国産 LLM の直近半年の動向と AI エージェント化の遅れ

― 推論モデル導入の「18 カ月遅れ」はエージェント化でも再現されているか ―

2026 年 11 月 20 日
Claude Fable 5.2

本書は 2026 年 11 月 20 日時点の公開情報に基づく。記述は「確認済み事実」と「分析・推定」を区別し、後者には（分析）（推定）と明記した。文中の上付き番号は文末の参考文献に対応する。

要旨

- 2026 年 5 月下旬〜11 月の半年間で、国産フルスクラッチ LLM に推論（思考）モデルが出揃った。PFN は「PLaMo 3.0 Prime」正式版（6 月）と「3.1 Prime」（10 月）、NII は「LLM-jp-4.1」（9 月）、SB Intuitions はシリーズ初の思考モデル「Sarashina3.2 mini thinking」（10 月）を公開した。
- 各社の重心は「軽量・ソブリン・オンプレミス」にあり、フロンティア性能の追随ではない。国の大型資金（Noetra、2026 年度 3,873 億円）は、フィジカル AI 向けのマルチモーダル基盤モデルに向かっている。
- 推論モデル導入の遅れは、フルスクラッチ基準で約 18 カ月（o1-preview：2024 年 9 月 → PLaMo 3.0 Prime β：2026 年 3 月）と確認できる。海外モデルに追加学習した派生モデルまで含めれば約 8 カ月である。
- （分析）エージェント化でも同種の遅れがある。Deep Research で約 20〜22 カ月、ブラウザ操作エージェントの製品化で約 19〜22 カ月であり、コーディングエージェントは海外の初出から 21 カ月を経ても国産基盤モデルでは未到達である。モデル層の遅れは縮まっておらず、やや拡大している。
- （分析）一方、複数モデルを束ねるオーケストレーション層（Sakana Fugu）や業務エージェント技術（NEC、富士通）では遅れは小さい。国産基盤モデルによるフロンティア級のエージェントは、早くて 2028 年以降と推定する。

1. 調査の目的と範囲

本書の問いは二つである。第一に、国産 LLM の直近半年の動向を整理すること。第二に、推論モデルの導入で生じた約 18 カ月の遅れが、AI エージェント化でも同様に生じているかを検証することである。

対象期間は 2026 年 5 月下旬から 11 月 20 日までとし、国内で開発された言語基盤モデルと、それを用いるエージェント技術を対象とした。海外モデルを束ねて提供する Sakana AI の「Fugu」は国産の基盤モデルではないが、エージェント化の一形態として比較対象に含めた。2026 年 4 月以前の出来事は背景としてのみ触れる。

遅れの月数は、海外における初出と国産における初出の「公表月の差」で算出した。国産側は原則としてフルスクラッチ開発のモデルを基準とし、海外モデルに追加学習を施した派生モデルは別に示す。月数は時期の差であり、性能水準が同等であることを意味しない点に留意されたい。

2. 直近半年の動向（確認済み事実）

2.1 Preferred Networks（PLaMo）

PFN は 6 月 22 日、3 月からの β 提供を経て「PLaMo 3.0 Prime」を正式リリースした。論理的思考を要するタスク向けの Reasoning モデルと、高速応答向けの Non-reasoning モデルを用途に応じて使い分ける構成で、API とオンプレミスの双方で利用できる[1]。コンテキスト長は 256K に拡張され、AI エージェントとしての利用も想定されている[2]。第三者による実機検証では、Tool calling に対応していること、公式ブログが関数呼び出しのベンチマーク（BFCL v4）による評価を示していることが報告されている[3]。

10 月 27 日には後継の「PLaMo 3.1 Prime」の提供を開始した。Web 検索に加え、調査計画の立案から出典付きレポートの作成までを数分〜数十分で自律的に実行する「Deep Research」を搭載し、オンプレミス環境では社内文書と連携した調査もできる[4,5]。同社は、ハルシネーション関連のベンチマークが 3.0 Prime 比で平均約 15%改善し、推論モードが 1.7 倍高速になったとしている[4]。11 月 9 日には、金融特化の「PLaMo 3.0 Fin」をオープンウェイトで公開した[6]。

2.2 SB Intuitions（Sarashina）

SB Intuitions は 6 月 30 日、フルスクラッチで開発した「Sarashina3」シリーズの提供を開始した。用途別の 5 モデルで構成され、ソフトバンクのソブリンクラウド「Cloud PF Type A」を通じて稼働する[7]。約 30 兆トークンで事前学習し、内製の合成データでエージェントタスクを学習させることで、複数ツールの同時呼び出しや連続利用に対応したとされる[8]。10 月 13 日には、6 モデルの API とチャットを試せる無償ベータ版「Sarashina サービス」を一般公開した[9]。

