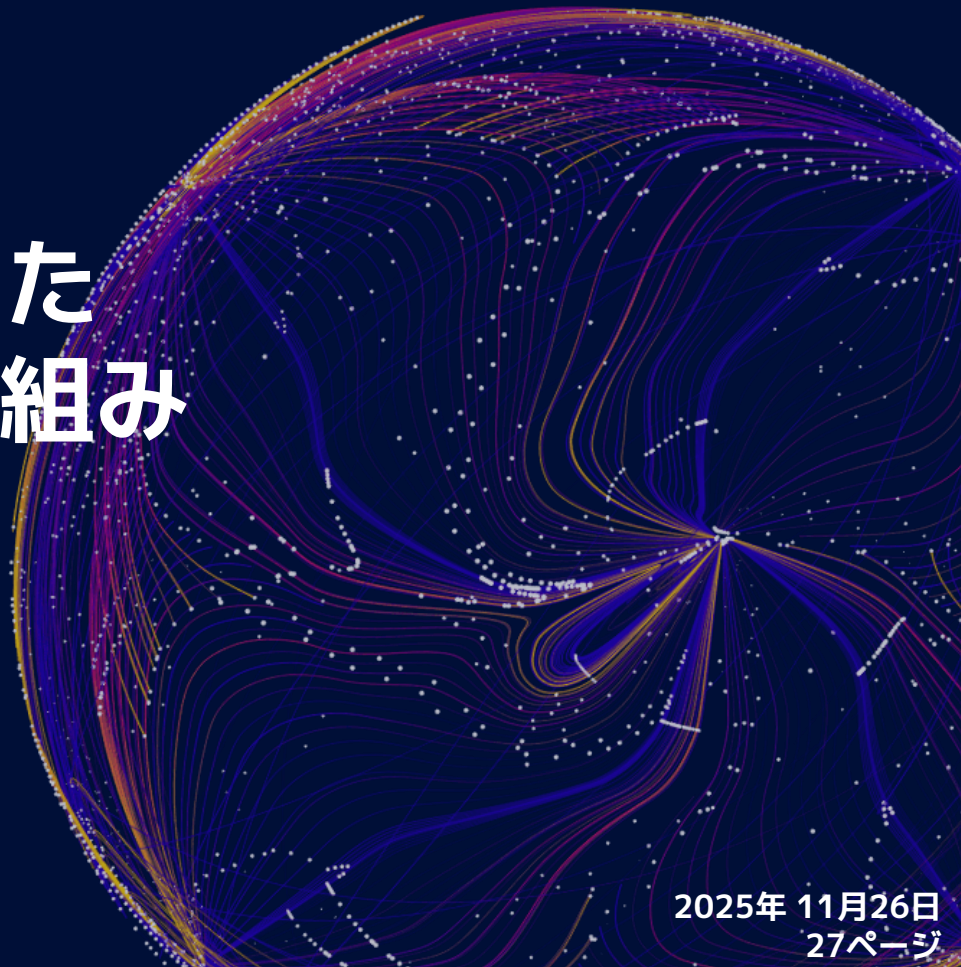
SoftBank

AIの社会実装に向けたソフトバンクの取り組み

ソフトバンク株式会社
法人統括 AIﾌﾟﾗｯﾄﾌｫｰﾑ開発本部
ｸﾗｳﾄﾞ・AIｻｰﾋﾞｽ第1統括部 統括部長
鈴木 邦佳



2025年 11月26日
27ページ

自己紹介

すずき　　くによし
鈴木 邦佳

【経歴】
大学卒業後、1997年に日本テレコム（当時）入社し、
法人営業、クラウドSaaS/PaaSなどのプロダクト部門を経て、
AIプラットフォーム開発本部クラウド・AIサービス第1統括部の統括部長に就任。
同統括部では、データセンター・クラウド接続ネットワークを含む
クラウドおよびAI事業の運営・事業企画・サービス企画を統括。

