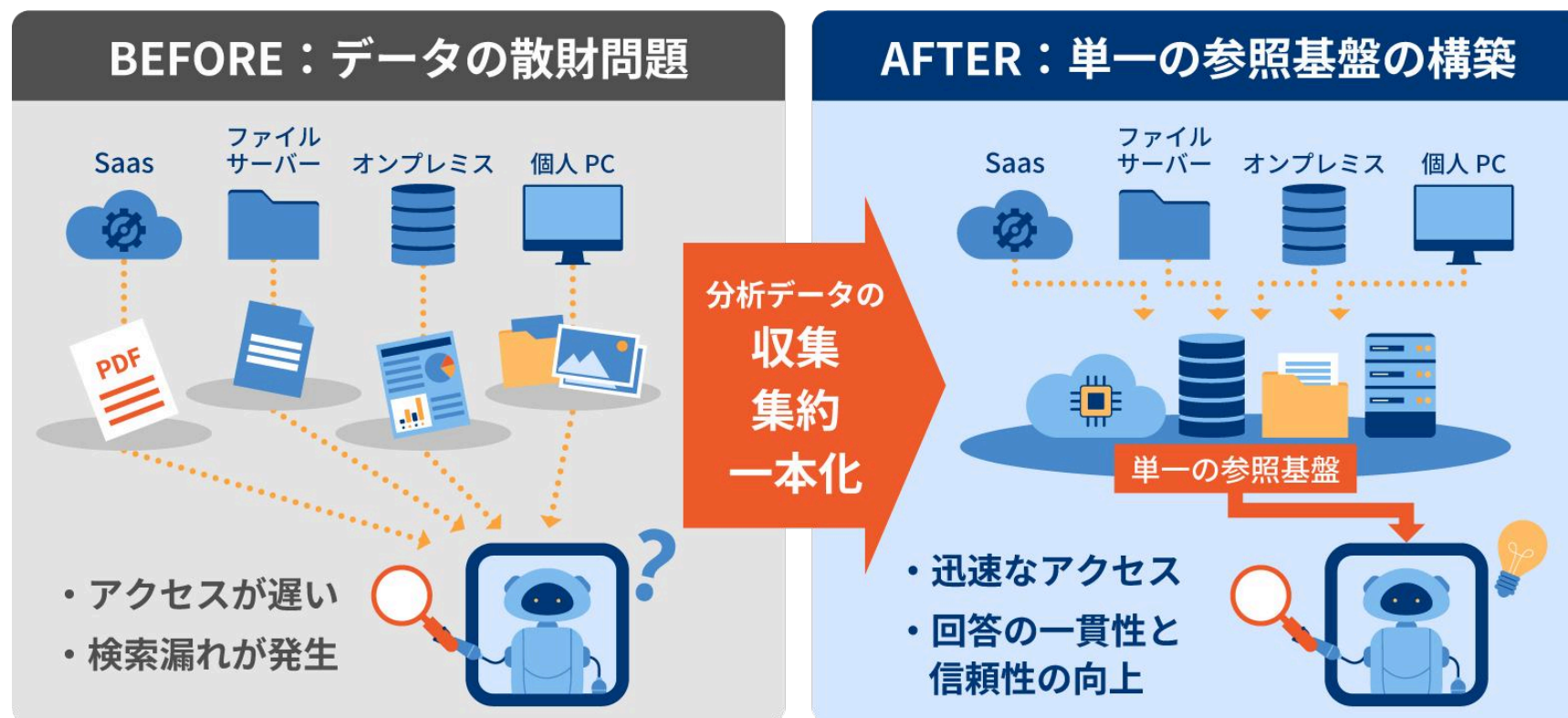
このような環境では、部門ごとにデータが管理される「サイロ化」が発生しやすく、必要な情報に一元的にアクセスすることが難しくなります。その結果、AIが参照すべき情報を網羅的に取得できず、検索漏れが発生してしまうのです。

さらに、同じ内容のデータが複数箇所に存在しているケースも少なくありません。こうした重複データはそれぞれが異なる更新状況にあることも多く、情報の不整合を引き起こす原因となります。



分散データを統合し「**単一の参照基盤**」を構築する

データの散在問題を解決するためには、分散したデータを統合し、**AIが参照する情報源を一本化することが重要**です。



具体的には、複数のシステムやサービスに散在するデータを横断的に収集・集約し、どこにどの情報があるのかを明確にすることです。これにより、必要な情報へ迅速にアクセスできる環境が整います。

さらに、統合されたデータを「**単一の参照基盤**」として整理することで、いわゆるSingle Source of Truth（信頼できる唯一の情報源）を確立できます。これにより、AIは常に同じ基準のデータを参照できるため、回答の一貫性と信頼性が大きく向上します。

フォーマットの不整合・情報の鮮度

データの散在に加えて、**データ形式の不統一**も大きな課題です。

企業内には、PDF、テキストファイル、Excel、データベースなどさまざまな形式のデータが混在しています。これらの多くは非構造データで、AIが内容を正確に理解・活用することが難しい状態にあります。十分に構造化されていないデータは検索精度を低下させ、結果としてAIの回答品質にも影響を与えるのです。