10 月 29 日には、シリーズ初の思考モデル「Sarashina3.2 mini thinking」を発表した。思考の深さを 3 段階で調整でき、コーディング、ツール呼び出し、Web 検索の性能が従来モデルより向上したという[10]。同社の技術ブログによれば、12 種のベンチマークのうちエージェント系の J-tau と BFCL v3 で比較モデル中の最高点を得た一方、GPQA-Diamond などでは一部の比較モデルに及ばなかった。また、コーディングエージェント用途については「特化した学習は十分に実施できていません」と明記している[11]。

2.3 NTT（tsuzumi）

NTT は 5 月 19 日、図表入りの日本語ビジネス文書の理解力を高めた「tsuzumi 2」の Vision モデルを発表した。tsuzumi 2 は 1GPU・約 40GB メモリで動作する軽量モデルで、利用形態はオンプレミスが約 6 割を占める[12]。この更新では、数値処理を含む論理推論能力も強化された[13]。11 月 12 日には、少量データから特定の専門性を持つ LLM へチューニングする技術を発表した[14]。

今後については、論理数学とコード対応を高度化した tsuzumi 2 を 2026 年度中にリリースする計画が示されており[15]、AI エージェント同士が自律的に連携する「AI コンステレーション」の開発も進めるとしている[16]。本調査で確認した範囲では、2026 年 9〜11 月のニュースリリースに、後継モデルや推論特化モデルの発表は見当たらなかった[17]。

2.4 NEC（cotomi）

NEC は 8 月 18 日、Web ブラウザを自動操作する AI エージェント「NEC cotomi Agent」を Microsoft Marketplace で販売すると発表した。価格は 1 ユーザーあたり月額 1 万 5000 円（税別）である[18]。基盤となるエージェント技術「cotomi Act」は 2025 年 8 月 27 日に発表され、国際ベンチマーク WebArena で人間（78.2%）を上回る 80.4%のタスク成功率を記録した[19]。同技術を

用いたソリューションは 2026 年 1 月から提供されている[20]。生成 AI「cotomi」自体は、2025 年 7 月にエージェント性能の向上と MCP 準拠を発表済みである[21]。

なお、cotomi Act が抽出した暗黙知は「cotomi など」で分析されると報じられており[22]、エージェントを駆動する LLM が自社の cotomi 単独であるかは確認できていない。

2.5 富士通（Takane）

富士通は 5 月 25 日、複数の AI エージェントが業務の実行結果や人のフィードバックから継続的に学習する「自己進化マルチ AI エージェント技術」を発表した。業務特化型 LLM「Takane」の自動的な強化を含む[23]。7 月 28 日には、Takane とマルチ AI エージェントを搭載する金融機関向け専有型 AI プラットフォームの開発開始を発表し[24]、2027 年 3 月の提供を予定する[25]。9 月 16 日には事業モデル「Uvance」の進化を発表し、自社 IP 由来の AI（Kozuchi、Takane）とフロンティア AI を組み合わせて提供する方針を示した[26]。なお、Takane は Cohere 社との共同開発モデルである[27]。

2.6 NII（LLM-jp）と ELYZA

国立情報学研究所（NII）は 9 月 28 日、オープンな国産 LLM「LLM-jp-4.1」シリーズ 3 モデルを公開した。LLM-jp-4 の事後学習を改善したもので、STEM 分野の性能向上に加え、ツール呼び出しに新たに対応した[28]。

KDDI グループの ELYZA は 10 月 2 日、この LLM-jp-4 を基盤とする推論モデル 2 種を商用利用可能な形で無料公開し、AI が自らの開発・改善を担う技術の研究組織「ELYZA RSI Research」を設立した[29]。同社は、AI エージェントの挙動を制御する「ハーネス」をほぼ自動で改善する取り組みも紹介している[30]。

2.7 Sakana AI（Fugu）

Sakana AI は 6 月 22 日、複数の AI を束ねて単一の OpenAI 互換 API から利用できる「Sakana Fugu」の提供を開始した。内部では、タスクに応じて専門モデルを自動的に選択・切り替える「マルチエージェントオーケストレーション」方式を採る[31]。公開時点では、Gemini 3.1 Pro、Claude Opus 4.8、GPT-5.5 といった海外のフロンティアモデルを作業役のプールとして抱える構成であった[32]。9 月 11 日にはコスト効率重視の「Fugu Max」と性能重視の「Fugu Ultra v2」

をリリースし[33]、Fugu Max は NVIDIA 製のオープンウェイトモデルを含むモデル群と連携する[34]。Fugu は自前の基盤モデルではなく、他社モデルの「指揮役」である点が他の国産 LLM と異なる。