【実績】
外部セミナーにおいてはMicrosoft、Google主催イベントでの講演や、
自社イベントでのクラウド・AIソリューションの紹介多数

AIの社会実装に向けた取り組み

SoftBank

AIの社会実装を進め、真のデジタル社会を実現

出所) ソフトバンク株式会社 第39回定時株主総会 事業戦略説明資料

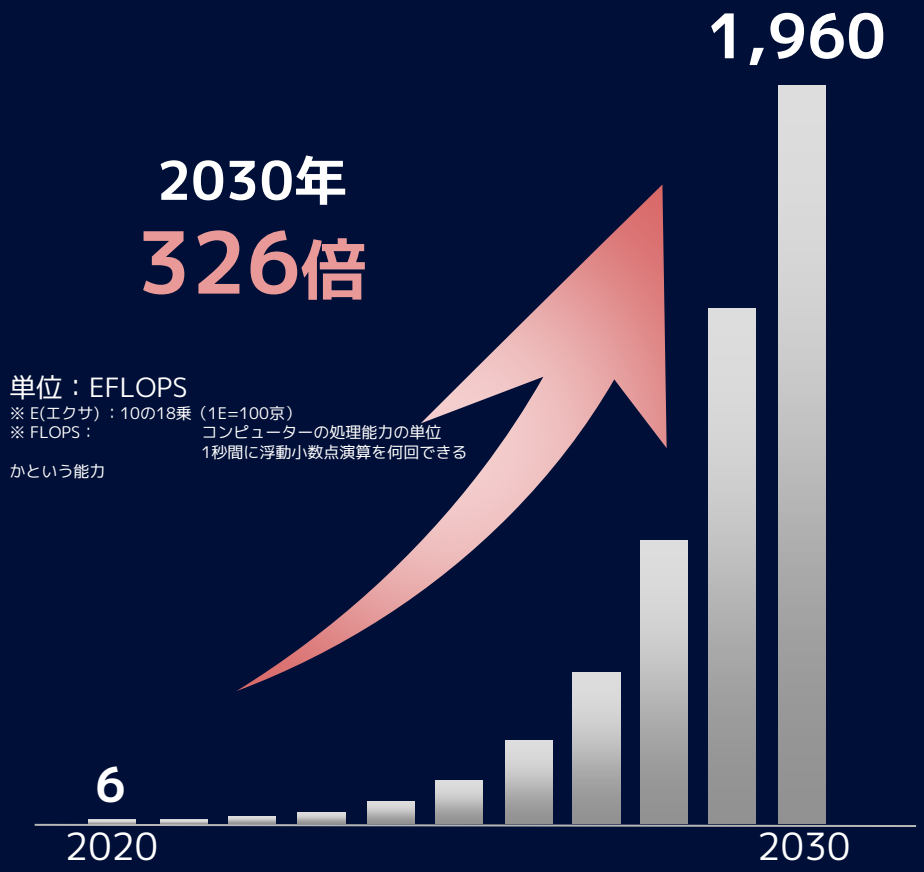
第４次産業革命がもたらす社会

SoftBank

AIが日常となり
膨大なデータの生成・処理が必要に

出所) ソフトバンク株式会社 2023年3月期 決算発表資料

AI社会に求められる計算能力



AI活用の拡がりによって

データ処理の需要が

急増

出所) 第5回 半導体・デジタル産業戦略検討会議（経産省）をベースに生成AIの普及を前提とした各種公開情報を元に当社推計

増大する計算需要に応えるために

SoftBank

日本国内における計算基盤設備が必要

学習用途での観点

開発環境の構築

- 世界のLLM開発競争において、日本独自のモデルを開発するには強力な計算基盤が必要
- 国内に大規模なデータセンターを整備しないと、海外クラウド企業に依存することになる

