

何ができるの？要点ぎゅぎゅっとまとめて解説！

AIの力を、すべての人に届ける



サービス紹介資料

目次

いま、求められる「安心・安全な国内基盤」	p.03
国産モデルを使うという選択	p.04
さくらのAI	p.05
さくらのAIの5つの価値	p.06
さくらのAI Engine のご紹介	p.07
さくらのAIソリューション のご紹介	p.18
お問い合わせ	p.28
会社概要	p.29

いま、求められる「安心・安全な国内基盤」

日本の未来を創るためには、安心・安全な利用環境の構築と経済安全保障の確立が何よりも不可欠なものとなっています。

国内法規制や業界基準に準拠したAI環境により、金融・医療・自治体などの高セキュリティ領域でも安心して活用できます。

国内のエコシステム構築を通じて、デジタル貿易赤字の縮小と持続可能な産業発展の実現を力強く着実に目指します。

日本の未来を創るために必要な要素1

安心・安全な利用環境

- 国内法規制や業界基準に準拠したAI環境
- 情報漏えいリスクを最小化し、金融・医療・自治体など高セキュリティ領域でも安心して活用
- 信頼性の高いデータ管理で、部内展開から全社導入へスムーズに拡大



日本の未来を創るために必要な要素2

経済安全保障の確立

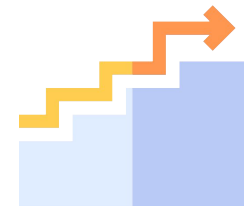
- 海外依存から脱却し資金の海外流出を抑制
- デジタル貿易赤字の縮小に貢献し、国内経済への投資循環を実現
- 国家レベルでデータ主権を確立し、政策や産業基盤を支える



日本の未来を創るために必要な要素3

国内産業の成長循環

- 国内のクラウドやAIベンダー、企業等が連携し、持続可能なエコシステムを築く
- 新しい産業分野での雇用創出やスタートアップの支援に貢献
- AIの価値を国内で循環させ、日本発のイノベーションを国際競争力に結びつける



国産モデルを使うという選択

AIモデルの選択肢が広がる中で、国産モデルが注目される理由は主に3つあります。

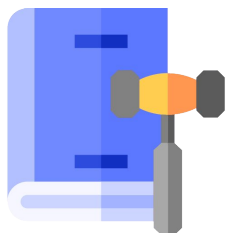
それは、「法的・ガバナンス対応の強さ」「日本語・業務文書との相性の良さ」「商用利用・運用のしやすさ」です。

安心して使える国内基盤と国産AIモデルを両立し、国内完結型で安心・安全にAIを活用できる環境を提供しているのが、さくらのAIです。

国産モデルが注目される理由1

法的・ガバナンス対応に強い

- 国内完結の構成により、日本の法律・基準に沿った運用が可能
- 公共・教育・金融など、ガバナンス要件が厳しい領域でも導入しやすい



国産モデルが注目される理由2

日本語・業務文書との相性がよい

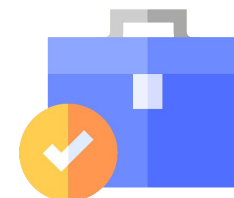
- 日本語に最適化されたモデルを使えるため、敬語・業務用語・表現の違和感が少なく、業務での応用がしやすい



国産モデルが注目される理由3

商用利用・運用がしやすい

- ライセンスや条件が明確で、商用・長期利用にも安心
- 日本語でサポートが受けられるため、導入や検証がスムーズ



さくらのAI

国内完結で生成AIをすぐに活用できるプラットフォームで、用途に応じて選べる2つのサービスラインアップをご用意しています。

API基盤を提供する「さくらのAI Engine」は、柔軟な構成により生成AIアプリ開発に最適なサービスです。

パッケージ型の「さくらのAIソリューション」なら、専門知識が不要で業務にAIを手軽にご活用いただけます。

1 さくらのAI Engine

生成AIアプリケーションの開発に最適な「API基盤」を提供します。
高品質なモデルをAPI経由で利用でき、柔軟な構成でアプリやサービスに組み込むことができます。

こんな人におすすめ



- AIやモデル構成を理解しているエンジニアの方
- 海外APIサービスを活用し、自社アプリに組み込んでいる方
- API連携で生成AIアプリをスムーズに開発したい方

おすすめポイント



- 複数の生成AIモデルを自由に選択可能
- 無料プラン & OpenAI API互換で手軽に試せる
- 共有型での提供で、コストを抑えて効率利用

2 さくらのAIソリューション

GPUクラウド×基盤モデル×業務アプリをパッケージで提供する、国内完結型の「生成AI業務支援サービス」です。開発および導入支援も用意しており、専門チームがモデル選定・構築・検証を一貫サポートします。

こんな人におすすめ



- AIを自分で構築するのは難しいが、業務に活用したい方
- 自社開発までは難しいが、AIを組み込みたい企業
- SaaS型のAIサービス利用を検討している方

おすすめポイント



- GPUベースで、専門知識がなくても直感的に利用可能
- App代も月額固定で、コスト管理がしやすい
- 専有環境と閉域網接続による、高いセキュリティを実現

さくらのAIの5つの価値

さくらのAIは、信頼性・計算資源・公正な料金・連携・簡易性という5つの確かな価値を提供しています。

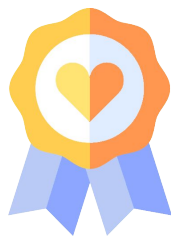
データ保護とプラットフォームの安定運用をしっかりと保証し、高性能GPUを幅広いユーザーの皆様へ確実にお届けいたします。