さらに見落とされがちなのが、**情報の鮮度**です。データが更新されないまま放置されていると、AIは古い情報をもとに回答してしまいます。業務においては最新の情報に基づいた判断が求められるため、これは大きなリスクとなるでしょう。

つまり、データが「存在している」だけでは不十分であり、「**AIが正しく読める状態で、かつ常に最新であること**」が求められるのです。

課題② フォーマットの不整合 (非構造データ)

データの形式が混在し、AIが内容を正確に理解・活用することが難しい



AIの検索精度を低下させる

AIの回答品質にも影響

課題③ 情報の鮮度 (古い情報)

データの更新がされないまま放置



AIは古い情報を参照し回答

業務における大きなリスク

データが「存在している」だけでは不十分であり、
「AIが正しく読める状態で、かつ常に最新であること」が求められる。

「AIが理解できる形」にデータを整備する

AIの精度を安定的に向上させるには、データを「AIが理解しやすい形」に整備することが基本です。

まず、不要なノイズや重複データを除去する「クレンジング」で、データの品質を高めます。次に、情報を意味のまとまりごとに分割する「チャンク化」で、検索精度の向上を図ります。

AI 精度向上と検索精度向上のために

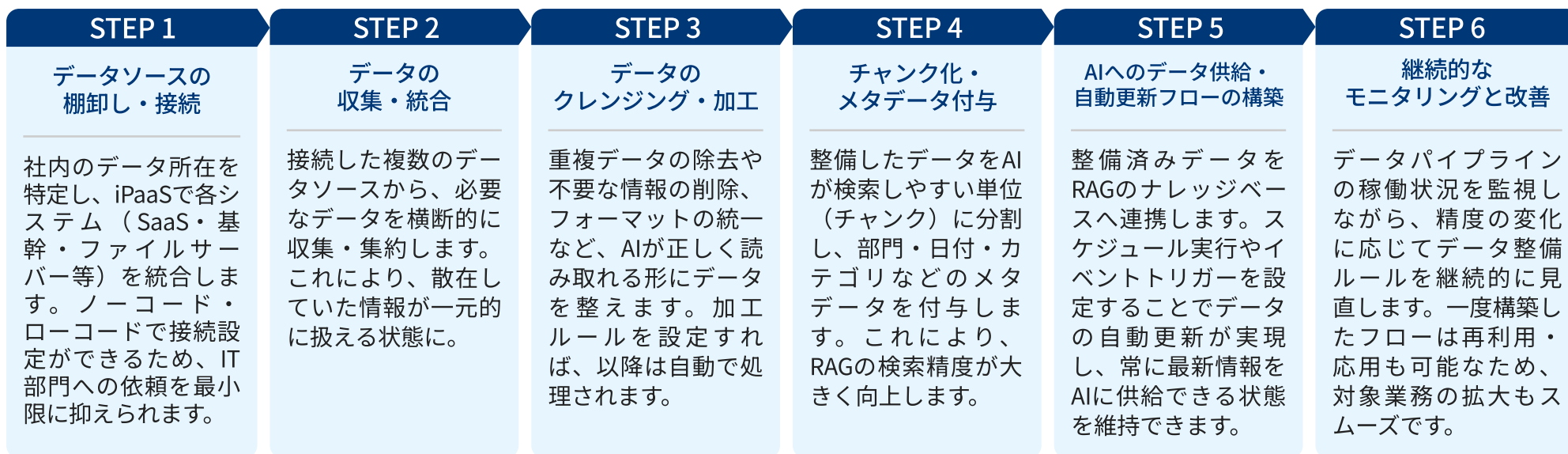
さらに、データに対して適切なメタデータを付与することで検索性を高め、必要な情報へ効率的にアクセスできるようになります。また、フォーマットの統一や構造化を進めることにより、AIがデータの内容を正確に理解できる状態を実現できるでしょう。



これらの処理は一度きりで終わらせるのではなく、データの更新フローを設計し、**継続的に**運用していくことが重要です。

iPaaSで実現する、データ課題解決の進め方

「データの統合・整備」と聞くと、大規模なシステム開発や長期にわたるリソースが必要なイメージを持たれる方も少なくないかもしれませんが。iPaaSを活用することで、専門的な開発スキルがなくても、段階的に・現場主導で進めることが可能です。



これらのプロセスを個別開発で実現しようとする、開発・運用コストが膨らみやすく、PoCで止まってしまうケースも少なくありません。iPaaSならそれらの問題を回避できます。

導入にかかる期間は、データ量よりも「対象範囲の広さ」と「データの複雑さ」に左右されます。部門単位のPoCであれば1~2ヶ月、複数部門・複数システムが絡む場合は3~6ヶ月、全社規模やレガシーシステムを含む場合は6ヶ月以上が目安です。

「HULFT Square」で課題を解決&工数削減。AIデータ連携を自動化するiPaaS

「HULFT Square」は、分散するデータを統合し、AIが活用できる形へと整備・供給するためのiPaaS（クラウド型データ連携基盤）です。データの収集・統合・加工・更新といった一連のプロセスを一元的に管理し、ノーコードで連携フローを構築できるため、データ整備の工数を短縮できます。また、システムごとの個別開発を避けられるため、大幅な負荷軽減にもつながります。

