Speed up your AI
FIXSTARS

LOCAL LLM GUIDE

5分でわかる
ローカルLLM入門

社外にデータを出さずにAIを使う、という選択肢

2026年11月　株式会社フィックスターズ

AIを開発に使いたい。
でも、社内のコードを社外に送ってよいのか。

生成AIによるコーディング支援は、開発の進め方を大きく変えつつあります。一方で、多くのAIサービスは、入力した内容を提供会社のサーバーに送って処理します。ソースコードや設計情報、お客様からお預かりした仕様を扱う開発現場では、この点が導入のブレーキになりがちです。

では、AIを社内で動かすことはできないのでしょうか。本資料では、その選択肢である**「ローカルLLM」**を基本から解説します。

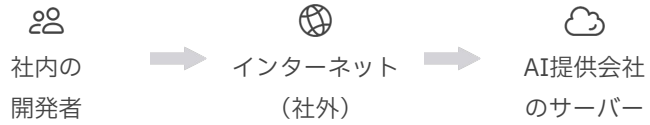


ローカルLLMなら、AIとのやりとりが社内で完結する

LLM（大規模言語モデル）は、ChatGPTなどの生成AIの中核にある技術です。
LLMをどこで動かすかによって、使い方は大きく2つに分かれます。

クラウドLLM

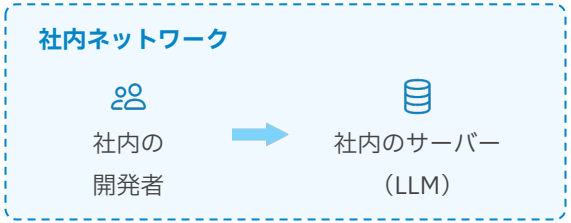
AI提供会社のサーバーで動くLLMを、インターネット経由で使う形態。本資料では「クラウドLLM」と呼びます。



入力したデータは社外に送られる

ローカルLLM

LLMを自社で管理するサーバーなどに置いて利用する形態。入力したデータが社外に出ないため、社内のセキュリティポリシーに沿ったAI活用がしやすくなります。



データは社内で完結する

COMPARISON

データを社内に留められる代わりに、動かす環境は自社で用意する

どちらが優れているかではなく、扱うデータと社内の体制に合わせて選ぶものです。

比較項目	クラウドLLM	ローカルLLM
動かす場所	AI提供会社のサーバー	自社のサーバーや、外部と切り離した専用ネットワーク
データの行き先	インターネット経由で社外に送られる	社内ネットワークの中で完結する
使うモデル	提供会社が用意したモデルから選ぶ。更新や提供終了の時期も提供会社が決める	公開されているモデルから自社で選ぶ。入れ替える時期も自社で決められる
費用	利用量に応じた課金や月額契約	サーバーなど機器の購入費と運用費
始めやすさ	契約すればすぐに使える	環境の構築と運用体制づくりが必要

ローカルLLMで動かせるのは「オープンウェイトモデル」

LLMの性能を決める中身のデータを「ウェイト（重み）」と呼びます。
このウェイトが公開されているかどうかで、LLMは2種類に分かれます。

クローズドウェイトモデル

ウェイトは非公開

提供会社のサービス経由でしか使えず、入力したデータは必ず社外に送られます。

例：ChatGPT、Claude、Geminiの各モデル

オープンウェイトモデル

ウェイトが公開

自社のサーバーに置いて動かせます。データを社内に閉じたまま使えますが、動かせるモデルの規模は手元の機器の性能に左右されます。

例：GLM、DeepSeek、Kimi、Qwenの各モデル

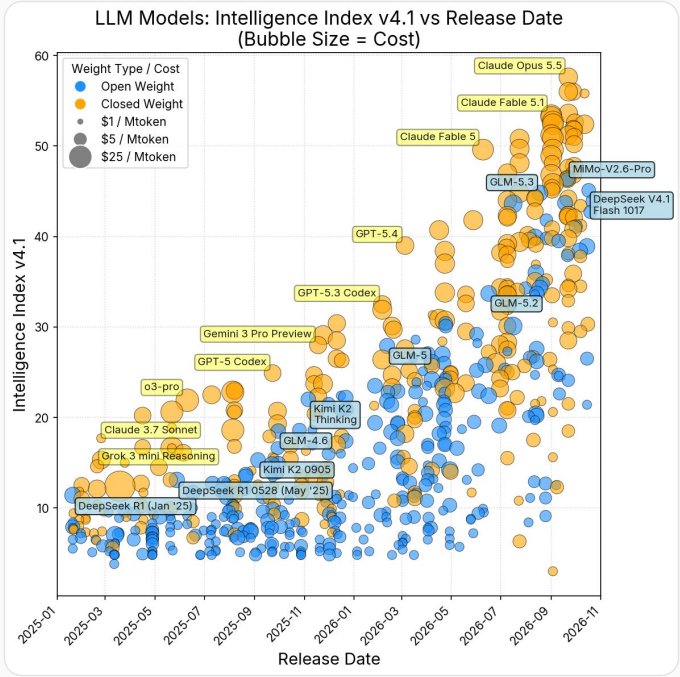
オープンウェイトモデルをクラウドで提供するサービスもあります。契約すればすぐに使え、料金は利用量に応じた従量課金です。ただしこの場合、データは社外に送られます。

