国産LLMの直近半年動向（2026年5月～11月）とAIエージェント化における「18カ月遅れ」仮説の検証

Gemini 4 Carbon

2026年5月から11月にかけての直近半年は、日本の国産大規模言語モデル（LLM）にとって「フルスクラッチ推論モデルの一斉実用化」と「対話補助から自律型AIエージェント・フィジカルAIへの産業構造シフト」が同時に進行した転換期となりました[1]。モデル開発面では、Preferred Networks（PFN）の「PLaMo 3.0 / 3.1 Prime」[4]、国立情報学研究所（NII）の「LLM-jp-4.1」[6]、KDDI傘下ELYZAの「ELYZA-Thinking-1.0（llm-jp-4版）」[7]、ソフトバンク傘下SB Intuitionsの「Sarashina3.2 mini thinking」[8]など、国内でゼロから事前学習された純国産基盤モデルが相次いで思考（Reasoning）機能とツール利用能力を獲得しています[9]。並行して、デジタル庁のガバメントAI「源内（GENAI）」による全府省庁約18万人規模の実証開始[13]や、ソフトバンク・ソニーグループ・NEC・本田技研工業を中核に出資を受け（報道によれば初年度約3,873億円の経産省支援）、国産マルチモーダル・フィジカルAI基盤を開発する新会社「Noetra（ノエトラ）」の始動など、国家規模の需要創出と投資も具体化しました[15]。
「推論モデルの導入で18カ月遅れた国産勢は、AIエージェント化でも同様に遅れているのではないか」という問いに対する結論は、「単一基盤モデルによる自律コーディングや長時間実行能力では依然12～18カ月規模の遅れが残る一方、AIエージェント全体としては一律18カ月遅れではなく、技術階層によって『12～18カ月遅れ』と『ほぼ同時並行（0～6カ月差）』に二極化している」というものです[9]。
実際に、OpenAI o1（o1-1217は2024年12月）やDeepSeek-R1（2025年1月公開）の登場から純国産フルスクラッチ推論モデルが商用・公開水準で出揃う（2026年6月～10月）までには、実態として約18カ月を要しました[4]。しかし、国産各社はこの半年間の推論モデル開発パイプラインにFunction Calling（関数呼び出し）やAgentic Search（自律検索）の強化学習を同時統合したため、「推論の18カ月遅れの上にエージェント化の18カ月遅れが単純加算される」事態は回避されています[9]。
それでも基盤モデル単体で海外フロンティアとの距離が縮まり切らない背景には、2026年秋にかけて海外主要ラボが単発のツール呼び出しを超え、数時間～数日単位でOSやターミナルを自律操作し続ける「長時間耐久実行」へと競争軸を一段引き上げたことがあります[18]。その一方で、モデルの巨大化のみに依存しない「エージェント・ハーネス（制御機構）」「人間の操作ログからの暗黙知抽出によるブラウザ自動操作（CUA）」「複数フロンティアLLMの動的オーケストレーション」の領域では、NECの「cotomi Agent」[1]やSakana AIの「Sakana Marlin / Fugu / Aquarium」[23]、富士通やELYZAの自己進化型ハーネス技術[26]など、海外勢と大きな時間差なく商用・実運用へ到達した事例が相次いでいます。ただし、国内勢の多くは最新の世界標準エージェント評価（SWE-bench ProやTerminal-Bench 4.0等）での統一比較スコアを公表しておらず、モデル単体の汎用自律性能とシステムインテグレーション（SI）込みの業務完遂力を切り分けて評価する視点が不可欠です[18]。

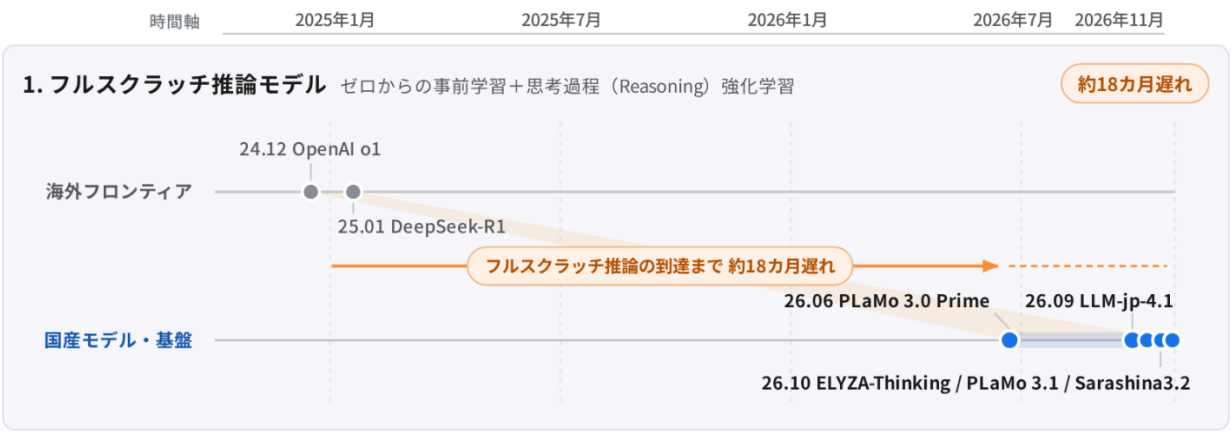
「推論18カ月遅れ」の構造とAIエージェント化における階層別タイムラグ

国産LLMの遅れを正確に診断するには、「海外オープンモデルの日本語追加学習」と「フルスクラッチ（ゼロからの事前学習）国産モデル」を区別したうえで、AIエージェントを構成する技術スタックを3つの階層に分解する必要があります[29]。

推論の18カ月遅れに対し、AIエージェント化は基盤層と実装層で二極化している

● 海外フロンティア　● 国産モデル（遅行領域）　● 国産ハーネス・CUA（並走領域）　— 機能到達ラグ

各ノードにカーソルを合わせると詳細を表示



2024年後半〜2026年11月の主要マイルストーン比較。純国産フルスクラッチの推論モデルは海外勢から約18カ月遅れで2026年夏〜秋に出揃った。AIエージェント化では、単体モデルの長時間・自律コーディング領域で**12〜18カ月規模の差**が続く一方、ブラウザ操作（CUA）やオーケストレーション、業務ハーネス層では海外勢とほぼ同時期に商用展開が進んでいる。

データ出典: Preferred Networks、SB Intuitions、NEC、カラクリ（PR TIMES）、ビジネス+IT（Sakana Marlin）、技術評論社（Sakana Fugu）、AIgent Lab（Sakana Aquarium）、IT Leaders（富士通）、AI Revolution（ELYZA）、Spike Studio、WEEL、NII LLM-jp-4.1解説、エージェント実装動向調査

第一に、推論モデル導入における「18カ月遅れ」の内実を振り返ると、海外のオープンウェイトモデルを土台にした日本語適応自体は極めて迅速でした。2025年1月20日に中国のDeepSeekが推論モデル「DeepSeek-R1」を公開した際、サイバーエージェントは7日後の1月27日に蒸留モデルの日本語追加学習版を公開し[21]、ELYZAも2025年5月にはアリババのQwenを土台とした「ELYZA-Thinking-1.0-Qwen-32B」を公開しています[29]。しかし、外部ライセンスや海外モデルの設計制約を受けないフルスクラッチの純国産基盤モデルにおいて思考能力を獲得するには、事前学習データの拡充だけでなく、思考過程データの蒸留・自動検証器（Validator）の構築・強化学習（RL）の安定化という事後学習（Post-training）パイプラインを一から確立する必要がありました[3]。公開知見が限られるなかでこの技術的ハードルを突破し[3]、PLaMo 3.0 Prime（2026年6月）[4]、LLM-jp-4.1（同年9月）[6]、ELYZA-Thinking-1.0-llm-jp-4およびPLaMo 3.1 Prime、Sarashina3.2 mini thinking（いずれも同年10月）[5]が出揃うまでに、OpenAI o1やDeepSeek-R1から数えて約17〜22カ月（約18カ月）を要したのが実情です[20]。実際、SB Intuitionsは2026年6月末に非推論型の「Sarashina3 mini / nano」を公開した時点では、「推論（Reasoning）モデルが主流になった今、エージェント性能などにはまだ課題が残る」と自ら認めていました[1]。