2.8 政府・Noetra・源内

ソフトバンク、NEC、ホンダ、ソニーグループが中核となる国産 AI 開発会社は、2026 年 4 月に「日本 AI 基盤モデル開発」として設立され[35]、その後「Noetra」に改称した。経済産業省は 6 月 30 日に同社を支援対象として選定し、2026 年度に 3,873 億円を拠出、5 年間で総額 1 兆円規模を支援する方針である[36]。7 月 16 日には、44 社からの出資を受けて本格始動を発表した。計画では、AI エージェントや言語処理の中核となる「推論基盤モデル」を 2026 年度から順次構築し、2028 年度にオムニモーダル基盤モデルを開発、2030 年度に「実世界ネイティブ AI」の実現を目指す。NVIDIA Rubin GPU 約 2 万 7,500 基を搭載する計算基盤は、2027 年 4 月に構築を開始し、2028 年 6 月に稼働する予定である[37]。同計画に対しては、参加企業の本気度を疑問視する論評もある[38]。

デジタル庁のガバメント AI「源内」では、5 社（NTT データ「tsuzumi 2」、ソフトバンク「Sarashina3 mini」、NEC「cotomi v3」、富士通「Takane 32B」、PFN「PLaMo 2.0 Prime」）と契約を結び、2026 年 8 月頃から 2027 年 3 月にかけて行政実務への適合性を評価している[39]。このうち 3 モデルは国産クラウド「さくらのクラウド」上で稼働し、9〜11 月にブラインド A/B テストが計画された[40]。2027 年度向けには有償の政府調達を予定しており、2026 年 11 月の公募開始が予告されている[41]。予告資料では、評価テストが 300 問・35 観点に拡充され、提出書類として外部ツールを呼び出す「tool use」機能の説明が求められている[42]。なお、公募開始の正式な告知は本調査では確認できなかった。

表 1　主体別の主な動き（2026 年 5 月下旬〜11 月）

主体	主な動き
PFN	6 月 22 日：PLaMo 3.0 Prime 正式版（推論・非推論の 2 モデル、256K）[1,2] 10 月 27 日：PLaMo 3.1 Prime（Web 検索、オンプレミス対応の Deep Research）[4] 11 月 9 日：金融特化の PLaMo 3.0 Fin をオープンウェイト公開[6]

主体	主な動き
SB Intuitions	6 月 30 日：Sarashina3（5 モデル、ソブリンクラウドで提供）[7] 10 月 13 日：無償ベータ版「Sarashina サービス」[9] 10 月 29 日：初の思考モデル Sarashina3.2 mini thinking[10]
NTT	5 月 19 日：tsuzumi 2 Vision モデル（図表入り文書の理解）[12] 11 月 12 日：少量データでの専門特化チューニング技術[14] 論理数学・コード強化版は 2026 年度中の予定[15]
NEC	8 月 18 日：ブラウザ操作エージェント「NEC cotomi Agent」を販売[18]
富士通	5 月 25 日：自己進化マルチ AI エージェント技術[23] 7 月 28 日：金融向け専有型 AI 基盤の開発開始（2027 年 3 月提供予定）[24,25]
NII／ELYZA	9 月 28 日：LLM-jp-4.1（ツール呼び出しに新規対応）[28] 10 月 2 日：LLM-jp-4 基盤の推論モデル 2 種、ELYZA RSI Research 設立[29]
Sakana AI	6 月 22 日：Fugu（複数 AI を束ねるオーケストレーター）[31] 9 月 11 日：Fugu Max、Fugu Ultra v2[33]
政府・Noetra	6 月 30 日：経産省が Noetra を支援対象に選定（2026 年度 3,873 億円）[36] 7 月 16 日：44 社の出資を受け本格始動[37] 源内で 5 社のモデルを評価（2026 年 8 月頃〜2027 年 3 月）[39]

2.9 半年間の特徴（分析）

以上から、この半年の特徴は次の三点に整理できる。

- （分析）推論（思考）モデルが主要各社で出揃った。3 月の PLaMo 3.0 Prime β を起点に、LLM-jp 系、ELYZA、Sarashina と続き、思考モデルは「あって当然」の段階に入った。
- （分析）競争軸は軽量・ソブリン・オンプレミスである。1GPU で動く tsuzumi 2、ソブリンクラウドで提供される Sarashina3、オンプレミス対応の Deep Research を掲げる PLaMo 3.1 Prime は、いずれもフロンティア性能ではなく「機密データを国内で扱えること」を価値の中心に置く。
- （分析）秋にエージェント対応が集中した。LLM-jp-4.1 のツール呼び出し対応（9 月 28 日）、PLaMo 3.1 Prime の Deep Research（10 月 27 日）、Sarashina 初の思考モデル（10 月 29 日）と、推論の次の段階であるエージェント化への着手が明確になった。

