

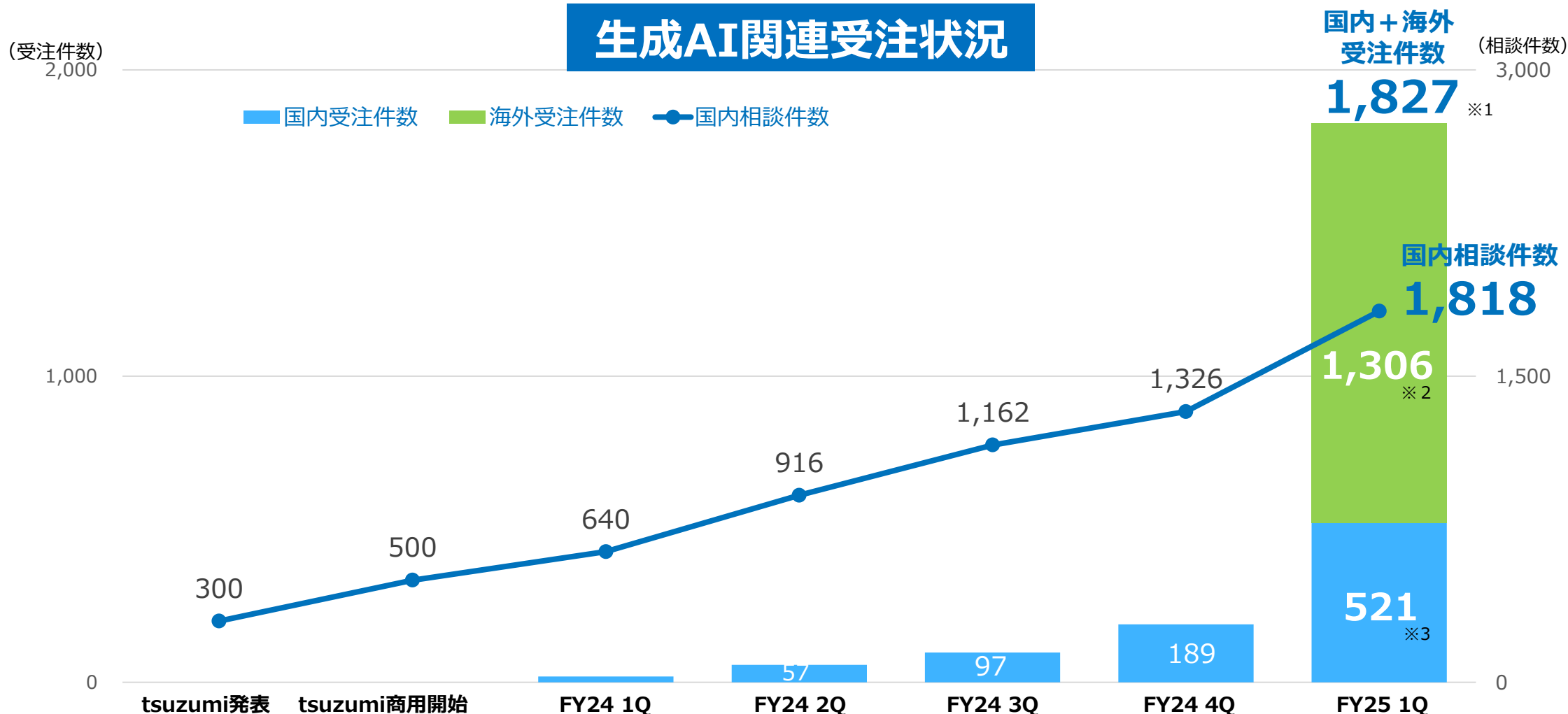
NTT版LLM  **tsuzumi**

AI For Quality Growthの実現へ **tsuzumi 2**のリリース

NTT株式会社
代表取締役社長 **島田 明**

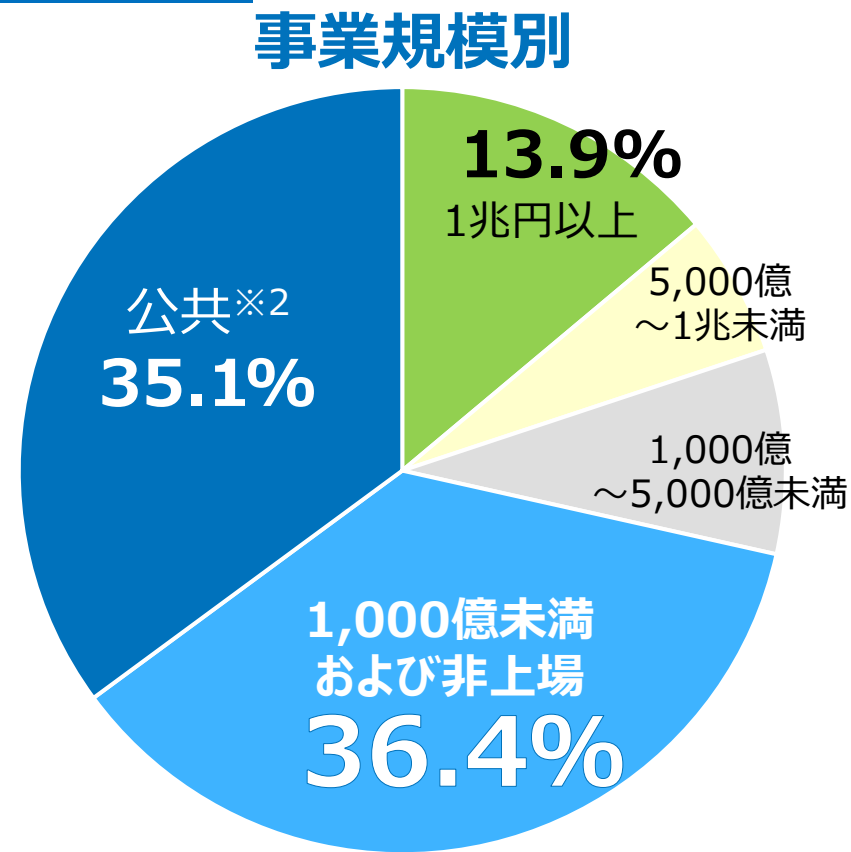
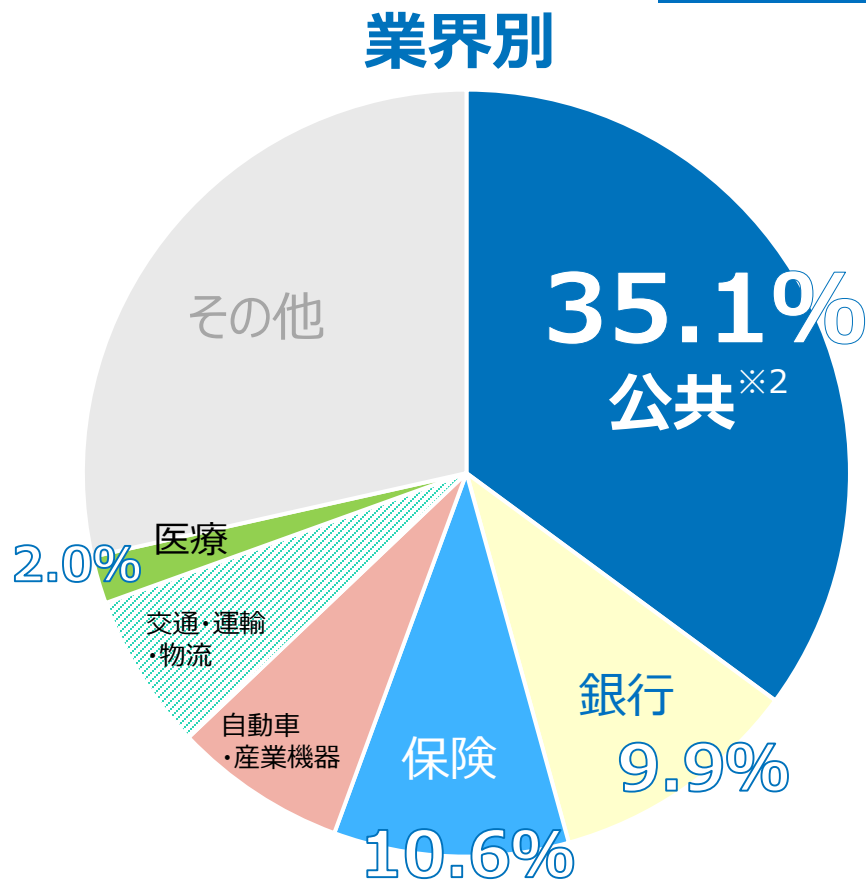
相談件数は継続的に増加、 受注件数は1,800件を超え堅調に進展

生成AI関連受注状況



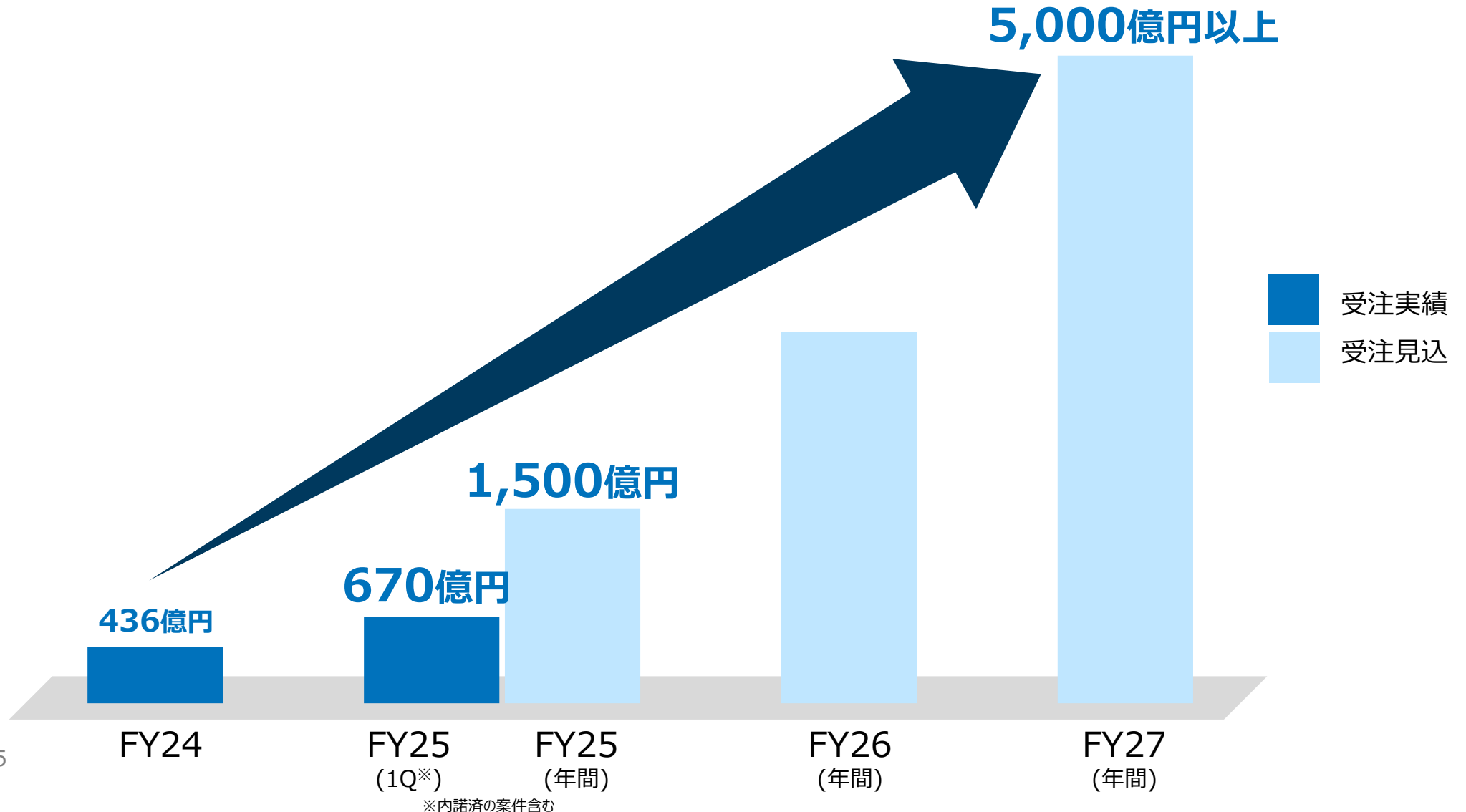
様々な業界、事業規模のお客さまから受注 AIがITソリューションビジネスを加速

受注件数割合(国内)^{※1}



※1 NTTグループ（ドコモビジネス・東日本・西日本・データグループ）生成AI関連受注件数より集計（累計）
McKinsey & Companyが定める業界区分および受注案件企業の売上高/経常収益に基づき作成
※2 中央省庁・地方自治体含む

