2026年11月20日

各位

株式会社フィックスターズ
代表取締役社長 CEO 三木 聡
(コード番号: 3687 東証プライム)
問合せ先 取締役管理本部長 堀 美奈子
(電話番号：03-6420-0751)

フィックスターズ、国内完結の高速推論AIサービス「Fixstars Inference Cloud」でGLM-5.3の提供を開始

期間限定！直接契約の法人に10万円分のトークン付与キャンペーン開催

パフォーマンスエンジニアリング技術のリーディングカンパニーである株式会社フィックスターズ（東証プライム：3687、代表取締役社長 CEO：三木 聡）は、国内のGPUサーバー上で動く高速なAIをAPIで利用できるサービス「Fixstars Inference Cloud（FIC）」にて、新たなAIモデルとしてGLM-5.3の提供を開始したことをお知らせします。
FICから使用できるGLM-5.3は、公式モデルをさらにフィックスターズが最適化して提供するもので、100万トークンあたり入力 10.49円、出力 210円と、主要な提供事業者の中でも最安水準の単価を実現しました。（＊）
あわせて、提供開始を記念し、フィックスターズと直接契約する法人アカウントへ10万円分のトークンを付与する期間限定キャンペーンを実施します。

GLM-5.3について

GLM-5.3は、中国のZ.ai社が2026年8月に公開したオープンウェイトモデルのLLMです。複雑なソフトウェア開発やエージェントとしての長時間タスクを得意とし、前バージョンのGLM-5.2から大きく性能を伸ばしています。

Fixstars Inference Cloud版GLM-5.3の特長

FICで提供するGLM-5.3は、開発元であるZ.ai社が公開したモデルをNVIDIAの4ビット浮動小数点形式（NVFP4）で量子化し、計算量とメモリ使用量を抑えたものです。
パフォーマンスエンジニアリングに強みをもつフィックスターズが、モデルをさらに最適化することで、より安価なトークン単価と短い応答時間を両立しました。
トークン単価は主要な提供事業者の中で最安水準での提供となり、応答時間も最速クラスを実現しています。

2026年11月20日時点では、FICの直接契約者向けの提供となります。後日、FastMetal経由でも提供を開始する予定です。

トークン単価	100万トークンあたり 　入力 10.49円 　出力 210円 　キャッシュ読み出し 10.34円
応答開始までの時間（TTFT）	中央値 1.127秒 測定条件： 　計測場所 東京、同時実行数64、 　エージェント型ワークロード（平均入力 約13.9万トークン・キャッシュヒット率約96%）、計測時期 2026年11月

＊ GLM-5.3の開発元であるZ.ai社の公式APIは入力$1.40・出力$4.40で提供されており、FICの価格は入力が約20分の1、出力が約3分の1相当となります。OpenRouterにおける各事業者の提供価格と比較しても最安水準の価格を実現しています。（2026年11月20日時点）

Fixstars Inference Cloudについて

FICは、フィックスターズが国内で管理するGPUサーバー上でオープンウェイトモデルAIを稼働させ、OpenAI互換APIで推論AIを利用できるようにしたサービスです。ソースコードや社内文書を海外のサービスに送れない企業でも、最新のオープンウェイトモデルを業務や開発に使えます。
FICはフィックスターズとの直接契約のほか、AIゲートウェイ「FastMetal」経由によるトークン課金制でも利用できます。

- データは日本国内で完結：推論サーバー・ストレージはすべて国内にあります。
- フィックスターズが自ら運用：東証プライム上場企業であるフィックスターズが、推論基盤の最適化から運用・保守まで担います。
- 入力も出力も保存しない（Zero Data Retention）：プロンプトと出力は推論の終了後に保存せず、ログにも残しません。AIの学習にも使いません。
- OpenAI互換APIに対応したIDE、CLI、コーディングエージェントは、接続先とAPIキーを変えるだけで利用できます。

直接契約の法人への特典

- 日本語の技術サポート：直接契約の利用者には、APIの組み込み方やレート制限の設計、モデル更新への追従などをフィックスターズのエンジニアが日本語で支援します。
- 有償セミナーに無償で招待：直接契約の利用者を、フィックスターズが定期開催する有償のAI活用セミナーに1回無償で招待します。

FIC提供開始記念キャンペーン

内容	FICで利用可能な、10万円分のトークンを登録アカウントに付与
対象	期間中にフィックスターズと直接契約で利用登録した法人 ＊ 登録には審査があります。 ＊ FastMetal経由でのご登録・ご利用は対象外です。 ＊その他、トークン利用の細則についてはお問い合わせください。
対象モデル	• GLM-5.3 • DeepSeek-V4.1-Flash-1017 （2026年11月20日時点。今後公開するモデルも対象となります。）
登録期間	2026年11月20日～12月15日
申込先	https://www.fixstars.com/ja/products-services/inference-cloud

株式会社フィックスターズについて

フィックスターズは、“Speed up your AI”をコーポレートメッセージとして掲げるテクノロジーカンパニーです。計算資源を最大限に活用するソフトウェア最適化技術を駆使し、AIモデルの推論処理と学習プロセスの両面で圧倒的な高速化を実現します。医療、製造、金融、モビリティをはじめ様々な分野で、次世代AI技術の進化を推進しています。
https://www.fixstars.com/

報道関係のお問い合わせ
株式会社フィックスターズ 広報担当
Email: press@fixstars.com
Tel: 03-6420-0751