透明性のある明瞭な料金体系と直感的な操作性により、初心者を含むすべての方が安心して快適に利用できます。

価値1

信頼

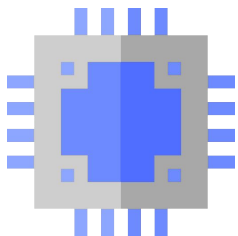
信頼性と安全性が最優先。
データの保護とプラットフォームの安定運用を保証



価値2

計算資源

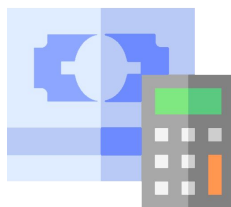
豊富で高性能なGPUを幅広いユーザーへ。計算リソースの利用を民主化



価値3

公正な料金

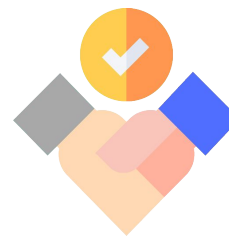
透明性と公平性を重視した明瞭な料金体系で、負担なく生成AIを利用できる環境を実現



価値4

連携

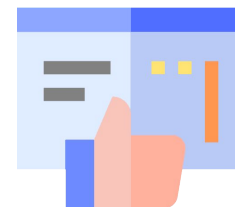
国内外の企業や技術者を集結。パートナーシップの力を最大限に活用



価値5

簡易性

初心者から専門家まで、すべてのユーザーが直感的に操作できる体験を提供



① さくらのAI Engine のご紹介

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。フルマネージドかつサーバレスな環境で提供されるため、インフラ構築の手間なく、すぐに利用を開始できます。さらに、RAG（Retrieval-Augmented Generation）を組み合わせることで、出力の最適化も容易に実現できます。

生成AIアプリ開発に最適なセキュアで柔軟なAPI基盤『さくらのAI Engine』

主な特徴

- **OpenAI互換**
OpenAI及びAnthropic互換APIで最新のLLMが使える、国産モデルも利用可
- **無償プランあり**
利用上限に達するまで従量プランと同様に使える
- **利用データの流出なし**
日本国内DCで完結で、モデル提供事業者へ利用データが送信されない
- **セキュアな閉域構成**
VPN/LGWAN/専有ネットワーク対応



さくらのAI Engineの特徴

さくらのAI Engineは、複数の基盤モデルから最適なものを選択できるフルマネージド型のAPIサービスです。
インフラ構築の手間が不要なフルマネージド環境で、RAGを手軽に組み込むことも実現していただけます。
また、国産インフラと国産基盤モデルの組み合わせにより、高度なセキュリティ要件にも万全に対応できます。

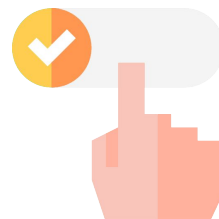
圧倒的低価格で提供

手軽な従量課金に加え、完全無料でご利用いただけるプランもご用意。コストを抑えて、気軽に始められます。



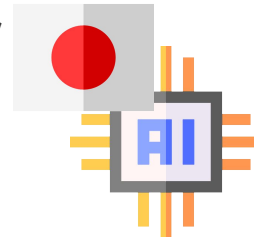
最適なモデルを選択可能

国内開発モデルを含む複数モデルをラインアップ。用途やコストに合わせて、最適なモデルを選べます。



国内完結型AI基盤で安心

フルマネージド環境で、データは国内で完結。モデル提供元にも送られないため、学習に利用される心配もありません。



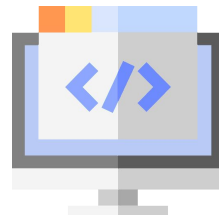
RAGを簡単に構築可能

必要に応じてRAG※を手軽に組み込み、出力を最適化。自社データを活かした高精度な回答を実現。



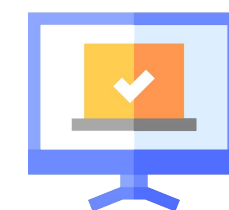
APIベースで開発しやすい

OpenAI / Anthropic互換APIで、既存システムからの移行もスムーズ。公式ライブラリをそのまま活用でき、スピーディーに開発できます。



Playground機能で複数モデルを試せる

ブラウザ上で複数モデルを比較検証。導入前の評価や動作環境をノーコードで行えます。



RAG（Retrieval-Augmented Generation）とは、検索（Retrieval）で見つけた外部知識を、生成AI（LLM）の入力に追加して回答を作る手法。

「モデルが知っていること」＋「手元のドキュメント/DBの最新情報」を都度合成できるのが特徴。幻覚（ハルシネーション）を抑え、根拠を示せる回答を目指す。

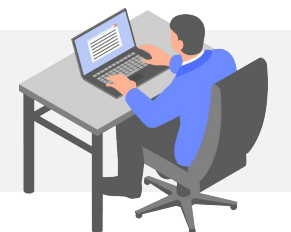
さくらのAI Engineの全体像

さくらのAI Engineは、モデルの実行・通信・保存までを国内クラウドで完結し、外部へのデータ送信なしで安心な生成AI環境を提供します。
高火力GPU上で基盤モデルを稼働させ、API経由でエンドユーザーへ確実に提供する仕組みを実現しています。
NVIDIA H100を搭載した高性能サーバーにより、ベアメタル型からコンテナ型まで幅広く柔軟に対応しております。