第二に、基盤モデル単体のAIエージェント能力（ツール呼び出し・情報検索・自律実行）については、「基礎機能の同時キャッチアップによる遅れの圧縮」と「最新フロンティアのさらなる先行」が同時に起きています。国産開発陣は2026年夏〜秋の推論モデル開発において、数学・論理推論にとどまらず、OpenAIのオープンモデル「gpt-oss-120b」による思考・行動軌跡の蒸留や、NVIDIAの「Nemotron-Agentic」系データセット、クラスター上のOpenSearchを用いた疑似検索環境での強化学習を駆使し、複数ターンのFunction Calling、並列ツール呼び出し、Web検索・Deep Research機能を同時に実装しました[9]。その結果、海外で先行普及した初期エージェント能力（BFCL v3やτ^2-bench水準のツール利用、出典付き調査レポート生成）については、推論モデル完成と同時にキャッチアップを果たしています[9]。ところが、2026年11月時点の国内技術動向レポートの整理によれば、2026年9月〜10月にかけて海外主要プレイヤーは「長時間タスクの自律実行」へとさらに世代を進めました[18]。同レポートでは、Anthropicの「Claude Opus 5.5」（2026年9月22日リリース）がPC操作のOSWorld 2.1で81.8%（partial）、ターミナル操作のTerminal-Bench 4.0で66.4%を記録し、進捗をテキストファイルとgit履歴に残してセッション横断で実行を維持する二段構成ハーネスを提示したこと、またOpenAIの「GPT-6.1 Terra」（同年10月26日リリース）が長時間評価Agents' Last Examで前世代比+15ポイント・コスト約3分の1・サンドボックス事業者7社連携の耐久実行を標準装備したことが報告されています[18]。これに対し、国産フルスクラッチ勢は300億（30B）パラメータ前後の軽量・中規模帯が中心であり、SB Intuitionsが2026年10月末の公式ブログで「コーディングエージェントとして活用する用途に特化した学習が十分に実施できていない」と課題に挙げているとおり[9]、自律的なソフトウェア開発や長時間ターミナル操作では依然として12〜18カ月規模の差が残っています[18]。

第三に、それにもかかわらず日本市場のAIエージェント実装が全面的な周回遅れに陥っていない要因は、ハーネス・オーケストレーション・業務特化**CUA**（**Computer-Using Agent**）層における独自の実装アプローチにあります。単一モデルの巨大化競争と距離を置いた国内企業は、①人間のPC操作ログから明文化されていない業務の「暗黙知」を自動抽出してブラウザ操作を代行する技術（NECが2025年8月に開発発表し、2026年8月に販売開始した「NEC cotomi Agent」）[1]、②8Bの軽量視覚言語モデルで日本語PCアプリ操作に特化した国産CUA（カラクリが2026年4月に公開した「KARAKURI VL2」）[33]、③海外を含む複数フロンティアLLMを動的に束ねて指揮する学習型オーケストレーター（Sakana AIが2026年6月に提供開始した「Sakana Fugu」および最大8時間自律調査を行う「

Sakana Marlin」）[24]、④AIエージェント自身が失敗原因を分析してハーネス（動作・ツール制御）を自動改修する自己進化技術（富士通の「Kozuchi MAAF」やELYZAの「RSI Research」）[26]へと注力しました。このアプリケーション制御・オーケストレーション層においては、海外トレンドと**0〜6**カ月差のほぼ同時並行で商用サービスが展開されています[18]。

直近半年（**2026**年**5**月〜**11**月）の主要国産**LLM**・プレイヤー動向

2026年春から11月にかけて、国内主要プレイヤーは基礎研究フェーズを越え、エンタープライズ実装・ソブリンクラウド提供・エージェント基盤統合へと明確に軸足を移しました[1]。主要各社の直近半年の動き（直前の3〜4月の前提マイルストーンを含む）は以下のとおりです。

開発元・陣営	直近半年（**2026**年**5**月〜**11**月中心）の主要リリース・動向	モデル規模・方式	推論（**Reasoning**）・**AI**エージェント対応の到達点（各社発表ベース）	ライセンス・提供形態
Preferred Networks (PFN)	・6月：PLaMo 3.0 Prime正式提供[4] ・8月：さくらのAI Engine対応[36] ・10月：Microsoft Foundry対応、PLaMo 3.1 Prime提供開始[1]	フルスクラッチ 256K文脈長[1]	3.1で単一モデルによる推論／非推論ハイブリッド化を実現（推論速度最大1.7倍）[10]。Web検索・**Deep Research**・構造化出力に対応し、幻覚関連8指標で平均約15%改善[3]。	商用API（Freeプラン有）、Microsoft Foundry、さくら基盤[1]
SB Intuitions (ソフトバンク)	・6月：Sarashina3 mini / nano / guard商用提供[1] ・10月13日：無償β版	フルスクラッチ （30Tトークン事前学習、パラメータ非公開）[1]	初の思考モデル3.2で3段階のreasoning_effort制御を導入[9]。検索・安全性の強化学習により開発元評価の**J-tau**・**BFCL**	商用API（ソブリンクラウドTypeA）、無償Web/API β版[9]

	Sarashinaサービス公開[9] ・10月29日：Sarashina3.2 mini thinking発表[8]		**v3**で比較モデル中最高値[9]。OpenCode / Hermes Agent接続対応[9]。	
NII（国立情報学研究所）	・（4月：LLM-jp-4 8B / 32B-A3B公開[39]） ・8月：LLM-jp-4 33B（Dense）公開[41] ・9月28日：LLM-jp-4.1シリーズ公開[6]	フルスクラッチ 8B / 33B (Dense)、32B-A3B (MoE)[12]	4.1で全モデルThinking対応[12]。5段以上の連続・並列Tool Callingを学習し、日本語マルチターン関数呼出（BFCL-nejumi）でQwen3-32Bを上回る[12]。Nejumi 4で国産オープン首位[42]。	オープンソース（学習データ・GGUF量子化版も公開）[12]
ELYZA (KDDIグループ)	・10月2日：ELYZA-Thinking-1.0-llm-jp-4（33b / 32b-a3b）無料公開[44] ・10月2日：ELYZA RSI Research設立、Voice Agent開発[1]	純国産LLM-jp-4ベース 33B (Dense) / 32B-A3B (MoE)[44]	Mid-training（約63Bトークン）＋SFT＋RLを人手1〜2人日で回す自動化基盤を構築[11]。RSI研究によりエージェントハーネスを自律改善し、応対完遂率+18.6ptまたは推論コスト67%削減を実証[27]。	Apache 2.0（Hugging Face無償公開・商用可）[13]
NTT / NTTデータ	・5月：tsuzumi 2 Vision発表、Microsoft	フルスクラッチ 30B（1GPU・	複雑な日本語図表・帳票読解を1GPUで実現	オンプレミス／専有環境、Microsoft

