調査報告　2026 年 5 月 20 日〜11 月 20 日

国産 LLM の進展と AI エージェントの国際比較

GPT-6.2 Astra

自律調査では約 21〜23 カ月、業務用エージェントの提供では約 3 カ月。どの段階を比べるかで、日本と海外の時間差は大きく変わる。

OpenAI の o1-preview 公開から、Preferred Networks（PFN）が PLaMo 3.0 Prime β 版のモニター企業を募集するまで、18 カ月と 7 日を要した。「国内フルスクラッチの推論モデル」に限れば、18 カ月という見立てには具体的な根拠がある。正式提供はさらに約 3 カ月後だった。［1–3］

同じ機能をいつ使えるようになったか

比較対象	海外側の節目	国内側の節目	暦上の差
推論モデル	OpenAI o1-preview 2024.09.12 公開	PLaMo 3.0 Prime 2026.03.19 β 募集／06.22 正式提供［1–3］	18 カ月 7 日／ 21 カ月 10 日
関数呼び出し	OpenAI API 2023.06.13	KARAKURI LM 8x7B 2024.06.20 公開［5, 6］	12 カ月 7 日
業務用エージェント	Salesforce Agentforce 2024.10.29 一般提供発表	NEC AI Agent 2025 年 1 月 提供開始［7–9］	約 3 カ月
自律的な調査と出典付きレポート	Google 2024.12.11 OpenAI 2025.02.02	PLaMo 3.1 Prime 2026.10.27 提供開始［10–12］	対 Google 22 カ月 16 日 対 OpenAI 20 カ月 25 日

比較の範囲

上表は公表された代表製品の機能と時期を照合したもの。国内初・世界初の認定、性能が同水準に達した日、各社への普及日は別の問いになる。Google の初回提供は英語、PFN の自律調査はオンプレミスにも対応する。［10–12］

海外モデルの国内追加学習まで含めると、2025 年 1 月 27 日にサイバーエージェントが DeepSeek-R1 系の日本語モデルを公開しており、o1-preview との差は約 4 カ月半になる。カラクリの上記モデルも Swallow-MX を基にし、その系譜は Mixtral にさかのぼる。［4, 33, 34］

最近半年の動向

企業と研究機関は何を公開したか

推論機能を商用化する動きと、日本語でツールを使う訓練、専門分野への適応、業務システムへの接続が並行して進んだ。

月	主な発表・公開	確認できる段階
5 月	富士通が、業務の実行結果や人のフィードバックを使って業務特化型 LLM「Takane」を継続的に改善するマルチ AI エージェント技術を発表。［14］	研究開発・自社評価
6 月	PFN が PLaMo 3.0 Prime を正式提供。推論・非推論モデル、256k の文脈長に対応。SB Intuitions が Sarashina3 を提供し、mini は外部 API を選んで呼び出せる。［3, 15］	API・オンプレミス、国内クラウドで提供
7 月	44 社が出資した Noetra が、エージェントの中核となる推論基盤モデルなどの開発に本格着手。ストックマークは Aconnect に技術的な解決策の比較根拠を文献で確認する機能を追加。［16, 30］	開発計画／業務向け機能の提供
8 月	NEC がブラウザ・業務システムを操作する cotomi Agent の販売経路を Microsoft Marketplace に拡大。源内では国産クラウド上で 3 モデルを試用する構成を公表。［17, 28］	販売経路の拡大／政府の試用
9 月	Sakana AI が複数のオープンモデルを組み合わせる Fugu Max を Sakana Chat で提供。国立情報学研究所はツール呼び出しを学習した LLM-jp-4.1 を 3 サイズで公開。［18, 19］	複数モデルの統合／モデル・学習データ公開
10 月	ELYZA が国産の LLM-jp-4 を土台にした 2 つの推論モデルと顧客応対の評価課題を公開。PFN は Deep Research を、SB Intuitions は Sarashina3.2 mini thinking を提供。［12, 20, 21］	公開モデル／商用機能／サービスへの追加
11 月	PFN が金融特化の PLaMo 3.0 Fin をオープンウェイトで公開。NTT は業務マニュアルをもとに学習データを生成し、tsuzumi 2 に知識を定着させる手法を発表。［22, 23］	モデル公開／研究成果