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 Engine



GPU基盤

NVIDIA H100をはじめとする、
高性能GPUを搭載したサーバーを、
ベアメタル型・VM型・コンテナ型で提供する
GPUサービス「高火力シリーズ」を採用



基盤モデル

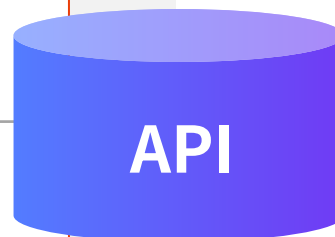


基盤モデル



基盤モデル

API



エンドユーザー

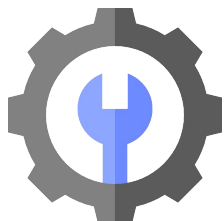
さくらのAI Engineでできること

「さくらのAI Engine」は、利用者がすぐに使い始められるよう設計された フルマネージドの生成AI（API）サービスです。
国内完結のセキュアな運用により、柔軟に構成できるLLM環境を提供しています。
また、APIベースでの導入が簡単なため、すぐに使い始められる生成AI環境を国内クラウドで実現できることも特徴です。

特徴1

柔軟に構成できる LLM環境

- 国産モデルやOSSモデルを使った構成に対応
- 柔軟なスケーリングと料金設計
- 予算や処理量に応じた柔軟なスケーリング



特徴2

国内完結の セキュアな運用

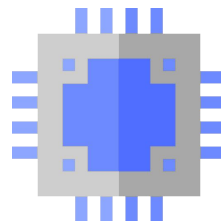
- 通信・実行・保管を含めてすべて国内クラウドで完結
- 閉域接続・暗号化など用途に応じたセキュリティ対策
- 機密性の高い業務データを使ったAI活用の相談も可能



特徴3

高性能GPUで 安定の実行環境

- NVIDIA H100 TensorコアGPUを採用
- 生成AI向けに最適化された高火力クラウド基盤を使用
- 高負荷なワークロードも安定かつ高速に処理



特徴4

本番導入へつなげる 支援体制

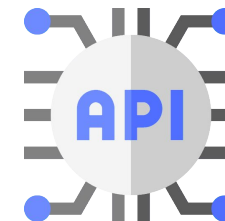
- 国内完結の生成AI基盤で安心利用が可能
- 専門メンバーによる実導入サポートにも対応
- パートナー企業との連携で高度な要件にも柔軟に対応



特徴5

APIベースでの 導入が簡単

- APIキー取得だけで、すぐにモデルを実行可能
- LLM APIと互換性が高く、既存コードも活かしやすい
- 開発ナシでも導入しやすい



さくらのAI Engineのユースケース

さくらのAI Engineは、社内情報を外に出さずにAI活用できる安心の環境を幅広く提供しております。

社内Chatbotから音声認識、プログラム支援、コンタクトセンター業務まで多様な用途に幅広く確実に対応します。

また、属人化の解消と業務効率化を実現し、高精度なAIによるオペレーション品質の向上を強力に支援いたします。

社内Chatbot

社内情報を 外に出さずAI活用

課題

- 社内FAQの検索に時間がかかる
- 問い合わせが属人化している
- 本来業務に集中できない

さくらのAI Engineによる成果

- 社内ドキュメントや規程を横断検索して即座に回答
- 標準化された応答で情報のブレ防止
- 問い合わせ対応を自動化し、工数を削減

音声認識

個人情報を含む データを文字起こし

課題

- 議事録作成に時間がかかる
- 電話対応の記録が属人化
- 現場負荷が大きい

さくらのAI Engineによる成果

- 高精度な音声認識で会話を自動でテキスト化
- セキュアな環境で録音データを処理し、情報漏洩リスクを低減

プログラム支援

知的財産を守る 開発支援

課題

- コード作成に時間がかかる
- 属人化によって品質がばらつく
- 初学者にとってハードルが高い

さくらのAI Engineによる成果

- コードを自動生成し、作業を効率化
- 複数言語・フレームワークに対応し、幅広い開発シーンを支援
- テストコードやサンプルコードも即時生成でき、開発スピードを加速

コンタクトセンター

AIによる オペレーション品質向上

課題

- 必要な情報を探すのに時間がかかる
- 対応品質にばらつきがある
- ナレッジが属人化して蓄積されない

さくらのAI Engineによる成果

- 音声の書き起こしや要約を自動化
- 対応内容を自動で整理・蓄積し、FAQやマニュアルへ即反映
- 属人化を防ぎ、一定品質の対応が可能になり、業務効率も大幅改善

