社内にデータを置いたまま AI を事業に組み込む

2026年9月14日

データを外に出せない。しかし AI は使いたい

論点は「AI を使うか使わないか」ではありません。 **データをどこに置いたまま使うか**です。

- 外部の AI サービスに問い合わせるということは、 **その都度、社外にデータが出る**ということです
- 原価・設計・検査画像・人事情報は、**出せないから使えない**—— この一点で止まっている業務が多くあります
- **止めている間に、使えている同業との差は開き続けます**
- 一方で「全部を自社で作る」も現実的ではありません。 **どこを自前にし、どこを借りるか**の線引きが要ります

よくある行き止まり

- 情報システム部門が止める
- 実証実験は通るが本番に上がらない
- 誰も「どこまでなら自前でできるか」を数字で持っていない

本日は、その線引きを当社が自分で測った数字でお示しします。

大手メーカーさまの基盤を、数年 運用しています

大手メーカーさま向けに、大型クラウド基盤を設計・構築し、数年にわたり運用まで担っています。

扱うのは、外に出せないデータです

この基盤で扱うのは、社外に出せないデータです。

本日のご提案の中心そのものを、すでに大手企業さまでお任せいただいています。

納めて終わりではありません

設計・構築だけでなく、数年にわたり運用まで担っています。

「導入した後どうなるか」に、実績でお答えできます。

この経験の上に、いま自社設備で回している運用を重ねてご提案します。

具体的な案件名・規模・時期は、守秘のため本資料には記載しておりません。ご関心に応じて個別にご説明します。

資産を守りながら回す、四層のフレームワーク

データも、学習した重みも、社内に置いたまま回します。 外に出るのは何もありません。

層	置くもの	選べるやり方
業務	現場が実際に使う画面と処理	原価計算・文書検索・社内問い合わせ など
ハーネス	モデルを業務に繋ぐ仕組み（道具の呼び出し・検索・評価・記録）	ここがいちばん時間のかかるところです。 モデルを入れ替えても、**ここは資産として残ります**
モデル	重み・学習データ・評価データ	**① 一から作る**／**② オープンモデルを継続学習・微調整**／③ オープンモデルをそのまま
基盤	GPU・データベース・運用	自社設置／専有環境。**外部へ出る通信を持たない**

守るのは「資産」です

- データだけでなく
- 学習した重みも
- ハーネス（業務に繋ぐ仕組み）も
- 業務のノウハウも

これらは一度 外に出すと、**戻せません。**

いちばん積み上がるのはハーネスです。

会社の運営を、自社の AI 基盤で回しています

作って納めて終わり、ではありません。会社の運営そのものを載せています。 当社自身が、毎日この基盤を運用している側です。

基盤の上で動いているもの	いまの状態
言語モデルの推論本番	複数のモデルを 24時間 稼働。社内利用と社外向けの両方
稼働中のサービス	同一基盤の上で 常時 数十。うち公開しているものも数十
社内での AI 活用	数十のセッションが並行して業務にあたっています （開発・運用・調査・制作）
扱っている種類	言語だけでなく、**画像生成・動画生成・音声合成・意味検索**も同じ基盤
必要であれば	言語モデルを**一から学習する**こともできます（次ページ）

提案するものを、自社でも使っています

ここに挙げたものは、**当社が自分の会社の運営に日々 使っている**ものです。

作った経験だけでなく、運用して分かったこと——うまくいかない所も含めて——を そのままお話しできます。

「導入をご支援します」ではなく、毎日これを運用している側からのご提案です。

2026年9月時点。設備構成・処理性能などの詳細は個別にご説明します。

どこまで自前で作れて、どこからは作らないか

できないことを先に申し上げます。 この線引きが、投資の判断に直結します。

作りたい規模	当社設備での所要	現実的な進め方
十億パラメータ級まで	数日	一から作れます。自社の言葉づかい・専門用語に合わせられます
同じ規模で学習量を増やす	数週間	計画すれば可能です
数十億パラメータ以上	数か月	一から作りません。オープンモデルを土台に育てます

この線引きを、当社は実測で持っています

- どこまでが自前で現実的かを、机上の計算ではなく自社設備で実際に回して確かめています
- 「何でもできます」とは申し上げません。できない規模は、できないとお伝えします

どちらを選んでも、データは外に出ません

- 一から作る——学習データも重みも自社内
- オープンモデルを育てる——公開された重みを土台に、社内データで仕上げます。仕上げも自社設備で行います

当社の実測にもとづく目安です。前提により幅がありますので、具体的な条件でのお見積もりは個別にお出しします。

モデルの作りで、要る機材が変わります

パラメータ数だけでは機材は決まりません。 同じ規模でも、作りが違えば構成の考え方が変わります。

	密（dense）	専門家混合（MoE）
毎回使う部分	全体を使います	一部だけを使います
必要なメモリ	規模なりに要ります	規模なりに要ります（使わない分も置いておく）
計算の量	規模なりに要ります	規模のわりに少なくて済みます
効いてくるもの	1枚あたりの速さ	総メモリ量・枚数・つなぎ方