データ主権の確保

- 海外クラウドに依存するとデータ流出や規制の影響を受けるリスクがある
- 日本国内のデータを安全に活用するには、日本国内の計算基盤が不可欠

推論用途での観点

応答速度

- 海外クラウドを利用すると、ネットワーク遅延が発生し、推論速度が低下する可能性がある
- 多数のユーザーからの同時リクエストを迅速に処理するには、物理的に近いデータセンターが必要

安全保障リスク

- CLOUD法や中国国家情報法の影響で、日本の機密データが外国政府にアクセスされるリスクがある
- 有事の際のクラウド遮断により、日本の重要インフラや防衛システムが制御不能になるリスクがある

ソブリン性確保の必要性

SoftBank

01
安全性観点
機密保持性、信憑性、公平性、頑健性…

現状、AI大手モデルは高い性能をすでに実現、
ただし今後のフェイクニュース等の拡散抑止の観点も含めて今後も必要

02
文化適合性･教育的観点
日本文化に関する深い常識･知識

現状:日本の歴史など深い知識や、各国の法令･慣習の違いなどに起因して文化的ハルシネーションの生成可能性、その他領土認識問題･政治的イデオロギーなどにおける文化非適合可能性

03
国内制度･法令準拠観点
コンプライアンス観点での正しい知識･情報提供

現状:各国の法制度の違いなどに応じて、文化的ハルシネーションを発生

04
説明性･透明性観点
データ･システム構造など、有事対応が国内で完結

現状:出力の説明性に加え、システムの透明性や
運用情報などに関する情報がなければ改善ができない

05
インフラの所在地的観点
サービス運用においての国内サーバー活用など

外国にサーバーがある場合、責任の所在が不明もしくは
追及が困難になるなど不都合が発生

06
データの越境性観点
サービス運用時のユーザーデータ･運用データが越境しないことなど

海外サーバー上にデータが存在する場合、
漏えいのリスクがあり、責任の所在が不明・追究が困難になるなど不都合が発生

07
責任観点･紛争解決観点
問題発生時にモデルを含む修正能力を保有するかどうか?
係争が生じた時に、国内で裁判 が実施できるか?

モデルなどに関わる不都合が国内で解決できず、
紛争になった場合も国内で解決できない枠組みになっている場合あり

08
国内産業振興観点
利益が出た時に、国内企業が十分な収益を上げられる
すなわち税金が国内に落ちる枠組みなっているかどうか?

コストとしてほとんどが海外企業に落ちる仕組みになっている場合もあり

コントローラビリティーの確保

SoftBank

ソブリンAI・ソブリンクラウドによる
安全性・透明性の担保が可能

国内規制・制度・文化に準拠

- ガードレール
- Red Teaming
- ハルシネーション低減
- 偽情報・盗用対策技術
- •••

国内で管理・制御・問題解決できる

- セキュア業務システム連携
- 暗号化・認証機密Computing
- オンプレ・エッジ推論（国内）
- 外部サービス依存性低減
- •••

国内ソブリンAIの発展に向けて

SoftBank

AIサービス開発に必要な国産技術基盤を一気通貫で提供

国・自治体・産業 など

SaaS

SaaS

SaaS

SaaS

SaaS

…

AIのもたらす価値を
日本全体が享受できるサービスを展開

デジタル
公共インフラ

AIモデル	ソブリンAI　Sarashina			
プラットフォーム	ソブリンクラウドプラットフォーム			
計算基盤	ソブリンAI向け大規模GPU基盤	CPU/GPU/ストレージ	クラウドプログラムDGX等	
データセンター	堺	苫小牧	…	

