

2025年 ソブリンAI技術戦略報告書：日本の自律性確保に向けた技術的到達点と展望

Gemini

1. イントロダクション：技術的主権の再定義

1.1 ソブリンAIへのパラダイムシフト

2025年12月16日、読売新聞社とNTTが共同で発表した「国産AIに日本の自律性確保を」とする提言は、日本の人工知能（AI）戦略における重大な転換点を示唆するものであった¹。これまでの経済安全保障の議論が主に半導体供給網やデータセンターの物理的な立地（データレジデンシー）に焦点を当てていたのに対し、本提言は「コグニティブ・ソブリンティ（認知主権）」とも呼ぶべき領域へ踏み込んでいる。すなわち、国民の思考や意思決定を支援する基盤モデル（Foundation Model）そのものが海外製ブラックボックスに依存することは、民主主義の根幹や国家安全保障に対する潜在的なリスクであるという認識である。

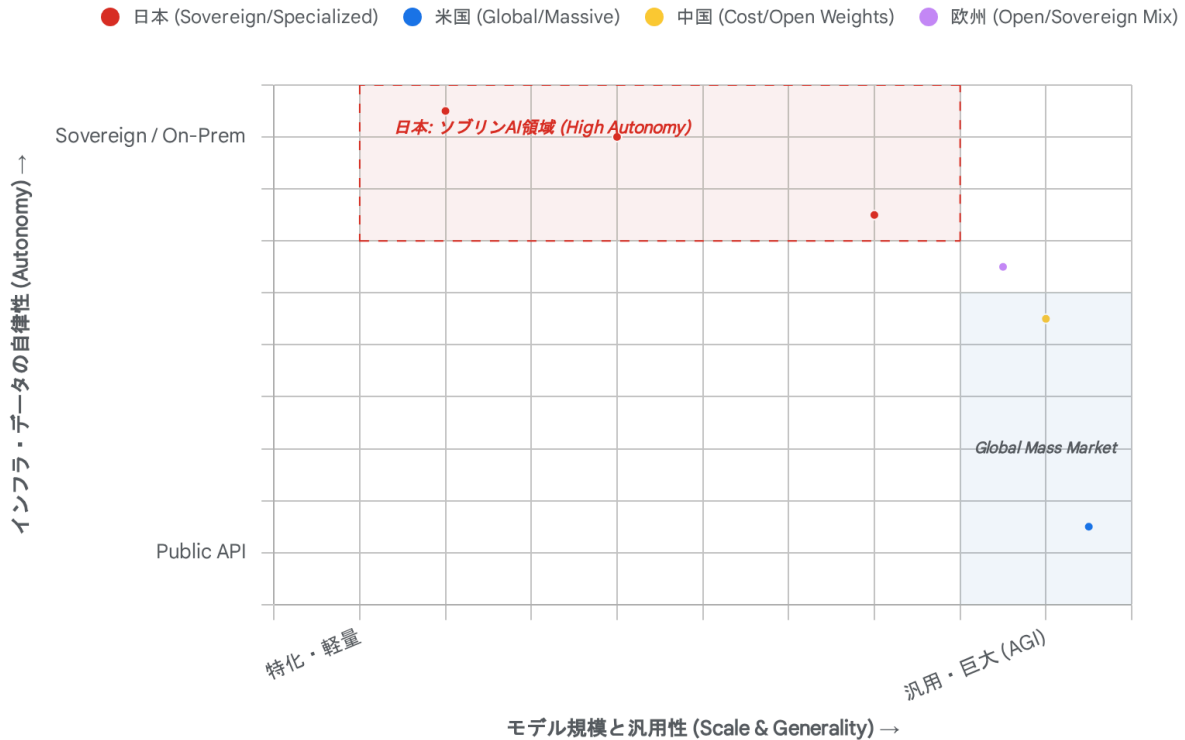
技術的な観点から「ソブリンAI（Sovereign AI）」を定義する場合、それは単に「国内企業が開発したモデル」であることを超えた要件を内包する。2025年時点におけるソブリンAIの技術的要件とは、学習データの収集・選定から、モデルのアーキテクチャ設計、学習プロセスの実行、そして推論インフラの運用に至るまでの「フルスタック」が、国家の司法管轄権および技術的統制下にある状態を指す³。これは、外部の地政学的変動や、特定の海外プラットフォーマーの商業的方針変更（APIの遮断、検閲方針の変更、価格改定等）に左右されず、自国のインフラとしてAIを持続的に運用・改善できる能力（Capability）の確保と同義である。