670億円の受注実績、今年度末には1,500億円 ©NTT 2027年には5,000億円を超える見込み



安全保障・産業競争力強化で各国が自国開発AI重視へ

日本政府もAI基本計画で国産AI開発強化の動き

各国の動き



"Winning the AI Race:
America's AI Action Plan"
を発表 (2025年7月)

AIイノベーションの加速、インフラ構築、
AIにおける米国のリーダーシップ確立

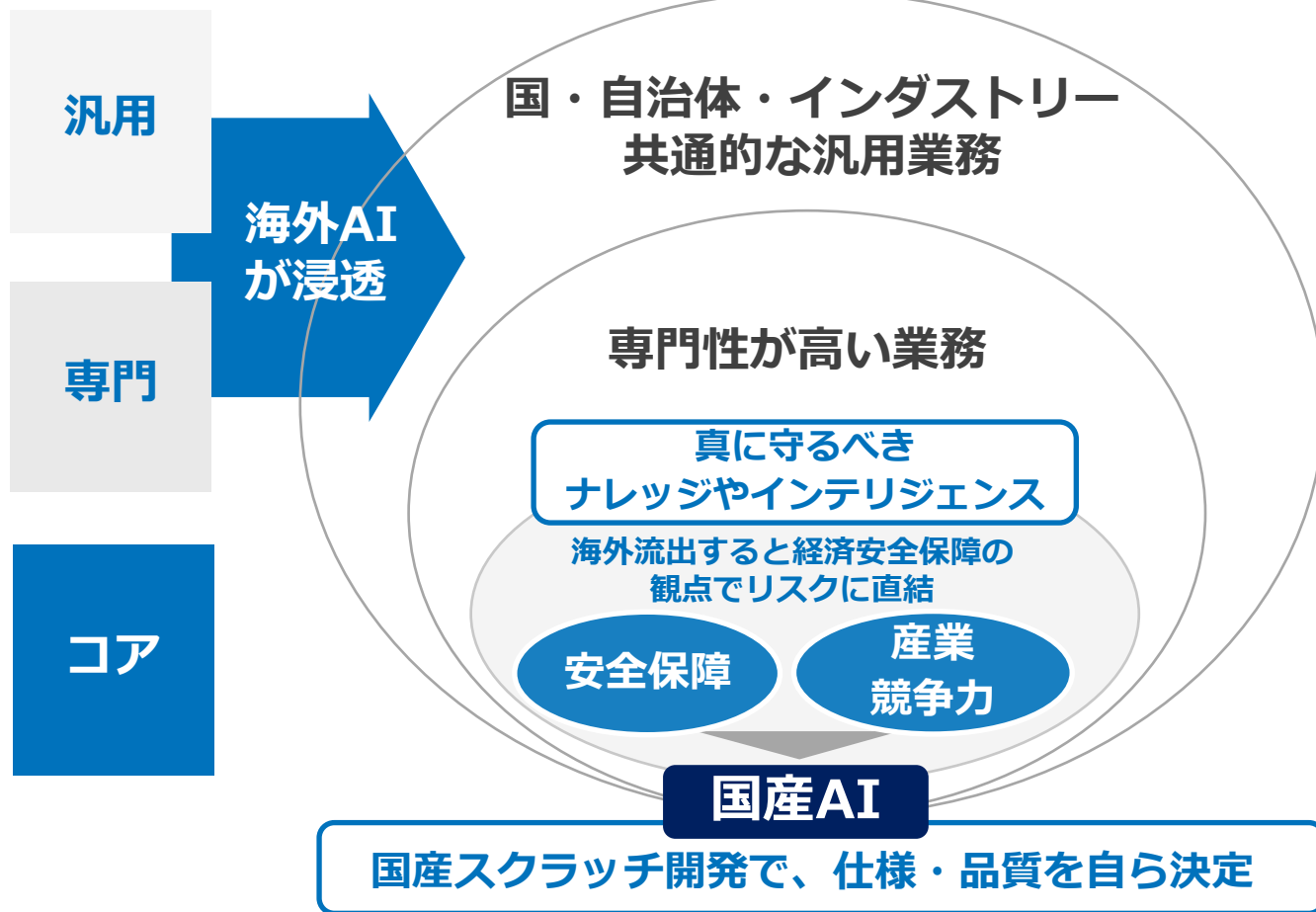


"グローバルAIガバナンス行動計画"
を発表 (2025年7月)