高性能な国産LLM
日本独自の価値観を備えた
国産の大規模言語モデル

**ソブリンクラウド
プラットフォーム**
国産GPUクラスタと国産PFを
統合的に運用/制御するマネージド基盤

国内大規模計算インフラ
国際競争力のある国産LLMを持続的に
学習/開発できる大規模計算インフラ

出所) SoftBank World 2025【基調講演】AX with SoftBank ～次の一手がここにある

国内ソブリンAIの発展に向けて

SoftBank

AIサービス開発に必要な国産技術基盤を一気通貫で提供

国・自治体・産業 など

SaaS　SaaS　SaaS　SaaS　SaaS　…

AIのもたらす価値を
日本全体が享受できるサービスを展開

デジタル
公共インフラ

AIモデル	ソブリンAI　Sarashina			
プラットフォーム	ソブリンクラウドプラットフォーム			
計算基盤	ソブリンAI向け大規模GPU基盤	CPU/GPU/ストレージ	クラウドプログラム DGX等	
データセンター	堺	苫小牧	…	

高性能な国産LLM
日本独自の価値観を備えた
国産の大規模言語モデル

ソブリンクラウド プラットフォーム
国産GPUクラスタと国産PFを
統合的に運用/制御するマネージド基盤

国内大規模計算インフラ
国際競争力のある国産LLMを持続的に
学習/開発できる大規模計算インフラ

出所) SoftBank World 2025【基調講演】AX with SoftBank ～次の一手がここにある

ソフトバンクの次世代社会インフラ構想

SoftBank

分散型デジタルインフラを構築しデジタル社会を実現

国内大規模計算インフラのアーキテクチャー

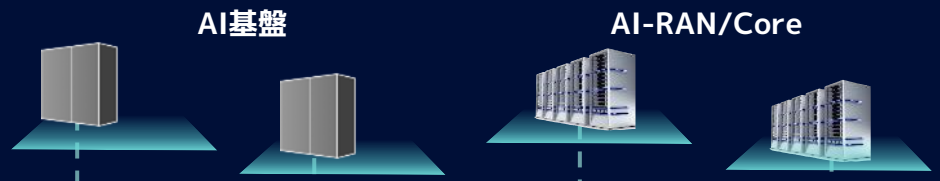
SoftBank

階層構造によるデータ分散処理
利用用途に応じてデータ処理場所を最適化

Brain DataCenter

中核拠点
（東阪＋第3極）

Regional Brain



地方都市/
都道府県単位など

Edge Computing（MEC）

全国に多数配置

現在の注力領域

大規模データセンターおよび国内最大級の計算基盤の構築に注力
数千億規模の投資を決定し今後も拡充に取り組む

大規模データセンター

苫小牧で大規模AIデータセンターを建設中
2026年度 稼働開始予定

国内最大級のAI計算基盤

各4,000基超の「NVIDIA Hopper GPU」
「NVIDIA Blackwell GPU」整備完了

SoftBank

AI計算基盤を活用したクラウドサービス

AIデータセンター GPUサーバー

2025年10月～ サービス開始

科学シミュレーション

大規模言語モデル (LLM)

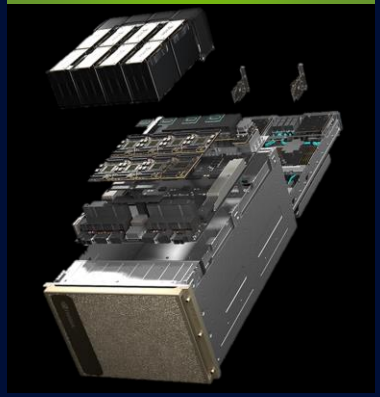
材料開発

システム仕様

SoftBank

NVIDIA推奨構成「NVIDIA DGX SuperPOD™」採用
高性能GPUクラスターを利用シーンに合わせてご提供

搭載GPUサーバー
(2025年11月時点)

NVIDIA DGX A100	NVIDIA DGX H100
	
約2,000GPU (0.7EFLOPS)	約6,000GPU (4.7EFLOPS)