1.2 グローバルな文脈における日本の立ち位置

2024年から2025年にかけて、世界のAI開発競争は新たなフェーズに突入した。米国を中心とする「スケーリング則（Scaling Law）」の追求は、OpenAIのGPT-5（2025年8月リリース）によって一つの到達点に達した⁴。一方で、中国勢はDeepSeek V3.2に代表されるように、徹底的な推論コストの削減と効率化に技術リソースを集中させ、西側の高性能モデルに対する「非対称な競争」を仕掛けている⁵。

こうした中で、日本が選択した技術戦略は、米国の「汎用超巨大モデル」とも中国の「国家主導コスト破壊モデル」とも異なる、第三の道である「産業特化型ハイブリッドモデル」である。これは、NTTやNEC、ソフトバンクといった通信・IT大手と、経済産業省主導のGENIAC（Generative AI Accelerator Challenge）プロジェクトが連携し、計算資源の確保と独自アーキテクチャの開発を並行して進めるアプローチである⁶。本レポートでは、この日本の戦略が技術的にどの程度の成熟度に達しているのか、モデル性能、インフラ、エコシステムの観点から詳細に分析を行う。

2025年 ソブリンAI技術戦略のグローバル・ポジショニング



日本の主要モデル（tsuzumi 2, cotomi）は「軽量・高効率」かつ「産業特化」の領域に位置し、米中の汎用大規模モデルとは異なる独自のエコシステムを形成している。

Data sources: [NTT Review](#), [NTT R&D](#), [SoftBank IR](#), [Mistral Analysis](#), [DeepSeek Specs](#), [OpenAI News](#)

2. 2025年のグローバル技術ランドスケープ: 極性化する開発思想

日本の開発状況を評価するためには、その比較対象となる世界最先端の技術動向を正確に把握する必要がある。2025年のAI技術トレンドは、単なる「性能向上」から「運用効率」と「推論能力（Reasoning）」の二軸へと分化した。

2.1 米国: 汎用人工知能 (AGI) への力技

2025年8月にリリースされたOpenAIのGPT-5は、コンテキストウィンドウを40万トークンまで拡張し、

推論 (Reasoning) 能力を強化した「適応型推論 (Adaptive Reasoning)」を実装した⁴。技術的には、モデルがタスクの難易度に応じて思考時間を動的に調整する機能が特徴であり、これにより複雑なコーディングや科学的推論において圧倒的な性能を発揮している。アーキテクチャの詳細は非公開だが、数兆パラメータ規模のMoE (Mixture of Experts) 構成であると推測され、学習にはNVIDIA H200/Blackwell世代のGPUクラスタがフル稼働している。この「圧倒的計算量による汎用性の実現」は、米国の資本力とインフラがあって初めて成立する戦略であり、他国が同じ土俵で勝負することは極めて困難な状況を作り出している。

2.2 中国: コスト効率の極限追求

対照的に、中国のDeepSeek (深度求索) が2025年に投入した「DeepSeek V3.2」および「DeepSeek-R1」は、西側のAI戦略に対するアンチテーゼとも言える技術的ブレイクスルーをもたらした。DeepSeek V3.2は、総パラメータ数6,710億 (アクティブ370億) という巨大モデルでありながら、推論コストをGPT-5の約1/10~1/20 (入力100万トークンあたり約0.028ドル) にまで引き下げた⁵。