	Foundry提供開始[13] ・8月：統合AI基盤LITRON機能拡充[48] ・11月：少量業務資料からの知識定着技術発表[49]	40GB動作）[1]	[13]。エージェント実行はNTTデータのLITRONと連携し、固定資産判定業務3,000件で正答率99.6%と報告[18]。	Foundry、さくら基盤[13]
NEC	・8月：NEC cotomi Agent販売開始[1] ・9月：cotomi Security for Industry発表[1] ・11月：cotomi Gateway提供開始[13]	フルスクラッチ （cotomi v3はエージェントプロトコル対応）[52]	人間の操作ログから暗黙知を学ぶcotomi Actを製品化（WebArena成功率80.4%で人間平均超え）[19]。11月には機密伏せ字・プロンプト圧縮・LLM振り分けを行うGatewayを投入[13]。	Microsoft Marketplace（月額1.5万円/人＋専用テナント費）等[1]
富士通	・5月：Takane自動強化・自己進化エージェント技術発表[54] ・7月：Kozuchi Multi AI Agent Framework (MAAF)発表[55] ・11月：Google Cloud提携拡大（Cohereは日本法人計画発表）[56]	Cohere共同開発 （Takane 32B等）[13]	専有型基盤Kozuchi Enterprise AI Factoryと自己進化マルチエージェント基盤MAAFを展開[55]。中央省庁パブコメ実証や法令整合性チェック（報道で精度8割超）に実績[1]。	プライベート環境、Nutanix Enterprise AI、さくら基盤[1]

Sakana AI	・6月：自律調査AISakana Marlin、オーケストレーターSakana Fugu提供開始[25] ・9月：住友商事・SCSKと3社提携[59] ・10月：Sakana Aquariumおよび独自基盤Sakana Sawara計画発表[60]	複合オーケストレーション＋初のフルスクラッチ（Sawaraは年末に開発・学習開始計画）[35]	最大8時間自律調査を行うMarlin、複数フロンティアLLMを束ねるFugu Max / Ultra v2、オントロジー基盤Aquariumを展開[25]。エージェント・ネイティブ設計のSawaraを2027年春公開目標で計画[59]。	商用プロダクト／API（Sawaraは2027年春ウェイト公開目標）[25]
リコー	・5月：セーフガードモデル・推論ベンチマーク無償公開[64] ・6月：リコーLMM_27B / 9B発表[65] ・9月：オンプレAIエージェント提供[64] ・10月15日：MoE型リコーLMM_35B-A3B開発発表[67]	マルチモーダル 27B / 9B (Dense)、35B-A3B (MoE)[65]	日本語の複雑な図表読解と段階的リーズニング、日本語エージェント性能（ツール利用・ループ抑制）を強化しつつ応答時間を約3分の1〜4分の1に短縮[64]。Hermes Agent搭載[66]。	オンプレミス（11月よりRICOH オンプレLLMスターターキット搭載）[69]
楽天グループ	・（3月：約700B MoERakuten AI 3.0公開[1]） ・7月：日本HPとRakuten AI for	DeepSeek-V3-0324継続事前学習 約700B MoE（稼働約40B）/	PCローカル動作へ展開する一方、11月の自主開示でベースがDeepSeek-V3	Apache 2.0（無償ダウンロード・商用可）、HP製PC搭載[1]

	Desktop（7B）提供開始[1] ・11月：モデルカード更新・3.0.1公開・ベースモデル自主開示[70]	7B[1]	であることを明記し、「推論と数学はベースのほうが強い」と公式記載（無害応答率は80.0%→93.2%に改善）[1]。	

1. フルスクラッチ純国産勢の「ハイブリッド推論×検索・ツール統合」への進化

直近半年の技術面で特に注目されるのは、フルスクラッチ開発を貫く国内勢が、思考モデルの実装にとどまらず、推論時の計算資源コストとレイテンシという実運用上の課題まで解消し始めた点です。PFNの公式技術ブログによれば、2026年6月の「PLaMo 3.0 Prime」時点では推論モデルと非推論モデルを別々に開発・デプロイしていたため、GPUサーバー配備に2倍の計算資源を要していました[10]。しかし、4カ月後の10月27日に提供開始した「PLaMo 3.1 Prime」では、NVIDIAのNemotron 3 Superの手法を参考に推論過程の有無を混ぜて事後学習させることで、単一モデルがリクエストに応じて思考の有無を切り替える「ハイブリッド推論モデル」を実現し、推論速度を最大1.7倍に高めています[10]。さらに、根拠不在時に推測せず「答えられない」と応答する合成データ学習により8つのハルシネーション関連ベンチマークで平均約15%の改善を達成し、Web検索を繰り返して出典付きレポートを生成する「Deep Research」機能も実装しました[10]。

オープンソース陣営では、NIIの「LLM-jp-4」とKDDI傘下ELYZAの連携が国産エコシステムの広がりを示す成果となりました。NIIが8月18日に約332億パラメータのDense型「LLM-jp-4 33B」を公開し[41]、9月28日に事後学習を見直した「LLM-jp-4.1」シリーズを公開すると[6]、その4日後の10月2日にはELYZAがLLM-jp-4（33B Denseおよび32B-A3B MoE）を基盤とした推論モデル「ELYZA-Thinking-1.0-llm-jp-4」2種をApache 2.0ライセンスで無償公開しています[44]。ELYZAの発表によれば、約63Bトークン規模のMid-training（事前追加学習）から数百万件規模の自動検証付きSFT、約10万件の強化学習（RL）に至る一連の工程を、人間の介在作業**1〜2**人日程度に抑えて実行する社内学習基盤が構築されました[11]。同様にSB Intuitionsも、6月の非推論モデル公開から4カ月後の10月29日に「Sarashina3.2 mini thinking」を発表し、オンポリシー自己蒸留（OPSD）と検索・安全性それぞれの強化学習を組み合わせることで、思考プロセスと最終出力（JSONや表形式など）を分離して出力できる構造を実現しています[9]。

2. SI・通信・OA大手による「現場の暗黙知・非構造化帳票」攻略と流通網の拡大

企業システムへの実装を担う国内大手各社は、1GPU〜少数のオンプレミス環境で稼働する省リソース性と、日本企業特有の「複雑な帳票・図表」および「属人化した業務手順」の自動化に集中投資しています。NTTは5月19日に「tsuzumi 2 Vision」を発表し、源泉徴収票からの金額抽出やチェックボックス付き問診票の構造化といった図表入り日本語ビジネス文書処理を1GPU（40GB）環境で実現しました[13]。さらに11月12日には業務マニュアルなどの限られた資料から特定業務知識をLLMに定着させるチューニング技術を発表しています[49]。また、NTTデータの「tsuzumi 2」（5月20日開始）[47]、NECの「cotomi Agent」（8月18日開始）[1]、PFNの「PLaMo 3.0 Prime」（10月6日開始）[1]が相次いでMicrosoftのクラウド基盤（Microsoft Foundry / Marketplace）や「さくらのAI Engine」（PLaMoは8月4

日開始）[36]上に並び、既存のクラウド調達経路から国産モデルやエージェントを導入できる流通環境が整いました。
セキュリティとガバナンスを軸とした周辺機能の製品化も進んでいます。NECは全社11万人規模での自社実践（「クライアントゼロ」）を経てブラウザ操作エージェント「NEC cotomi Agent」を外販したのに続き[1]、9月2日には脆弱性分析や仮想パッチ生成を支援する「cotomi Security for Industry」[1]、11月5日には機密情報の自動伏せ字化・プロンプト圧縮・問い合わせ先LLMの振り分けでトークン費用を抑える「cotomi Gateway」の提供を発表しました[13]。リコーも10月15日にMoE型マルチモーダルモデル「リコーLMM_35B-A3B」を発表し、図表読解・日本語リーズニング・エージェント応答のループ抑制を両立させつつBF16版比で応答時間を約3分の1（MTP併用時は約4分の1）に短縮し、11月から順次オンプレミス向けスターターキットへ搭載しています[64]。