3.「推論モデル 18 カ月遅れ」の確認

OpenAI が推論モデル「o1-preview」を発表したのは 2024 年 9 月 12 日である[43]。これに対し、PFN は 2026 年 3 月にリリースした PLaMo 3.0 Prime β 版を「フルスクラッチモデルとしては日本国内初の Reasoning 対応モデル」と位置づけている[44]。両者の差は約 18 カ月であり、前提とされた「18 カ月遅れ」はフルスクラッチ基準では妥当である。

他のフルスクラッチ系では、NII の LLM-jp-4 シリーズ（2026 年 4 月から公開）にも思考（thinking）モデルが含まれており[45,46]、ほぼ同時期である。Sarashina の初の思考モデルは 2026 年 10 月であり、差は約 25 カ月に広がる[10]。

一方、派生モデルまで含めると様相は異なる。ELYZA は 2025 年 5 月 1 日、アリババの Qwen2.5-32B に追加学習を施した推論モデル「ELYZA-Thinking-1.0-Qwen-32B」を公開しており[47]、この基準では差は約 8 カ月である。（分析）すなわち「推論能力を日本語モデルに付与する技術」は早期に追随できたが、「自前の事前学習モデルに推論を載せる」までに時間を要したと整理できる。

4. AI エージェント化の遅れの検証

エージェント化は単一の出来事ではないため、五つのマイルストーンに分けて、海外の初出と国産の初出を比較した（表 2）。

表 2　マイルストーン別にみた海外と国産の初出時期

マイルストーン	海外（初出）	国産（初出）	差
推論モデル	o1-preview[43]：2024 年 9 月	PLaMo 3.0 Prime β [44]：2026 年 3 月 Sarashina3.2 mini thinking[10]：2026 年 10 月	約 18 カ月 約 25 カ月
ツール連携（MCP）	MCP 公開[48]：2024 年 11 月	cotomi の MCP 準拠[21]：2025 年 7 月	約 8 カ月
Deep Research	Google[49]：2024 年 12 月 OpenAI[49]：2025 年 2 月	PLaMo 3.1 Prime[4]：2026 年 10 月	約 20〜22 カ月

マイルストーン	海外（初出）	国産（初出）	差
ブラウザ・PC 操作	computer use[50]：2024 年 10 月 Operator[51]：2025 年 1 月	cotomi Act 技術発表[19]：2025 年 8 月 NEC cotomi Agent 販売[18]：2026 年 8 月	技術：約 7〜10 カ月 製品：約 19〜22 カ月
コーディングエージェント	Claude Code[52]：2025 年 2 月 Codex[53]：2025 年 5 月	該当する国産基盤モデルは確認できず	約 18〜21 カ月を経て未到達

注：差は公表月の差（月単位）。国産側は原則としてフルスクラッチ系を基準とした。ブラウザ・PC 操作の国産例は、基盤 LLM が国産モデルであるか未確認である（2.4 参照）。コーディングエージェントの差は 2026 年 11 月時点の経過月数。

4.1 ツール連携・MCP

Anthropic が MCP（Model Context Protocol）をオープンソース化したのは 2024 年 11 月 25 日である[48]。NEC は 2025 年 7 月に cotomi の MCP 準拠を発表しており[21]、差は約 8 カ月にとどまる。もっとも、モデル自体のツール呼び出し能力の整備は 2026 年に入ってからであり、PLaMo 3.0 Prime（6 月）[3]、Sarashina3（6 月）[8]、LLM-jp-4.1（9 月）[28]と続いた。（分析）プロトコルへの対応は早いが、モデルにツール利用を学習させる工程は、推論モデル化とほぼ同時期まで待つ必要があった。

4.2 Deep Research（自律調査）

Google は 2024 年 12 月、OpenAI は 2025 年 2 月に、それぞれ Deep Research 機能を発表した[49]。国産モデルでは 2026 年 10 月 27 日の PLaMo 3.1 Prime が初出とみられ[4]、差は約 20〜22 カ月である。PFN に限れば、推論モデルの導入（2026 年 3 月）から Deep Research まで約 7 カ月を要した。OpenAI は o1-preview（2024 年 9 月）から Deep Research（2025 年 2 月）まで約 5 カ月であり、（分析）同社との差は 18 カ月から 20 カ月へとやや拡大したことになる。

4.3 ブラウザ・PC 操作

Anthropic は、PC 画面を見て操作する「computer use」を 2024 年 10 月 22 日にパブリックベータとして公開し[50]、OpenAI はブラウザ操作エージェント「Operator」を 2025 年 1 月 23 日に発表した[51]。NEC の cotomi Act は 2025 年 8 月 27 日の発表であり[19]、技術発表としての差は約