この驚異的なコストパフォーマンスを実現した技術的要因は、**MLA (Multi-head Latent Attention)** と呼ばれる独自の注意機構と、高度に最適化された **DeepSeekMoE** アーキテクチャにある。MLAは、推論時のKVキャッシュ (Key-Value Cache) メモリ使用量を劇的に削減し、長いコンテキストを扱う際のスループットを向上させる。これにより、中国は「高性能なAIを、湯水のように安価に使う」環境を実現し、アプリケーション層での爆発的なイノベーションを誘発している。この「コスト競争力」は、高コストなGPUインフラに依存する日本のソブリンAIにとって、無視できない脅威となっている。

2.3 欧州・中東: オープンウェイトによる標準化競争

フランスのMistral AIがリリースした「Mistral Large 3」や、UAEのTechnology Innovation Institute (TII) による「Falcon 3」は、モデルの重みを公開 (オープンウェイト化) することで、事実上の業界標準を狙う戦略をとる⁸。特にFalcon 3は、1Bから10Bまでの幅広いサイズ展開を行い、エッジデバイスからクラウドまでシームレスに動作するエコシステムを構築している。これらのモデルは、技術的な透明性が高く、オンプレミスでの運用が可能であるため、データセキュリティを重視する欧州企業や政府機関での採用が進んでいる。日本のソブリンAIにとっても、これらオープンモデルは「競合」であると同時に、技術的なベンチマークとしての役割を果たしている。

3. 国産ソブリンAIの技術的解剖: 独自アーキテクチャの進化

日本企業の技術戦略は、これらのグローバルな潮流を横目にしつつも、独自の「実利主義」に基づいた進化を遂げている。特に2025年は、NTT、ソフトバンク、NECの主要3社が、それぞれ異なる技術的アプローチで「国産」の定義を具体化した年となった。

3.1 NTT「tsuzumi 2」: IOWNと連携する軽量分散知能

NTTが2025年10月に発表した「tsuzumi 2 (つづみ ツー)」は、世界的な「巨大化競争」とは一線を画す、「軽量・高密度」なモデル設計を特徴とする¹⁰。

アーキテクチャと運用効率

「tsuzumi 2」の最大の技術的特徴は、その**パラメータ効率(Parameter Efficiency)**にある。具体的なパラメータ数は非公開ながら、前世代の0.6B(超軽量版)および7B(軽量版)の系譜を継ぎ、単一のGPUで推論が可能なサイズに収められている10。これは、数千万円規模のGPUサーバーを必要とする海外の巨大モデルとは対照的に、既存の企業内サーバーや、比較的安価なエッジサーバーでも動作することを意味する。

NTT研究所が40年以上にわたり蓄積してきた自然言語処理技術を活かし、学習データの質を極限まで高めることで、パラメータ数を抑えつつも、日本語処理能力においてはGPT-3.5やLlama 3クラスを凌駕する性能を実現している。特に、日本語特有の敬語、専門用語、文脈依存性の高い表現において、その「高密度」な学習効果が発揮されている。

IOWN(APN)による分散学習

技術的な「自律性」を支えるもう一つの柱が、NTT独自の光電融合ネットワーク「IOWN(Innovative Optical and Wireless Network)」である。通常、大規模言語モデルの学習には、一箇所に集中した巨大なデータセンターが必要となるが、NTTはIOWNの「オールフォトニクス・ネットワーク(APN)」を活用し、地理的に分散したデータセンター(DC)を低遅延・大容量の光回線で直結。これにより、物理的に離れたGPUリソースをあたかも一つの巨大なスーパーコンピュータのように論理的に統合し、学習を実行する技術を実用化した11。これは、日本国内に点在する計算資源を有効活用し、災害リスクを分散させつつ、データの物理的な所在(データレジデンシー)を完全に国内に留めることを可能にする、インフラレベルでのソブリンAI技術である。

