2026年
国産AIモデルの
到達点と未来展望

主権型AIの確立から「エージェント実装」の時代へ

Noto Serif JP
Based on Domestic LLM Current Status and Future Prospects 2026 Comprehensive Report

2026年：キャッチアップから「差別化と実装」のフェーズへ

2023-2024	2025	2026
キャッチアップ期 (Catch-up)	差別化期 (Differentiation)	実装フェーズ (Implementation)

1. Sovereign AI（データ主権）

地政学的リスクと**経済安全保障**の観点から、金融・医療・行政などの**コアデータを国内で完結させる**必要性が増大。米国テック巨人への全面依存からの脱却。

2. Strategic Bifurcation（戦略的分岐）

「**規模追求**（Scale/MoE)」のSoftBank・楽天と、「**効率特化**(Efficiency/Edge)」のNTT・NECに戦略が**二極化**。各社が独自の勝ち筋を明確化。

3. The Agent Shift（エージェント化）

単なるチャットボットから、**自律的にタスク**を遂行する「**エージェント**」へ。2026年の競争軸は「**対話**」から「**行動**」へとシフト。

なぜ今、「国産」なのか：データ主権と日本的商習慣

Security & Sovereignty

- **コアデータの保護**（金融、医療、住民情報）
- **カントリーリスクの低減**（米国政策や規制変更への対応）
- 学習と推論のサイクルを**国内で完結**