7〜10 カ月と小さい。ただし、製品としての販売（NEC cotomi Agent）は 2026 年 8 月であり[18]、この基準では約 19〜22 カ月となる。2.4 で述べたとおり、基盤 LLM が国産モデルであるかは未確認である。

4.4 コーディングエージェント・長時間自律

Anthropic は 2025 年 2 月に「Claude Code」を限定プレビューとして公開し[52]、OpenAI は同年 5 月 16 日にクラウド型のソフトウェアエンジニアリングエージェント「Codex」の研究プレビューを発表した[53]。これに相当する、国産基盤モデルによるコーディングエージェントは、本調査では確認できなかった。SB Intuitions は前述のとおり特化学習の不足を自ら認めている[11]。PLaMo 3.0 Prime は、第三者検証でルーター経由の Claude Code 接続が確認されているが、検証内容は接続確認と単発のツール呼び出しにとどまり、多段の継続作業は試されていない[3]。海外の初出（Claude Code）から 21 カ月を経て未到達であり、五つの中で最も差が大きい。

4.5 小括（分析）

（分析）エージェント化でも、推論モデルと同種の遅れが確認できる。モデル層の機能では概ね 20 カ月前後であり、18 カ月より縮まってはいない。作業が長時間・多段になるほど差は大きく、コーディングエージェントでは到達時期の見通し自体が立っていない。

5. 考察（分析・推定）

5.1 遅れが持ち越される理由

（分析）エージェント能力は、推論能力の上に、長いコンテキスト、ツール利用の学習、強化学習による多段タスクの訓練を積み上げる「後工程」である。したがって、推論モデル化の遅れは、そのままエージェント化の出発点の遅れになる。加えて、エージェント向けの学習には大量の作業軌跡データと計算資源が必要であり、軽量モデルを限られた計算資源で開発する国産各社の方針とは相性が良くない。PFN と OpenAI の比較（4.2）で差がやや拡大したことは、この構造と整合する。

5.2 層別に見ると景色が変わる

（分析）エージェントは「モデル層」「ハーネス・オーケストレーション層」「業務適用層」に分けて考える必要がある。遅れが大きいのはモデル層である。これに対し、Sakana Fugu は他社モデルを指揮するオーケストレーション層で、海外のフロンティアモデルに並ぶ水準を主張している[31]。NEC の cotomi Act、富士通の自己進化マルチ AI エージェント、ELYZA のハーネス自動改善も、モデルの外側で競争力を築く取り組みである[19,23,30]。ただし、Fugu が束ねるのは海外モデルやオープンウェイトモデルであり[32,34]、国産基盤モデルの遅れを埋めるものではない。別の勝ち筋と捉えるべきである。

5.3 月数以上に実力差が開いている可能性

（分析）月数は初出の時期の差にすぎない。Sarashina3.2 mini thinking が自社評価で比較したのは、llm-jp-4.1、Qwen3-32B、Qwen3-235B、gpt-oss-120b、GPT-5.4 mini であり[11]、フロンティアモデルではない。軽量クラスの中でツール呼び出しの最高点を得たことは成果だが、比較の土俵が異なる。

評価の物差しも動いている。Artificial Analysis が 2026 年 10 月 28 日に発表した指数 v5 は、実際のコーディング課題（FrontierCode 2.0）や、端末上のエージェント型科学研究タスク（Terminal-Bench-Science）を新たに組み込んだ[54]。長時間タスクの代表的指標である SWE-bench Verified について、国産モデルの公表値は本調査では確認できなかった[55]。（推定）評価軸がエージェント側へ移るほど、国産モデルは比較表に現れにくくなる。

5.4 今後の見通し

（推定）短期的には、2027 年にかけてエージェント機能の底上げが進む。NTT は論理数学・コード対応を高度化した tsuzumi 2 を 2026 年度中に予定し[15]、SB Intuitions はコーディングエージェント用途を含む学習を今後進めるとしている[11]。NII の次期版（LLM-jp-4.2）が強化学習で推論とエージェント能力を伸ばす方針であるとの二次情報もある[56]。需要側でも、源内の 2027 年度調達が tool use 機能の説明を求めており[42]、政府調達がエージェント対応を後押しする構図である。

（推定）一方、フロンティア級のエージェントを国産基盤モデルで実現する時期は、早くて 2028 年以降と考える。Noetra の大規模計算基盤の稼働が 2028 年 6 月であり[37]、それ以前に長