- ・ 政府策定のAI基本計画
「質の高い日本語を整備し、日本文化
や慣習を理解した信頼できるAI開発」
- ・ 経済産業省 GENIAC

ソブリンAIで守るべき領域





プライベートLLMとして活用可能な純国産モデル NTT

NTTがゼロから開発 純国産モデル



tsuzumi



ソブリン対応プライベートクラウド、
オンプレミス

データ・システム・運用の観点で **自国・自社** コントロール

質の高い日本語学習データ、日本文化、慣習を理解

自然言語処理研究の蓄積による良質な日本語コーパス、辞書を学習

学習データのコントロールにより権利保護

新聞等のデータを学習データから自主的に削除

米国での新聞社からの提訴の実態等を踏まえ

権利者の保護も意識し、学習データをコントロール

仕様・品質を自ら決定し、開発プロセスをフルコントロール

リリース・ライセンスをコントロールして安定的に提供

tsuzumi をアップグレード

2025年 10月 ● **tsuzumi 2 提供開始**

生成AI、機械学習等の関連研究の継続、各種学会での発表・論文賞等の受賞
AI関連ビジネスにおけるお客様ニーズに応じた活用、導入

2024年 3月 ● **tsuzumi 商用開始**

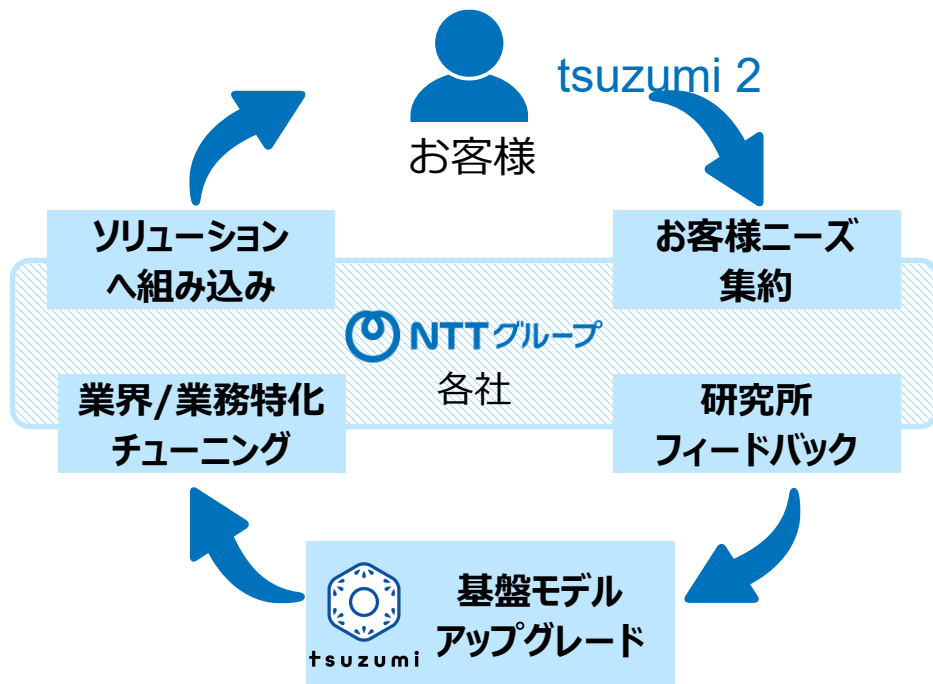
2023年 11月 ● **NTT版LLM tsuzumi 発表**

2020年 ● **LLM「NTT版BERT」をリリース**

1980年 ● **自然言語処理の研究開始**

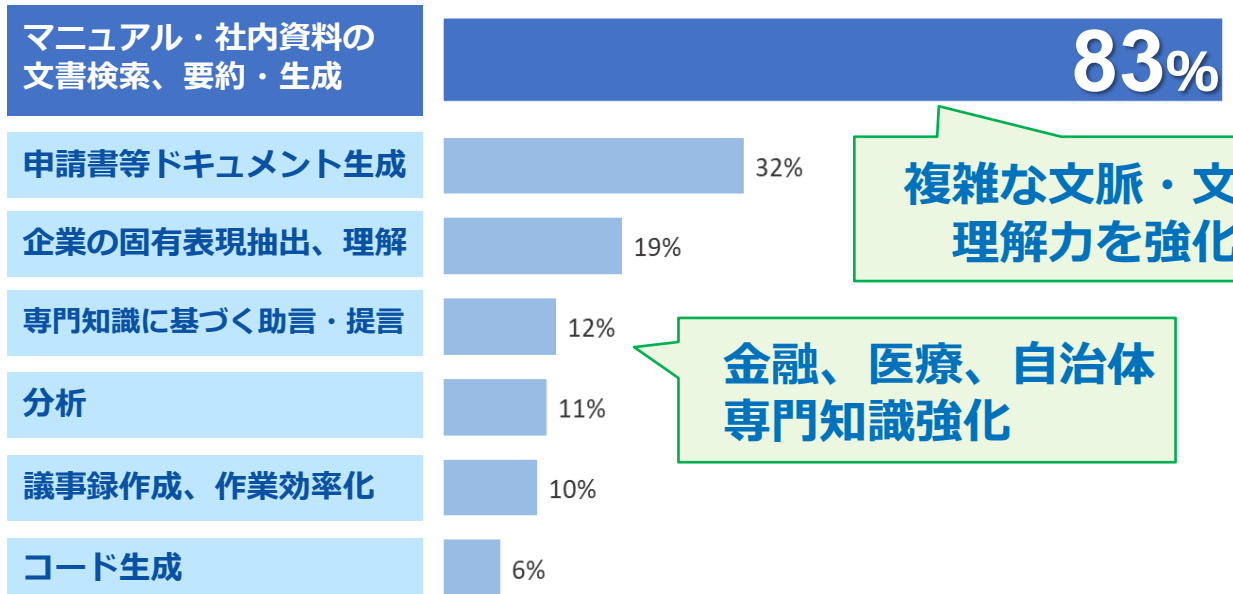
お客様ニーズ ▶ 研究開発のサイクルでアップグレード

開発 – 顧客のサイクルで
純国産AIモデルをアップグレード



お客様が

tsuzumi を活用したい業務 ▶ tsuzumi 2



複雑な文脈・文意
理解力を強化

金融、医療、自治体
専門知識強化

tsuzumi 2の進化について

業務適応力の進化

- 企業内文書の複雑な文脈や意図理解
- 長文読解や指示遂行能力強化



1

特化性能の進化

- 特定業界知識の強化
- カスタマイズ用データ量の省コスト化



2

基礎能力の更なる進化



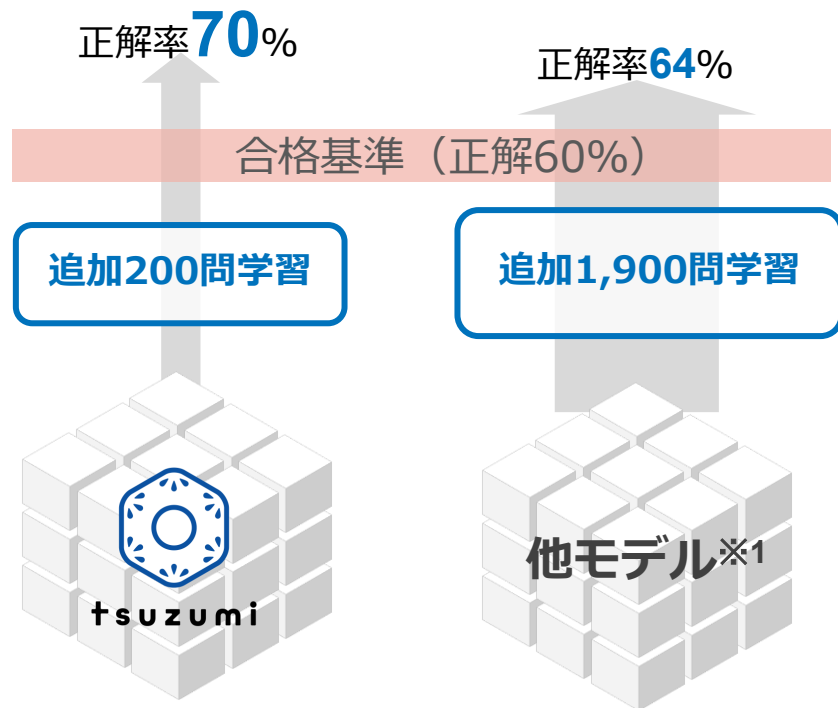
- 1 GPUで動作する省コストを維持し性能向上
- パラメータサイズアップ

3

特化モデルの開発効率を大幅に向上

強化した「金融」分野の実績

FP（ファイナンシャルプランニング）2級試験で検証



合格基準への到達が他モデルと比較し
1/10の追加学習データ量で実現※2

tsuzumi2

金融・公共・医療
 の業界を強化

その他の業界・領域についても、
今後市場ニーズに応じ強化

※1 Gemma-2 27B

※2 LoRAチューニングの学習データ量（ファイナンシャル・プランニング技能検定2級テスト問題数）

東京通信大学 × tsuzumi 2

「**学びの質**」を高める生成AI：**安全性・品質・運用性**で教育現場をサポート

- データを**学内に安全に保持**+**クラウド依存のない国産LLM**を中核とした学内LLM基盤整備
- tsuzumi 2の**高精度な文脈理解**と**長文処理**を強みに生成AI活用を加速

tsuzumi 2 採用のポイント

主権性・セキュリティ

ローカル運用・国産LLM

- ▶ 個人識別情報・学修データを**安全に学内で保持**

高度な指示理解・応用力

生成AIによるリアルタイム返信・高度な素案作成

- ▶ **質問回答、講義・テストや評価の品質向上**

教育現場に即した設計

学内システム連携を含むソリューション提供

- ▶ **教育品質維持と職員の負担軽減を両立**

***Innovating a Sustainable Future
for People and Planet***

NTT版LLM  **tsuzumi**

tsuzumi 2 進化のポイント

NTT株式会社
執行役員 研究企画部門長
木下 真吾

tsuzumi 2 進化のポイント



① 日本語性能のさらなる向上

② 特化型モデル開発効率の向上

③ 低コスト・高セキュアの維持、国産AI

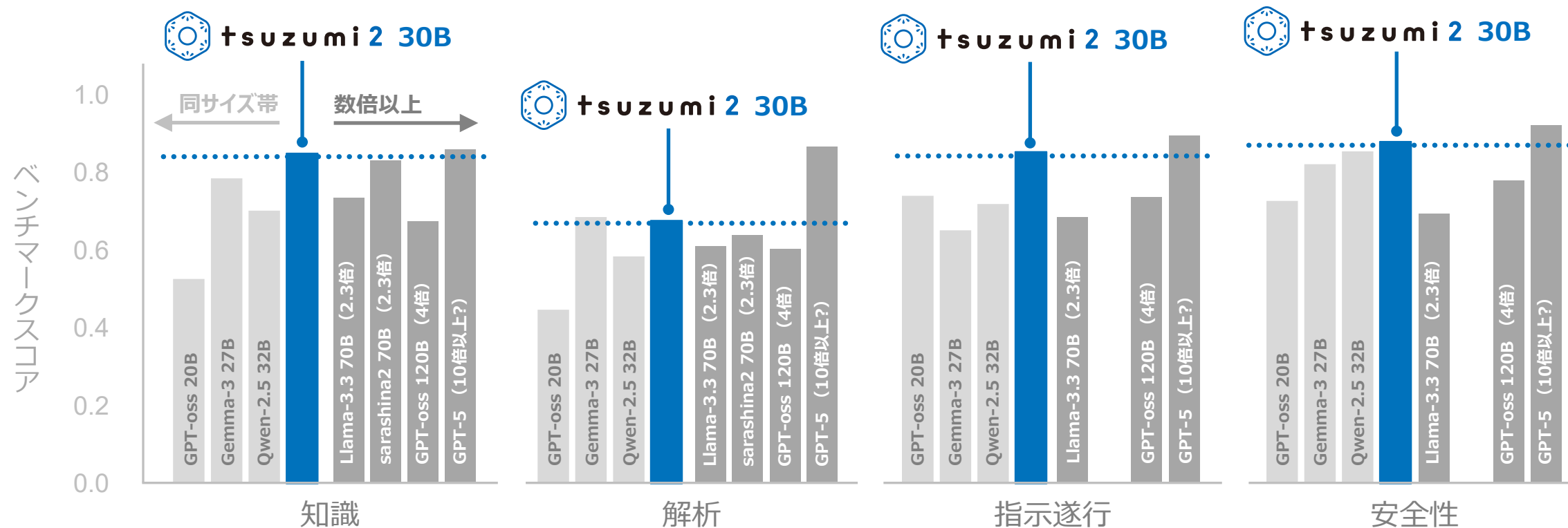
① 日本語性能のさらなる向上

② 特化型モデル開発効率の向上

③ 低コスト・高セキュアの維持、国産AI

特長① 日本語性能のさらなる向上

- 同サイズ帯のモデル（Gemma-3等）では世界トップクラスの日本語性能
- 数倍以上大きなモデル（GPT-oss 120B, GPT-5等）にも引けを取らない性能でコストパフォーマンスに優れる



※評価条件

- 知識: llm-jp-eval (JAQKET (日本語QAタスク), NIILC (質問応答), JCommonsenseQA (常識推論能力)), pfgn-bench (PFN日本知識ベンチ) の平均値
- 解析: llm-jp-eval (Wikipedia読み推定, Wikipedia係り受け解析) の平均値
- 指示遂行: M-IFEval_Ja
- 安全性: AnswerCarefully (x0.2)
- 知識と解析においては、事前学習モデルを使用。ただし、GPT-oss 20B/120B, GPT-5は事前学習モデルが公開されていないため事後学習モデル (Reasoningモデル) を使用
- 指示遂行と安全性において、全て事後学習モデルを使用。sarashina2 70Bは事後学習モデルが公開されていないため対象外

特長① 日本語性能のさらなる向上

契約書案を社内チェックリストを使って確認。条項の不足や問題点を的確に指摘

〇〇〇の共同研究に関する契約書

本契約は、A株式会社(以下「甲」という。)とB株式会社(以下「乙」という。)が、〇〇〇の開発及び実用化に関する研究を共同で行うことを目的として、必要な事項を定めるものである。

■第1条(定義)

本契約において、以下の用語はそれぞれ以下の意味を有するものとする。

- 「本契約」:本契約書に基づき共同研究に関する契約をいう。
- 「甲」:A株式会社をいう。
- 「乙」:B株式会社をいう。
- 「共同研究」:甲及び乙が本契約書に基づき共同で行う〇〇〇に関する研究をいう。
- 「成果物」:共同研究において得られた発明、技術、ノウハウ、データ、論文、報告書その他の成果をいう。
- 「機密情報」:甲又は乙が相手方に関する情報であって、文書、口頭、電子データその他の形態を問わず、かつ、秘密である旨の表示が付されたもの、又は相手方が合理的に秘密であると認識すべき情報をいう。

第2条(共同研究の目的と範囲)

- 目的
甲及び乙は、大規模言語モデルの開発及び実用化に関する研究を共同で行い、社会に貢献することを目的とする。
- 範囲
共同研究の具体的な範囲は、別紙「共同研究計画書」に定めるものとする。

4. 目的の達成方法
甲及び乙は、共同研究の目的を達成するために、以下の方法を採用することができる。↓

- 定期的なミーティングの開催↓
- 共同研究の進捗状況の報告↓
- 必要な設備、資材、ソフトウェア等の提供↓
- 専門家の招聘↓
- その他甲乙が協議の上、決定する方法

第3条(成果と外部発表)

- 成果物の所有
共同研究において得られた成果物は、甲及び乙が共同で所有するものとする。
- 外部発表
 - 成果物の外部発表については、甲乙協議の上、決定するものとする。
 - 甲乙は、成果物の外部発表に際し、事前に相手方に通知し、かつ、相手方の承諾を得るものとする。
 - 外部発表に係る費用は、甲乙が協議の上、決定するものとする。
- 成果物の管理
甲乙は、成果物の管理について、別紙「成果物管理規程」に従うものとする。
- 成果物の評価
甲乙は、共同研究の成果物について、甲乙協議の上、評価を行うものとする。

第4条(本共同研究の遂行)

- 誠実義務
甲及び乙は、本契約書に基づき、誠実に共同研究を遂行するものとする。
- 遂行方法
 - 甲乙は、共同研究の遂行に必要な人員、設備、資金等を適切に提供するものとする。
 - 甲乙は、共同研究の進捗状況を定期的に報告するものとする。
 - 報告の頻度及び形式については、甲乙協議の上、決定するものとする。
- 変更の手続
共同研究の範囲、内容、方法等に変更が生じた場合は、甲乙協議の上、書面により合意するものとする。