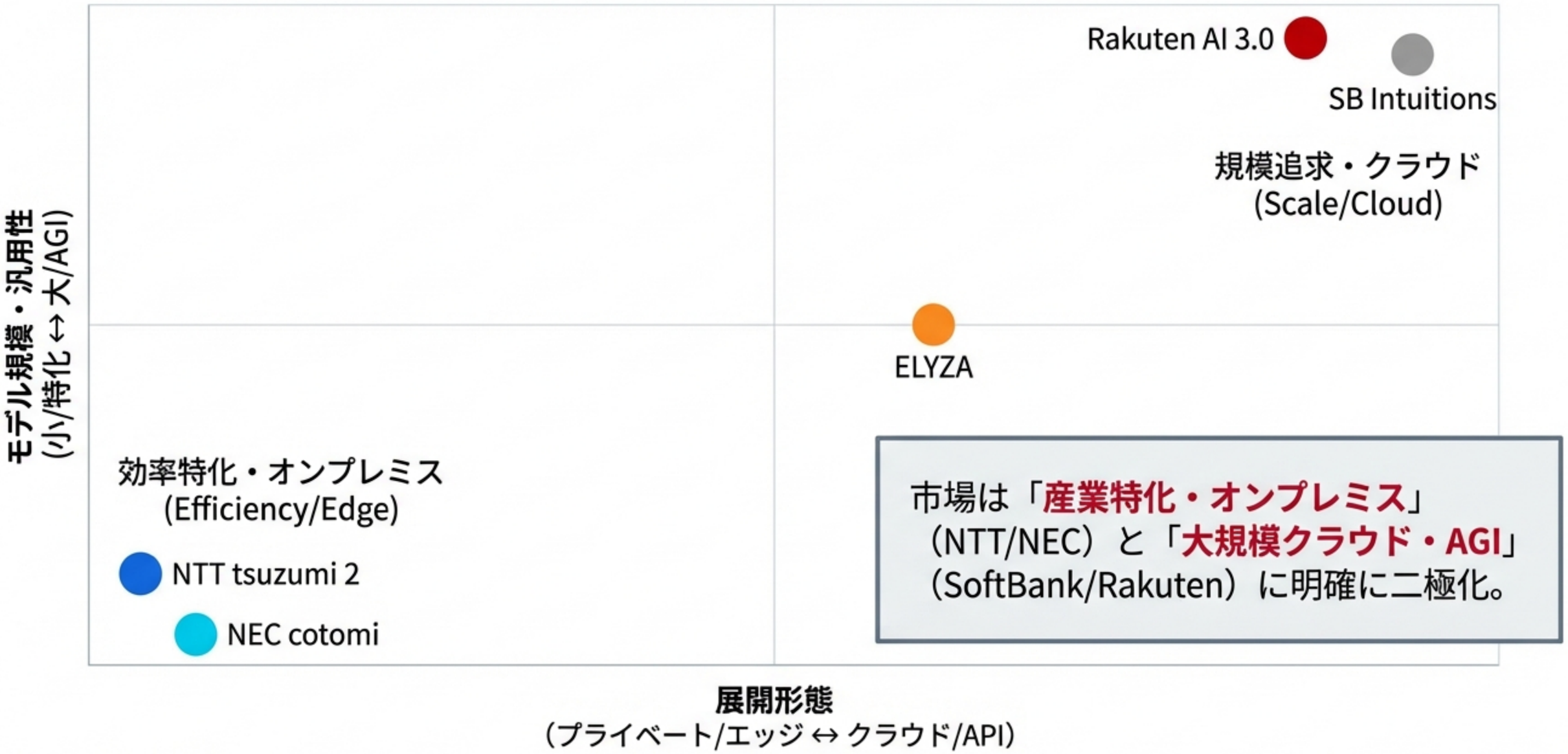
Cultural Fit

- 翻訳を超えた**文化的適応**（根回し、稟議、複雑な敬語）
- **日本法令**への準拠（著作権法、景表法など）
- **商習慣**に即した出力形式

Cost & Energy

- ドル建てAPI**コストの変動リスク回避**（円安対策）
- **省電力・高効率**（資源制約に適した軽量モデル）
- **オンプレミス**環境での運用コスト最適化

2026年 国産LLM 戦略ポジショニングマップ

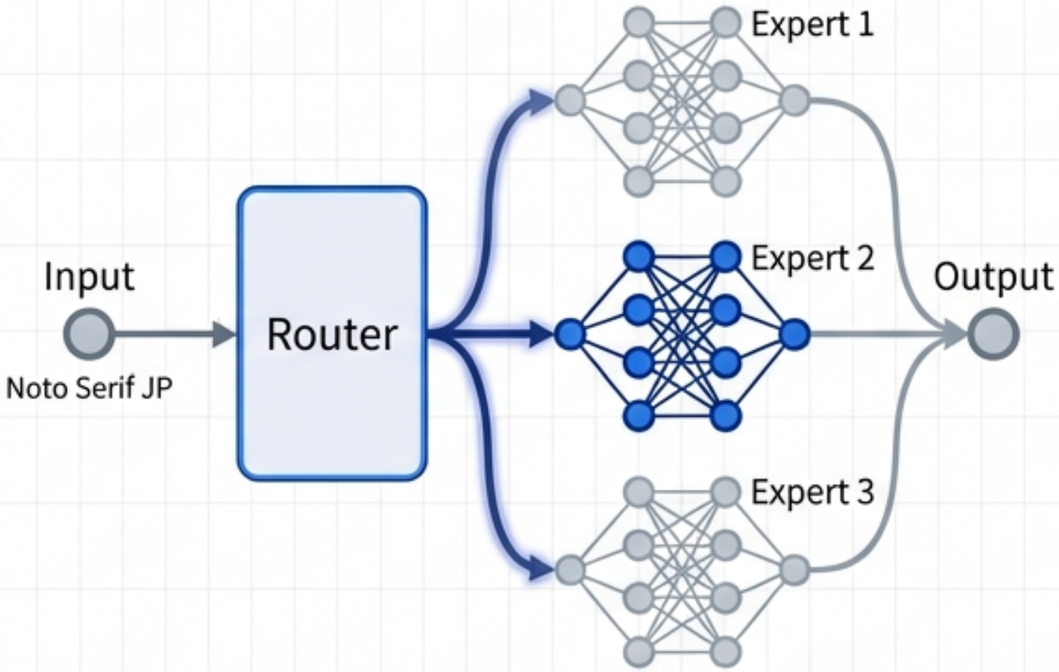


市場は「**産業特化・オンプレミス**」（NTT/NEC）と「**大規模クラウド・AGI**」（SoftBank/Rakuten）に明確に二極化。

技術アーキテクチャの分岐：規模追求 vs 効率特化

規模追求型 (MoE)