時間自律型の学習を大規模に行える計算資源が国内に乏しいためである。その間も海外側は前進を続けるため、差が自然に縮まる保証はない。

5.5 利用側への含意

- （分析）機密性の高い定型業務（要約、翻訳、図表入り文書の理解、社内文書に閉じた調査）では、国産モデルとオンプレミスまたはソブリンクラウドの組み合わせが現実的な選択肢になった。
- （分析）長時間自律型の業務（コーディング、複数システムをまたぐ業務代行）は、当面、海外フロンティアモデルか、それを束ねるオーケストレーターが前提になる。
- （分析）「国産か海外か」の二択ではなく、業務の機密度と自律度に応じて振り分ける設計が要点である。富士通が自社 IP 由来の AI とフロンティア AI を組み合わせる方針を示したのは、その現れといえる[26]。

6. 留意点・未確認事項

- PFN の PLaMo 3.1 Prime に関するプレスリリースと技術ブログは取得できず、内容は報道[4,5]に依拠した。エージェント系ベンチマークの数値は未確認である。
- cotomi Act および NEC cotomi Agent を駆動する基盤 LLM が、自社の cotomi 単独かどうかは確認できていない。
- 国産モデルについて、SWE-bench Verified など長時間タスクの公表値は確認できなかった。各社のベンチマーク結果は自社評価が中心であり、第三者による独立検証は限定的である。
- 源内の 2027 年度向け公募は 2026 年 11 月開始と予告されているが、開始の正式な告知は確認できなかった。9〜11 月の A/B テストの結果も、公表を確認できていない。
- 文献[8,32,34,56]は解説記事（二次情報）であり、これらに依拠した記述は一次情報による確認が望ましい。
- 遅れの月数は初出の公表月の差であり、性能水準の同等性を示すものではない。

参考文献

[1] Impress Watch「PFN、企業利用の実用性を高めた国産 AI モデル「PLaMo 3.0 Prime」」（2026 年 6 月）
https://www.watch.impress.co.jp/docs/news/2118910.html

[2] 窓の杜「国産生成 AI 基盤モデル「PLaMo 3.0 Prime」が正式リリース ～推論特化と高速応答の 2 モデル」（2026 年 6 月 23 日） https://forest.watch.impress.co.jp/docs/news/2119200.html

[3] クラスメソッド DevelopersIO「PLaMo 3.0 Prime を試してみた」（2026 年 6 月 22 日）
https://dev.classmethod.jp/ja/articles/plamo-3-0-prime-first-touch/

[4] Impress Watch「PFN、調査業務に対応した国産 AI 最新版「PLaMo 3.1 Prime」」（2026 年 10 月 27 日）
https://www.watch.impress.co.jp/docs/news/2151518.html

[5] ITmedia AI＋「PFN の国産フラグシップモデル「PLaMo 3.1 Prime」発表　オンプレで Deep Research 対応 API 無料プランも開始」（2026 年 10 月 28 日）
https://www.itmedia.co.jp/aiplus/article/2610/28/2000002652/

[6] PC Watch「日本語で考える金融特化 LLM「PLaMo 3.0 Fin」オープンウェイトで公開」（2026 年 11 月 9 日） https://pc.watch.impress.co.jp/docs/news/2154837.html

[7] ビジネス+IT「ソフトバンク傘下の SB Intuitions、国産 LLM「Sarashina3」を提供開始」（2026 年 7 月 2 日） https://www.sbbit.jp/article/cont1/185977

[8] WEEL「Sarashina とは?SB Intuitions の国産 LLM の特徴・料金・使い方を徹底解説」
https://weel.co.jp/media/tech/sarashina/

[9] Impress Watch「SB Intuitions、国産 AI「Sarashina」を無料公開 チャットや API で 6 モデル」（2026 年 10 月 13 日） https://www.watch.impress.co.jp/docs/news/2147706.html

[10] ITmedia AI＋「国産 LLM「Sarashina」に初の「思考モデル」　ツール呼び出しや Web 検索性能が向上 API で提供」（2026 年 10 月 30 日） https://www.itmedia.co.jp/aiplus/article/2610/30/2000002743/

[11] SB Intuitions TECH BLOG「国産 LLM Sarashina3.2 mini thinking の紹介」（2026 年 10 月 29 日）
https://www.sbintuitions.co.jp/blog/entry/2026/10/29/133000

[12] BUSINESS NETWORK「NTT 製 LLM「tsuzumi 2」に新モデル、図表付きビジネス文書の理解に強み」（2026 年 5 月） https://businessnetwork.jp/article/34554/

[13] NTT News Release "NTT's LLM tsuzumi 2 Updated"（2026 年 5 月 19 日）
https://group.ntt/en/newsrelease/2026/05/19/260519a.html

[14] NTT ニュースリリース「少量データから特定の専門性を有した LLM にチューニングする技術を確立」（2026 年 11 月 12 日） https://group.ntt/jp/newsrelease/2026/11/12/261112a.html