3. Sakana AIの「エージェント・ネイティブ」転換と、楽天モデルが示した「国産」の透明性課題

スタートアップやネット大手の動向からは、次世代アーキテクチャと「国産AI」の定義をめぐる対照的な動きが見られます。創業以来「進化的モデルマージ」や既存モデルの事後学習（Namazu等）、「単一巨大モデル競争とは異なる集合知アプローチ」を掲げてきたSakana AIは[59]、2026年6月に自律型リサーチAI「Sakana Marlin」と複数フロンティアLLMを動的に束ねる「Sakana Fugu」を提供開始し[25]、8月の防衛省実証受注や9月の住友商事・SCSKとの3社包括業務提携を通じて産業・安全保障領域への実装体制を固めました[59]。そのうえで10月23日にオントロジー（業務概念の関係性定義）を基盤とするエージェント運用プラットフォーム「Sakana Aquarium」を発表し[61]、続く10月26日には同社のRSI Labにおいてゼロから事前学習を行う初のフルスクラッチ「エージェント・ネイティブ」基盤モデル「Sakana Sawara（仮称）」の開発計画を公表しました[59]。既存モデルへの後付けでは限界があるエージェント動作を事前学習段階から織り込む設計であり、2026年11〜12月に開発・小型モデルの事前学習を開始し、順調にいけば2027年春にウェイトを公開する条件付き目標を掲げています（現時点でパラメータ数やライセンスは非公表）[59]。
一方で、「国産AI」の情報開示と透明性に議論を呼んだのが楽天グループの事例です。同社は経済産業省GENIACプロジェクトの一環として、2026年3月17日に国内最大規模の約7,000億（700B）パラメータMoEモデル「Rakuten AI 3.0」をApache 2.0で公開し、当初はベースモデルへの言及がないまま日本語版MT-Bench（8.88）での「GPT-4o（8.67）超え」をアピールしていました[1]。しかし情報公開の姿勢や公的支援の意義をめぐる批判が起き[70]、専門メディアの整理によれば2026年11月にモデルカード更新・「3.0.1」公開・公式ブログによる自主開示が行われました[70]。これによりベースモデルが「DeepSeek-V3-0324」であったことが明示され、公式ベンチマークの比較相手もGPT-4oからベースモデル比へと改められたうえで、楽天評価による無害な応答率（80.0%→93.2%）や日本文化知識が向上した一方、「推論と数学はベースのほうが強い」と自ら記載しています[70]。継続事前学習アプローチ自体は、東京科学大学・産総研の「Swallow」プロジェクト（2026年2月にQwen3版・GPT-OSS版をApache 2.0公開、5月にFIXERと社会実装連携、11月時点で累計323万ダウンロード達成）[76]のように最初からベースと手法を明示すれば高く評価される一方[79]、「フルスクラッチ開発」と「海外オープンモデルの日本語適応」の境界を曖昧にした訴求は厳しい検証を受ける段階に入っています[30]。なお、このほかフルスクラッチで1,000億パラメータの「Stockmark-LLM-100b」を開発するストックマークと産総研の共同発表（6月）[80]や、AI insideと橋本ホールディングスによる独自開発LLMを基盤とした建設・地域産業DX合弁会社設立検討のMOU締結（11月17日）[82]など、産業現場向けの展開も広

がっています。

ベンチマークと実運用データが示す現在地：海外フロンティアとの距離

定性的なリリース動向の裏にある定量的な性能を、公開リーダーボードおよび各社公表のベンチマークから検証すると、国産LLMと海外フロンティアの「得意領域の分岐」と「中規模モデルの容量トレードオフ」が浮かび上がります。

国産勢は業務実行と検索レポートで急伸する一方、総合・コーディング頂上決戦では海外勢が先行

LAYER 1 実業務適用・自律操作エージェントの到達精度

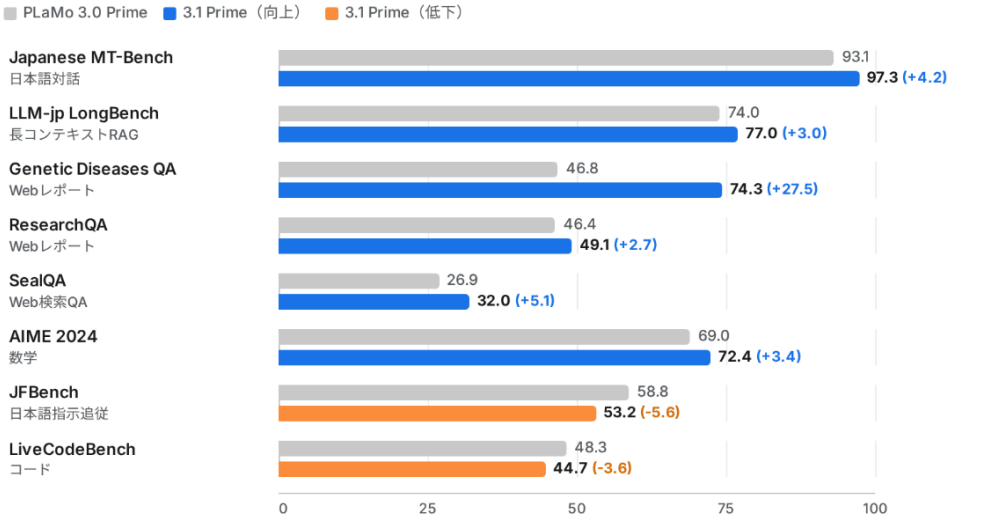
単位：タスク成功率・業務正答率（%）

国内商用エージェント	国内エンタープライズ実運用	海外最前線ベンチマーク参照
80.4%	**99.6%**	**81.8%**
NEC cotomi Agent	**NTTデータ LITRON**	**Claude Opus 5.5**
WebArena成功率・人間平均78.2%超え	固定資産判定3,000件の業務正答率	OSWorld 2.1 PC自律操作スコア（partial）

LAYER 2 国産フルスクラッチの進化とトレードオフ：PLaMo 3.0 vs 3.1 Prime（推論モード）

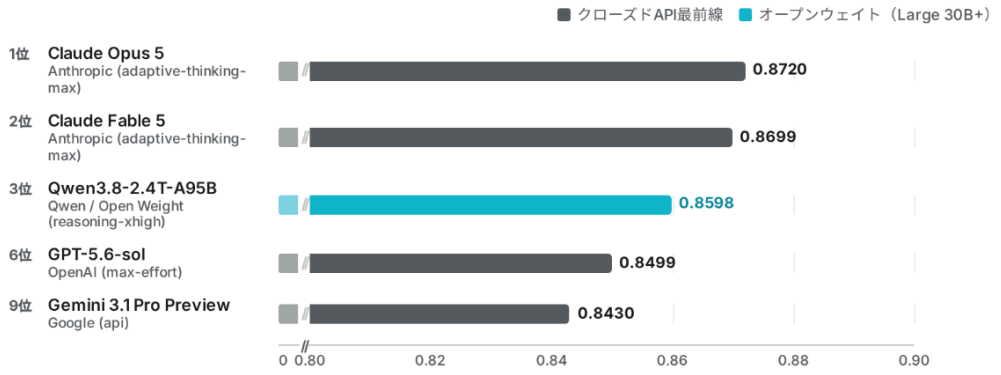
スコア範囲：0〜100

すべて表示 | 対話・長文RAG | Web検索・Deep Research | STEM・コード・指示追従



LAYER 3 日本語総合性能の最高到達点：Nejumi Leaderboard 4 上位モデル（2026年10月）

総合スコア（0.80〜0.90 拡大表示）



国産モデルの立ち位置：国立情報学研究所（NII）の「llm-jp-4.1（33B版）」が日本語ランキングに初登場し国産オープンモデル首位を獲得したが、総合TOP10（0.8425以上）は海外フロンティア・超大規模オープン勢が占める。

上段はエージェント実行・業務適用の代表指標、中段はPFN公表のPLaMo 3.0 Prime（2026年6月）と3.1 Prime（同年10月・推論モード）のカテゴリ別比較、下段は2026年10月時点のNejumi Leaderboard 4総合上位スコア。国産フルスクラッチ勢はWeb検索統合や対話で大幅に伸びたが、単一モデルでのコーディングや超大規模フロンティアとの総合力には依然として差がある。

Data sources: NEC 技術・ソリューション公開情報, AIエージェント動向調査（note）, Preferred Networks Tech Blog, Qualiteg Blog（Nejumi Leaderboard 4解説）