SoftBank / Rakuten



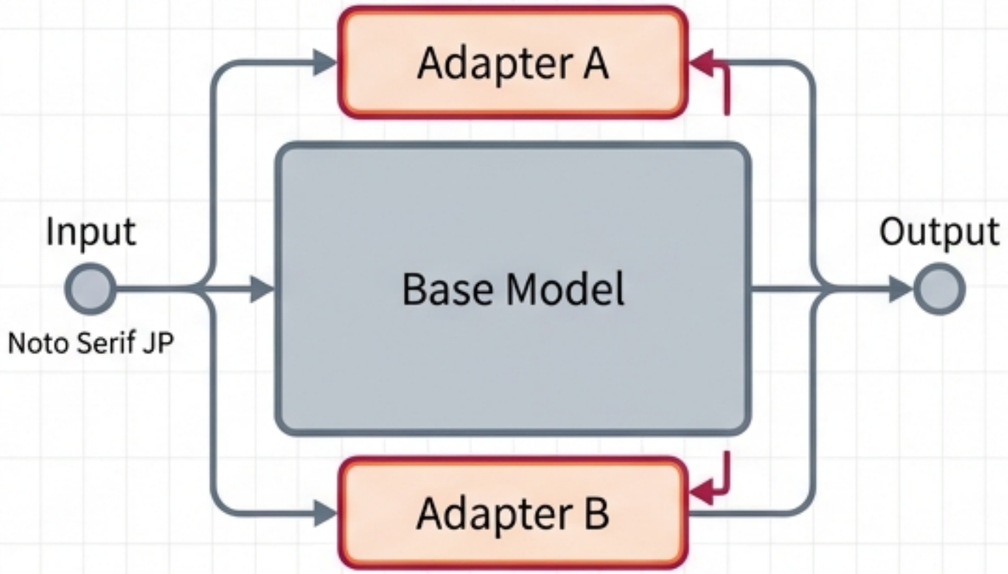
Noto Serif JP　Roboto Mono

ルーターが入力を適切な専門家に振り分け。巨大な知識と推論効率を両立。

推論コスト: 高

効率特化型 (Adapter/Edge)

NTT / NEC



Noto Serif JP　Roboto Mono

軽量なベースモデルに特化型アダプタ（カセット）を追加。低コストで専門業務に対応。

推論コスト: 低

NTT "tsuzumi 2"：IOWN構想と連動する「軽量・高性能」の極致

Theme: Lightweight & Secure

Key Data: 7B / 0.6B Parameters
(Runs on 1 GPU)

Use Case:
東京通信大学 (Secure Education Data)
Finance (FP Grade 2)

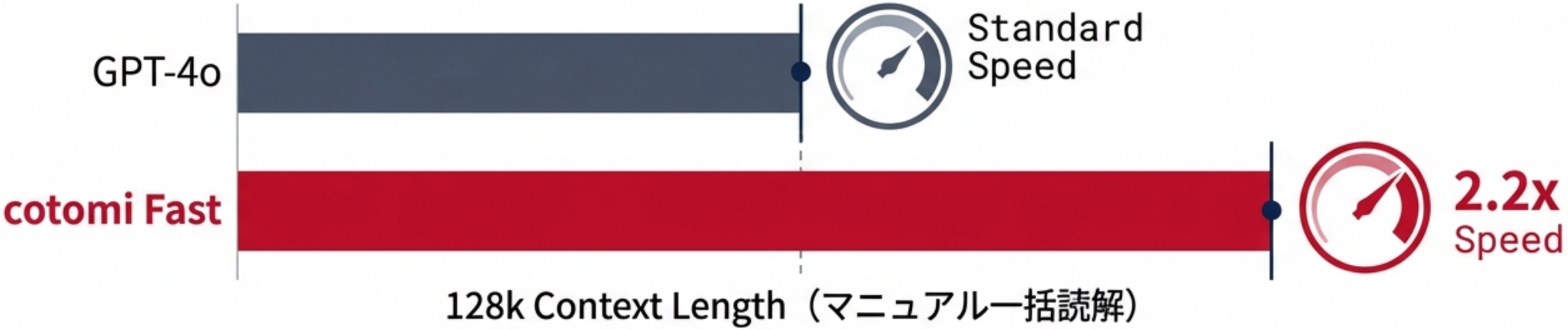
Feature:
Multimodal (Charts/Documents)