特長① 日本語性能のさらなる向上

ニュースリリースの改善提案。重要な改善点（簡潔な表現や全角半角の混在）を指摘



tsuzumi 2 進化のポイント

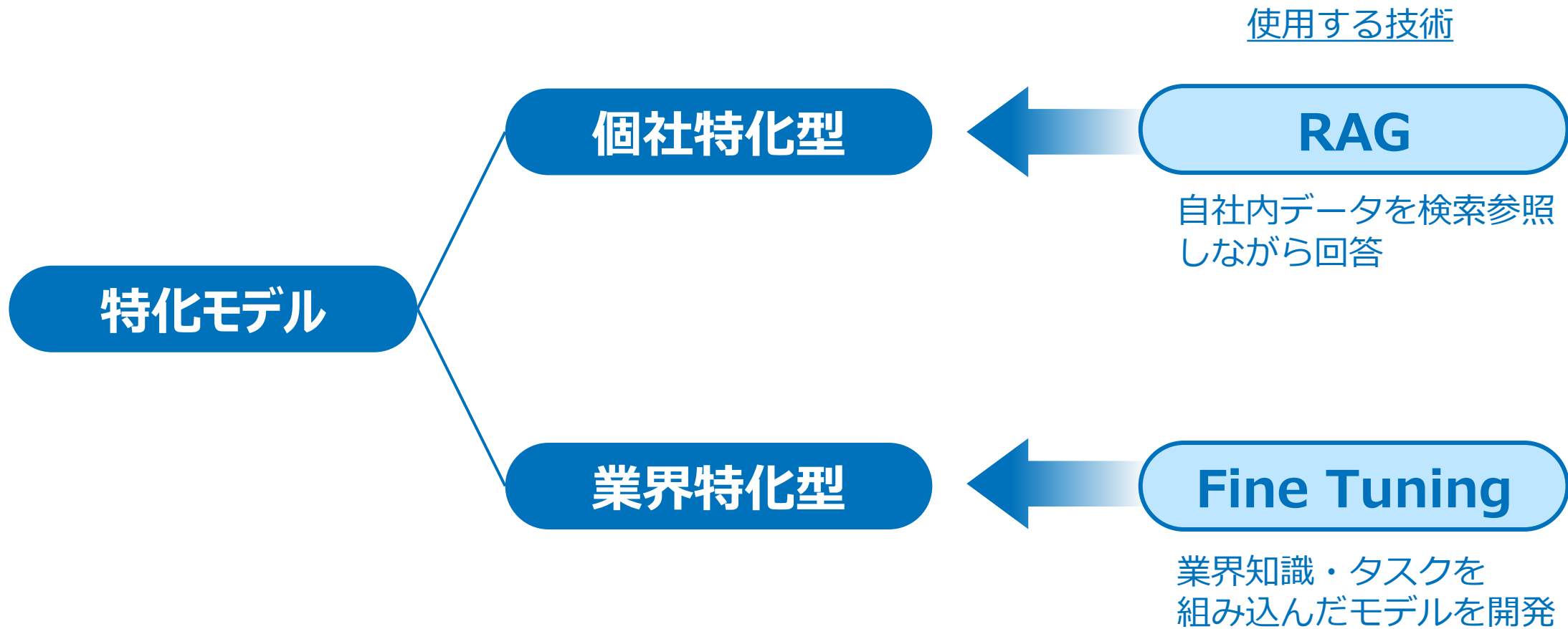


① 日本語性能のさらなる向上

② 特化型モデル開発効率の向上

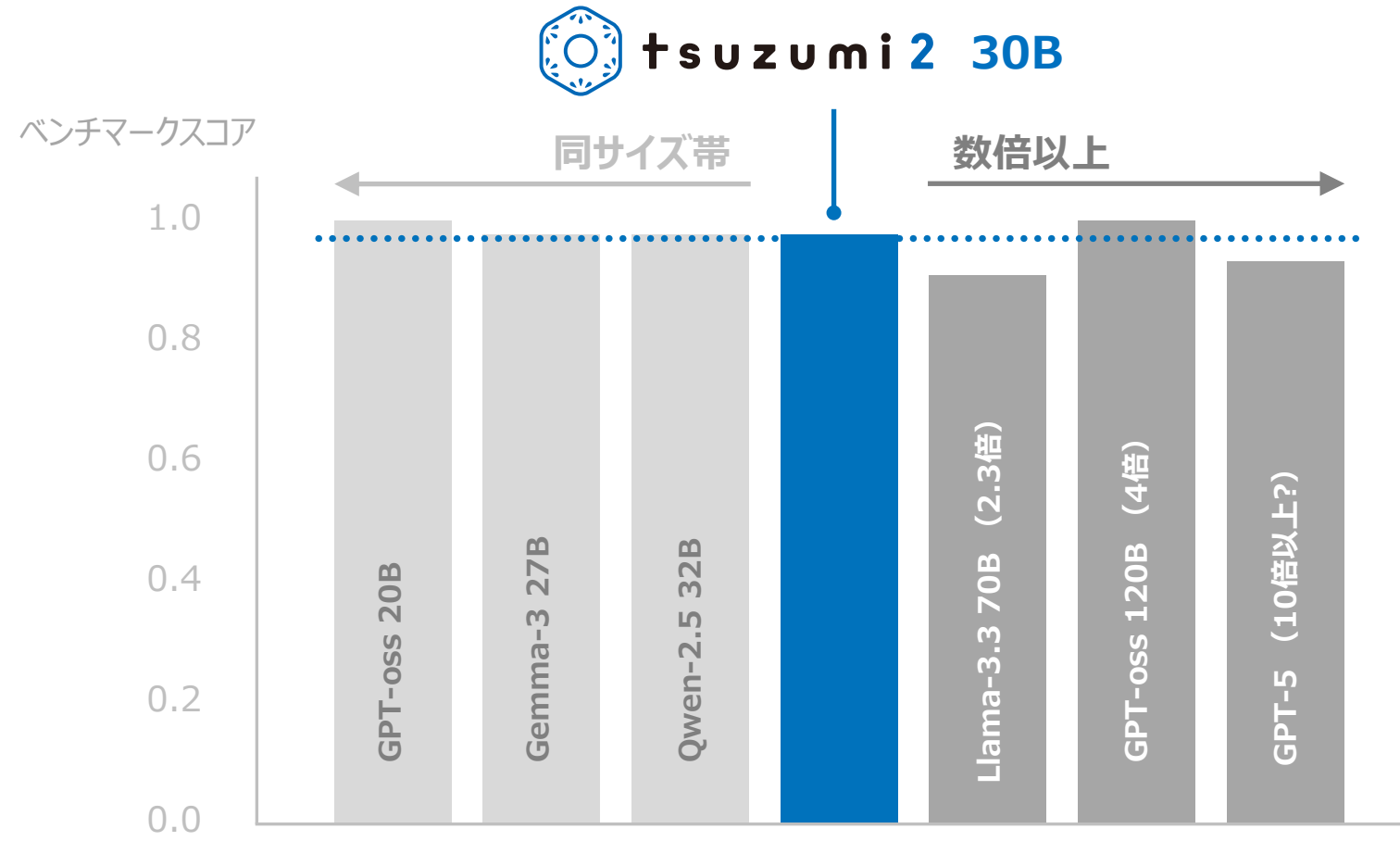
③ 低コスト・高セキュアの維持、国産AI

特長② 特化モデル開発効率の向上



特長② 特化モデル開発効率の向上 (RAG)

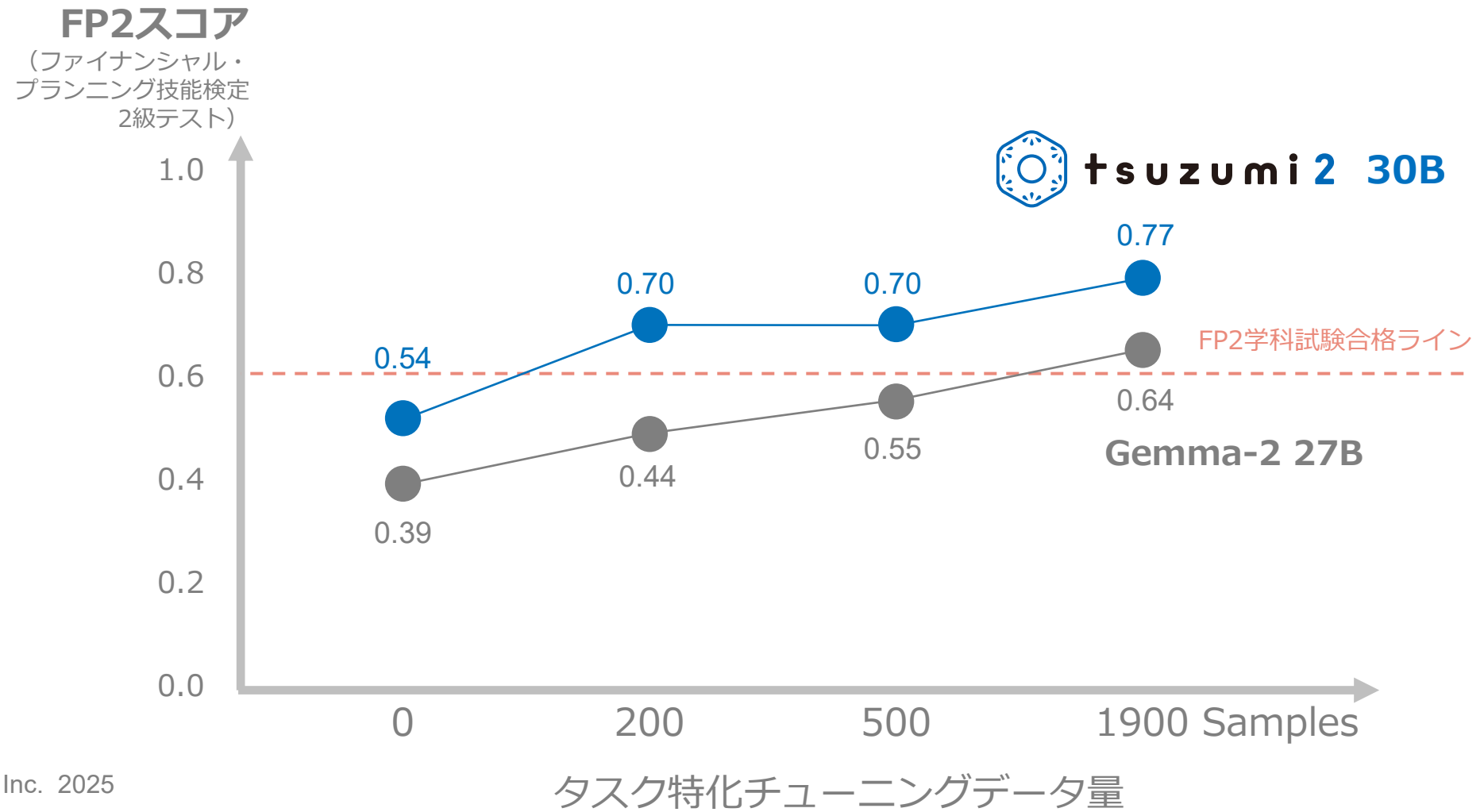
実システムへの適用評価において世界トップクラスのRAG性能を実現 (財務システムに関する社内ヘルプデスク)



※評価条件 NTT社内業務 (財務システムに関する社内ヘルプデスク) における
トライアル案件において、RAGによる問い合わせ回答正答率を独自に評価。

特長② 特化モデル開発効率の向上 (F.T)

