

2024年3月15日

SB Intuitions

日本語LLMの「いま」 そして未来に向けたSB Intuitionsの挑戦

SB Intuitions株式会社 取締役 兼 CTO

折原 大樹



オリハラ ダイキ
折原 大樹

SB Intuitions株式会社
取締役 兼 CTO

ソフトバンク株式会社
テクノロジーユニット統括 データ基盤戦略本部 副本部長
デジタル社会基盤整備室

サービスプラットフォーム開発の責任者として、アプリケーションサービスのプラットフォーム開発、ユニファイドコミュニケーションやボイスサービス、セキュリティサービスの開発を担うのち、現在はxIPF(次世代データ連携基盤)構想構築に向けての研究開発、及び事業化開発の推進責任者のほか、IoTを活用した先端ビジネスモデルの実現を推進。また、SB Intuitions株式会社 技術本部本部長 兼 CTO、一般社団法人日本IT団体連盟を兼務。

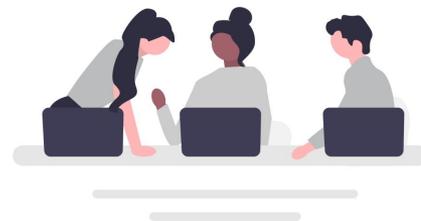
SB Intuitions | 生成AIモデルの開発を推進

SoftBank

出資(100%)

SB Intuitions (エスビー インテューイションズ)

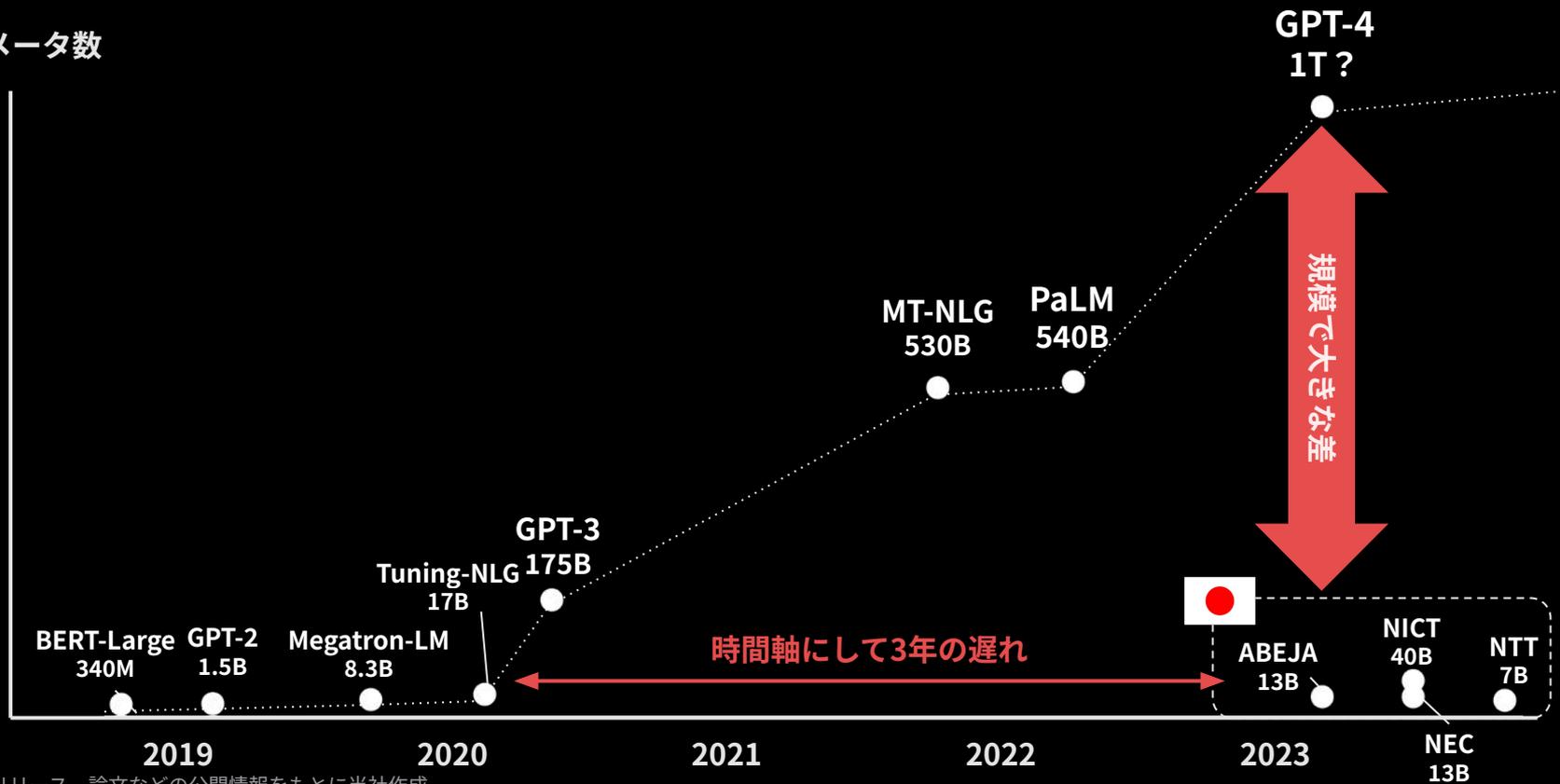
- 日本語LLMに関する研究開発
- 日本の文化・商習慣にあった安心安全なAIサービスを提供