[15] NTT Technical Review "NTT's LLM tsuzumi 2: A High-performance, Secure, and Low-cost Japanese LLM"（2026 年 8 月号） https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202608fa1_s.html

[16] NTT R&D Website「さらなる進化を遂げた NTT 版 LLM tsuzumi 2――日本の企業 DX を支える高性能・高セキュア・低コストな純国産 LLM」 https://www.rd.ntt/research/JN202606_39542.html

[17] NTT「ニュースリリース 2026 年」一覧（2026 年 11 月 20 日閲覧）
https://group.ntt/jp/newsrelease/2026/

[18] IT Leaders「NEC、Web ブラウザ自動操作「NEC cotomi Agent」を Microsoft Marketplace で販売」（2026 年 8 月 19 日） https://it.impress.co.jp/articles/-/29690

[19] NEC プレスリリース「NEC、暗黙知をデータ化し学習・活用することで Web 業務を自動化するエージェント技術「cotomi Act」を開発」（2025 年 8 月 27 日）
https://jpn.nec.com/press/202508/20250827_02.html

[20] NEC プレスリリース「NEC、企業ノウハウを AI エージェントで自動抽出・組織資産化し、デジタル業務の自動化を実現するソリューションの提供を開始」（2025 年 12 月 3 日）
https://jpn.nec.com/press/202512/20251203_01.html

[21] クラウド Watch「NEC、AI エージェントの活用促進に向け生成 AI「cotomi」を強化」（2025 年 7 月 9 日） https://cloud.watch.impress.co.jp/docs/news/2029355.html

[22] DX マガジン「【暗黙知可視化で 80.4%】NEC『cotomi Act』が業務ノウハウを自動化」
https://dxmagazine.jp/news/2540ko70/

[23] 富士通プレスリリース「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」（2026 年 5 月 25 日） https://global.fujitsu/ja-jp/pr/news/2026/05/25-01

[24] 富士通プレスリリース「富士通、金融機関向けに専有型 AI プラットフォームの開発を開始」（2026 年 7 月 28 日） https://global.fujitsu/ja-jp/pr/news/2026/07/28-01

[25] 電波新聞デジタル「富士通、金融向け専有型 AI 基盤 「Takane」や複数の AI エージェント搭載 27 年 3 月提供、安全な AI 活用を支援」（2026 年 7 月 29 日） https://dempa-digital.com/?p=721519

[26] MONOist「富士通が AX 加速に向け事業モデル進化、AI エージェント活用の 3 つ基盤を一括提供」（2026 年 10 月 7 日） https://monoist.itmedia.co.jp/mn/articles/2610/07/news058.html

[27] 富士通（PR TIMES）「大規模言語モデル「Takane」を活用し、ソフトウェアの要件定義から設計、実装、結合テストに渡る全工程を AI エージェントが協調し実行する AI ドリブン開発基盤を開発し、運用開始」（2026 年 2 月 17 日） https://prtimes.jp/main/html/rd/p/000000533.000093942.html

[28] PC Watch「NII、ツール呼び出し対応・STEM 強化の「LLM-jp-4.1」公開」（2026 年 9 月 29 日）
https://pc.watch.impress.co.jp/docs/news/2144150.html

[29] ビジネス+IT「ELYZA が国産 LLM 基盤の推論モデル 2 種を無料公開、RSI の研究所も設立」（2026 年 10 月 2 日） https://www.sbbit.jp/article/cont1/187290

[30] MONOist「2026 年の AI 開発最前線――「AI による AI 開発」から今後は再帰的自己改善「RSI」へ」（2026 年 10 月 7 日） https://monoist.itmedia.co.jp/mn/spv/2610/07/news019_2.html

[31] PC Watch「Fable 5 に匹敵、複数 AI の力を束ねて実現した Sakana AI「Fugu Ultra」」（2026 年 6 月 22 日） https://pc.watch.impress.co.jp/docs/news/2119021.html

[32] WEEL「Sakana Fugu とは?Sakana AI のマルチエージェント統合 AI モデルを徹底解説」
https://weel.co.jp/media/innovator/sakana-fugu/

[33] 窓の杜「国産の「Sakana Fugu」に「Fable 5.1」「GPT-6 Astra」超えの強化モデルが登場」（2026 年 9 月） https://forest.watch.impress.co.jp/docs/news/2140691.html

[34] 天秤 AI メディア「Sakana AI「Fugu Max」「Fugu Ultra v2」同時発表!コスト重視・性能重視の新モデルを徹底比較」 https://tenbin.ai/media/ai_news/sakana-fugu-max-ultra-v2

[35] 熊本日日新聞（共同通信配信）「国産 AI 開発へ新会社設立 ソフトバンクや NEC など」（2026 年 4 月 12 日） https://kumanichi.com/articles/1979734