- 金融業界の追加知識 + タスク特化チューニングによって、世界トップクラスの金融タスク性能を実現
- 他モデルに比べて少ないチューニングデータで高性能を実現



tsuzumi 2 進化のポイント



① 日本語性能のさらなる向上

② チューニング（個社・業界特化）性能の向上

③ 低コスト・高セキュアの維持、国産AI

特長③ 低コスト・高セキュリティの維持



 **tsuzumi 2**

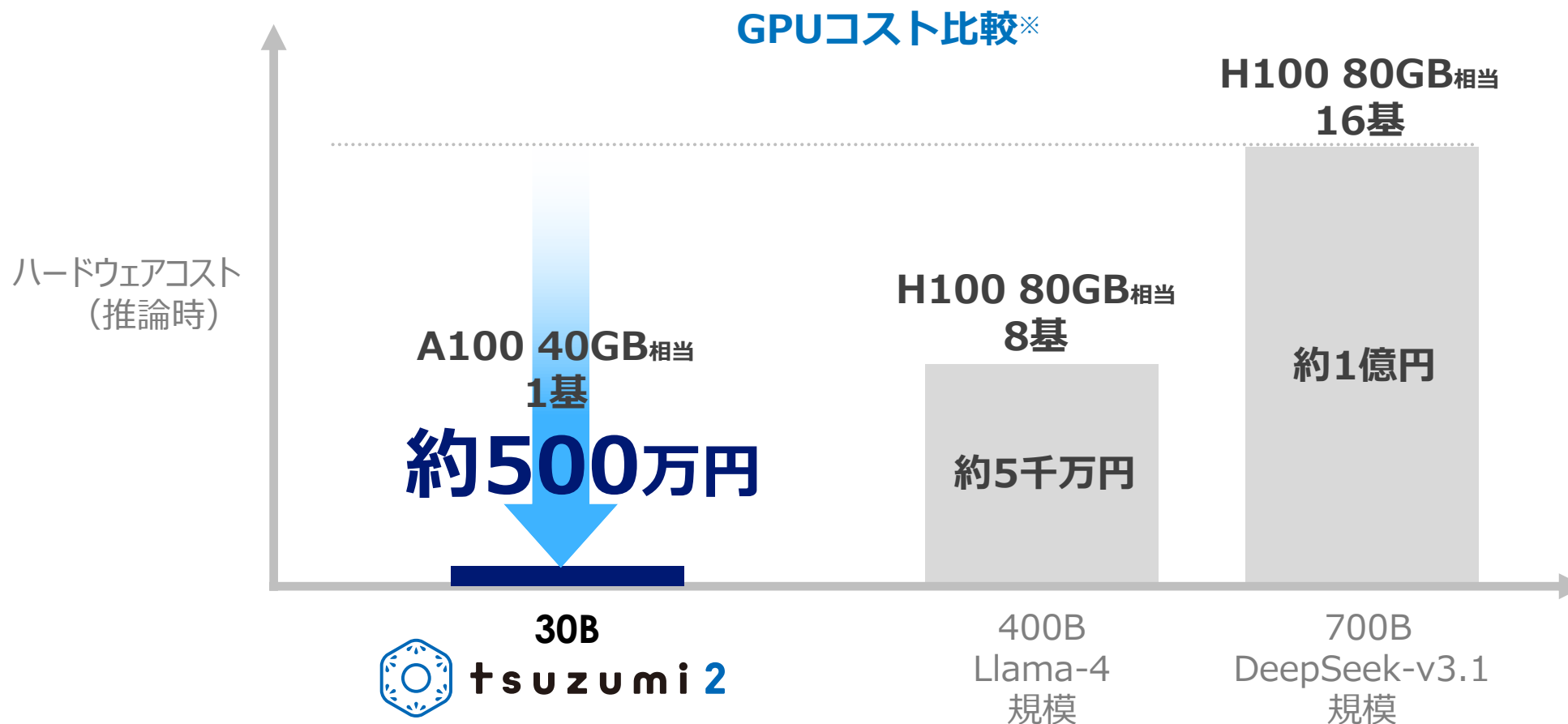
30B

Why 30B? → Only 1 GPU → On-premise

→ 低コスト&高セキュリティ AI

特長③ 低コストの維持

大規模クラスと比べて、推論コストを約10～20分の1に低減可能



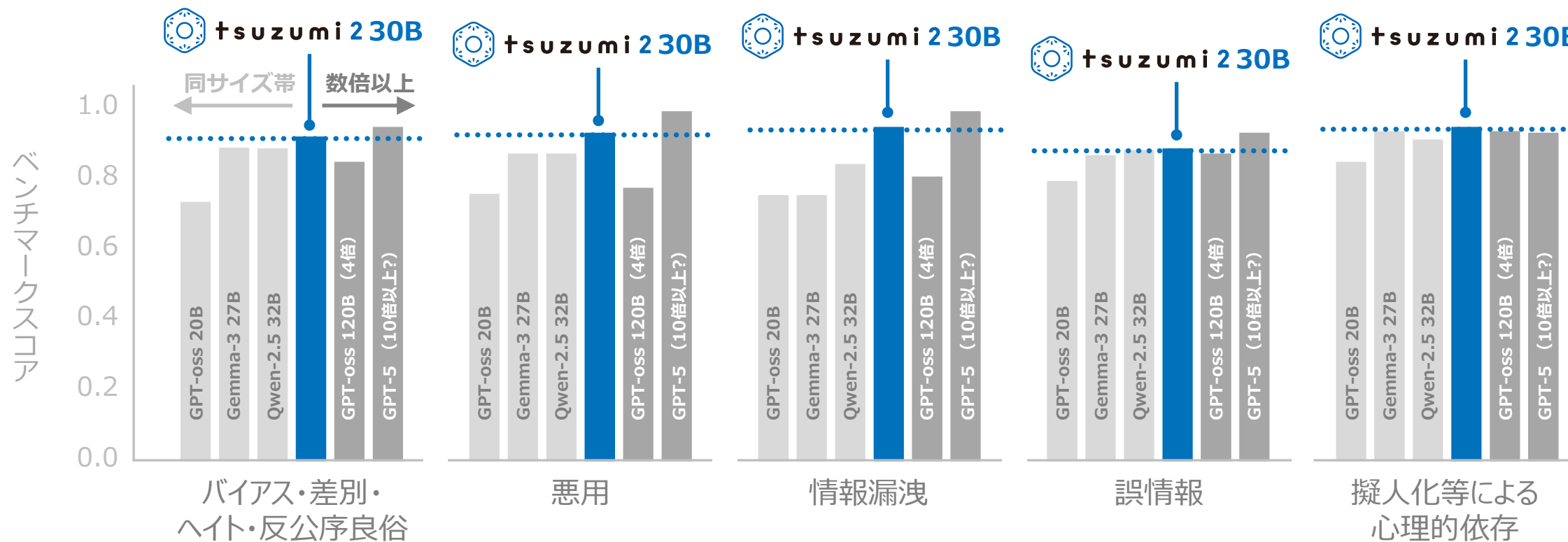
※試算条件

- 量子化: 8ビット
- 必要GPUメモリサイズ: パラメタ数 x 量子化サイズ/8bit (30Bは30GB、400Bは400GB、700Bは700GB)
- Llama-4(Llama 4 Maverick)は、MoE, H100 x 8 (1ノード)動作、DeepSeek-v3.1は、MoE, H100 x 16 (2ノード)動作を前提
- ハードウェアコストは、上位GPU H100 80GB: 1,000万円/台, 下位GPU A100 40GB: 500万円/台として換算、その他の運用などの費用は含まず

特長③ 高セキュリティの維持

モデル自身の安全性も主要モデルと比較して高いスコアをマーク

日本語安全性比較※



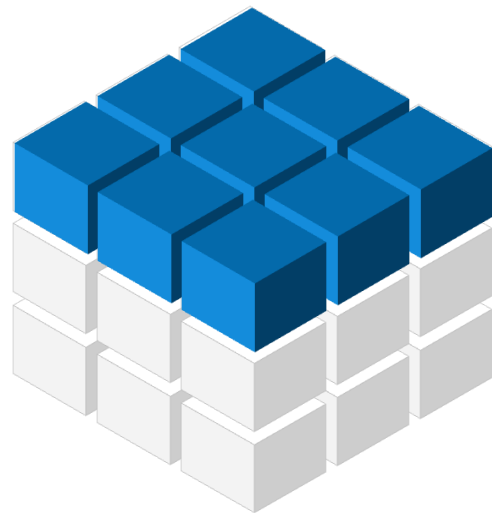
特長③ 国産AI

海外製オープンAIに頼ることなく、スクラッチで開発

LLMの開発アプローチ

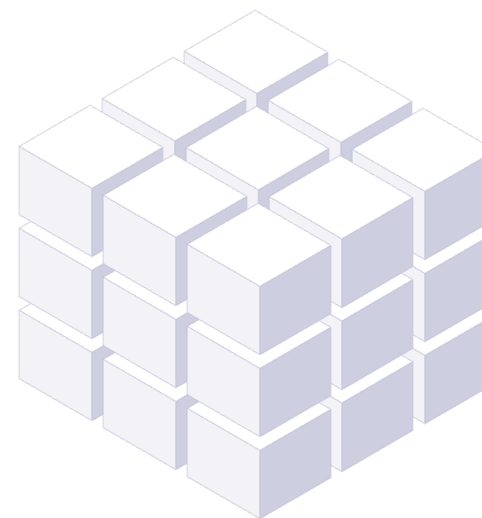
海外製オープンAIをベースに
日本語データで追加学習

- ELYZA, RICOHなどはLlamaベース



基盤モデルを一から作成

- NTT tsuzumi



NTT版LLM  **tsuzumi**

NTTのAI戦略について

NTT株式会社

常務取締役 研究開発マーケティング本部長
大西 佐知子

AI For Quality Growth

1. エンドエンドのフルスタック

～コンサルティング×AIソリューション設計、実装～

2. 豊富なAIラインナップ

～パートナー連携 & 自社研究開発AI～

3. AIをサステナブルに実現するインフラ

～GPU、コネクティビティ、電力～

AI For Quality Growth

1. エンドエンドのフルスタック

～コンサルティング×AIソリューション設計、実装～

2. 豊富なAIラインナップ

～パートナー連携 & 自社研究開発AI～

3. AIをサステナブルに実現するインフラ

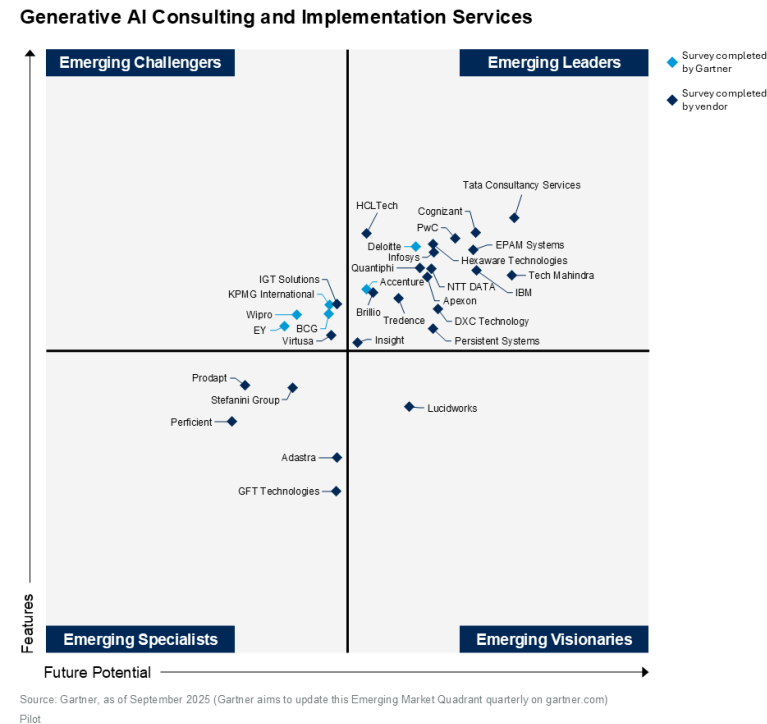
～GPU、コネクティビティ、電力～

AI コンサルティング～実装領域でリーダポジション

新技術活用を含めた世界最高水準の実装力、顧客・業界理解



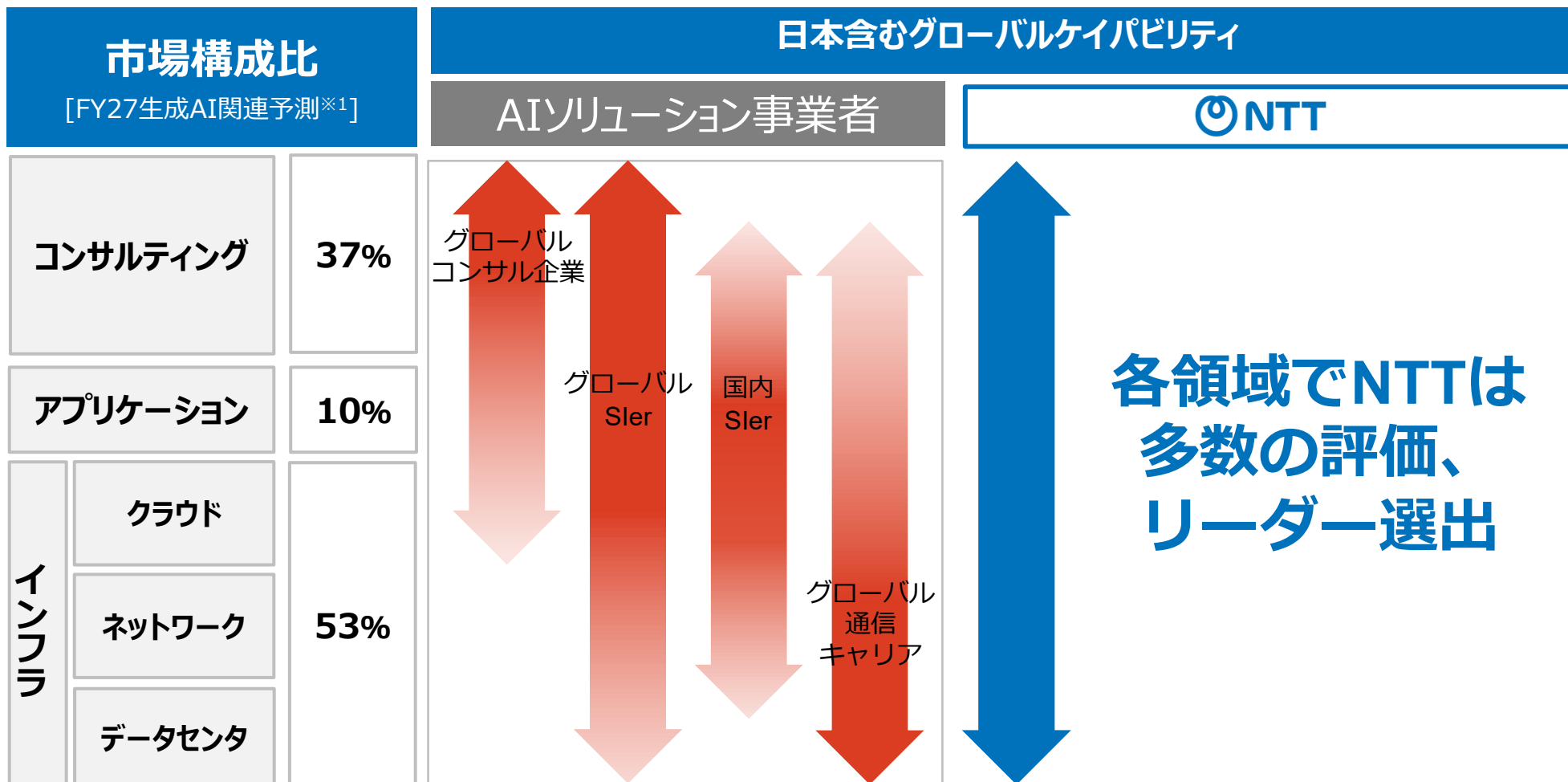
Gartner® Emerging Market Quadrant for Generative AI Consulting and Implementation Services



