

AI開発エンジニアの方・生成AIを使って自社開発をしたい方

生成AI開発に最適なAPI基盤を提供

さくらのAI Engine

- ✓ 国産LLMを含む複数の生成AIのAPIを提供
- ✓ インフラ構築の手間なくすぐに利用を開始
- ✓ 無料枠あり・従量課金制

さくらのAI Engineとは

「さくらのAI Engine」は、基盤モデルを搭載したGPUサーバーで推論を実行し、結果を返すAPIサービスです。
お客様側でGPUなどのインフラ構築の手間なく、AIアプリケーションの開発やコーディング支援などに活用できます。



最適コストで
各種LLMを
利用できる



国内完結の
セキュアな
運用が可能



インフラ構築不要で
安定の実行環境
をご提供



OpenAI/Anthropic
互換のAPIで
すぐに使える

さくらのAI Engineの特徴



「さくらのAI Engine」は、モデルの実行・通信を、すべて国内クラウドで完結。
データは学習に使われることも外部に送信されることもなく、安心・高速な生成AI環境を提供します。



さくらのAI Engine

NVIDIA H100をはじめとする高性能GPUを搭載したAI実行基盤



基盤モデル



基盤モデル



基盤モデル

入力データを
利用される心配なし



基盤モデル開発元

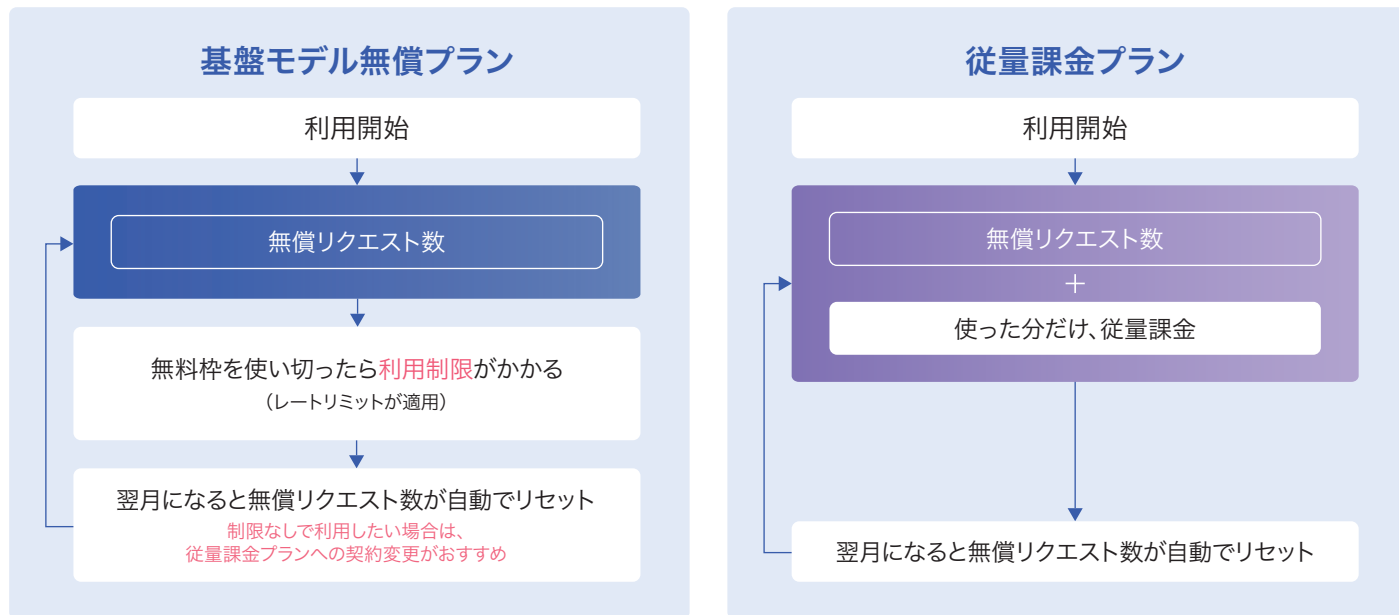
さくらのAI
API



エンドユーザー

料金体系

さくらのAI Engineは「基盤モデル無償プラン」と「従量課金プラン」の2つで構成されています。いずれのプランも翌月に無料枠が自動でリセットされます。円建て決済のため、為替変動の影響を受けず、安定した料金でお得にご利用いただけます。



※「無償リクエスト数」はご利用のモデルにより異なります。詳しくはサービスサイトをご覧ください。
※基盤モデル無償プランと従量課金プランは別プランとなります。自動的に従量課金プランに移行することはありません。

ユースケース

開発・コーディング支援に！
人気のコーディングエージェントと組み合わせての活用

コーディングエージェントを組み合わせることで、社内データを活用した安全なコード生成や開発自動化を実現。日本国内完結のセキュア環境で、次世代の開発スタイルを手軽に導入できます。

Point 全てのプランでデータは学習に一切使用されず(オプトイン/アウトなし)データが外部送信されることもありません



ログ解析&障害対応エージェントとしての活用

システムログやエラーメッセージをAIに解析させれば、原因特定から修正コードの提案までを自動化することも可能。運用負荷を大幅に削減します。

仕様理解・影響分析エージェントとしての活用

仕様書や設計書、API定義をAIが読み取り、変更時の影響範囲や修正が必要な箇所を自動抽出。安全な改修計画の立案やレビュー支援に活用できます。

提供中のモデル・詳しい料金はサービスサイトをご確認ください

さくらのAI Engineを気軽に試せる！
サービスサイト/playgroundはこちら



ブラウザからずんだものの読み上げ音声を生成してみたいのだ！



 **SAKURA internet**

さくらインターネット株式会社
〒530-0011 大阪府大阪市北区大深町 6-38 グラングリーン大阪 北館 JAM BASE 3F

資料ダウンロードはこちら

<https://www.sakura.ad.jp/document/ai-platform/>



マニュアルサイト

<https://manual.sakura.ad.jp/cloud/manual-ai-engine.html>



※2026年6月現在の情報です。内容は事前の予告なしに変更する場合があります(202606)