さくらのAI Engineの導入事例: 株式会社miibo

株式会社miibo様は、ノーコードで構築できる対話型AIチャットボットサービスを展開している企業です。

自然な言葉でのスムーズな対話を実現し、企業・自治体・教育機関におけるFAQ対応や業務支援、情報提供の自動化を支援されています。

“ノーコードでつくれるAI”を安心して動かすための基盤として、さくらのAIを採用いただきました。

会話型AIサービスの基盤に、さくらのAI Engineを採用

miiboのノーコード技術とさくらのAI Engineを組み合わせた、
安心・安全な“国産完結型”の生成AIアプリ基盤として、
「miibo 国産基盤パッケージ」がリリースされました。

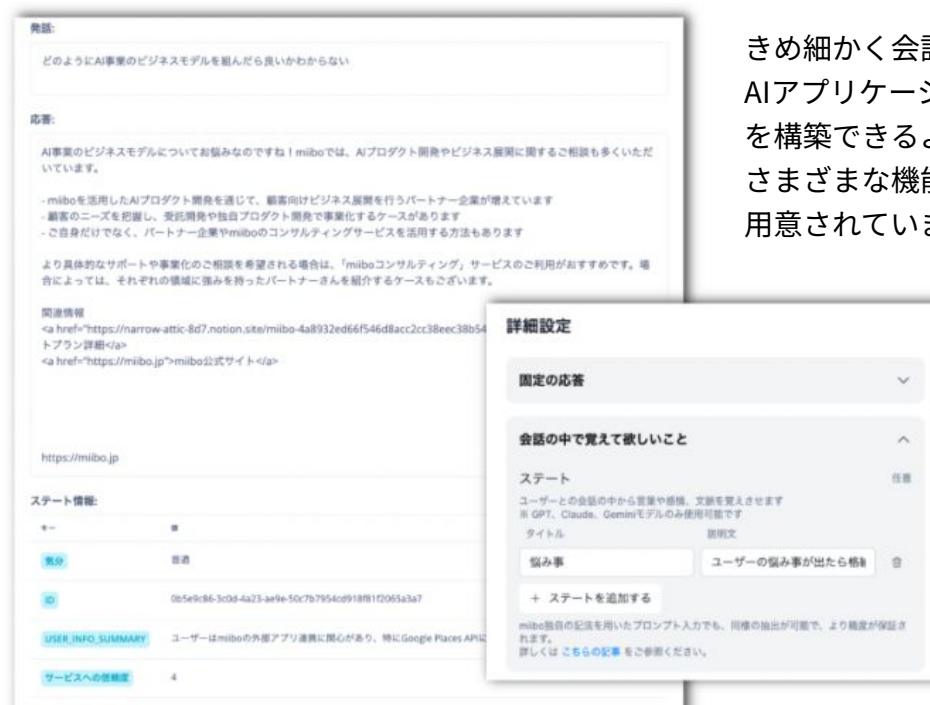
主な特徴

- チャットボット構築から実行まで、すべて国産構成
- 日本語に強い国産LLMに対応
- データ送受信・処理がすべて日本国内で完結
- 公共・教育機関などの高いガバナンス要件にも対応
- 「契約書不要・見積り不要」のパッケージ提供が可能

詳しくは、サービスサイトをご確認ください。



きめ細かく会話型
AIアプリケーション
を構築できるよう、
さまざまな機能が
用意されています



さくらのAI Engineの提供モデルと料金(1)

カテゴリー	提供モデル	基盤モデル無償プラン (リクエスト制限内での利用が無料)	従量課金プラン		注 意 点
			無 償 枠	無償リクエストを超過した場合	
1 Chat Completions	gpt-oss-120b	1ヶ月あたり 3,000リクエストまで	1ヶ月あたり 3,000リクエストまで	Input：0.15円 / 10,000トークン Output：0.75円 / 10,000トークン	- 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）。 - 従量課金プランでも基盤モデル無償プランと同等の無償利用枠が存在します。ドキュメント（RAG）については無償利用枠はなく、両プランとも利用に応じて最小単位（100チャンク）ごとに料金が発生します。 - 基盤モデル無償プランで無償利用枠を超過した場合、APIリクエストに対してレート制御がかかります。従量課金プランで無償利用枠を超過した場合、各基盤モデルの最小単位（10,000トークン / 60秒）ごとに料金が発生します。 - 基盤モデル無償プランには申し込み数に上限があり、上限に達した場合は新規利用の受付を停止いたします。 - クローズドモデルとは、提供元が商用利用を前提に提供するAIモデルです。 - モデル提供元の利用ポリシーに基づく審査・許諾が必要となるため、当社環境でのご利用には申請が必須となります。AI engineのコントロールパネルより申請が可能です。詳しくはマニュアルサイトをご確認ください。 - VOICEVOX提供モデルに関するライセンス表示および利用条件については、マニュアルサイトをご確認ください。モーラとは、かな 1 字分の音の長さを表す単位です。
	llm-jp-3.1-8x13b-instruct4			Input：0.15円 / 10,000トークン Output：0.75円 / 10,000トークン	
2 Chat Completions クローズドモデル	cotomi v3 ※1（NEC）	対象外	本モデルの価格は、申請が承認されたお客さまにのみご案内しています。		
	PLaMo 2.0-31B Instructedモデル (Preferred Networks)				
3 Audio Transcription	whisper-large-v3-turbo	1ヶ月あたり 50リクエストまで	1ヶ月あたり 50リクエストまで	0.5円 / 60秒	
4 Embeddings	multilingual-e5-large	1ヶ月あたり 10,000リクエストまで	1ヶ月あたり 10,000リクエストまで	Input：2円 / 10,000トークン Output：無料	
5 ドキュメント（RAG）		3円 / 100チャンク			
6 Text-to-Speech	VOICEVOX:ずんだもん VOICEVOX:東北ずん子 VOICEVOX:東北きりたん VOICEVOX:東北イタコ VOICEVOX:四国めたん VOICEVOX:あんこもん VOICEVOX:冥鳴ひまり VOICEVOX:春日部つむぎ	1ヶ月あたり 50リクエストまで	1ヶ月あたり 50リクエストまで	3円 / 10,000モーラ	