Gartner

Gartner®, Innovation Guide for Generative AI Consulting and Implementation Services, Luis Pinto et al., 18 September 2025
 GARTNERは、Gartner Inc.または関連会社の米国およびその他の国における登録商標およびサービスマークであり、同社の許可に基づいて使用しています。All rights reserved.
 この図表は、Gartner, Inc.がリサーチの一部として公開したものであり、文書全体のコンテキストにおいて評価されるべきものです。Gartnerは、Gartnerリサーチの発行物に掲載された特定のベンダー、製品またはサービスを推奨するものではありません。また、最高のレーティング又はその他の評価を得たベンダーのみを選択するようにテクノロジーユーザーに助言するものではありません。Gartnerリサーチの発行物は、Gartnerリサーチの見解を表したものであり、事実を表現したものではありません。Gartnerは、明示または黙示を問わず、本リサーチの商品性や特定目的への適合性を含め、一切の責任を負うものではありません。

AI導入環境をフルスタックでワンストップ提供





コンサルティング

Emerging Leaders
(東アジア唯一)

2025 Gartner® Emerging Market Quadrant for Generative AI **Consulting and Implementation Services**※1

アプリケーション

A Leader
(世界7,500社以上導入)

IDC MarketScape: Asia/Pacific **SAP Implementation Services 2024**
(Doc #AP51467624, August 2024)

インフラ

クラウド

A Leader
(東アジア唯一)

IDC MarketScape: Worldwide **Cloud Professional Services 2024**
Vendor Assessment (Doc #US51406224, June 2024)

ネットワーク

Leaders
(12年連続)

2025 Gartner® Magic Quadrant™ for Global **WAN Services**※2

Winner
(日本国内に本社を置く企業として初めて)

2025 Gartner® Eye on Innovation Awards **CSP(Communication Service Provider)分野**※3

Named
(日本国内に本社を置く企業で唯一)

2025 Gartner「Competitive Landscape: NaaS Communications Service Providers」**Network as a Service分野**※4

データセンタ

A Leader
(世界シェア第3位)

IDC MarketScape: Worldwide **Datacenter Colocation Services 2025**
Vendor Assessment (Doc #US52983725, August 2025)

※1 Gartner® Innovation Guide for Generative AI Consulting and Implementation Services, Luis Pinto et al., 18 September 2025 (NTT Dataとして東アジアに本社を置く企業で唯一位置づけ)

※2 Gartner® Magic Quadrant™ for Global WAN Services, Danellie Young et al., 24 March 2025 (NTTとして位置づけ)

※3 Gartner®, 2025 Gartner Eye on Innovation Awards for Communications Service Providers <https://www.gartner.com/en/about/awards/eye-on-innovation/communications> (NTT DOCOMO Businessとして、APAC地域のWinnerに選出)

※4 Gartner®, Competitive Landscape: NaaS Communications Service Providers, Dennis Stevens et al., 11 July 2025 (NTT Docomo Businessとして)

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and MAGIC QUADRANT is a registered trademark of Gartner, Inc. and/or its affiliates and are used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

The Gartner content described herein (the "Gartner Content") represents research opinion or viewpoints published, as part of a syndicated subscription service, by Gartner, Inc. ("Gartner"), and is not a representation of fact. Gartner Content speaks as of its original publication date (and not as of the date of this Investor Presentation), and the opinions expressed in the Gartner Content are subject to change without notice.