最新のオープンウェイトモデルは、クローズドモデルとの差を約3か月以内まで縮めている

「社内で動かせるAIは、性能が物足りないのでは」と思われるかもしれません。確かに2023年頃は、オープンウェイトモデルがクローズドモデルの性能に追いつくまで、1年以上かかっていました。

しかし、その差は急速に縮まっています。2026年8月に公開された**GLM-5.3**は、約2か月前に公開されたClaude Fable 5に迫る性能に達しています。データを社内に閉じたまま、第一線に近い性能のAIを使える段階に来ています。

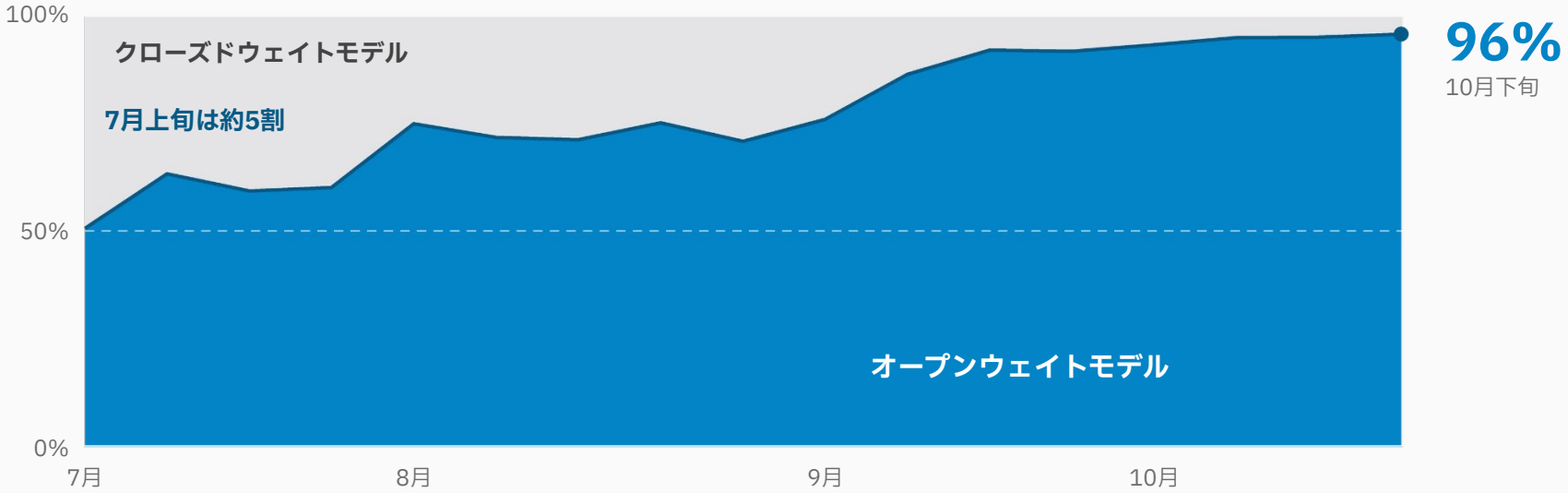
右のグラフ　青：オープン　オレンジ：クローズド　上ほど高性能



* 2026年10月時点。グラフは当社作成。Source: Artificial Analysis（artificialanalysis.ai）

日々コードを書く当社のエンジニアも、AI利用の9割超がオープンウェイトモデル

最新のオープンウェイトモデルは、ソフトウェア開発の実務で十分に使えるレベルに達しています。



出典：当社の社内AI基盤の利用ログ（2026年7月6日～11月1日、週次集計。7月6日の週とそれ以降で集計システムが異なる）。各週のトークン数（AIが処理した文字量の単位）に占める、オープンウェイトモデルの割合。