Vision: AI Constellation on IOWN

個々は軽量でも、光ネットワークで連携し巨大な知能として機能。

NEC "cotomi": スピードスターと「組織知」の自動化

Speed：圧倒的な推論速度 (cotomi Fast)



Act：From Chat to Act (2026.01〜)

組織内の暗黙知（トップセールスのメールなど）を自動抽出し、全社で共有可能な資産へ変換。

Reading Deck

ELYZA（KDDIグループ）：世界最高性能の「日本語化」と実用主義

High Performance

GPT-4 class score in Japanese benchmarks.
（日本語ベンチマークでGPT-4クラスのスコア）

Post-Training
(Japanese Injection)

ポストトレーニング（日本語注入）

Global Open Model
(Llama 3)

Noto Serif JP （#4A5568）
Technical data: Roboto Mono

Llama-3-ELYZA-JP

Noto Serif JP （#FF7E00）
Technical data: Roboto Mono

WAKONX Platform

Integration with KDDI’s business platform
(Contact Centers/Altius Link).
（KDDIのビジネス基盤（コンタクトセンター/Altius Link）との連携）

Swing-by IPO

Utilizing KDDI assets to accelerate
independent listing.
（KDDI資産を活用して独立上場を加速）

Reading Deck

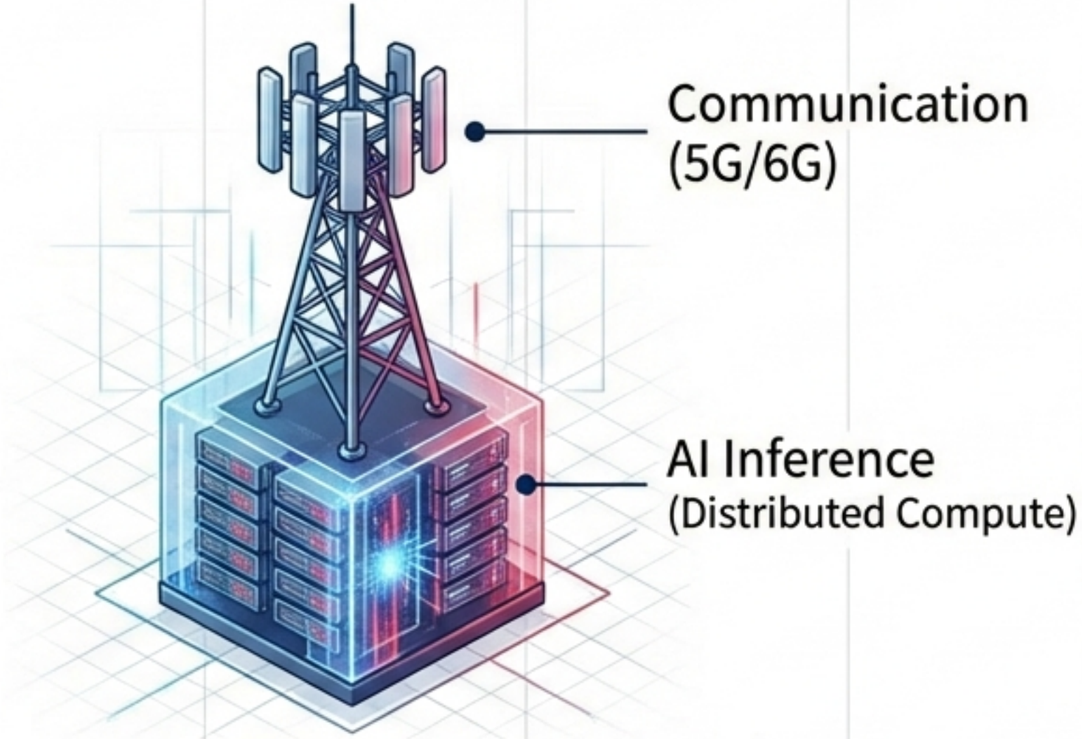
SoftBank "Sarashina"：圧倒的資本によるインフラ融合とAGI

Model
Sarashina2 (MoE) / Sarashina mini
Noto Serif JP Roboto Mono

Infrastructure
Massive H100 Clusters
Noto Serif JP Roboto Mono

Goal
1 Trillion Parameters & ASI (2035)
Noto Serif JP Roboto Mono

AI-RAN (Radio Access Network)