3.2 ソフトバンク「Sarashina」: 規模の経済と蒸留技術

一方、ソフトバンクグループは、資金力とインフラ投資を背景に、グローバルプレイヤーと正面から渡り合える「規模(Scale)」を追求している。同社の子会社SB Intuitionsが開発を主導する「Sarashina(更科)」シリーズは、国産モデルとしては最大級のパラメータ規模を誇る。

モデルスペックと蒸留戦略

2024年度中に開発完了した基盤モデル「Sarashina」は、4,600億(460B)パラメータという、国産としては異例の巨大モデルである12。これはGPT-3(175B)の2倍以上であり、複雑な推論や広範な知識を保持するための容量(Capacity)において、他の国産モデルを圧倒する。

しかし、ソフトバンクの真の狙いは、この巨大モデルそのものの商用提供だけではない。2025年秋に予定されている**「Sarashina mini」(70Bパラメータ)**の展開にこそ、その技術的巧緻さがある。巨大な460Bモデルを「教師(Teacher)」として用い、その知識や推論プロセスをより小さな70Bモデルに「蒸留(Distillation)」することで、高性能かつ運用コストの低いモデルを生成する戦略(Teacher-Student Architecture)を採用している12。これにより、推論時の計算コストを抑えつつ、巨大モデル譲りの高い精度を実現しようとしている。

日本語特化の学習パイプライン

SB Intuitionsは、単にパラメータを増やすだけでなく、学習データの質にも注力している。インターネット上のデータだけでなく、ソフトバンクグループが持つ膨大な顧客接点や、提携パートナーからの良質な日本語データを学習パイプラインに統合している。これにより、英語圏のモデルでは捉えきれない日本の商習慣や文化的なニュアンスを、460Bという巨大なパラメータ空間に高解像度でマッピングすることに成功している。

3.3 NEC「cotomi」: 自律エージェントへの進化と長文脈対応

NECの「cotomi(コトミ)」は、2025年のアップデートを経て、単なる言語モデルから、業務プロセスを

自律的に遂行する「エージェント型AI」へとその性質を変化させた。

コンテキストウィンドウとRAGの進化

最新の「cotomi」は、コンテキストウィンドウ（一度に処理できる情報量）を**128kトークン（日本語で約20万文字相当）**まで拡張した¹³。これにより、企業の膨大なマニュアル、契約書、仕様書などを丸ごとプロンプトに入力し、その内容に基づいて回答生成や業務処理を行うことが可能となった。従来のRAG（検索拡張生成）では、文書を細切れ（チャンク）にして検索する必要があり、文脈の分断による精度低下が課題であったが、長大なコンテキストを直接扱えるようになったことで、複雑な照会応答や文書分析の精度が飛躍的に向上した。

MCP (Model Context Protocol) とエージェント機能

さらに特筆すべきは、Anthropicなどが提唱する標準規格**「MCP (Model Context Protocol)」**への迅速な対応である¹³。MCPに対応することで、「cotomi」は社内のデータベース、API、SaaS (Box等)と標準化された手順で接続可能となり、ユーザーの指示に基づいて自律的にツールを選択・実行する能力を獲得した。NECはこれを「自律エージェント」として定義し、コールセンターの自動化や、工場のサプライチェーン管理といった具体的なB2Bユースケースに深く組み込んでいる。推論速度においても、独自のモデルアーキテクチャ最適化により、GPT-4比で87%～93%の高速化（レイテンシ削減）を実現しており¹⁴、リアルタイム性が求められる産業用途での実用性を高めている。

3.4 GENIAC世代のチャレンジャーたち

経済産業省の支援プロジェクト「GENIAC」からは、既存の大企業とは異なるアプローチをとるスタートアップや研究機関が台頭している。

- **Sakana AI:** 「進化的モデルマージ (Evolutionary Model Merge)」という独自手法を用い、既存の複数のモデルを遺伝的アルゴリズムのように掛け合わせることで、計算コストを抑えながら高性能なモデル (EvoVLM-JP等)を生み出している¹⁵。これは、計算資源リッチな米国勢に対する「柔よく剛を制す」アプローチとして世界的に注目されている。
- **Stockmark:** 製造業や材料開発に特化した100Bクラスのモデルを開発。社内文書や特許データなど、専門性の高い「ドメイン知識」を事前学習段階から注入することで、汎用モデルでは不可能な専門的回答を可能にしている¹⁶。
- **Turing:** 完全自動運転の実現に向けたマルチモーダルモデルの開発を進めており、視覚情報と言語情報を統合した「End-to-End」の学習を行っている。