生成AIの開発経験のあるエンジニアがグループ各社から集結

日本のLLM開発は海外に遅れを取っている

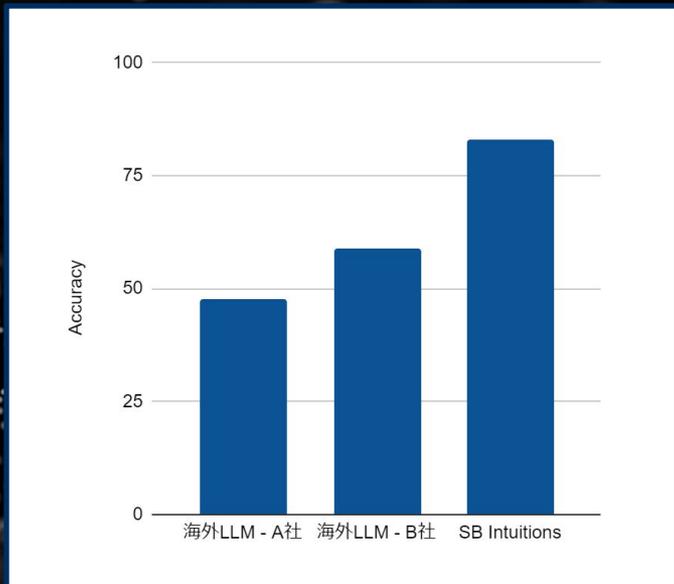
パラメータ数



※ プレスリリース、論文などの公開情報をもとに当社作成

一方、日本文化における知識レベルでは海外に見劣りしない

▶ 日本の文化や歴史に深く立ち入る問題や込み入った知識を問う問題が多く含まれる評価指標「AI王」では、SB Intuitionsにて開発途中のモデルがすでに海外モデルの性能を超える結果も