パラメータ数だけで決めると外します

・専門家混合は「大きいのに速い」。ただし置いておくメモリは大きいままです

・密は「小さくても重い」。1枚の速さが効きます

・専門家混合なら、使わない部分を主記憶側に置く構成も取れます

・どちらを使うかは、業務の要求（応答の速さ・同時に何人が使うか）から決めます

当社は、どちらの作りのモデルも自社設備で動かしています。機材の構成は、使うモデルを決めてからご提案します。

「考えさせるか」でも、速さと捌ける量が変わります

モデルを替えなくても効く設定があります。 じっくり考えさせると精度は上がり、速さと同時に捌ける人数は下がります。

	考えさせない	考えさせる
応答の速さ	速い	遅くなります
同じ機材で捌ける量	多い	少なくなります
向く仕事	分類・抽出・定型の下書き。件数の多い処理	難しい判断・手順の要る作業。件数の少ない処理
切り替え	モデルはそのままで、業務ごと・要求ごとに切り替えられます	

ここが機材の見積もりに直結します

・全部を「考えさせる」にすると、機材が足りなくなります

・逆に、本当に必要な業務だけ考えさせれば、同じ機材で足ります

・どの業務でどちらを使うかを決めることが、そのまま機材の見積もりになります

・当社は自社基盤で両方を使い分けています

応答の速さ・同時に使う人数・1日の件数を伺えれば、必要な構成をお出しできます。

業務の「型」が違うものを、同じ基盤で横断しています

一つの業務のデモではありません。 外部への通信を持たない形で、型の違う業務を並行して動かしています。

業務の型	実際に動いているもの
計算・突き合わせ	数値と原価の管理（**現行の Excel 計算と一致することを検証済み**）／請求と会計の突き合わせ
文書の読み取り	契約書・社内文書の読み取りと検索／**出典つきで引ける議事録検索**
意味での検索	在庫や表データを、言葉の意味で引く
点検・監視	セキュリティの外形点検

よくいただく3つのご質問

・**権限分離と監査ログ**は、外部に頼らず**自前で持っています**

・**段階的に導入できます。**全社一斉である必要はありません

・**最終的にお客さま側で回せるようにします。**内製化までご一緒します

本ページは当社自身の業務システムと、お客さまとの試作です（前掲の納入実績とは別のものです）。企業名・製品名・単価は記載していません。

blueqat がご一緒できること

当社は、自社の設備で日々これを回しています。 運用して分かったことを、そのままお持ちします。

領域	ご提供する内容
GPU の運用	機材の選定と構成、複数枚をつなぐ設計、学習と推論の同居、 **止めて再開できる運用**
データ基盤の最適化	学習データの前処理と保管、データベースの設計、 **権限分離と監査**
性能の出し方	GPU 稼働率の測り方と上げ方、前処理の高速化、 **「測って直す」手順そのもの**
業務への接続	自社のインターフェイスを通じて、現場が実際に使う形まで。 **モデルを作って終わりにしません**

当社は、自分で運用しています

自社の **GPU クラスタ**を保有し、そこで**推論の本番を24時間**動かしながら、 **日々 学習と検証**を行っています。

他社の設備を借りて説明しているのではありません。

いちばん根気が要るのは、ハーネスと聞き取りです

モデルは選べます。選べないのは、現場ごとに違う業務を 細かく聞き取って組み上げる作業です。

- **ハーネスは分野ごとに違います。汎用の正解がなく、 業務ごとに組み直します**
- **聞き取りが要ります。どの判断を、誰が、何を見て行っているか―― 現場と一緒に1つずつ確かめていきます**
- **ここは短縮できません。 お客さまと一緒に、根気よく進める作業です**

➡だからこそ、当社が現場に入って、お客さまとご一緒します

当社が、現場に入ります

・聞き取りもハーネスの作り込みも、当社が担います

・分野ごとの業務を、お客さまと一緒に1つずつ確かめながら進めます

・最終的にお客さま側で回せるように、進めながらお渡ししていきます

時間はかかりますが、ここで積み上げたものは、モデルを入れ替えても残ります。

費用の目安

規模によって幅があります。まず機材だけ、から始めることもできます。

内容	目安	備考
機材（GPU サーバー）	数百万円	構成によります
構築込み	数百万円 ～ 数億円	規模と、対象業務の数によります
PoC（小さく試す）	数百万円 ～	この段階では、当社のノウハウのご提供は限定的です（範囲を絞り、動くことの確認に集中します）

ご購入が難しい場合

・機材の貸し出しができます。 固定資産に計上せずに始められます

・お客さまのサーバールームへ設置し、専用回線で繋ぐ形も可能です

・どちらの形でも、データは外に出ません。

機材だけ・PoC だけから始めて、広げながらご判断いただけます。

いずれも目安です。具体的なお見積もりは、対象業務と規模を伺ったうえで個別にお出しします。

進め方と、次の一歩

大きく始める必要はありません。 出せないデータが、実際に外へ出ないことを確かめるところからです。

段階	やること	目安
1. 見極め	どの業務のどのデータが外に出せないかを整理し、 **自前でやる範囲を決めます**	2〜4週間
2. 小さく実証	ひとつの業務で、**外部通信のない形**で動くものを作ります	1〜2か月
3. 設備と運用	GPU 構成を決め、学習と推論の同居を含めた運用に載せます	案件により
4. 広げる	モデルを社内データで継続学習し、業務を増やしていきます	継続

次の一歩のご相談

・**いま外に出せずに止まっている業務**はどれですか
・**すでにお持ちの設備**はありますか
・**いつまでに**何を示す必要がありますか

この3点が分かれば、**具体的な構成と期間**をお出しします。

以　上

お問い合わせ

info@blueqat.co.jp