※1「cotomi」は、日本電気株式会社の登録商標です。※表示価格は税込です。
※提供モデルは随時追加・更新されます。最新の対応モデル一覧は [サービスサイト](#) をご確認ください。

さくらのAI Engineの提供モデルと料金(2)

カテゴリー	提供モデル	基盤モデル無償プラン (リクエスト制限内での利用が無料)	従量課金プラン		注意点
			無償枠	無償リクエストを超過した場合	
7 Chat completions パブリックプレビュー	Qwen3-0.6B-cpu	1ヶ月あたり 3,000リクエストまで	1ヶ月あたり 3,000リクエストまで	Input : 0.01円 / 10,000トークン Output : 0.03円 / 10,000トークン	- パブリックプレビューモデルはユーザーの皆様に実際の利用環境での動作検証およびフィードバックの収集を目的とした提供となり、予告なく提供を終了する場合があります。 - 提供期間中も安定性や応答品質について保証するものではなく、動作や仕様が変更となる可能性があります。 - パブリックプレビュー終了後、継続して提供する際には料金に変更になる場合があります。
	Phi-4-mini-instruct-cpu			Input : 0.01円 / 10,000トークン Output : 0.03円 / 10,000トークン	
	Qwen3-VL-30B-A3B-Instruct			Input : 0.1円 / 10,000トークン Output : 0.3円 / 10,000トークン	
	Phi-4-multimodal-instruct			Input : 0.1円 / 10,000トークン Output : 0.3円 / 10,000トークン	
	Kimi-K2.6			Input : 0.6円 / 10,000トークン Output : 3円 / 10,000トークン	
	Qwen3.6-35B-A3B			Input : 0.3円 / 10,000トークン Output : 1.5円 / 10,000トークン	
	gemma-4-31B-it			Input : 0.24円 / 10,000トークン Output : 0.96円 / 10,000トークン	
8 Embeddings パブリックプレビュー	Qwen3-Embedding-4B(FP16)	1ヶ月あたり 10,000リクエストまで	1ヶ月あたり 10,000リクエストまで	Input : 3円 / 10,000トークン Output : 無料	

※表示価格は税込です。
※提供モデルは随時追加・更新されます。最新の対応モデル一覧は [サービスサイト](#) をご確認ください。

さくらのAI Engineのご利用の流れ

さくらのAI Engineを利用する前の準備

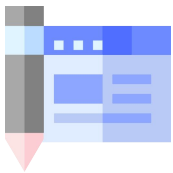
新規会員登録

会員IDをお持ちでない方は、
はじめに会員IDを作成
詳細は [さくらインターネット 会員IDの取得](#)
をご参照ください



プロジェクトの作成

コントロールパネルにログ
イン後、さくらのクラウド
のプロジェクトを作成
詳細は [さくらのクラウド プロジェクト登録](#)
を参照ください



コントロールパネルからご利用開始

コントロールパネルにアクセ
スし、さくらのAI Engineをご
利用ください



さくらのAI Engine利用手順

「さくらのAI Engine」を選択

さくらのクラウドホームから
「さくらのAI Engine」を選択



約款を確認

約款同意の必要があるため、同
意して利用を開始



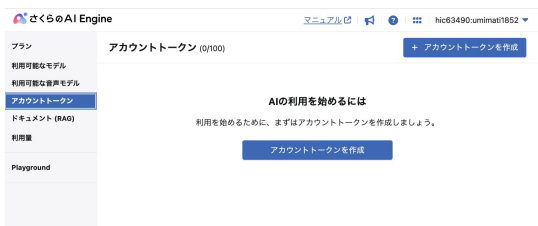
プランを選択

さくらのAI Engineのコントロー
ルパネルでプランを選択



アカウントトークンを発行

アカウントトークンを発行する
ことで、推論APIの利用開始



さくらのAI Engine Playground機能を利用する前の準備

アカウントを用意

GitHub または さくらID で利用できます。すでにお持ちのアカウントでそのままログイン可能です。さくらIDをお持ちでない場合も、その場で簡単に作成できます。

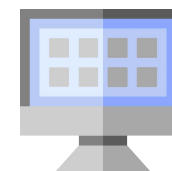
さくらIDの登録方法は [さくらIDに新規登録したい](#) をご参照ください



ブラウザでご利用開始

専用URLにアクセスし、さくらID または GitHub・X（旧 Twitter）で認証すると、Playgroundの画面が表示されます。インストールや環境構築などの事前準備は不要です。