SUMMARY

ローカルLLMは現実的な選択肢になった。残るハードルは「動かす環境」

ここまでのポイントは3つです。

社内で完結できる

ローカルLLMなら、AIとのやりとりを社内で完結できる

性能は数か月差まで迫る

その中身であるオープンウェイトモデルは、クローズドモデルに数か月差まで迫っている

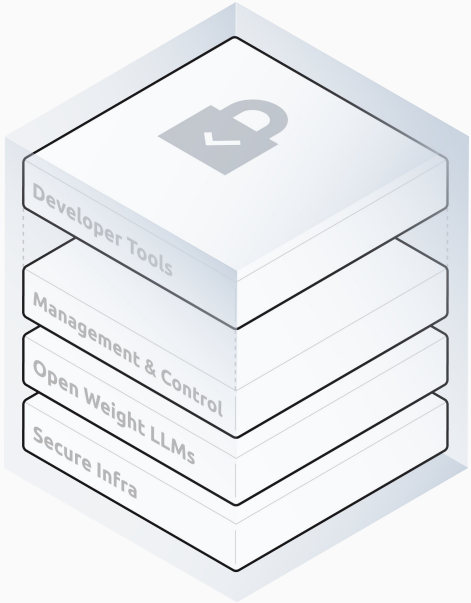
環境は自社で用意する

ただし、サーバーの調達から構築・運用まで、動かす環境は自社で用意する必要がある

見落とされがちなのが最後の点です。高性能なモデルほど大きな計算資源が必要になります。また、チームで同時に使っても待たされない速度を出すには、専門的な調整も欠かせません。

その環境を、すぐ使える形でまとめて提供するのが「Fixstars Vega」

Fixstars Vegaは、組込み開発チームがコードや設計情報を外部に出さずにAIを活用するための「セキュアAIアプライアンス」です。



開発者ツール
チャットUI、コーディングエージェント、性能最適化のためのハーネスなど、開発を支援するツール

管理・コントロール
AIモデルへのアクセスを制御するゲートウェイ、利用状況を可視化するダッシュボード、Skillsやナレッジを管理するリポジトリ

オープンウェイトLLM
推論環境の最適化が施された、検証済みオープンウェイトLLM

セキュアなインフラ
オンプレミスのGPUサーバー、またはセキュリティ対策の施されたクラウド環境

当社が検証したオープンウェイトモデルを、お客様のハードウェア構成に合わせて、入れ替えながら継続的にご提供します。
提供モデル：GLM-5.3、DeepSeek-V4.1-Flash（2026年10月時点・随時更新）

導入検討の段階に合わせて、3つのセミナーを開催しています

セキュアなAI活用の考え方を知りたい方から、Fixstars Vegaを実際に触ってみたい方、自社の開発課題でチーム導入を進めたい方まで、検討段階に応じたセミナーを開催しています。

まず検討材料を集めたい方に

ウェビナー

対象	AI導入の要件・構成・コストを検討する方
期間	1時間程度
開催形式	オンライン（Zoom）
費用	無料

セキュアなAI活用の選択肢から、構成の決め方、コストの試算まで。当社が実運用するセキュアAIの実測データをもとに、毎回テーマを変えて解説します。

詳しくはこちら →

まず体験したい方に

1Dayセミナー

対象	AI駆動開発を実際に試したい開発者・リーダー
期間	1日
開催形式	フィックスターズ会場で開催
費用	5万円／人

Fixstars Vegaの環境とツールを使い、AI駆動開発の一連の流れをハンズオンで体験します。受講後は評価環境をご利用いただけます。

詳しくはこちら →

自社課題で実践したいチームに

2Daysプライベートセミナー

対象	導入を具体的に検討するチーム
期間	2日間
開催形式	お客様先で個別開催
費用	50万円／開催

自社の開発課題を題材に、チームでAI駆動開発を実践します。導入後の運用イメージまで具体化できます。

詳しくはこちら →

CONTACT

まずはご相談ください

セキュアなAI環境の構築から、AI駆動開発のチーム定着まで。組込み開発のAI導入をワンストップでご支援します。

問い合わせ先

会社名	株式会社フィックスターズ
部署	AIプラットフォーム事業部

Fixstars Vega 製品情報

製品の特長や導入形態、導入までの流れは製品ページでご紹介しています。

製品ページを見る →

Techblog /proc/cpuinfo

Zenn

当社のエンジニアが執筆しています。「ソフトウェアの高速化・最適化」と「最新ハードウェア／AI技術の活用」に焦点を当てた、専門性の高い記事が中心です。

Tech Blogを見る →

公式サイト

Speed up your AI FIXSTARS

サービス情報、企業情報、採用情報、IR情報は公式サイトをご覧ください。

公式サイトを見る →

お問い合わせフォーム https://www.fixstars.com/ja/contact

ほかの「5分でわかる」シリーズも、あわせてご覧ください

本資料のほかに、5つのテーマを5分で読める資料にまとめています。検討中のテーマに合わせてご覧ください。

社内ルールを決めたい方に

機密度に応じたAIの使い分け

「全面禁止」でも「全面解禁」でもない、
AI活用のルールづくり

資料をダウンロードする →

AIの費用を見積もりたい方に

AIは従量課金か、定額か

使うほど増える費用と、予算の読み方

資料をダウンロードする →

ローカルLLMを導入したい方に

ローカルLLMの自社構築

「動いた」から「チームで使える」までに
必要なこと

資料をダウンロードする →

組込み開発でAIを使いたい方に

組込み開発の課題とAI

AIで「書く」が速くなったあと、
何が詰まるのか

資料をダウンロードする →

AIエージェントを使いこなしたい方に

AIエージェントとハーネス

AIエージェントに任せて、
開発を速くするための環境の整え方

資料をダウンロードする →