NVIDIA DGX™ A100・H100サーバー

- NVIDIA A100•H100 80GB GPUを8基搭載するGPUサーバー
- NVIDIA GPU利用に最適化されたハードウェア・ソフトウェア

サービスの特長

専有環境・柔軟な利用期間・開発支援により
効率的かつ迅速なモデル開発を実現

専有利用

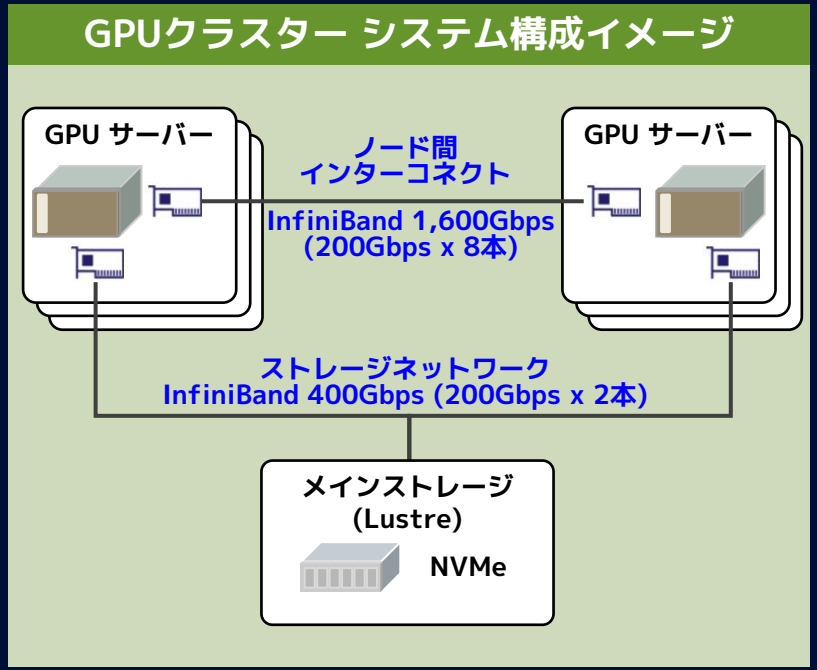
短期利用

開発支援

サービスの特長①

SoftBank

GPUサーバーを1台から複数台のクラスター構成で専有提供



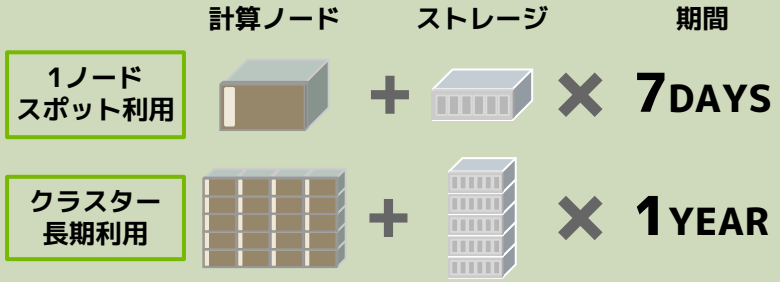
- **クラスター構成で占有利用可能**
  - GPUサーバーを1台から複数台クラスター構成でサービス提供
  - 全てのGPUサーバーはノード間1,600Gbps の高速回線で接続
- **高速ストレージ**
  - 計算処理データ保存にLustreファイルシステムの共有ストレージを提供
  - ストレージにはNVMeを採用しファイル保存の高速化を実現