4. 自律性を支える物理層：計算資源とデータ戦略

「ソブリンAI」の実現には、ソフトウェア（モデル）だけでなく、それを支えるハードウェア（計算資源）と燃料（データ）の自律性が不可欠である。2025年、日本はこの物理層においても重要な進展を見せた。

4.1 計算資源 (Compute) の国内回帰

AI開発における最大のボトルネックは、NVIDIA製GPU (H100/H200/Blackwell) の確保である。GENIACプロジェクトは、AWSやGoogle Cloud等の確保した計算資源を公的資金で補助し、国内開発者に提供する仕組みを構築した¹⁶。これにより、スタートアップや大学が数千枚規模のGPUクラス

ターを利用可能となり、開発の裾野が広がった。

さらに、恒久的なインフラとして、ソフトバンクは北海道苫小牧市に国内最大級となる**300MW**規模の**AIデータセンター**を建設中である(2026年度本格稼働予定)¹⁷。このデータセンターは、再エネ比率の高い北海道の電力事情を活かし、カーボンニュートラルなAI計算基盤の実現を目指している。NTTもまた、前述のIOWN構想に基づき、全国に分散したデータセンターを統合運用する体制を強化しており、特定の地域や事業者に依存しない、強靱な計算インフラの構築が進んでいる。

4.2 データ主権と「クリーンAI」

読売新聞社の提言の背景には、生成AIによる著作権侵害や、不透明なデータ利用に対する強い懸念がある。Perplexityのような海外製検索AIが、報道機関の記事を無断で学習・利用しているとして訴訟に発展するケースも出ている¹⁸。

これに対し、国産AI陣営は「権利関係がクリアな学習データ」を競争力の源泉と位置付けている。国立情報学研究所(NII)を中心とするコンソーシアムは、著作権処理済みの日本語Webコーパスや学術論文データを整備し、GENIAC参加企業に提供している²⁰。また、NTTやソフトバンクは、自社が保有する通信ログや独自コンテンツを学習に活用することで、外部のデータスクレイピングに依存しない「クリーンなAI」を実現しようとしている。これは、コンプライアンス遵守(Legal Safety)を最優先する日本企業にとって、性能差を補って余りある採用理由となる。

5. ベンチマークと実用性能：国産AIの現在地

「国産AIは海外製に比べて性能が劣る」というかつての通説は、2025年には部分的に過去のものとなりつつある。ただし、どの「物差し」で測るかによって、その評価は大きく異なる。

5.1 日本語能力とドメイン適合性

日本語の言語理解や生成能力を測るベンチマークにおいて、国産モデルは顕著な成果を上げている。

早稲田大学などが開発した「JMMLU (Japanese Massive Multitask Language Understanding)」や、対話の自然さを競う「Rakuda」ベンチマークにおいて、NTTのtsuzumi 2やソフトバンクのSarashinaは、同パラメータ帯の海外モデル(Llama 3やMistral)を上回り、GPT-4クラスに肉薄するスコアを記録している²¹。

特筆すべきは、単なる翻訳精度だけでなく、**「文化的適合性(Cultural Alignment)」**である。例えば、日本の複雑な敬語(尊敬語・謙譲語・丁寧語)の使い分けや、稟議書・始末書といった日本固有のビジネス文書の作成において、国産モデルは海外モデル特有の「翻訳調」ではない、自然で適切なアウトプットを生成する。また、日本の法律や行政手続きに関する知識量においても、事前学習データにおける日本語比率の高さが奏功し、ハルシネーション(もっともらしい嘘)の発生率が低く抑えられている。