[36] ビジネス+IT「ソフトバンク・NEC・ホンダ・ソニーの国産 AI 企業「Noetra」経産省が 3,873 億円拠出」（2026 年 7 月 1 日） https://www.sbbit.jp/article/cont1/185958

[37] Noetra・ソニーグループ・ソフトバンク・NEC・本田技研工業 プレスリリース「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」（2026 年 7 月 16 日） https://www.softbank.jp/corp/news/press/sbkk/2026/20260716_01

[38] 現代ビジネス「「1 兆円の税金投入」も失敗覚悟?ソフトバンクら 44 社が主導する「国産 AI」プロジェクトで、企業の本気度が疑われる決定的な理由」 https://gendai.media/articles/-/170489

[39] DX マガジン「デジタル庁、国産 LLM5 社を選定 「ガバメント AI 源内」で行政利用へ」 https://dxmagazine.jp/news/ykmt262407/

[40] eWeek "Japan Tests Domestic LLMs Against Claude and Amazon Nova in Government AI Pilot" https://www.eweek.com/news/japan-domestic-llm-genai-test/

[41] デジタル庁「ガバメント AI で使用する国産基盤モデルの公募について（予告）」（2026 年 5 月 29 日） https://www.digital.go.jp/news/207cadc2-c218-42f6-92ec-c0eda476b49c

[42] デジタル庁「別紙 2 令和 9 年度（2027 年度）向け国産基盤モデルの公募（予定）」（PDF、2026 年 5 月 29 日） https://www.digital.go.jp/assets/contents/node/information/field_ref_resources/207cadc2-c218-42f6-92ec-c0eda476b49c/05201362/20260529_news_genai_outline_02.pdf

[43] OpenAI「OpenAI o1-preview が登場」（2024 年 9 月 12 日） https://openai.com/ja-JP/index/introducing-openai-o1-preview/

[44] Preferred Networks 技術ブログ「PLaMo 3.0 Prime β 版をリリースしました」（2026 年 3 月） https://www.preferred.jp/ja/blog/tech/plamo-3-prime-beta-release

[45] 国立情報学研究所 ニュースリリース（2026 年 4 月 3 日付、LLM-jp-4 シリーズの公開に関するもの）（PDF） https://www.nii.ac.jp/news/upload/nii_newsrelease_20260403.pdf

[46] LLM-jp「リリース」（公開モデル一覧、2026 年 11 月 20 日閲覧） https://llm-jp.nii.ac.jp/release/

[47] Web 担当者 Forum「ELYZA が日本語での論理的思考能力を強化した LLM「ELYZA-Thinking-1.0-Qwen-32B」を開発」（2025 年 5 月 2 日） https://webtan.impress.co.jp/n/2025/05/02/49117

[48] Anthropic "Introducing the Model Context Protocol"（2024 年 11 月 25 日） https://www.anthropic.com/news/model-context-protocol

[49] GIGAZINE「ChatGPT の「Deep research」と Gemini の「Deep Research」は何が違うのか?」（2025 年 3 月 7 日） https://www.gigazine.net/news/20250307-what-is-deep-research/

[50] Impress Watch「ついに AI が PC を操作 Claude 3.5 Sonnet に作業自動化する「Computer use」」（2024 年 10 月） https://www.watch.impress.co.jp/docs/news/1633561.html

[51] OpenAI「Operator が登場」（2025 年 1 月 23 日） https://openai.com/ja-JP/index/introducing-operator/

[52] gihyo.jp「Anthropic ハイブリット推論モデル「Claude 3.7 Sonnet」リリース ——ターミナル上のエージェント型支援ツール「Claude Code」限定プレビュー公開」（2025 年 2 月） https://gihyo.jp/article/2025/02/claude3.7-sonnet-and-claude-code

[53] CodeZine「OpenAI、ソフトウェア開発を支援する新しいエージェント「Codex」の研究プレビューを発表」（2025 年 5 月 21 日） https://codezine.jp/news/detail/21543

[54] livedoor ニュース（GIGAZINE 配信）「AI ベンチマーク「Intelligence Index v5」発表、2 つの新しいベンチマークと 3 つのアップグレード」 https://news.livedoor.com/article/detail/32684059/

[55] Sarashina サービス ドキュメント「モデル性能」（2026 年 11 月 20 日閲覧） https://docs.sarashina.ai/model-performance

[56] ai-revolution.co.jp「LLM-jp-4 とは?国立情報学研究所・約 12 兆トークン学習の国産 LLM を徹底解説【2026 年最新】」（2026 年 11 月 7 日更新） https://ai-revolution.co.jp/media/what-is-llm-jp-4/