サービスの特長②

SoftBank

最低利用期間7日からの短期スポット利用が可能

台数 x 期間 計算ニーズに応じて提供



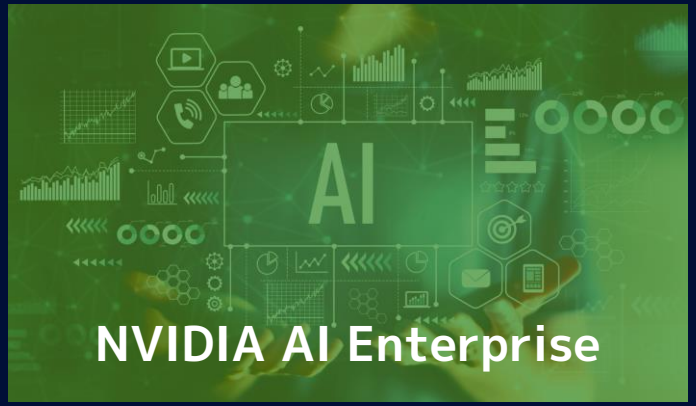
専用線提供によるセキュアな閉域網接続

- **台数 x 期間 計算ニーズに応じて提供**
  - ノード台数は1台(8GPU)から最大20台(160GPU)を1ユニットとしてクラスターでの提供可能
    ※20台以上は応相談
  - 高速通信可能なLusterメインストレージは100GB単位で1PB超の大容量ストレージも提供可能
  - 最低利用期間は7日間から計算処理のニーズに合わせてご利用頂けます

- **その他契約メニュー**
  - インターネット回線以外に統合VPNサービスであるSmartVPNを利用し、専用線での閉域回線として接続可能

サービスの特長③

お客さまの計算ニーズを支援するソフトウェアを標準付帯



- **NVIDIA AI Enterprise**
  - NVIDIAが提供する一元管理されたプラットフォーム
  - AIデータセンターで実行可能な50を超える NVIDIA AI ソフトウェア フレームワークと 事前トレーニング済みモデルで開発を効率化
  - お客様の計算プログラムをNVIDIA Private Registoryで管理

- **JOB管理システム**
  - Slurmジョブスケジューラをセットアップ済みでご提供 面倒なノード管理設定不要で、契約後すぐに計算処理に 利用可能

AIデータセンター活用事例

SoftBank

AIデータセンターを活用し、日本独自のAIモデルを開発

国・自治体・産業 など

SaaS　SaaS　SaaS　SaaS　SaaS

AIのもたらす価値を
日本全体が享受できるサービスを展開

デジタル
公共インフラ

AIモデル	ソブリンAI　Sarashina			
プラットフォーム	ソブリンクラウドプラットフォーム			
計算基盤	ソブリンAI向け大規模GPU基盤	CPU/GPU/ストレージ	クラウドプログラムDGX等	
データセンター	堺	苫小牧	…	

高性能な国産LLM
日本独自の価値観を備えた
国産の大規模言語モデル

**ソブリンクラウド
プラットフォーム**
国産GPUクラスタと国産PFを
統合的に運用/制御するマネージド基盤

国内大規模計算インフラ
国際競争力のある国産LLMを持続的に
学習/開発できる大規模計算インフラ

日本語国産LLM「Sarashina」

SoftBank

日本語に特化した大規模言語モデル(LLM)