AI & Data Integrationソリューション

データ連携のプロフェッショナルとして数多くの企業のシステム基盤を支えてきたセゾンテクノロジーが、その知見を凝縮して開発したのが「AI & Data Integrationソリューション」です。社内外のさまざまなAI・データをワンプラットフォームで統合し、RAGの精度向上から業務自動化までを一気通貫で実現します。生成AIの導入が進む一方で、「精度が安定しない」「現場で使えるレベルに達しない」「システム部門の工数が増える一方で成果が出ない」といった声は後を絶ちません。本ソリューションは、こうした課題に対してデータ連携の観点から包括的にアプローチし、AIが実際に業務で機能する状態を実現します。

技術的な課題

経営層

- ✔ 事業の成長につながらない
- ✔ PoCのまま止まってしまっている
- ✔ 導入までに時間が掛かる

AI推進担当

- ✔ ハルシネーションが懸念される
- ✔ モデルの変化に追従できない
- ✔ データをつなぐのに手間がかかる

業務現場

- ✔ 精度が低く実務で使えない
- ✔ 参照できるデータが限定的
- ✔ 業務での使い方が分からない

システム担当

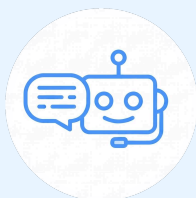
- ✔ 社内にAIがバラバラに乱立している
- ✔ 現場の依頼が多くて手が回らない
- ✔ データの整備に時間が掛かる

AI活用における課題は、立場によって異なります。経営層が「PoCのまま止まっている」と感じる一方、業務現場では「精度が低くて実務で使えない」、AI推進担当は「データをつなぐのに手間がかかる」、システム担当は「現場からの依頼が多く対応しきれない」と、それぞれが異なる壁に直面しているのが実情です。本ソリューションは、これらすべての課題に対し、データ連携とAI基盤の両面から一体的に対応します。

AI & Data Integrationソリューションの機能と特徴

本ソリューションは「社内のデータをつなぎ、知識に変え、業務を動かす」をコンセプトに設計されたエージェント型AI基盤です。既存のシステム環境を大きく変えることなく導入でき、現場主導でのAI活用を段階的に拡大していくことができます。AIエージェントの構築からデータパイプラインの自動化まで、AI活用に必要な機能をワンストップで提供します。

AIエージェント機能



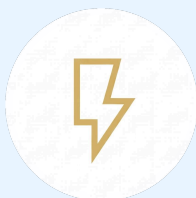
部門やユースケースごとにAIエージェント（チャットUI）を作成。
モデル選択、関連ドキュメント、アクションなどをノーコードで設定できます。

関連ドキュメント機能



FAQ・マニュアル・社内規定などを関連ドキュメントとして登録。
ベクトル検索用に自動インデックス化され、AI応答の根拠として活用されます。

アクション機能



HULFT Squareのスクリプトや外部APIを「アクション」として登録。
自然文からの指示で業務処理（例：CSV出力、通知）が実行できます。

モデルプロバイダー機能






Claudeなどの生成AIモデルを管理画面から登録・選択可能。
AIエージェントごとに利用モデルを切り替えられます。

4つの機能が連携することで、「データを集める→AIが理解できる形に整える→エージェントが業務で活用する」という一連の流れを、単一のプラットフォーム上で完結させることができます。既存の仕組みを最大限活かしながら、着実にAI活用を前進させることが可能です。

HULFT Square導入事例

AI・データ活用の全社浸透を目指すセブン銀行。
データ連携基盤・生成AI活用環境の実装にHULFT Squareが貢献。



-  オンプレミスからSaaSまであらゆるシステム環境に対応、**手
軽なデータ連携**を実現
-  日本の習慣にあった使いやすいインターフェースで**内製開発も
容易**
-  Azure OpenAI Serviceと連携し、**自然言語によるデータ分析検
証に貢献**

銀行勘定系システムや各種SaaSに分散したデータ環境により、手作業による連携や属人化、未整備データを起因とした検索精度・回答精度の課題を、HULFT Squareによるノーコードでのデータ統合と前処理・構造化、生成AI環境との連携によって解消。内製主導での生成AI活用基盤構築と自然言語による高度な分析活用を実現。

[詳細をみる](#) ▶

AI活用の成否は、 データの質と構造に 大きく依存します。

.....

まずは小規模なPoCから始め、
対象を限定しながら検証と改善を重ねることで、
段階的な活用拡大の検討を
おすすめします。

SAISON
TECHNOLOGY

株式会社セゾンテクノロジー

<https://www.saison-technology.com>

お問い合わせ

● HULET、その他HULFT関連製品は、株式会社セゾンテクノロジーの登録商標または商標です。● 記載されている会社名及び製品名は、各社の商標または登録商標です。
● 記載されているシステム名、製品名などには、必ずしも商標表示（TM、®）を付記していません。このカタログの記載内容は、2026年5月現在のものです。本カタログの記載内容は予告なく変更することがあります。