日本語総合性能の指標である「Nejumi Leaderboard 4」の推移を見ると、調査レポートが引用する2025年12月時点では上位50モデル中に国産LLMが1モデル（rinnaのqwq-bakeneko-32b、50位・0.6910）しか入らず、当時の首位GPT-5.2（0.8285）と約0.13ポイントの開きがありました[83]。それから10カ月後の2026年10月22日時点の同ランキング分析記事によれば、首位「Claude Opus 5」（0.8720）、2位「Claude Fable 5」（0.8699）、3位に2.4兆パラメータ級のオープンモデル「Qwen3.8-2.4T-A95B」（0.8598）が続き、6位「GPT-5.6-sol」（0.8499）、9位「Gemini 3.1 Pro Preview」（0.8430）と、海外勢の天井自体が0.84〜0.87台へ切り上がっています[42]。このなかでNIIの「LLM-jp-4.1（33B版）」が初登場で国産オープンモデルの首位に立ったものの（ランキング概要記事では全体の具体的な順位番号・総合スコア値は非掲載）[42]、超大規模な海外フロンティア群との総合スコア差は依然として残っています[42]。

他方で、AIエージェントの実務に直結する「日本語でのツール呼び出し（Tool Calling）」と「検索統合レポート生成」に焦点を絞ると、30Bクラスの純国産モデルが同規模の海外モデルを上回る局所的な逆転も見られます。NIIの公式ブログによれば、英語の関数呼び出し評価（BFCL）ではQwen3シリーズと同等以下だった「LLM-jp-4.1-33B」が、日本語のマルチターン関数呼び出し（BFCL-nejumi）ではQwen3-32BおよびQwen3.5-9Bを上回る成績を記録しました[12]。SB Intuitionsの公式ブログでも、「Sarashina3.2 mini thinking」がエージェント評価の「J-tau」と「BFCL v3」、およびマルチターン安全性の「JMT-Safety」で比較モデル中の最高スコアを達成したと報告されています[9]。またPFNの発表数値では、「PLaMo 3.1 Prime」（推論モード）がWeb検索レポート生成（Genetic Diseases QAで46.8→74.3、ResearchQAで46.4→49.1）やJapanese MT-Bench（93.1→97.3）で大きくスコアを伸ばしました[10]。

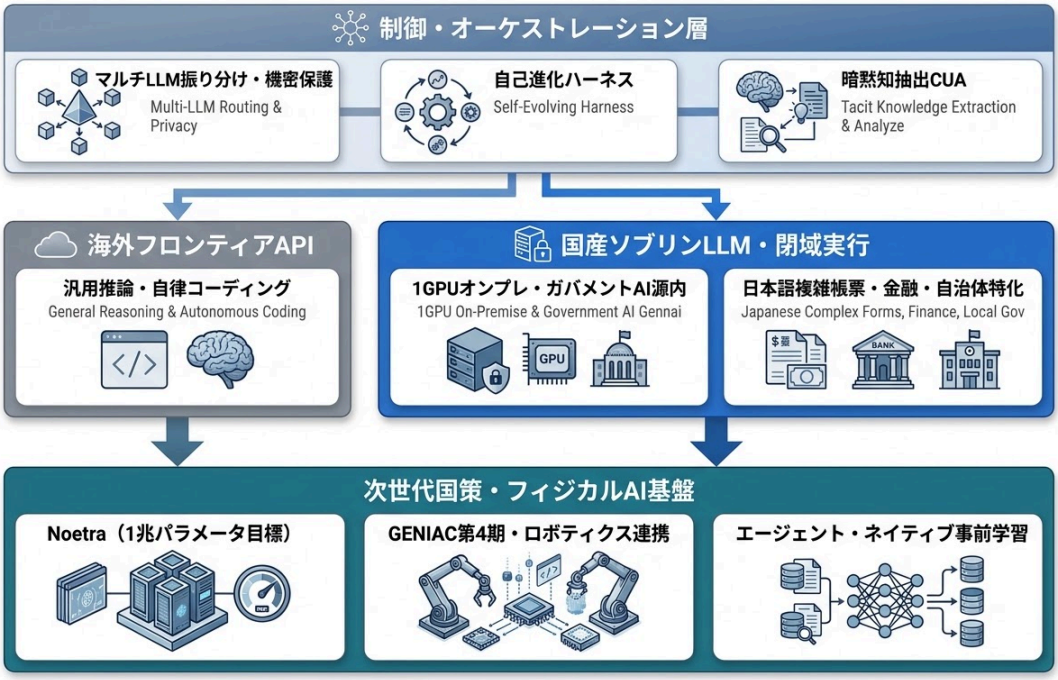
しかし同時に、開発各社の技術報告からは「中規模パラメータゆえの能力トレードオフ」という構造的制約も確認できます。PFNの公表値では、PLaMo 3.1 Primeがハイブリッド化と事実性・Web検索強化学習を優先した結果、推論モードにおける日本語指示追従（JFBench：58.8→53.2、IFBench：72.2→64.8）やコーディング（LiveCodeBench：48.3→44.7）のスコアが3.0 Primeから低下しています[10]。SB Intuitionsの評価でも、Sarashina3.2 mini thinkingは前世代から日本文化QA（JamC-QA）のスコアが低下したほか、STEM推論（GPQA-Diamond）やマルチターン対話（J-MultiChallenge）では「GPT-5.4 mini」や「Qwen3-235B-A22B-Thinking-2507」に及ばないことが開示されています[9]。またコーディングエージェント領域では、2026年2月に国産AIエージェント基盤「JAPAN AI Code」がSWE-bench Verifiedでの成果を発表した例はあるものの[85]、2026年秋時点の国際的な最前線はすでに旧ベンチマークの飽和を経て「SWE-bench Pro」や「Terminal-Bench 4.0」といった高難度評価へ移行しており[22]、8B〜33B中心の純国産基盤モデル単体では検索・安全性・ツール利用・高度コーディングのすべてを同時に極大化しにくいのが実情です[9]。

これに対し、視覚・**GUI**操作や業務システム統合を伴う国内特化エージェントでは高い実用精度が報告されています。カラクリの発表によれば、GENIAC第3期で開発しオープンソース公開した8Bの国産CUAモデル「KARAKURI VL2」は、独自の日本語PC操作ベンチマーク「OSWorld-JP v0.2」（100タスク）において、画像編集（GIMP操作）で「Claude Sonnet 4.6」の約2倍のスコアを記録し、メール操作でも同モデルを上回ったほか、複数アプリの横断操作スコアをベースモデル比2.8倍に向上させました[34]。また、NECの公式発表では「cotomi Act / NEC cotomi Agent」が国際ベンチマークWebArenaで人間平均（78.2%）を上回る80.4%のタスク成功率を達成したとされ[19]、国内の技術動向レポートでも、NTTデータの統合AI基盤「LITRON」による固定資産の耐用年数判定3,000件での正答率99.6%（年間作業時間を352時間から10時間へ削減）や、三菱UFJ銀行が2026年4月に導入した

ノーコードエージェント基盤「Jinba」による海外送金データチェック精度90%・帳票処理時間の短縮（90分→15分）といった業務成果が挙げられています[18]。なお、海外では2026年8月にエージェント間連携規格「A2A（Agent2Agent）」が外部ツール接続規格「MCP（Model Context Protocol）」と同じLinux Foundation内の「Agentic AI Foundation」に合流し標準化が進んでいますが[91]、国内でもNECが公式発表で「エージェントプロトコルへの対応」を掲げ（2026年の関連資料でもMCPサーバやAgent API・BYOMを含む統合基盤を提示）[52]、SarashinaやリコーがOpenCodeやHermes Agentといった標準エージェント実行系への接続を整えるなど、プロトコル面の追随は速やかに進んでいます[9]。ただし、国内主要エージェントの多くはOSWorld 2.1、Terminal-Bench 4.0、SWE-bench Proといった2026年秋時点の国際標準ベンチマークでの直接比較スコアを公開しておらず、独立した第三者評価が限られる点には留意が必要です[18]。