Sarashina

SB Intuitions株式会社が開発するAIモデルシリーズ。
日本語の理解や生成において高い性能を発揮する。

日本語で生成AIを作る意義

言語特性への対応

漢字・ひらがな・カタカナの共存』や
『敬語・方言のニュアンス』を学習し、
汎用モデルでは捉えられない
日本語の複雑さを自然に反映できる。

文化・文脈への理解

日本語の表現には、
文化的背景や暗黙の了解が
(「忖度」「空気を読む」など)密接に関わる。
こうしたニュアンスをより正確に反映できる。

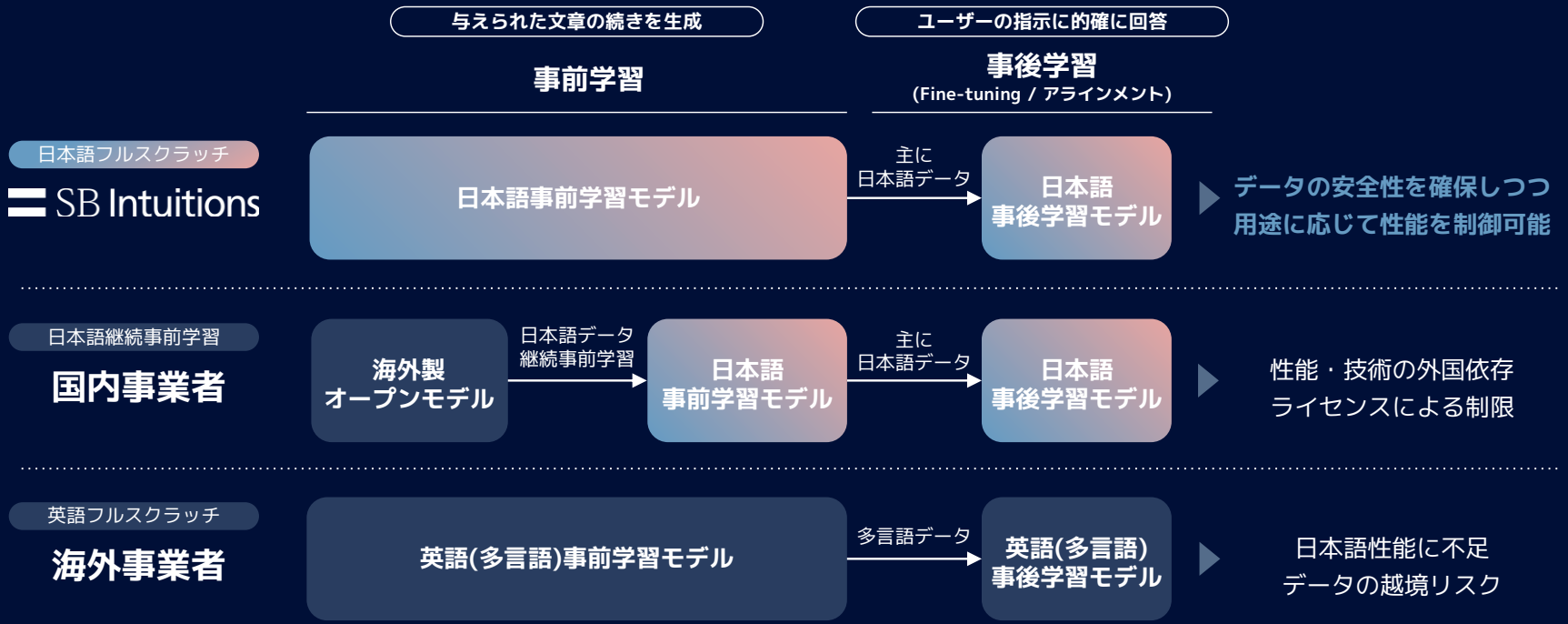
規制・倫理への対応

各国の法規制や倫理観は異なる。
日本語モデルを国内で開発することで、
日本の法律や社会的規範に沿った
運用が可能になる。

計算資源を活かした基盤モデル構築アプローチ

SoftBank

学習データをフルコントロールすることで高い安全性を確保



学習するデータによる性能の違い

日本語データの学習量に応じて日本語性能が向上

- 海外製モデルは学習データのほとんどが英語データであり、日本語データの割合は数パーセントにとどまる
- 学習させるデータ（日本語/英語）によって、モデルの性能に差異が生じる※