「AI王」を用いた評価結果

※ SB Intuitions社内で実施した評価

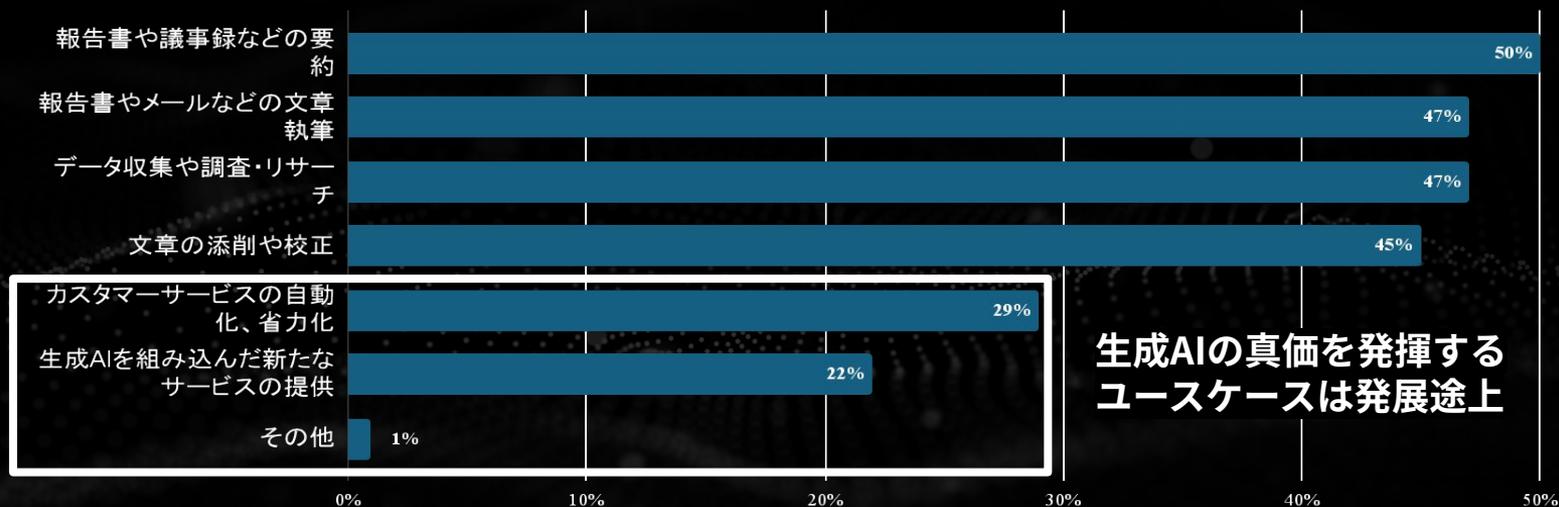
※ 海外LLMについては2024年2月に公開されているモデル / APIを用いて実施した評価結果を記載

問題例	回答
玉ねぎを炒める時の炒め具合の目安となる明るい褐色のことを、あるお菓子にたとえて何色という？	飴色
銀行などの金融機関で、長期間預け入れや払い戻しのない預貯金口座のことを何という？	休眠口座
複数のプレイヤーがコンピュータゲームで対戦する競技のことを、アルファベット1文字を使って「何スポーツ」という？	eスポーツ

日本市場での生成AIのユースケース

要約や添削などのシンプルな利用は拡大している一方で、生成AIのシステムへの組み込みは途上

Q. 現在検討されている（あるいは既の実現できている）生成AIのユースケースとして、当てはまるものをお答えください。（n=799）



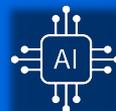
生成AIの真価を発揮するユースケースは発展途上

SB Intuitions の Vision

SB Intuitionsの生成AIが
あらゆる産業の競争力の源泉になり、世の中を成長させる



大規模モデル



AI Suite



国内最大規模のAI計算基盤

SB Intuitionsの大規模モデルとAI Suite

日本語データで構築した 生成AI基盤モデル

2024年度内に3,900億パラメータの生成
AI基盤モデルを達成予定

公開モデルを利用せず
日本語性能の高い国産LLM

利用環境を提供する AI Suite

生成AIを活用したシステムを
構築するための機能群

国内最大規模のAI計算基盤

生成AI開発向け大規模計算基盤の整備

- 2023年10月、生成AI開発向けの国内最大級※の計算基盤の稼働を開始
- NVIDIA Tensor コア GPUを2,000基以上搭載したAIスーパーコンピュータ-NVIDIA DGX SuperPOD™、NVIDIA ネットワーキング、NVIDIA AI Enterpriseソフトウェアで構成された大規模クラスター
- まずソフトバンクとSB Intuitionsで段階的に利用しながら、2023年度中に全ての投資と構築を完了させ、早期に大学や研究機関・企業などに提供する予定