基地局をAIコンピュータ化し、
日本全国を分散推論基盤へ変える。

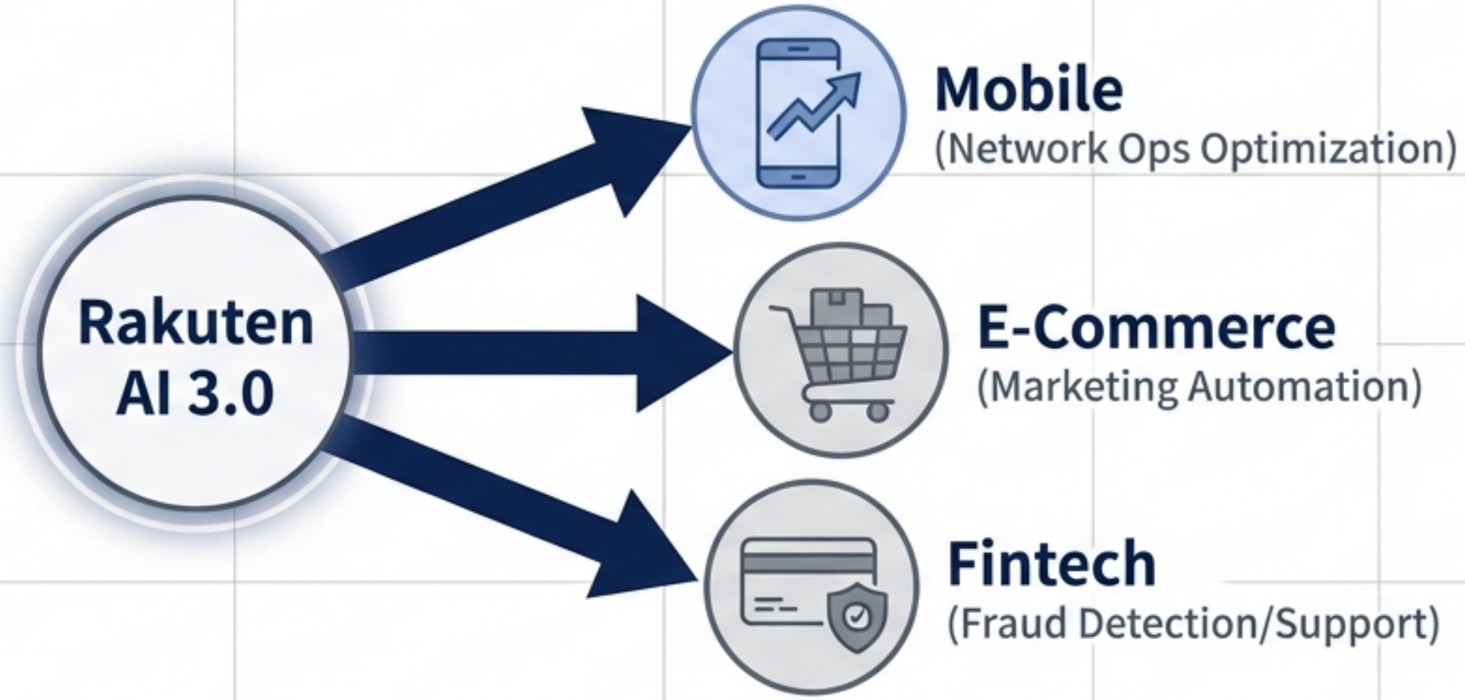
Reading Deck

楽天 "Rakuten AI 3.0"：7000億パラメーターとエコシステム

700B Parameters (Active 40B)
Domestic Largest Scale

Open Weights Release (Spring 2026)

Rakuten Ecosystem Triple Play



楽天の業日や、営業による経減、90%の企業・90%にとる提程器を衝撃度。
Internal Goal: 90% operational cost reduction by replacing external vendors.

加速装置としての政府支援：GENIACプロジェクト

Compute Democratization

Providing GPU resources to Startups & Giants (METI/NEDO).

→

Evaluation Standards

Establishment of JGLUE & Rakuda benchmarks.