日本語データの学習効果が顕著 ▶ 日本の知識に関する質問応答、日英翻訳

英語データからでも学習可能 ▶ 日本語の一般教養、算術推論、コード生成

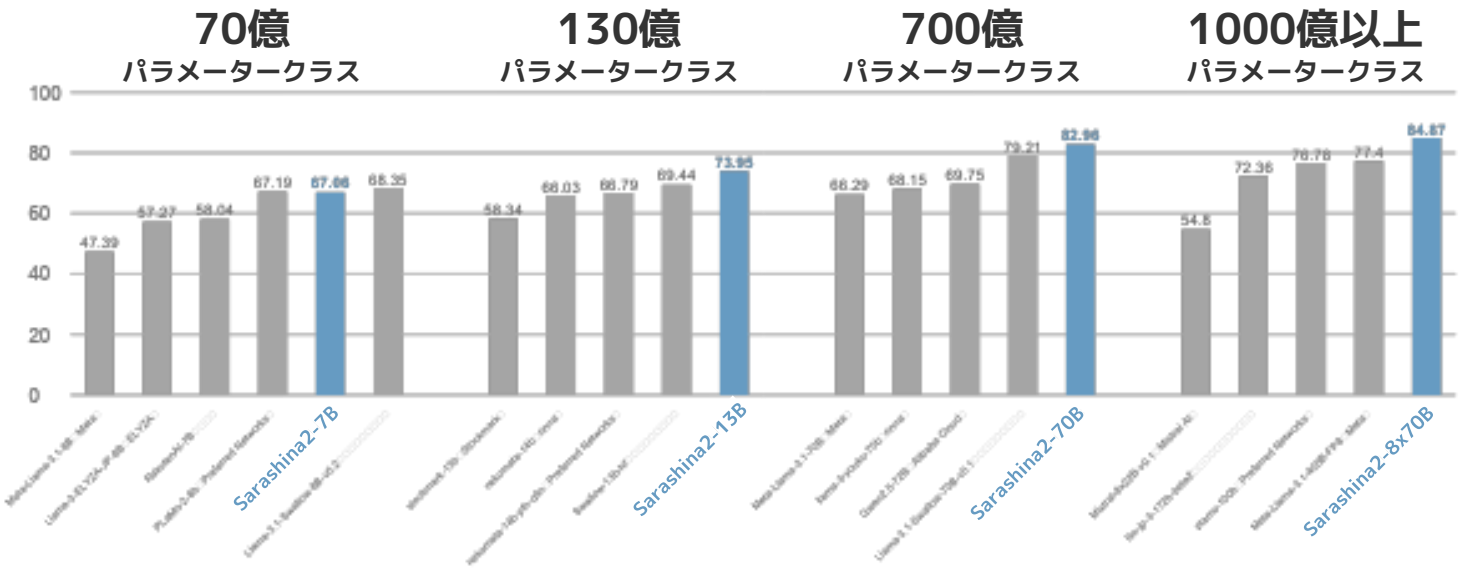
※第261回自然言語処理研究発表会での東京科学大学と産業技術総合研究所に所属する研究者らが発表した「LLMに日本語テキストを学習させる意義」より

事前学習モデル｜日本語性能に強みを持つ

SoftBank

「Sarashina2」が各パラメータークラスにおいて トップクラスの日本語性能を記録

日本語質問応答評価セットによる評価（平均スコア）



※2025年3月時点。当社内での比較による。

今後の取り組み

SoftBank

AI SaaSの開発・提供環境となる ソブリンクラウドプラットフォームを準備中

国・自治体・産業 など

SaaS　SaaS　SaaS　SaaS　SaaS

デジタル公共インフラ

AIモデル	ソブリンAI	Sarashina		
プラットフォーム	ソブリンクラウドプラットフォーム			
計算基盤	ソブリンAI向け大規模GPU基盤	CPU/GPU/ストレージ	クラウドプログラムDGX等	
データセンター	堺	苫小牧	…	

高性能な国産LLM
日本独自の価値観を備えた
国産の大規模言語モデル

ソブリンクラウド プラットフォーム
国産GPUクラスタと国産PFを
統合的に運用/制御するマネージド基盤

国内大規模計算インフラ
国際競争力のある国産LLMを持続的に
学習/開発できる大規模計算インフラ

出所) SoftBank World 2025【基調講演】AX with SoftBank ～次の一手がここにある


SoftBank
国産技術基盤により
AIの社会実装を加速させる

SoftBank

AIデータセンター お問い合わせフォーム

https://www.softbank.jp/biz/contact-us/demand/ai/gpu-inquiry/