SB IntuitionsでのLLM開発計画



※LLMの学習向けの計算基盤において国内最大級。2023年10月31日時点での公開情報に基づく。弊社調べ。

SB Intuitionsの大規模モデルとAI Suite

日本語データで構築した
生成AI基盤モデル

公開モデルを利用せず
日本語性能の高い国産LLM

利用環境を提供する
AI Suite

生成AIを活用したシステムを
構築するための機能群

国内最大規模のAI計算基盤

「日本語性能」の高い生成AI基盤モデルの構築

生成AIの性能

パラメータ数

×

大規模で良質な学習データ

×

適材適所なモデル学習

×

安全性の向上

×

外部知識活用・ツール連携

専用モデルより高い性能を狙うパラメータ数

2024年度内に3,900億パラメータの国産LLMを構築予定
その後、さらなるパラメータ数の増加を検討

生成AIの命となる、学習データの質 × 量

内製もしくは研究機関や企業との共同研究により、
良質な日本語の学習データを大規模に構築

様々な場面に適応したモデル学習

大規模モデルからの蒸留による小規模モデルの構築
医療などのドメインへのモデル適応

高度な安全性と利便性を兼ね備えたサービスを提供

有用性を保ちつつ日本の文化・習慣における安全性向上
外部知識を活用したRAGやツール連携、エージェント機能を実現

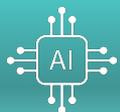
日本語データでゼロから生成AI基盤モデルを構築

豊富な計算基盤と独自の技術力を用いて、日本語性能の高いモデルをhomegrownに構築して提供

	公開モデルへの継続学習	公開モデルを利用しない構築
デリバリー	<p>◎ <u>少ない計算資源でも可能</u> 公開されたモデルを利用できるため素早く必要なモデルを開発できる</p>	<p>△ <u>計算資源、技術人材が必要</u> 開発にかかるコストは時間、人的リソースともに膨大である</p>
データセット	<p>× <u>コントロールできない</u> 公開モデルが利用した学習データセットに著作権物が含まれる等信頼性に課題</p>	<p>◎ <u>用途に合わせて自由自在</u> モデルに学習させるデータをすべて管理でき、要求性能に合わせて構築可能</p>
発展性	<p>△ <u>提供元の意思決定に依存</u> 無償でモデルが提供され続ける確証はなく継続的な安定提供に課題がある</p>	<p>◎ <u>再現性のある安定提供</u> 自社でモデル、データを構築するため、恒常的に研究開発を推進</p>
モデル品質	<p>○ <u>既に性能のよいモデル</u> 汎用的に性能の良いモデルが公開されているため、その品質を基準に利用できる</p>	<p>◎ <u>用途に合わせて特化</u> 不要なデータの影響を受けず、モデル用途に合わせて品質を特化させられる</p>