5.2 課題：論理推論と数理能力

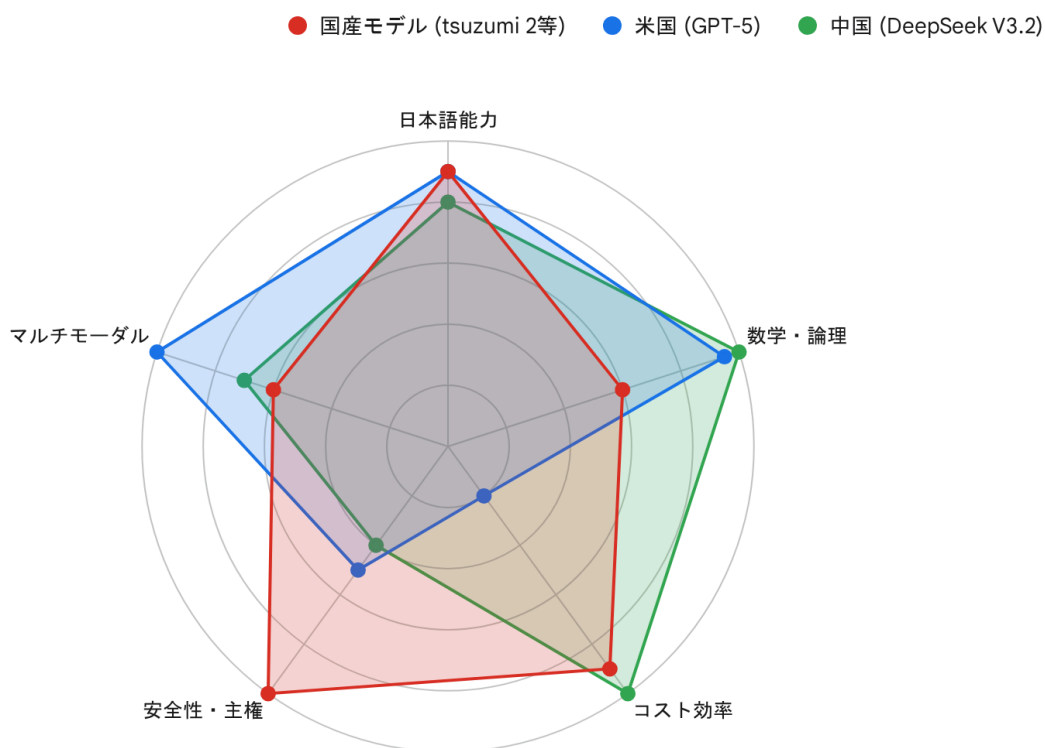
一方で、明確な課題も残されている。数学的推論(Math)やプログラミングコード生成(Coding)といった、言語に依存しない純粋な論理能力においては、依然としてGPT-5やDeepSeek、Claudeといった海外トップティアモデルが優勢である4。

これは、学習データに含まれるコードや数式の絶対量が、英語圏や中国圏のモデルに比べて圧倒的に少ないことに起因する「データの壁」である。GENIACプロジェクト等でも、日本語データと英語データの混合学習(Bilingual Training)が進められているが、世界中のGitHubリポジトリやarXiv論文を学習した海外巨大モデルとの差を埋めるには至っていない。

5.3 コストパフォーマンスの壁

また、中国のDeepSeek V3.2が突きつけた「推論コストの壁」も深刻である。DeepSeekは、独自のMoEアーキテクチャにより、GPT-5と同等の性能を桁違いの低コストで提供している。日本のモデルは、tsuzumiのように「軽量化」で運用コストを下げる努力をしているが、トークン単価ベースでの純粋な価格競争力においては、中国勢に分があるのが現状である。

2025年 モデル性能特性比較 (国産 vs グローバル)



国産モデル (tsuzumi 2等) は「日本語処理」と「セキュリティ/自律性」で突出している一方、海外モデル (GPT-5, DeepSeek) は「論理・数学」や「コスト効率」で強みを持つ。