既存の業務プロセスからAI活用の17タスクを抽出 既存のシステム更改+AI環境（GPU等）構築まで

導入事例：医療機関様

AI活用により効率化可能なタスク抽出



- ★カルテの作成
 - ・退院サマリの作成
 - ・医療保険用診断サマリ
- ・入院計画書作成
- ・当直シフト作成
- ・看護師のシフト作成

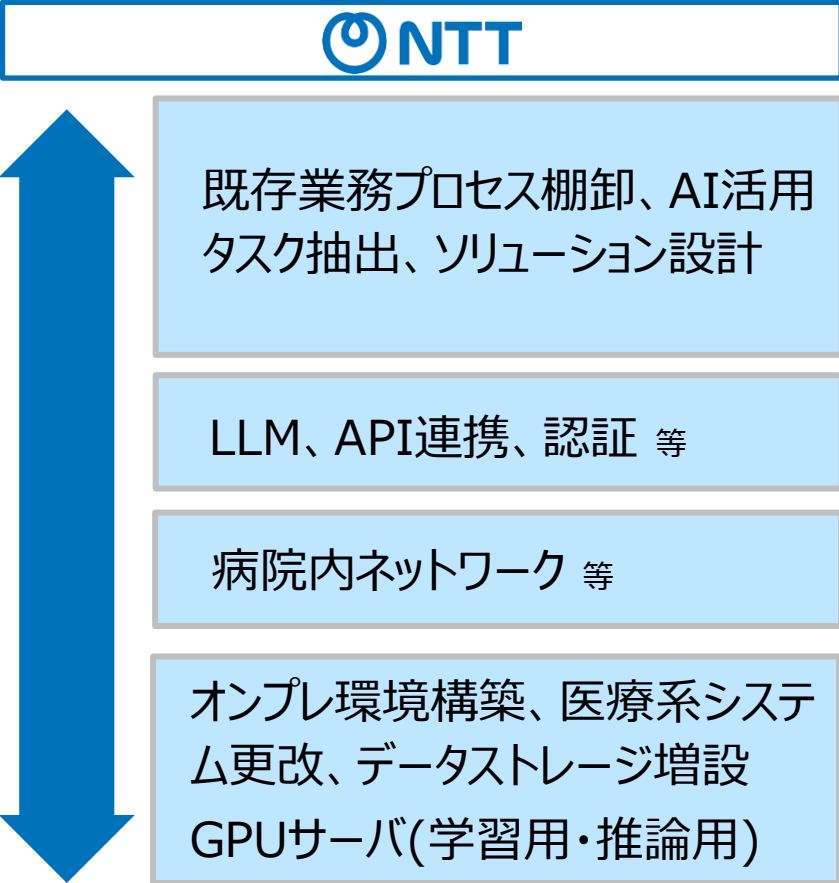


★例えば

想定効果 **6,000時間/年**※

医師の電子カルテ入力時間削減

※全職員が活用できる精度が実現した場合



AI For Quality Growth

1. エンドエンドのフルスタック

～コンサルティング×AIソリューション設計、実装～

2. 豊富なAIラインナップ

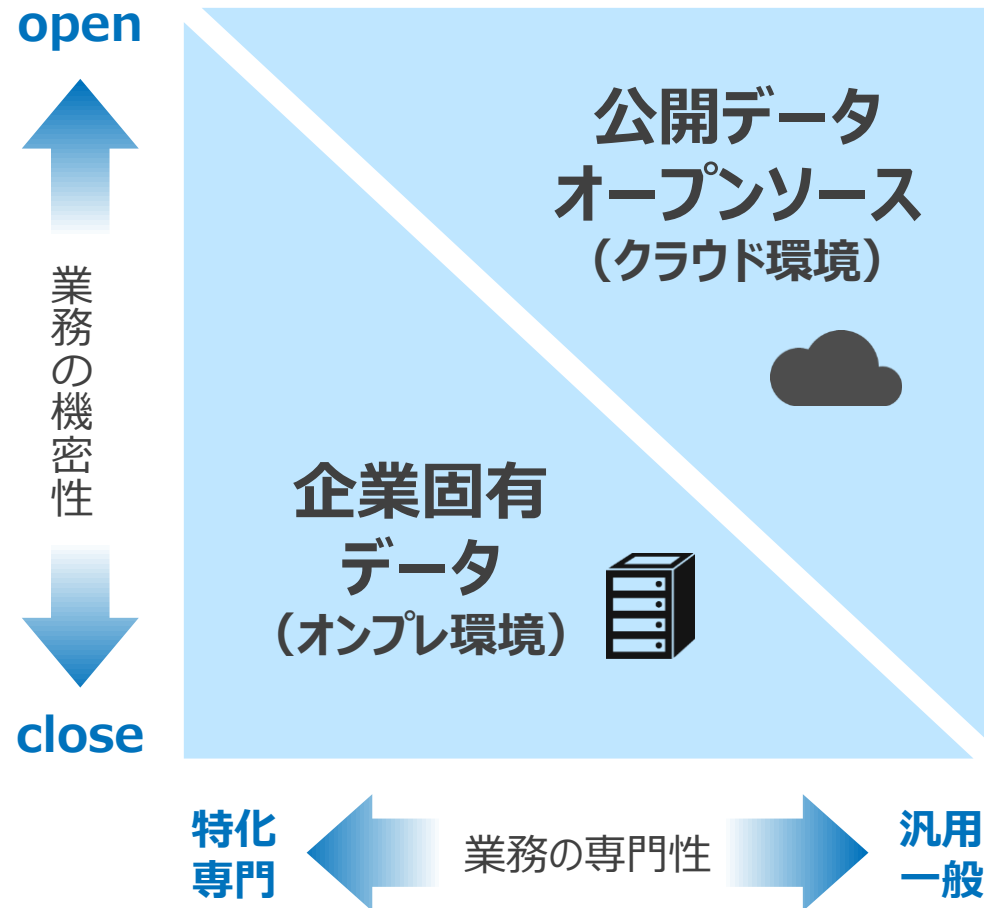
～パートナー連携 & 自社研究開発AI～

3. AIをサステナブルに実現するインフラ

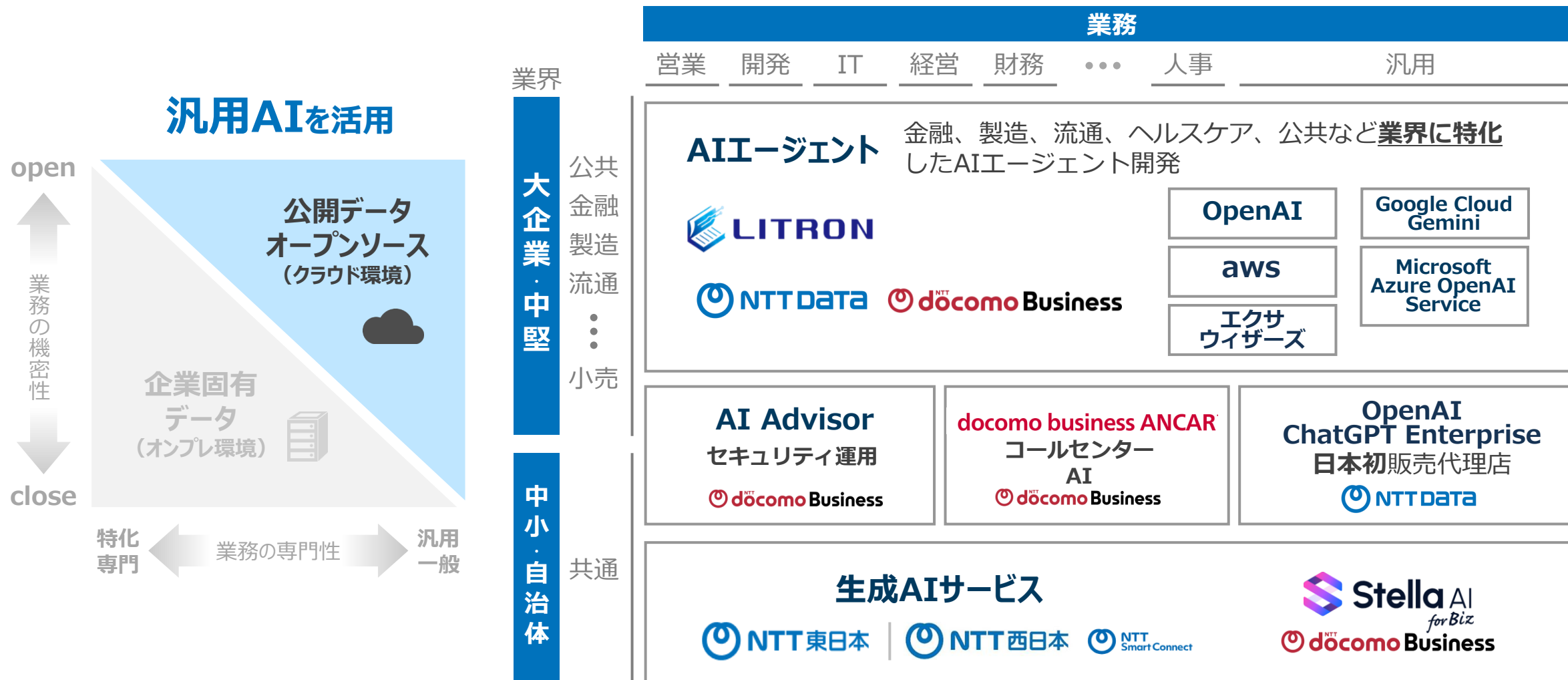
～GPU、コネクティビティ、電力～

多様なお客様ニーズに応じた最適なAIモデル、AIEージェント

企業における業務・環境ごとのデータ

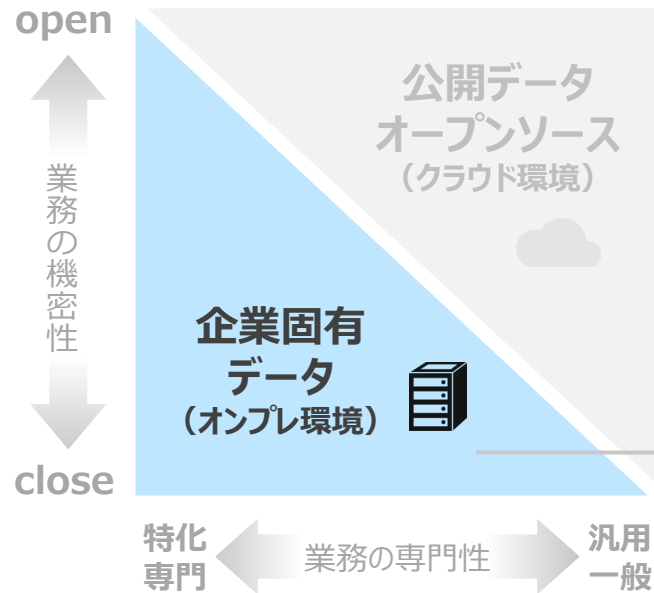


多様なお客様ニーズに応じた最適なAIモデル、AIEージェント

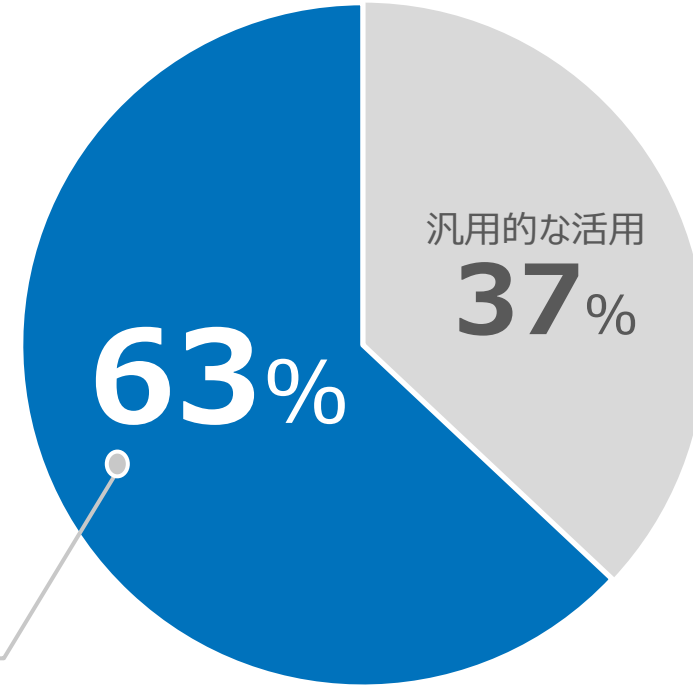


AI活用の領域は、Open、汎用な領域から、企業固有のノウハウ、インテリジェンス、Close領域へ

企業における業務・環境ごとのデータ

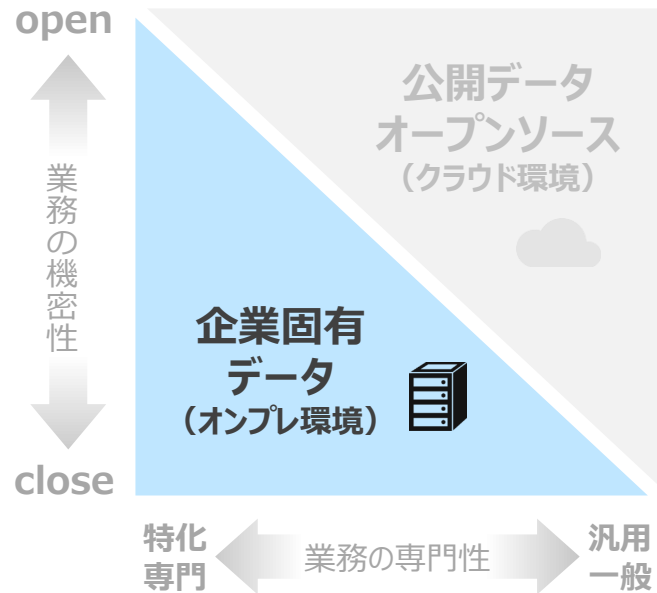


個人情報や機密性が高いデータをお客さま環境でクローズドにセキュアに学習させたい



tsuzumiへのお問合せにおける活用シーン (2024年3月時点)

機密データ、クローズドニーズにも対応



日本国内は

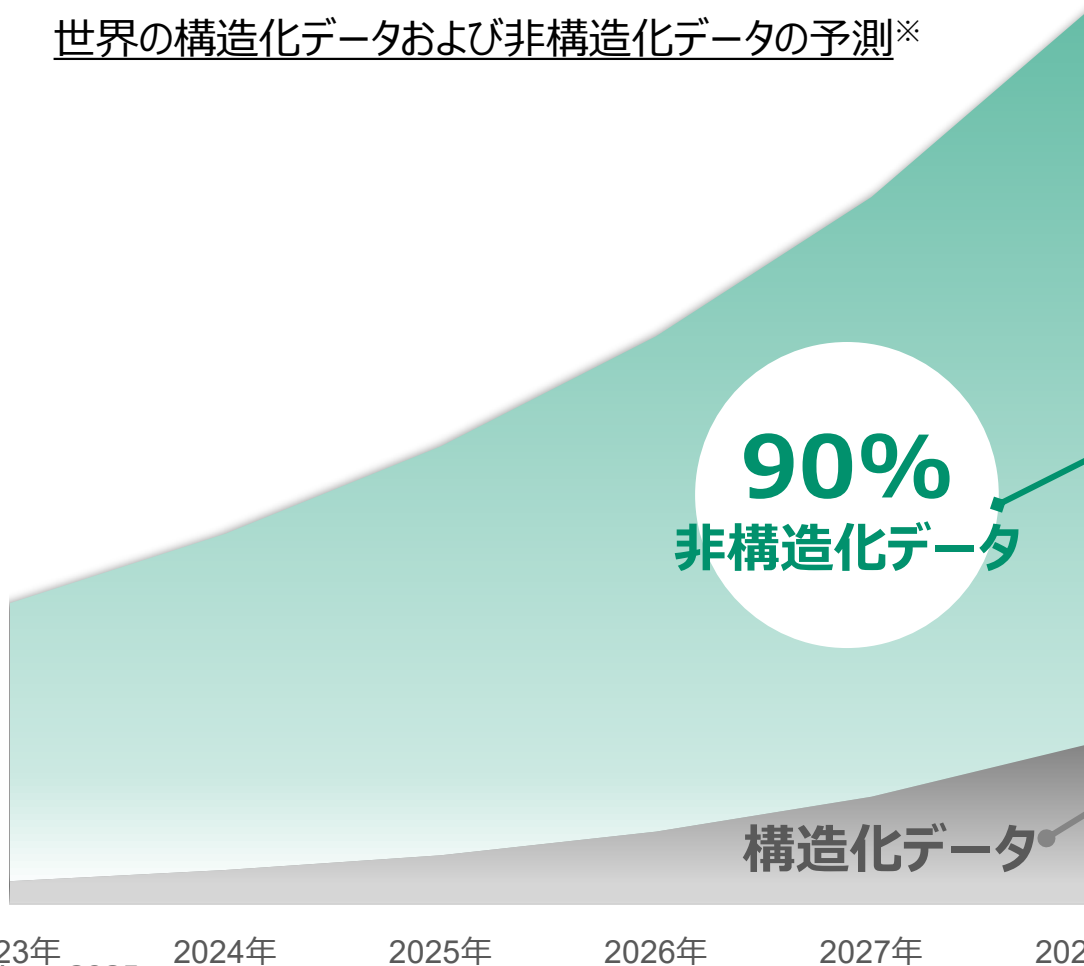
純国産モデル **tsuzumi** で対応



tsuzumi

企業活動を通じて蓄積される固有の情報の90%は非構造化データ

世界の構造化データおよび非構造化データの予測※



90%は非構造化データ

- ✓ メール文
- ✓ 提案書
- ✓ 手書きアンケート
- ✓ 契約書
- ✓ 請求書/納品書 ...etc.

構造化データは僅か10%

- ✓ 業務システム
- ✓ 基幹システム
- ✓ システムログ ...etc.