<https://playground.aipf.sakura.ad.jp/>



さくらのAI Engine Playground利用手順

モデルを選択

複数のAIモデルから、利用したいモデルを自由に選択できます。音声生成も体験可能です。



プロンプトを入力し送信

入力欄にプロンプト（指示文）を入力し、送信します。



結果を確認

出力結果はブラウザ上に即時表示され、モデルごとの挙動を手軽に確認できます。



② さくらのAIソリューション のご紹介

さくらのAIソリューションとは

生成AIを業務に活用するための、AI開発から導入までをトータルで支援するソリューションです。

生成AI向けクラウドサービス「高火力」上で、基盤モデルと業務アプリケーションを組み合わせた生成AIパッケージを提供しています。

お客さまのニーズや環境に応じて、開発から導入支援までを一気通貫でサポートし、スムーズかつ確実なAI導入を実現します。

国内完結型 生成AI業務支援サービス『さくらのAIソリューション』

GUIベースで専門知識不要

音声文字起こし・議事録生成、コーディング支援、RAGを活用した社内文書検索など



AI導入～運用までワンストップでご相談いただける伴走支援

PoC設計から本番導入まで専門チームがサポートし、
既存のオンプレ環境や他クラウドとの閉域接続も柔軟に対応



さくらのAIソリューション 生成AIパッケージ

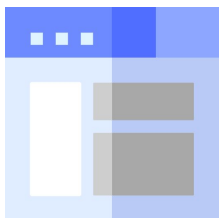
GPUクラウド×基盤モデル×業務アプリケーションを組み合わせ、生成AIパッケージとして提供するサービスです。

GUIベースで直感的に操作でき、専門知識がなくても利用可能で、国内完結型の閉域環境で、機密データも安心して扱えます。

AI導入に必要な「インフラ構築」「モデル接続」「アプリケーション開発」を省き、スピーディーに業務やサービスへAIを組み込みめます。

GUIベースで 専門知識不要

GUIで直感的に利用でき
AI開発の知識がなくても
すぐに利用可能



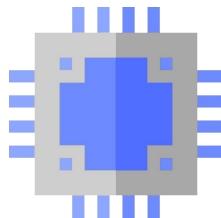
AI開発に必要な要素を ワンパッケージに

インフラ構築や、モデル
接続の手間を省き、AI導
入をスピーディーに実現



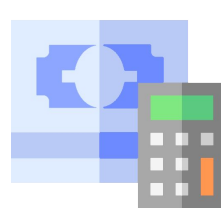
専有GPU×国内完結で 高セキュリティを実現

各社GPUは専有型構成に
より、「さくらのAI
Engine」よりも高いセ
キュリティレベルの環境
を実現



月額固定で コスト管理も安心

月額固定で費用が明確。
長期利用でも予算が立て
やすく、管理・運用コス
トも把握しやすい設計



手軽に、 すぐ始められる

導入前に操作感や機能を
お試しいただける無料期
間があるので、安心して
導入の検討が可能



さくらのAIソリューション 生成AIパッケージの詳細

生成AIパッケージは、社内文書検索や音声文字起こし、プログラム支援など多彩な機能を幅広くしっかりと提供しております。

「パッケージ単位」での月額料金設定でコストを管理しやすく、お客様の予算に合わせた柔軟な運用を確実に実現していただけます。

無料トライアルもご用意しておりますので、導入をご検討中の方はまずはお気軽にお問い合わせください。

提供パッケージ例

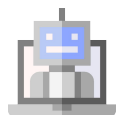
社内文書検索チャット/RAG



プログラム支援



RAG/AIエージェント
作成プラットフォーム



提供中のパッケージ

InfiniCloud® AI パッケージ

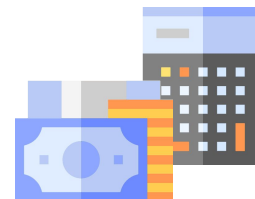
neoAI Chat for さくらインターネット

さくらのAI議事録

料金について

「パッケージ単位」で契約・月額料金が設定。コストを明確に管理でき、予算に合わせた運用が可能。

また、一部のパッケージには1か月の無料トライアルもご用意しています。（β版については無料でご提供しています）



ご利用の流れ

本サービスは、オンラインサインアップでの提供は行っておりません。まずは問い合わせフォームからお気軽にご相談ください。

※neoAI Chat for さくらインターネットについては、ご利用前に審査が必要となります。審査の結果やご協力状況によっては、サービスをご利用いただけない場合がございますので、あらかじめご了承ください。

InfiniCloud® AI パッケージ (Ver2)

InfiniCloud® AIとは？

企業の業務や目的に合わせて、最小限の構成から導入できる国産Private AI基盤です。RAGやFine-tuningに加え、AI Agent機能やAgentic RAG、証拠駆動型思考（EDTMモード）を標準搭載。企業独自の知識や業務ルールを活用しながら、高精度な検索・回答から業務支援まで実現します。さらに、OpenAI互換APIによる既存システムとの連携にも対応し、企業のAI活用を幅広く支援します。