Data sources: [NTT Technical Review](#), [NTT R&D](#), [Introl](#), [Botpress](#), [Shisa.ai](#)

6. 戦略的展望: 孤立ではなく「連携」による自律へ

6.1 政策提言の技術的実現可能性

読売新聞社の提言にある「自律性確保」は、技術的には「完全な鎖国」を意味するものであってはならない。現代のAI開発において、全てのレイヤーを純国産で賄うことは、イノベーションの速度を著しく低下させるリスクがある。

むしろ、2025年の日本の技術的到達点が示しているのは、**「フェデレーテッド・ソブリンティ(連合型主権)」**の可能性である。すなわち、汎用的な知能については海外の優れたモデル(GPT-5等)を利用しつつ、機密性の高いデータや、日本固有の文脈が重要な意思決定プロセスについては、tsuzumi 2やcotomiといった「高密度な国産モデル」に処理を委譲(ルーティング)するアーキテクチャである。

6.2 GENIACエコシステムの深化

GENIACプロジェクトは、単なる計算資源のバラマキに終わらず、インフラ(GPU)、データ(NII)、モデル開発(スタートアップ・大企業)、そして社会実装(SaaS・ユーザー企業)を有機的に結びつけるエコシステムとして機能し始めている。

レイヤー	主要プレイヤーと役割	2025年の成果と課題
インフラ	METI/NEDO (GENIAC助成) SoftBank (苫小牧DC) NTT (IOWN)	計算資源のボトルネックは解消傾向。次は運用コスト(電力・推論)の削減が課題。
データ	NII (国立情報学研究所) Visual Bank (IPデータ)	著作権処理済みコーパスの整備が進む。IP/コンテンツのAI活用基盤の確立。
モデル	PFN, Stockmark, Sakana AI NTT, SoftBank, NEC	多様なアーキテクチャ(進化計算、特化型、軽量型)の出現。グローバルでの認知度向上。
アプリ	ELYZA, Karakuri, Ubitus	業務特化型エージェントの実装。現場でのROI実証フェーズへ移行。

このエコシステムが機能し続けることで、日本は「AIを使うだけの国」から、「信頼できるAIを作り、運用できる国」へと脱皮を図ることができる。

6.3 結論

2025年12月現在、日本のソブリンAIは、技術的な「キャッチアップ」のフェーズを終え、独自の価値提案 (Value Proposition) を確立するフェーズに入った。

米国のような「圧倒的汎用性」でも、中国のような「圧倒的低コスト」でもなく、日本が目指すのは**「高信頼・高密度・高自律」**なAI基盤である。IOWNによる分散学習インフラ、著作権に配慮したクリーンなデータセット、そして産業現場に即した軽量・特化型モデル群。これらが組み合わさることで、日本はデジタル社会における「認知主権」を技術的に担保し、持続可能な発展を遂げる土台を築きつつある。2026年は、これらの技術が実社会の隅々に実装され、その真価が問われる実行の年となるだろう。