国家プロジェクトと産業構造の転換：「源内」「GENIAC第4期」「Noetra」

直近半年のもう一つの大きな変化は、AI新法（人工知能関連技術の研究開発及び活用の推進に関する法律）や2025年12月閣議決定の「人工知能基本計画」（5年間で約1兆円規模の公的支援構想）を背景とした政策群が、2026年5月～8月にかけて具体的な政府実証・大型開発事業として一斉に動き出したことです[2]。これにより、国内におけるAIエージェントの実装アーキテクチャは「海外フロンティアとのハイブリッド共存」と「フィジカルAIへの長期投資」を組み合わせた三層構造へと移行しつつあります[16]。



1. デジタル庁ガバメントAI「源内（GENAI）」の大規模実証と「世代ギャップ」の教訓

政府自らが初期需要をつくるアンカーバイヤーとして動き出したのが、全府省庁職員約18万人を対象とする生成AI環境「源内」です[14]。公募に応じた15件から7件の国内モデルが選定された後、2026

年5月にNTTデータ、ソフトバンク、NEC、富士通、PFNの5社と契約して大規模実証が始まり、7月10日の発表を経て8月頃から「さくらのクラウド」上で「tsuzumi 2」「Takane 32B」「PLaMo 2.0 Prime」の3モデルの先行試用が開始されました（ソフトバンクとNECのモデルは一部省庁で試用）[13]。2026年度中は各社が無償提供しガバメントクラウド・推論費用をデジタル庁が負担したうえで、2027年1月頃に評価結果の一部を公表し、同年4月から実証されたモデルを有償で政府調達する予定です[98]。
もっとも、ここには行政調達のサイクルとモデル進化速度のギャップという課題も表れています。「源内」で契約・試用されているPFNモデルは2025年5月リリースの「PLaMo 2.0 Prime」である一方、民間企業は2026年10月時点でハイブリッド推論・Deep Research対応の「PLaMo 3.1 Prime」を利用可能となっており、Sarashinaも同様に官民で利用できる版に差が広がりました[13]。また、初回選定時の評価テストが「当日初見の50問」だった点についても見直しが進み、報道（ITmedia AI+、2026年5月29日）によれば、次回選定では「法律・制度や行政課題など8分野・35能力領域にわたる300問（上限15時間）」へ拡大し、初回契約5社以外の応募も認める方針へ改められています[13]。

2. GENIAC第4期と「Noetra」によるフィジカルAI・1兆パラメータ級への投資

経済産業省・NEDOの支援策は、テキスト中心のLLM開発支援から「現場データ（暗黙知）の構造化」と「ロボティクス（フィジカルAI）」へと明確に対象を広げました[97]。2026年3月に懸賞金総額約8億円・参加企業150社超の「GENIAC-PRIZE」から応用エージェント事例が創出されたのに続き[101]、2026年6月4日に計算資源支援枠16件が発表された「GENIAC第4期」では、7月2日採択のデータエコシステム構築等9件や製造業データのAI-Ready化枠、ロボット基盤モデル枠と連動した開発が始まっています[100]。業界メディアの解説でも、産業用ロボットに強みを持つ日本の産業構造を背景に、味の素や伊藤忠商事などの現場に蓄積された熟練者の「暗黙知」をAIエージェントが扱える形式知へ変換し、将来のフィジカルAIへ還元する狙いが指摘されています[97]。
その国策フラッグシップとして始動したのが、2026年4月に「日本AI基盤モデル開発」として報じられ、6月30日のNEDO事業採択を経て7月16日に研究開発体制を発表した新会社「Noetra（ノエトラ）」です[2]。ソフトバンク、ソニーグループ、NEC、本田技研工業の4社が中核出資し産業技術総合研究所（産総研）と連携する同事業は、報道によれば初年度（2026年度）約3,873億円の公的支援を受け（5年間で約1兆円支援構想の一環だが確約ではなく、毎年度のステージゲート審査で継続判断）、1兆パラメータ規模を目標とする国産オムニモーダル・フィジカルAI基盤モデルの開発を目指します[15]。ただし、公式発表に基づくNoetra自身の計算基盤は2027年4月構築開始・2028年6月稼働予定とされており、2026〜2027年は国内事業者の既存計算基盤を借りて開発を進める計画です[104]。海外フロンティアが数カ月単位で世代交代するなか、専用計算基盤の本格稼働までに約2年を要するスケジュールは、稼働時点で再び世代差が開くリスクも伴うため、毎年度のステージゲート審査における進捗検証が重要となります[15]。

結論と実務への示唆：日本企業が採るべきAIエージェント戦略

以上の調査結果を踏まえると、2026年11月時点における国産LLMおよびAIエージェントの現在地と、ユーザー組織が採るべき実務上の意思決定指針は以下の3点に集約されます。

1. 「モデル単体の汎用自律性能」と「業務システム完遂力」を切り分けて評価する 純国産フルスクラッチLLMは直近半年で推論モデル化と基本的なツール利用・検索統合へのキャッチアップを果たしましたが、ソフトウェア開発の自律実行や数時間に及ぶ汎用マルチステップ操作においては、海外フロンティアモデルに対し依然12〜18カ月規模の遅れが残っています[9]。一方で、既存SaaSや社内レガシーWebシステムの自動操作、日本語の複雑な非定型帳票読解、コー

ルセンター応対といった業務領域では、NEC（cotomi Agent）、NTTデータ（LITRON＋tsuzumi 2 Vision）、富士通（Kozuchi MAAF＋Takane）、カラクリ（KARAKURI VL2）、リコー（LMM_35B-A3B）などがハーネス設計と暗黙知抽出によって高い実用精度を報告しています[13]。「国産モデルは総合ランキングで海外最上位に届かないためエージェント活用も時期尚早」と判断するのは早計であり、対象業務の性質に応じた適合性で評価すべき段階にあります[3]。

2. ゲートウェイ・オーケストレーション前提の「ハイブリッド共存」を標準設計とする 単一のモデルですべての社内エージェントを賄うのではなく、ルーティング・ガバナンス層（cotomi GatewayやSakana Fugu / Aquarium、Microsoft Foundry等）を上位に配置し[1]、高度な汎用推論や自律コーディングには海外フロンティアAPIを割り当て、機密性の高い閉域データ処理・高頻度の定型判断・日本語固有の帳票処理・行政関連実務には1〜2GPUや国内ソブリンクラウドで専有運用できる国産モデル（tsuzumi 2、PLaMo 3.1 Prime、Sarashina3.2、LLM-jp-4.1 / ELYZA-Thinking等）を割り当てる「適材適所のマルチモデル構成」が合理的な選択肢となります[1]。特にPLaMo 3.1 Primeの単一ハイブリッド推論化やSarashina3.2・LLM-jp-4.1のreasoning_effort制御の実装により、タスク難度に応じて思考トークン消費を動的に抑える運用が国産モデルでも現実的になっています[9]。
3. 「国産」の内実（学習経路の透明性）と自社業務**Evals**（評価基盤）を保持する 2026年11月のRakuten AI 3.0をめぐる自主開示が浮き彫りにしたとおり、「国産LLM」と称されるモデルであっても、事前学習データまで管理されたフルスクラッチ開発（LLM-jp、PLaMo、Sarashina、tsuzumi、cotomi、Stockmark等）なのか、海外オープンウェイトの継続事前学習・事後学習なのかによって、法的透明性・ライセンス条件・推論特性は異なります[30]。また、ベンダーが提示する旧世代の海外モデル（GPT-4o等）との比較数値のみに依拠せず[70]、自社の実データと業務手順に基づく評価セット（Evals）や異常時に人間へ処理を戻す停止・承認フローをユーザー企業側で設計・保持することが、AIエージェント導入の成否を左右します[19]。