生成AI関連のオープンな環境づくりの分類

モデルだけが公開の対象ではない。「オープンソース」の中にも、異なる目的のリポジトリがある。



公開モデル

訓練済みのモデルを公開しているもの
これらのモデルを用いて独自のモデルを開発することができる

例) Llama 2 / Gemma



データセット

モデルの学習に利用できるデータセットを公開しているもの
構築するモデルに合わせて必要なデータセットを利用できる

例) Common Crawl / The pile



周辺ツール
ソースコード

モデルの学習や利用に必要なコード類を公開しているもの
構築や利用時に車輪の再発明を防ぎ、コミュニティで開発できる

例) Megatron-LM / DeepSpeed / LangChain

一定の計算基盤
があれば、モデル
の再現が可能
なもの

例) RedPajama /
Bloom

データ獲得における競争

訴訟文化のビックテック。法整備を待っているアプローチでは差が開くばかり。
コンテンツホルダーとAI開発者のwin-winな関係構築が望ましい。



- 著作権法 30条の4
- 文化庁「AI と著作権に関する考え方について」
- 広島AIプロセス
- 新事業者ガイドライン

など



データの利用におけるルールや仕組み

生成AIはデータ獲得の競争へ

データが圧倒的に不足。早急なデータ流通の整備が必要



データ流通までのハードル

個人情報の扱い

個人情報の秘匿するなど、エンドユーザーに納得感のある仕組み

トレーサビリティ

学習・生成において、データがどのように利用されたのか追跡可能

著作権などの整備

各データをどのように扱うことが最もよいのかルールメイキング

アライアンス戦略

公開データを使った汎用モデルを各社で作るのか、差別化するのか

データホルダー還元

データ提供者に経済的なメリットを還元し、長期的な成長

利用管理の徹底

流通データが制定されたルールの中で使われる保証となる仕組み

SB Intuitionsの大規模モデルとAI Suite

日本語データで構築した
生成AI基盤モデル

公開モデルを利用せず
日本語性能の高い国産LLM

利用環境を提供する
AI Suite

生成AIを活用したシステムを
構築するための機能群