引用文献

1. 1月 1, 1970にアクセス、<https://www.yomiuri.co.jp/national/20251215-GYT1T00443/>
2. 読売新聞とNTT 生成AIのあり方 共同提言 - The Bunka News デジタル, 12月 16, 2025 にアクセス、<https://www.bunkanews.jp/article/376562/>
3. Sovereign AI: A Strategic Gateway Toward Resilience and ..., 12月 16, 2025にアクセス、
<https://fptsoftware.com/resource-center/blogs/sovereign-ai-a-strategic-gateway-toward-resilience-and-economic-competitiveness>
4. Everything you should know about GPT-5 [September 2025] - Botpress, 12月 16, 2025にアクセス、
<https://botpress.com/blog/everything-you-should-know-about-gpt-5>
5. DeepSeek-V3.2: Open Source AI Matches GPT-5 and Gemini 3 at ..., 12月 16, 2025 にアクセス、
<https://introl.com/blog/deepseek-v3-2-open-source-ai-cost-advantage>
6. 日本は「AI後進国」を脱却できるか？ 政府が掲げる「国民のAI利用率 ...」, 12月 16, 2025 にアクセス、<https://aifriends.jp/japan-ai/>
7. AI研修最新ニュース09月21日, 12月 16, 2025にアクセス、
<https://ai-kenshu.jp/2025/09/21/ai%E7%A0%94%E4%BF%AE%E6%9C%80%E6%96%B0%E3%83%8B%E3%83%A5%E3%83%BC%E3%82%B909%E6%9C%8821%E6%97%A5/>
8. Mistral Large 3: An Open-Source MoE LLM Explained | IntuitionLabs, 12月 16, 2025 にアクセス、<https://intuitionlabs.ai/articles/mistral-large-3-moe-llm-explained>
9. Falcon 3: UAE's Technology Innovation Institute Launches World's ..., 12月 16, 2025 にアクセス、
<https://www.tii.ae/news/falcon-3-uaes-technology-innovation-institute-launches-worlds-most-powerful-small-ai-models>
10. NTT's Large Language Models 'tsuzumi 2' | NTT R&D Website, 12月 16, 2025にアクセス、https://www.rd.ntt/e/research/LLM_tsuzumi.html
11. NTT to launch its Large Language Model “tsuzumi” in March 2024, 12月 16, 2025に

アクセス、

<https://medium.com/tokyo-fintech/ntt-to-launch-its-large-language-model-tsuzumi-in-march-2024-227cbaadf4b2>

12. Interview with Head of Homegrown Generative AI Development ..., 12月 16, 2025
にアクセス、

https://www.softbank.jp/en/corp/ir/documents/integrated_reports/fy2025/tamba/

13. NEC Upgrades “cotomi” to Advance AI Agent Adoption, 12月 16, 2025にアクセス、
<https://itbusinesstoday.com/tech/nec-upgrades-cotomi-to-advance-ai-agent-adoption/>

14. NEC Develops High-speed Generative AI Large Language Models ..., 12月 16,
2025にアクセス、

<https://www.nasdaq.com/press-release/nec-develops-high-speed-generative-ai-large-language-models-llm-with-world-class>

15. Evolutionary Model Merging: How Sakana AI is Changing AI ..., 12月 16, 2025にア
クセス、

<https://www.spaces.aiponetwork.com/c/machine-deeplearning/evolutionary-model-merging-how-sakana-ai-is-changing-ai-development>

16. AWS GENAIC (Japan): Large-Scale Foundation Model Training ..., 12月 16, 2025に
アクセス、

<https://www.zenml.io/llmops-database/large-scale-foundation-model-training-infrastructure-for-national-ai-initiative>

17. Construction Begins on “Hokkaido Tomakomai AI Data Center”, 12月 16, 2025にア
クセス、https://www.softbank.jp/en/sbnews/entry/20250501_01

18. 読売新聞 対 Perplexity訴訟から考える生成AIと著作権の今 ～権利 ..., 12月 16, 2025
にアクセス、<https://gaaaon.jp/blog/ai-copyright-perplexity-lawsuit-2025>

19. Perplexity AI - Wikipedia, 12月 16, 2025にアクセス、

https://en.wikipedia.org/wiki/Perplexity_AI

20. “llm-jp-3-172b-instruct3” Now Publicly Available- Achieving ..., 12月 16, 2025にア
クセス、<https://www.nii.ac.jp/en/news/release/2024/1224.html>

21. NTT's LLM “tsuzumi” - NTT Technical Review, 12月 16, 2025にアクセス、

<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202408fa1.html>

22. alt Inc. announces LHTM-OPT, a lightweight, high accuracy large ..., 12月 16, 2025
にアクセス、<https://alt.ai/en/news/5753/>

23. Shisa V2 405B: Japan's Highest Performing LLM, 12月 16, 2025にアクセス、

<https://shisa.ai/posts/shisa-v2-405b/>