高機密な文書や画像などの非構造化データを構造化し、価値ある情報へ

FUJIFILM
富士フイルム ビジネス イノベーション株式会社



REiLI
知の構造化

NTT

docomo Business



tsuzumi
純国産モデル

高機密ニーズが高いお客様

官公庁

金融

医療（病院）

製造／物流

⋮

グローバルでAIリスクへの対応、AIガバナンスを整備・運用 安全性ガードレール機能サービス“chakoshi”

各国の動きにあわせた ガバナンス活動

グローバルでAIリスクを共通的に整備



G7/OECD

✓ 広島AIプロセス
総務省
国際戦略室
と対応。



日本政府

✓ 内閣府
総務省等、
様々な組織
からの要請に
応じ対応中



米国政府

✓ CSIS※を
中心に
連携中



EU政府

✓ EUの
AI-ACTの
ガイドライン
のWGに参加中

安全性ガードレールAPI “chakoshi”

自社開発

- 悪用を図る問合せへの回答を遮断
- 機微情報抽出するような悪意あるプロンプトをブロック

AI For Quality Growth

1. エンドエンドのフルスタック

～コンサルティング×AIソリューション設計、実装～

2. 豊富なAIラインナップ

～パートナー連携 & 自社研究開発AI～

3. AIをサステナブルに実現するインフラ

～GPU、コネクティビティ、電力～

AI活用に必要なインフラ

AIを動作させる

コンピューティング
リソース

+

拠点と
データと
AIをつなぐ

コネクティビティ

+

AIを駆動させる

安定的な
電力供給

AIを動作させるコンピューティングリソース

大規模クラスタから1GPUまで、AI利用状況に応じ
必要なコンピューティングリソースをご提供

AI向け基盤サービス
GPU as a Service



日本でトップクラスの計算量

NVIDIA Blackwellアーキテクチャーを採用した
「NVIDIA DGX™ B200システム」をクラスター構成

大規模クラスタの構築・運用実績

「NVIDIA Partner Network」パートナーに贈られる

国内最高のアワード

『Best NPN of the Year』を受賞

拠点とデータとAIをつなぐコネクティビティ

必要な分だけ利用できるNaaS (Network as a Service)

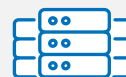
インターネット



クラウド



データセンター



docomo business RINK®

NaaS (Network as a Service)

必要な分を
分単位、10M単位で利用

日本初!
ネットワークがセンサー
脅威を検知・遮断



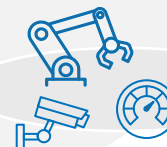
オフィスでも



店舗でも



自宅・外出先でも



エッジ

世界で評価される **docomo business RINK®**

米Gartner®社が

クラス最高の技術革新を推進する世界中のCSP※を表彰する
「Eye on Innovation Awards」のAPAC地域において

NaaS (Network as a Service)

NTTドコモビジネスが

日本に本社を置く企業として

初めて“Winner”に選出

(2025年10月発表)



<出典・免責>

Gartner®, 2025 Gartner Eye on Innovation Awards for Communications Service Providers <https://www.gartner.com/en/about/awards/eye-on-innovation/communications>

Gartnerは、Gartnerリサーチの発行物に掲載された特定のベンダー、製品またはサービスを推奨するものではありません。また、最高のレーティング又はその他の評価を得たベンダーのみを選択するようにテクノロジーユーザーに助言するものではありません。Gartnerリサーチの発行物は、Gartnerリサーチの見解を表したものであり、事実を表現したものではありません。Gartnerは、明示または黙示を問わず、本リサーチの商品性や特定目的への適合性を含め、一切の責任を負うものではありません。

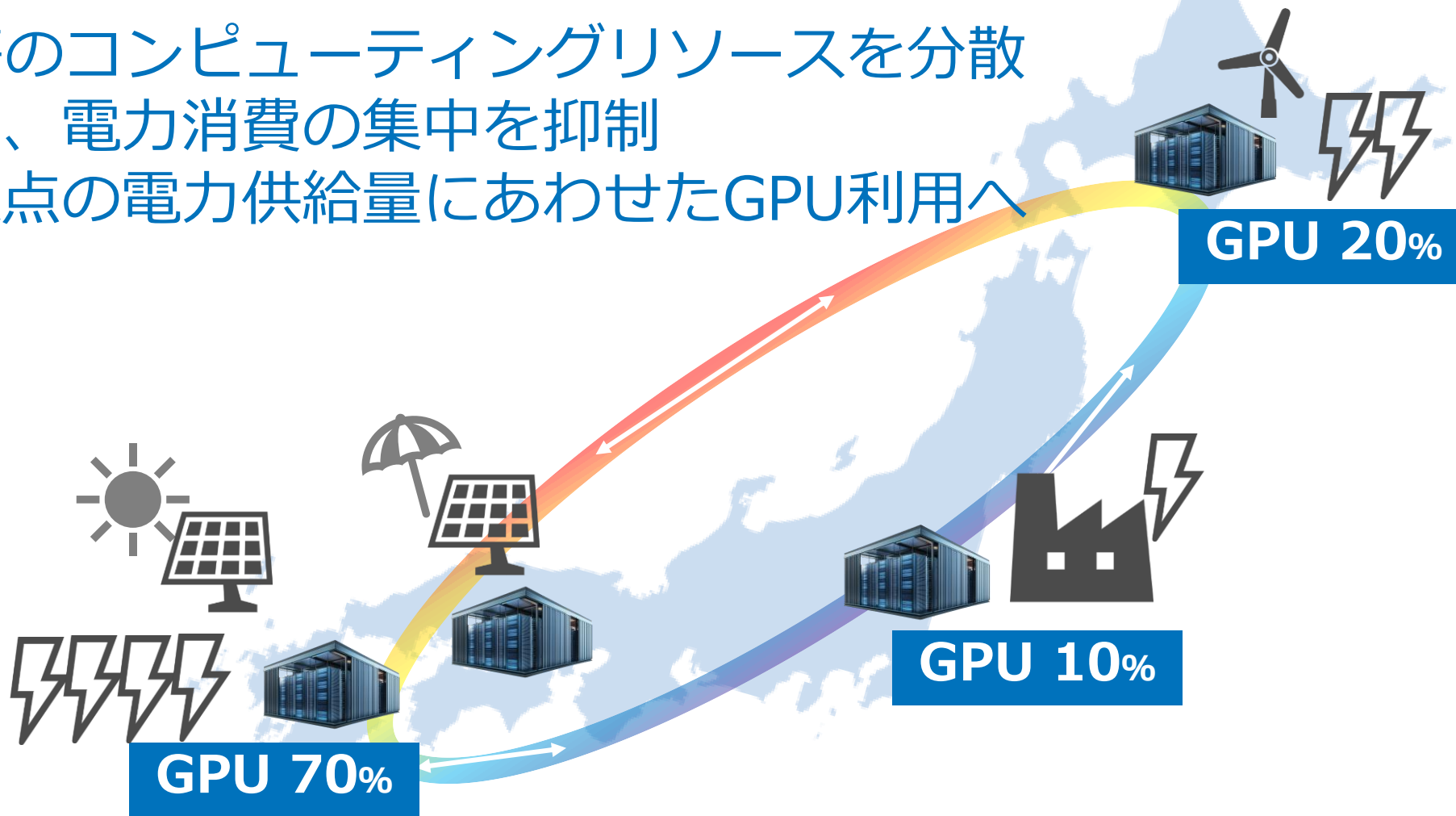
GARTNERは、Gartner Inc.または関連会社の米国およびその他の国における登録商標およびサービスマークであり、同社の許可に基づいて使用しています。All rights reserved.

※ Communication Service Provider

AIを駆動する安定的な電力供給

電力消費の分散とリソース効率利用の両立

GPU等のコンピューティングリソースを分散
設置し、電力消費の集中を抑制
設置拠点の電力供給量にあわせたGPU利用へ





AI For Quality Growth を 企業からインダストリーへ

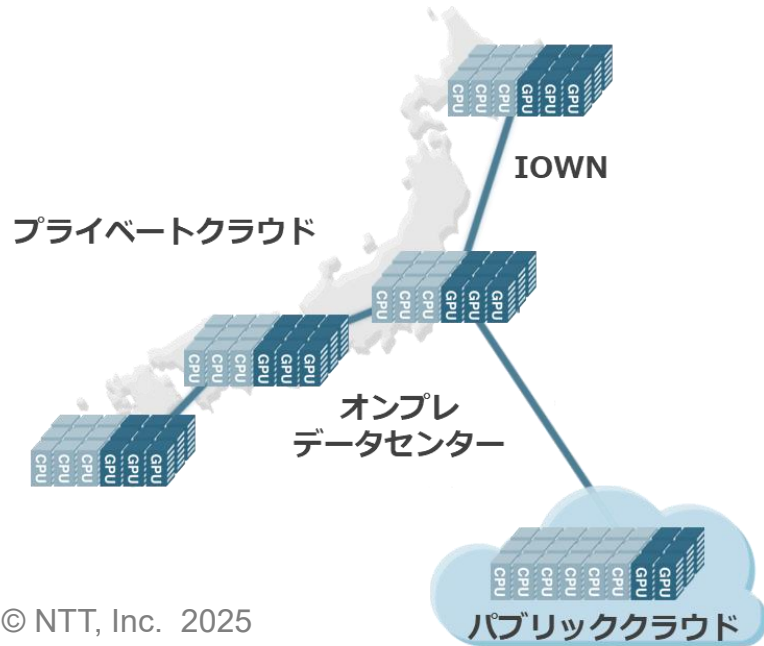
交通事故ゼロ社会の実現にむけ「モビリティAI基盤」を TOYOTA社と共同で開発・運用



モビリティAI 基盤

①分散型計算基盤

データセンターに
分散型計算基盤を構築



②インフラ協調AIモデル

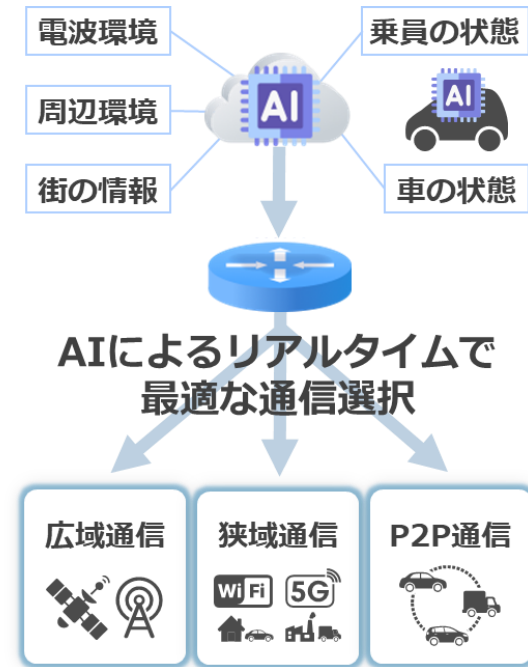
業界特有のデータを
学習したモビリティAI

LMM (Large Mobility Data Model)



③インテリジェント通信基盤

インテリジェントで
高信頼な通信の手段



連鎖型AIにより製造・流通業界のサプライチェーン全体を最適化

TRIAL HLDGS

Retail Ai

× NTT
 NTT AI-CIX

流通業における連鎖型AI活用

従来の自動発注



在庫が減少したタイミングで自動発注

発注作業時間の削減

連鎖型AI導入後



棚上がスッキリ

需要予測に基づき発注量とタイミングを最適化

欠品せず、納品回数・在庫を減らして業務を平準化

流通サプライチェーン



生産・出荷計画の合理化
製造現場の作業平準化
返品率・廃棄率の低減

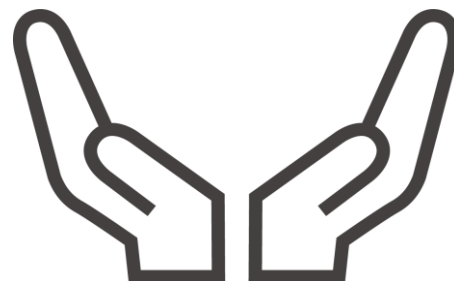
発注・棚割最適化
店舗作業コスト低減
在庫圧縮・廃棄削減

Retail-CIX ※の

連鎖型AI により、川上工程へ効果が波及

お客様の様々なご要望に
担当させていただいている各社がお応えいたします。

お客さま



AI



GPU as a Service

NaaS



IOWN



データセンター

The image features a central title 'AI For Quality Growth' in a bold, blue, sans-serif font with a slight drop shadow. The text is set against a white background that is framed by a decorative border. The top and bottom borders consist of a complex, multi-layered geometric pattern of overlapping squares and lines in shades of blue and green, creating a sense of depth and digital connectivity. On the right side, a large, abstract graphic of blue dots of varying sizes is arranged in a pattern that suggests a network or data flow, extending from the top right towards the center.

AI For Quality Growth