国内最大規模のAI計算基盤

生成AI環境を構築するためのAI Suite

生成AIの利用に必要なツール群をフルサポートし、システムへの生成AIの組み込みをスムーズに

AI Suite

AI Solution Tools

AgentやRAGをはじめ、AI活用に欠かせないツールセット

Model Store

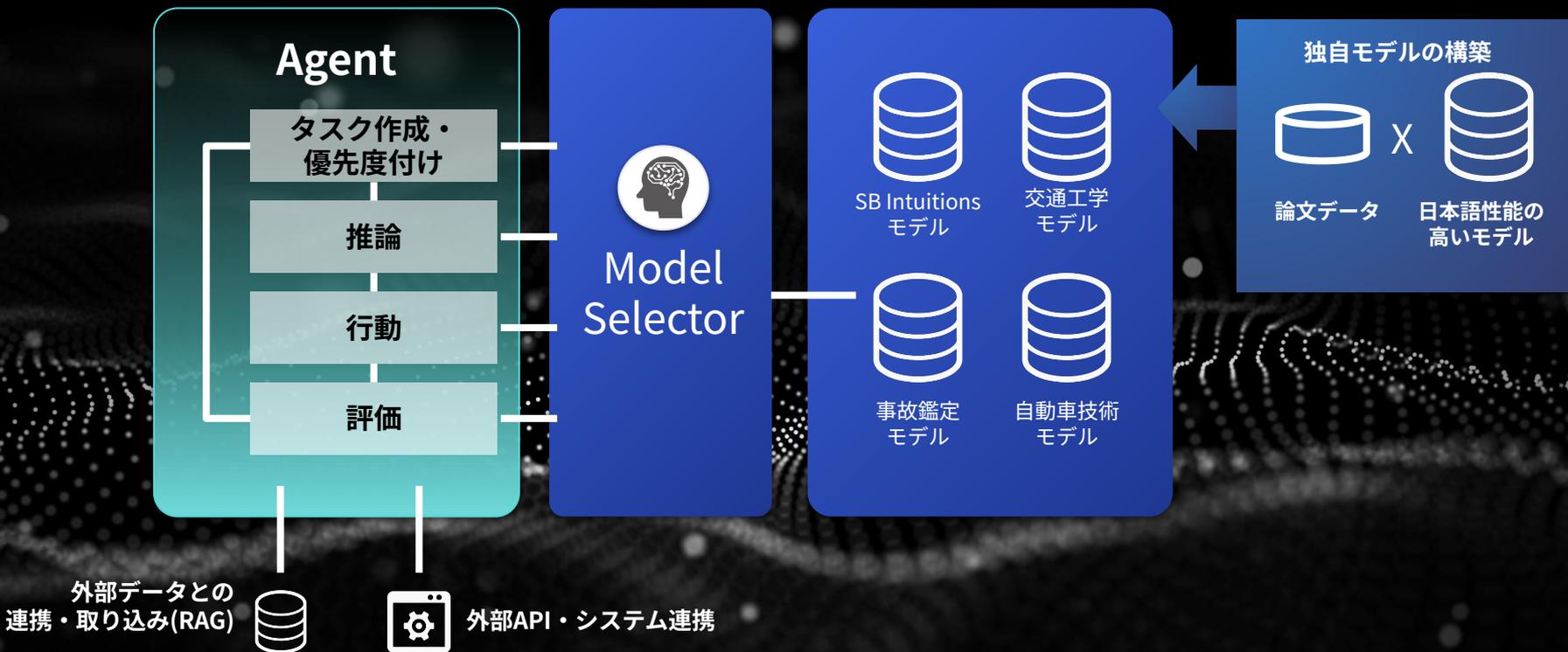
日本語性能が高い自社モデルなど、追加学習の基盤となるモデル群の提供

ML Ops Tools

堅牢なデータ管理のもと、モデルの追加学習・運用を容易にする仕組み

複数のモデルを組み込んだシステム構想

複数の専門家モデルの意見をファシリテートして結論を導き出す



SB Intuitionsの大規模モデルとAI Suite

日本語データで構築した
生成AI基盤モデル

公開モデルを利用せず
日本語性能の高い国産LLM

利用環境を提供する
AI Suite

生成AIを活用したシステムを
構築するための機能群

国内最大規模のAI計算基盤

生成AIにまつわる競争軸



計算基盤の競争



クラウド
例) Azure/AWS



オンプレミス
例) お客様環境



生成AI開発・活用基盤の競争



例)
Amazon Bedrock



例)
Google Vertex AI



例)
NVIDIA AI Enterprise



モデルの競争



Open Model



Open Model



Open Model



Closed Model

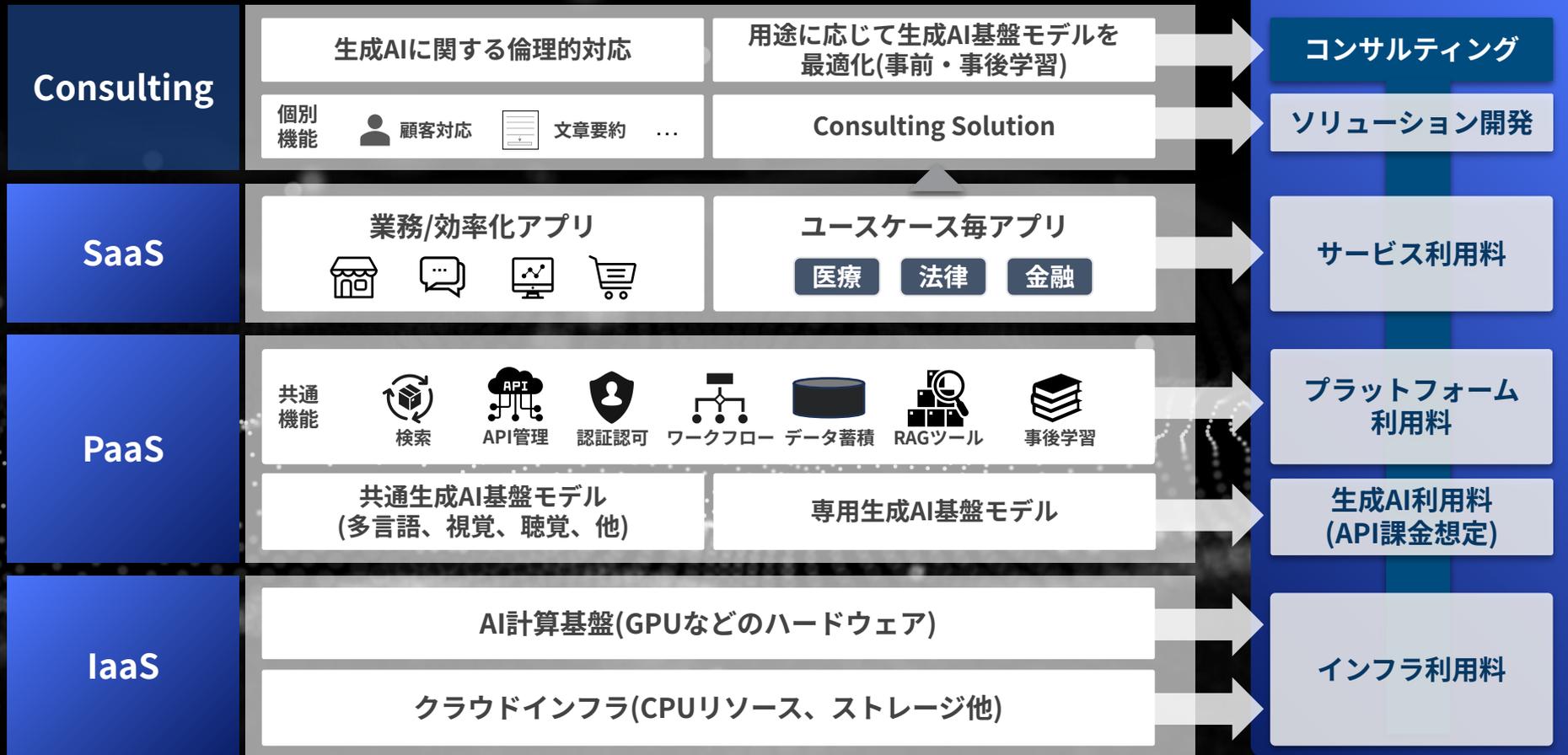


Closed Model



Closed Model

生成AIのビジネスモデル構造

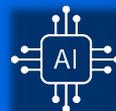


SB Intuitions の Vision

SB Intuitionsの生成AIが
あらゆる産業の競争力の源泉になり、世の中を成長させる



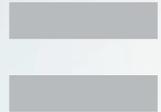
大規模モデル



AI Suite



国内最大規模のAI計算基盤

 **SB Intuitions**