引用文献

1. 国産LLMを徹底比較！各モデルの特徴と選び方【2026年最新版】, https://spikestudio.jp/blog/japan-llm-companies
2. Noetraとは？1兆円規模の国産フィジカルAI連合【2026】- Arpable, https://arpable.com/artificial-intelligence/robot/noetra-physical-ai-guide/
3. https://www.aib.vote/ja/news/preferred-networks-plamo-3-prime-interview
4. 国産生成AI基盤モデルPLaMo 3.0 Primeを正式リリース, https://www.preferred.jp/ja/news/pr20260622
5. 国産生成AI基盤モデルPLaMo PrimeにWeb検索・Deep Research搭載, https://www.preferred.jp/ja/news/pr20261027
6. 【解説】日本語LLM「LLM-jp-4.1」シリーズが公開 | Yoshi_it - note, https://note.com/natty_toucan386/n/nb4d0b4183ff3
7. ELYZA-Thinking-1.0とは？LLM-jp-4ベースの国産推論モデル | 33B, https://ai-revolution.co.jp/media/what-is-elyza-thinking/
8. ソフトバンクの国産LLM・AIモデル全45件を読む——Sarashinaの進化, https://note.com/kagawatomo/n/nb9b2a21a29b4
9. 国産LLM Sarashina3.2 mini thinking の紹介 - SB Intuitions, https://www.sbintuitions.co.jp/blog/entry/2026/10/29/133000

10. https://www.preferred.jp/ja/blog/tech/plamo-3-1-prime-release
11. https://zenn.dev/elyza/articles/79a3d4ed4be915
12. https://llm-jp.nii.ac.jp/blog/llm-jp-4-1/
13. 国産LLM 7選 徹底比較【2026年版】デジタル庁「源内」選定モデル, https://ai-revolution.co.jp/media/japan-llm-7-comparison/
14. 【2026年版】国産LLM7選 | デジタル庁「源内」が選んだAIモデルを, https://www.fidx.co.jp/%E3%80%902026%E5%B9%B4%E7%89%88%E3%80%91%E5%9B%BD%E7%94%A3llm7%E9%81%B8%EF%BD%9C%E3%83%87%E3%82%B8%E3%82%BF%E3%83%AB%E5%BA%81%E3%80%8C%E6%BA%90%E5%86%85%E3%80%8D%E3%81%8C%E9%81%B8%E3%82%93%E3%81%A0ai/
15. 日本AI基盤モデル開発（現Noetra）とは？SoftBank・NEC・Honda, https://ai-revolution.co.jp/media/what-is-japan-ai-foundation-model/
16. ソフトバンク/ソニー/NEC/ホンダら出資の「Noetra」が始動 産総研, https://robotstart.info/article/2026/07/01/382103.html
17. 国産AI開発を担うNoetra、マルチモーダル基盤モデルの ... - IT Leaders, https://it.impress.co.jp/articles/-/29594
18. https://note.com/tomo06103280/n/nb80956e8c32a
19. https://jpn.nec.com/rd/technologies/202609/index.html
20. 論文解説：DeepSeek-R1 - Zenn, https://zenn.dev/upgradetech/articles/0a546fa188fc26
21. DeepSeek-R1とは？料金・使い方・性能や日本語対応モデル, https://weel.co.jp/media/tech/deepseek-r1/
22. 2026年のコーディングに最適なLLM：9モデルをランキング, https://www.datacamp.com/ja/blog/best-llm-for-coding
23. Sakana Aquariumとは | サカナAIのAIエージェント基盤の仕組み, https://aigentlab.tech/articles/sakana-aquarium-ai-agent-platform-2026-10/
24. Sakana Fuguとは？Sakana AIのマルチエージェント統合AIモデルを, https://weel.co.jp/media/innovator/sakana-fugu/
25. Sakana AIが自律型リサーチエージェント「Sakana Marlin」の提供, https://www.sbbit.jp/article/cont1/185772
26. 富士通、自己進化マルチAIエージェント技術を組み込んだFujitsu, https://global.fujitsu/ja-JP/technology/research/article/topics/202607-multi-al-agent-framework
27. https://elyza.ai/news/2026/10/02/rsiresearch
28. SWE-Bench Proランキング 2026年10月 - QCode.cc, https://qcode.cc/ja/swe-bench-pro-2026-ranking
29. ELYZA-Thinkingとは？社外に出せない文書を無料の国産AIで扱う4条件, https://smart-eight.jp/blog/elyza-thinking-guide
30. 国産生成AI基盤モデル「PLaMo 3.0 Prime」とは | 2026年最新の, https://www.issoh.co.jp/tech/details/12701/
31. DeepSeek-R1に日本語で追加学習を施したLLMを公開 - gihyo.jp, https://gihyo.jp/article/2025/01/deepseek-r1-japanese
32. Sarashinaとは？SB Intuitionsの国産LLMの特徴・料金・使い方を, https://weel.co.jp/media/tech/sarashina/
33. カラクリ株式会社 の求人 | LLMソリューションアーキテクト / PM,

https://herp.careers/careers/companies/karakuri/jobs/8ncY4xpPGZ9O
34. カラクリ、日本企業初のComputer-Using Agent「KARAKURI VL,
https://prtimes.jp/main/html/rd/p/000000122.000025663.html
35. Sakana AI、複数のAIモデルを協調させて高いパフォーマンスを発揮,
https://gihyo.jp/article/2026/06/sakana-fugu
36. さくらインターネット、Preferred Networks、情報通信研究機構 - NICT,
https://www.nict.go.jp/press/2026/10/14-1.html
37. PLaMo 3.0 Primeとは？性能や料金、国産LLMとしての立ち位置を解説,
https://www.ai-souken.com/article/what-is-plamo-3-0-prime
38. PFN、Deep Researchを搭載した「PLaMo 3.1 Prime」を提供開始,
https://codezine.jp/news/detail/30154
39. 約12兆トークンの良質なコーパスで学習した新たな国産LLM「LLM,
https://www.nii.ac.jp/news/release/2026/0403.html
40. 国産LLM「LLM-jp-4」公開 - コプラス株式会社,
https://coplus-one.jp/llm-jp-4-release/
41. 新たな大規模言語モデル（LLM）「LLM-jp-4 33B」を公開,
https://current.ndl.go.jp/car/283665
42. https://blog.qualiteg.com/llm-ranking-2026/
43. K2 Horizonとは？MBZUAI/IFMの完全オープンAIモデル6種（0.9B,
https://ai-revolution.co.jp/media/what-is-k2-horizon/
44. ELYZA LLM-jp-4 とは | 商用利用と自社ホストの条件 - FIXIT,
https://fixit.co.jp/tips/elyza-thinking-llm-jp-4/
45. ELYZA、国産「LLM-jp-4」基盤の推論モデル 2種を無料 ... - Ledge.ai,
https://ledge.ai/articles/elyza_thinking_llm_jp_4_rsi_research
46. LLM「tsuzumi 2 Vision」にアップデート 世界トップクラスの図表,
https://k-tai.watch.impress.co.jp/docs/news/2109436.html
47. Microsoft Azure上で「tsuzumi 2」の提供開始、生成AIの検証環境を,
https://www.nttdata.com/global/ja/news/topics/2026/052001/
48. 業務プロセス全体の実行を実現する統合AI基盤「LITRON®」を拡充,
https://www.nttdata.com/global/ja/news/topics/2026/082501/
49. NTT、少量データで業務特化LLMを調整 | AIB,
https://www.aib.vote/ja/news/ntt-llm-tuning-limited-business-data
50. NTT版LLM tsuzumi 2アップデート ～世界トップレベルの図表入り,
https://group.ntt/jp/newsrelease/2026/05/19/260519a.html
51. NTTデータ、AIが複数システム横断で業務実行できる環境の構築,
https://enterprisezine.jp/news/detail/24554
52. NEC、「ガバメントAIで試用する国内大規模言語モデル(LLM)」に選定,
https://jpn.nec.com/press/202603/20260309_03.html
53. 国産マルチモーダル基盤モデルの開発に向けて本格始動 - ソフトバンク,
https://www.softbank.jp/corp/news/press/sbkk/2026/20260716_01/
54. 業務とともに学び続ける自己進化マルチAIエージェント技術を開発,
https://global.fujitsu/ja-JP/pr/news/2026/05/25-01
55. 富士通、“自己進化型”のマルチAIエージェント開発・運用基盤,
https://it.impress.co.jp/articles/-/29578
56. Cohere、日本法人の設立計画とソブリンAIへの投資拡大を発表,