こんな人におすすめ

部署ごとに専門性の高いAIを活用したい

だれでも簡単に使えるAIを導入したい

既存システムとも簡単に連携したい

社内ドキュメントやナレッジを有効活用したい

誤答のない自社専用AIを使いたい

機密情報を守りながら安全にAIを活用したい

高性能GPUで安定運用

社内に蓄積された文書、PDF、画像、音声データをそのまま業務に活用できます。また、AutoDGISによりドキュメントに基づく学習データをAIが自動生成し、ハルシネーションを抑えた安定した回答を実現。既存資産を活かしながら、業務に即したAIを効率的に構築できます。

チャット型WEBインターフェイス搭載

利用者にとって分かりやすいチャット型のWebインターフェイスを採用し、直感的な操作でAIを活用できます。スレッド管理やゲストモードに対応しているほか、管理画面の機能強化により、ユーザー招待やサインアップ設定、通知メール管理も簡単に行えます。



※本パッケージでは、1か月間の無料トライアルを実施しております。
トライアルをご希望の方は、お問い合わせフォームよりお申し込みください。
▶ <https://ai.sakura.ad.jp/sakura-ai/ai-solution/infinicloud-ai/>

neoAI Chat for さくらインターネット

neoAI Chat for さくらインターネットとは？

neoAI Chat for さくらインターネットは、さくらインターネットが提供する高火力サービスを基盤とした、日本国内完結のエンタープライズ向けAIチャットサービスです。機微な社内データを扱う企業でも安心して利用できるセキュリティ設計により、短期間で導入し、業務で使い続けられるAI活用を実現します。

こんな人におすすめ

問い合わせ対応の
属人化を解消したい

FAQやナレッジ検索の
手間を減らしたい

AIチャットの回答品質
や使い勝手に課題あり

外部AIの利用にセキュリ
ティ面で不安がある

社内ドキュメントを
活かしたAIを使いたい

専門知識がなくても
AIを活用したい

企業データを活かした、高精度な回答

企業が保有する社内データをAIに活用することで、一般的なAIでは難しい社内情報に基づいた高精度な回答を生成できます。社内問い合わせ対応や顧客サポートなど、業務に直結する用途で効果を発揮します。

現場で使いやすい生成AI基盤

業務に特化した「AIアシスタント」を、専門知識不要で簡単に作成できます。AIモデルやRAG（検索拡張生成）の設定をカスタマイズでき、多様な業務ニーズに対応。社内資料を参照しながら根拠のある回答を生成します。

企業ナレッジを学習

社内規定

マニュアル

営業資料

過去の稟議書

高精度の検索と回答

neoAI
Chat
for SAKURA internet



独自 AI を迅速に作成

営業担当 AI

稟議書作成 AI

※本パッケージは、ご利用にあたっては審査がございます。

審査の結果により、又は審査にご協力いただけないことにより、サービスをご利用いただけない場合がございますのでご了承ください。

▶ <https://ai.sakura.ad.jp/sakura-ai/ai-solution/neoai/>

さくらのAI議事録

さくらのAI議事録とは？

このパッケージは、AIが会議中のやり取りを整理し、話の流れ・合意事項・次に取り組むべきアクション(ToDo・期限など)を議事録の形で可視化するサービスです。音声データからの要約生成にも対応し、議事録作成の時間と属人性を大幅に削減します。



こんな人におすすめ

会議内容を
すばやく共有したい

専門用語や固有名詞も
しっかり認識してほしい

録画データも活用して
議事録を作成したい

議事録作成の
負担を減らしたい

セキュリティ面も
重視しながら運用したい

決定事項や次のアク
ションも整理してほしい

高セキュリティで安心

他社と共有しないプライベート環境に保存でき、データが外に出ない設計になっています。国内完結のセキュアな運用で、機密情報や個人情報を扱う会議や公共分野の打ち合わせでも安心です。

現場ですぐに活用しやすい

AIが会議内容をスピーディーに要約、辞書登録で専門用語も正確に認識し、安定した品質の議事録を生成。また、シンプルでわかりやすいUIで、迷うことなくすぐに使い始められます。



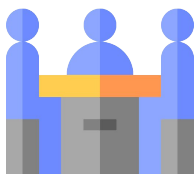
※本パッケージでは、1か月間の無料トライアルを実施しております。
トライアルをご希望の方は、お問い合わせフォームよりお申し込みください。
▶ <https://ai.sakura.ad.jp/sakura-ai/ai-solution/meeting-ai/>

さくらのAIソリューション 開発導入支援とは？

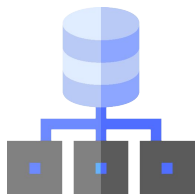
さくらのAIソリューションは、GPUクラウド×基盤モデル×業務アプリケーションを一体化した「パッケージ」だけにとどまりません。基盤モデルの選定から環境構築、効果検証、導入後の運用サポートまで、あらゆるフェーズでのご支援が可能です。豊富な知見をもとに、PoC導入支援チームが、お客様の課題や要望に合わせた最適なAI活用をご提案します。

相談から運用サポートまでワンストップでご提供

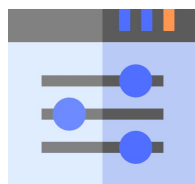
コンサルティング
(導入相談)



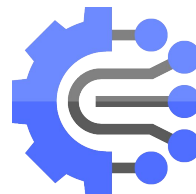
データ整備・
アノテーション



モデル選定・
チューニング



システム開発・
連携



導入支援・検証
(PoC)