Creating objective “rulers” for Japanese cultural understanding.

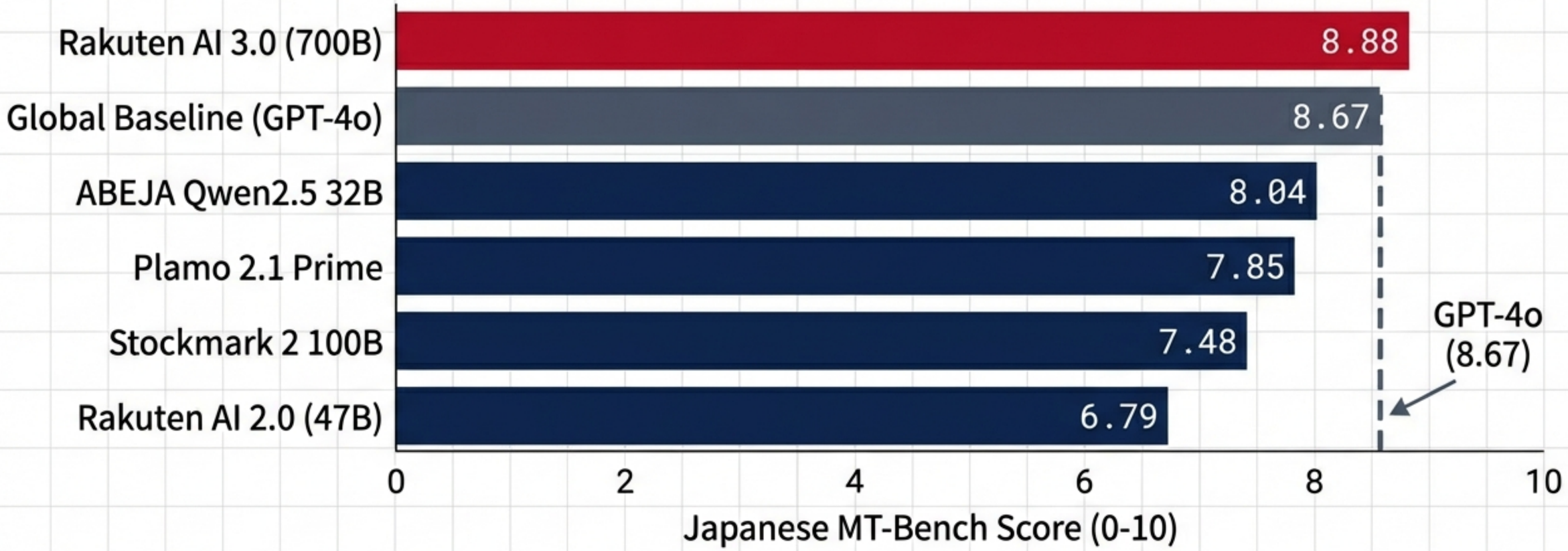
→

Social Implementation (2025-2026)

Shift from “Development” to “Real-world Application”.

Key Beneficiaries: Rakuten, PFN, Turing, Stockmark.

日本語性能ベンチマーク比較（MT-Bench Japan 2025）



特定の日本語評価基準において、国産トップモデルがGPT-4oを逆転。

Data sources: Rakuten Group, Tech in Asia, Hugging Face, Note.com.
(注：スコアは各社の発表値および公開リーダーボードに基づく)

Reading Deck

2026年の焦点：「チャット」から「エージェント」へ

2025: Chatbot

Human Prompt
(Noto Serif JP, grey)

AI Answer
(Noto Serif JP, grey)

Passive Information Retrieval.