https://jichitai-dx-navi.com/articles/98109ace-4ce9-4ae7-8816-9b8a1ac27fa8
57. 富士通×Google Cloud協業を拡大 - AI・セキュリティなど4領域で連携, https://www.excite.co.jp/news/article/Cobs_3248765/
58. 富士通、業務の変化に合わせて進化するAIエージェント技術を開発, https://www.itmedia.co.jp/enterprise/articles/2605/27/news028.html
59. 上場期待で沸く「Sakana AI」：独自技術から純国産AI開発まで注目, https://note.com/mecha_rhythm/n/n20ed8111f80c
60. 国産生成AI「Sakana Sawara」の開発を発表。2027年春に小規模, https://luminous-core.net/2026/10/sakana-sawara/
61. Sakana AIは日本のAI産業を変えるのか—2027年IPO観測 - note, https://note.com/kagawatomo/n/n4f59afa11247
62. Sakana AI、国産モデル「Sawara」開発へ Fuguとの違いと防衛, https://www.ai-toptier.com/blog-post/sakana-sawara-fugu
63. Sakana AIとは？M2N2・Namazu・Fuguと学術成果を解説【2026年】, https://arpable.com/artificial-intelligence/sakana-ai-innovative-ai-technology/
64. リコー、高速・省リソースを特徴とするMoE型マルチモーダルLLM, https://news.mynavi.jp/techplus/article/20261015-5132583/
65. MoE型マルチモーダル大規模言語モデル「リコーLMM_35B-A3B, https://jp.ricoh.com/release/2026/1015_1
66. 新着情報 | RICOH オンプレLLMスターターキット - リコーのAI, https://promo.digital.ricoh.com/ai/service/ricoh-on-premises-llm-starter-kit/topics/
67. リコー, https://jp.ricoh.com/-/media/Ricoh/Sites/jp_ricoh/release/2026/pdf/1015_1.pdf?rev=627e121392264054928de64f5f20df7f&sc_lang=ja-JP
68. “はたらく”を支えるリコーの大規模言語モデル（LLM）, https://jp.ricoh.com/technology/ai/LLM
69. リコー MoE型マルチモーダル大規模言語モデル「リコーLMM_35B, https://www.oalife.co.jp/new_p/9405/
70. 楽天Rakuten AI 3.0＝DeepSeek V3リブランド炎上のその後, https://ai-revolution.co.jp/media/rakuten-ai-3-deepseek-scandal/
71. 楽天、「GENIACプロジェクト」の一環として国内最大規模の新たな, https://corp.rakuten.co.jp/news/press/2025/1218_01.html
72. 日本から、フロンティアAIを - Sakana AI, https://sakana.ai/ja/
73. hardmaru - Tilt, https://tilt.now/profiles/hardmaru
74. Rakuten AI 3.0から考えたこと | kajitaka - note, https://note.com/kajitaka1201/n/n9465b1e782a7
75. Rakuten AI 3.0の性能と導入手順 | 日本語に強い商用利用モデルの実力, https://aiagent-navi.com/ai-agent/rakuten-ai-3-capabilities/
76. Swallow LLM — 日本語に強いオープンなLLMを、アカデミックな, https://swallow-llm.github.io/index.ja.html
77. FIXER、東京科学大学との共同研究で、日本語に最適化されたLLM, https://prtimes.jp/main/html/rd/p/000000142.000009536.html
78. 東京科学大学が日本語特化LLM「Swallow」シリーズをリリースした, https://note.com/humble_bobcat51/n/n07d1d5187874

79. Qwen3-Swallow & GPT-OSS-Swallow - Zenn, https://zenn.dev/tokyotech_lm/articles/fa56f10e51fd7d
80. ストックマーク、Noetra等が進めるフィジカルAI向け「国産マルチ, https://prtimes.jp/main/html/rd/p/000000424.000024407.html
81. LLMによるビジネスアイデアの評価の個別最適化手法を開発 - 産総研, https://www.aist.go.jp/aist_j/press_release/pr2026/pr20260625/pr20260625.html
82. なぜ、焼津の老舗ゼネコンが「日本発のLLM」を持とうとするのか。, https://prtimes.jp/main/html/rd/p/000000077.000134835.html
83. 2026 年 日本の国産 LLM 採用状況, https://yorozuipsc.com/uploads/1/3/2/5/132566344/d6a6cdf2a223a0433091.pdf
84. PFN、PLaMo 3.1 Primeを公開 | AIB, https://www.aib.vote/ja/news/pfn-plamo-3-1-prime-release
85. コード生成AI比較10選！Cursor・Copilot・Claudeから国産まで, https://japan-ai.co.jp/media/6907/
86. 国産AIエージェント基盤「JAPAN AI Code」、ソフトウェア開発, https://prtimes.jp/main/html/rd/p/000000142.000124536.html
87. "2026年のおすすめAIコーディングエージェント: 自律度と適合性で, https://resources.rework.com/ja/tools/ai-agents/best-ai-coding-agents-2026
88. KARAKURI LMは中小企業に向く? 国産AIチャットボットの料金, https://ai-keiei-lab.jp/ai/karakuri-lm/
89. 経産省「GENIAC」第3期で国産CUAモデル「KARAKURI VL2」を開発, https://karakuri.ai/news/karakuri_vl2
90. ローカル環境で動作する軽量CUAモデル「KARAKURI VL2」を, https://codezine.jp/news/detail/28681
91. 「2026年、AIはここまで来た」エンジニア視点の最前線, https://www.idnet.co.jp/column/page_433.html
92. 「賢いAI」を追う時代は終わった。2027年に向け - note, https://note.com/betaitohuman/n/n9a587e3341f0
93. 【2026年8月第4週】週刊AIニュースハイライト | Aidotters, https://www.aidotters.com/blog/weekly-ai-news/2026-08-23/
94. NEC国内AIモデル cotomi の知財実務適用評価, https://yorozuipsc.com/uploads/1/3/2/5/132566344/684cc675adc03ded5da6.pdf
95. 国産LLM、ここからが本番。| 小田 志門 - note, https://note.com/shimonoda/n/n683ed98ac514
96. AI for Science の推進に向けた 基本的な考え方について - 文部科学省, https://www.mext.go.jp/content/20251006-mxt_jyohoka01-000045188_04.pdf
97. GENIACとは？第4期が目指す「暗黙知の資産化」とAI戦略, https://aiagent-navi.com/generation-ai/what-is-geniac/
98. 国産LLMにはどんなものがある？主要モデルの比較と企業の選び方, https://exawizards.com/column/article/ai-article/japanese-domestic-llm/
99. ガバメントAIで試用する国内大規模言語モデル（LLM）の公募結果, https://www.digital.go.jp/news/10d55c63-b3e1-42b9-9cc5-93a06943ae0e
100. GENIACとは？第4期の採択企業16件・予算を解説【2026年版】, https://arpable.com/artificial-intelligence/geniac-japan-ai-sovereignty/
101. 経産省AIエージェント | GENIAC-PRIZE受賞と3条件 - vottia,

https://vottia.jp/meti-ai-agent-geniac-prize/
102. GENIAC-PRIZE, https://geniac-prize.nedo.go.jp/
103. GENIACとは？5つの支援枠や採択企業,
https://www.ai-souken.com/article/what-is-geniac
104. Noetra（ノエトラ）とは？国産フィジカルAI基盤モデルの全貌・44,
https://ai-revolution.co.jp/media/what-is-noetra/