社内教育・
運用サポート



どのフェーズからでも、最適なAI活用に向けてご支援いたします。

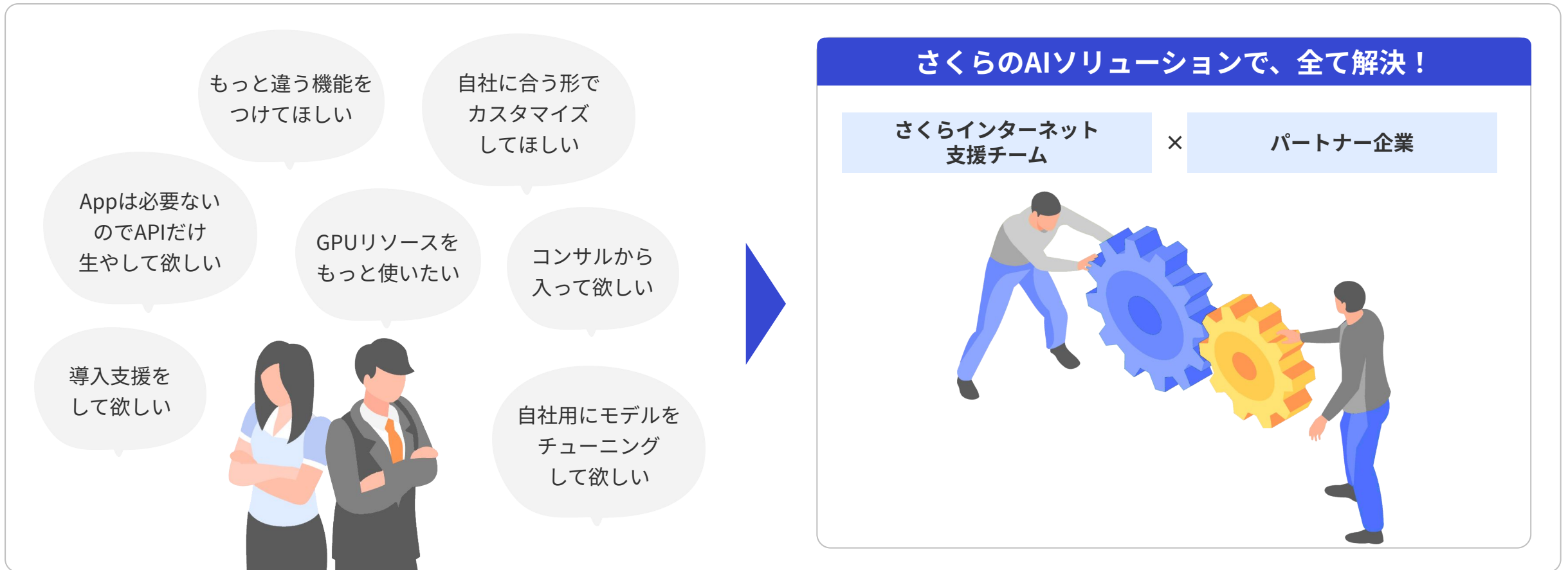
チューニングのみなど各社ごとの悩みに応じたサポートも可能ですので、お気軽にご相談ください。

さくらのAIソリューション 開発導入支援サービスでできること

さくらのAIソリューションは、お客様の多様なニーズにワンストップで柔軟かつ確実に丁寧に親身に対応いたします。

機能追加、カスタマイズ、GPUリソース拡張、API提供、コンサルティングまで幅広く柔軟に支援いたします。

さくらインターネットとパートナー企業の専門支援チームが、お客様のAI導入を全面的にサポートいたします。



さくらのAIソリューション 開発導入支援サービス 利用方法

本サービスは、オンラインサインアップでの提供は行っておりません。

お客様ごとのご要望に応じて契約を締結のうえご提供しております。導入をご希望の方は、まずはお気軽にお問い合わせください。



お問い合わせ



サービスの新規導入に関するご質問や契約前のご相談を受け付けております。
ぜひ、お気軽にお問い合わせください。

▶ お問い合わせはこちらから



すぐにモデルをお試しいただけます

多彩なモデルをコントロールパネルからPlaygroundですぐに試すことができます。組み込み前に動作を直感的に確認でき、利用イメージの検証から本格導入までをスムーズに進めることができますので導入検討中の評価にお役立ていただけます。

商号	さくらインターネット株式会社（SAKURA internet Inc.）
代表	代表取締役社長 田中 邦裕
本社	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
創業	1996年12月23日
設立	1999年8月17日
資本金	112億8,316万円
事業内容	・クラウドコンピューティングサービスなどの提供 ・データセンター運営
上場証券取引所	東京証券取引所プライム市場（証券コード：3778）
CSR	・ISMAP・ISMS・ISMSクラウドセキュリティ認証 ・PCI DSS・プライバシーマーク・SOC2 SOC3
所属団体	・特定非営利活動法人 日本データセンター協会 ・一般社団法人日本インターネットプロバイダー協会 ・一般社団法人ソフトウェア協会 ・日本MSP協会 他



「やりたいこと」を「できる」に変える

私たちは「熱量」を持って挑戦する人々の夢を全力で応援します。

新しい社会のインフラを支えながら、
最先端のサービスを構築していく。

さくらインターネット株式会社 代表取締役社長
田中 邦裕