2026: Agent

Human sets Goal
(Noto Serif JP, Japan Indigo)

AI Plans
(Noto Serif JP, Japan Indigo)

AI uses Tools
(API/DB via MCP)
(Roboto Mono, crimson)

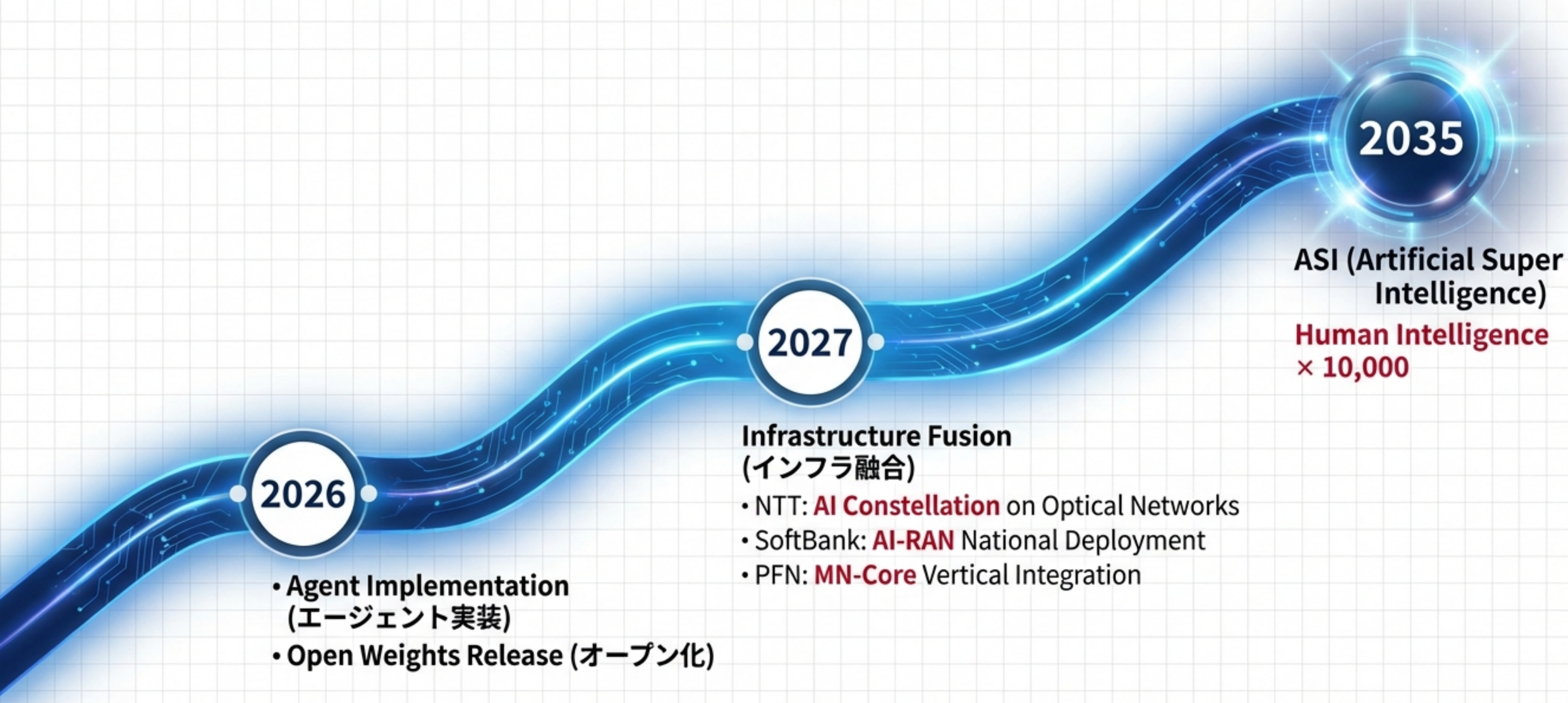
AI Executes
(Noto Serif JP, Japan Indigo)

AI Reports
(Noto Serif JP, Japan Indigo)

Active Task Execution (e.g., cotomi Act).

Reading Deck

ロードマップ 2027+：ハードウェア融合とAGIへの道



Data sources: Ministry of Internal Affairs and Communications, NTT, SoftBank, PFN, Industry Analysis.

Reading Deck

結論：適材適所の「マルチモデル戦略」

Global Models (GPT-5 / Claude)

- Creative Tasks
- English/Global Research
- General Purpose Coding

Domestic Models (tsuzumi / cotomi / Sarashina)

- Confidential Data (Personal/Finance)
- High-speed Industrial Automation
- Cultural Nuance (Ringi/Keigo)

国産AIはもはや「コピー」ではない。日本のデジタル主権を支える「社会インフラ」である。

Adopt a Hybrid Strategy.