推論とツール利用が実装された順序

PFN は 2025 年 10 月の PLaMo 2.1 Prime で、指示に応じた検索先や外部 API の自動選択を実装した。Sarashina も 2026 年 6 月にツール利用、10 月に思考モデルを提供している。両社で、ツール利用は思考モデルの登場に先行した。［2, 13, 15, 21］

日本語のツール利用と長い仕事の完遂

エージェントに一つの関数を正しく選ばせることと、途中の失敗や情報不足に対処しながら成果物を仕上げさせることでは、必要な能力が異なる。測定対象を合わせると、国産モデルの到達点が見える。

日本語での複数ターンのツール利用

国立情報学研究所の測定では、日本語版の関数呼び出し評価 BFCL-nejumi の複数ターン課題で、LLM-jp-4.1-33B は Qwen3-32B と Qwen3.5-9B を上回った。一方、英語の対話型業務ベンチマーク τ^2-bench では Qwen3.5 系との差が残った。比較モデルの世代に加え、言語・対話の長さ・業務ルールで結果が変わる。［25］

SB Intuitions の評価では、Sarashina3.2 mini thinking が、日本語の業務対話評価 J-tau と関数呼び出し評価 BFCL v3 で、GPT-5.4 mini などを含む比較対象中の最高スコアを記録した。複数ターンの指示追従や科学推論などでは改善の余地も報告している。［21］

自律調査では検索からレポート作成までを学習

PFN は、モデルが複数回検索し、情報を選び、出典付きの長文レポートにまとめる過程を学習させた。3.1 Prime の推論モードは旧 3.0 と比べて Web 検索 QA で平均 2.4 ポイント、Web 検索レポートで平均 11.1 ポイント改善した。比較相手は自社旧世代や同規模帯の公開モデルが中心である。［24］

長時間の自律作業に残る比較上の空白

海外の評価では、タスクの長さと完遂率まで測る。第三者機関 METR は、人間の専門家に要する時間を尺度として、ソフトウェア課題を 50%または 80%の確率で完遂できる難度を推定する。OpenAI も 11 月に長時間の知識労働を対象とする社内評価 GDPval v2 を公表した。対象業務も判定方法も異なる。［26, 32］

国産モデルと海外の最新最上位モデルを、同じツール・権限・時間と費用の上限で動かし、長い業務の完遂率を時系列で比較した資料は、今回確認した公表資料では限られる。長時間の汎用能力差を「何カ月」と換算するには、この比較が必要である。

製品化の早さと企業への普及

日本の大企業では導入の進み方に差がある

PwC の 2026 年春調査は、売上高 500 億円以上の企業・組織に勤務し、生成 AI 導入に関わる課長職以上が対象である。「AI エージェントについて理解しており、導入済み／導入を進めている」とした割合は、日本が 6 カ国で最も低かった。［27］

回答者の所在	日本	米国	英国	中国	ドイツ	韓国
導入済み・導入中	33%	59%	55%	48%	37%	36%
回答者数	932	670	412	412	309	309

日米差は 26 ポイント。回答は勤務先の状況についての自己申告で、導入の途上にある企業も含む。基盤モデルの国籍や本番での成功件数を尋ねた調査ではなく、現時点の普及差を 18 カ月の時間差に換算することはできない。［27］

提供・検証・運用を読み分ける

デジタル庁は源内で 5 社の国産モデルを試用する方針を示し、うち tsuzumi 2、Takane 32B、PLaMo 2.0 Prime をさくらのクラウドで動かす構成を示した。これとは別に、10 月にはエージェントチャットの順次導入を発表した。「約 18 万人」は政府職員の対象規模で、国産モデルでエージェントを常用している人数を示す数字ではない。［28, 29］

楽天は 2026 年第 3 四半期の決算説明資料で、消費者向け 20 サービスへのエージェント提供と、社内の Rakuten AI Workforce の月間利用者 7,100 人以上を公表した。国内企業での運用実績を示すが、全処理で国産基盤モデルを使うことを意味しない。ELYZA Voice Agent は、発表時点では現場導入に向けた検証段階にある。［31, 35］

知財の現場なら何を測るか

例えば、既知の正解文献がある先行技術調査を用いる。国内外のモデルに同じ特許データベースと同じ工数上限を与え、検索式の修正から文献選択、請求項と根拠段落の対応付けまで任せる。既知文献の回収率、根拠段落の正確さ、人が修正した時間、完遂 1 件当たりの費用を比べれば、実務上の差を判断できる。

未公開の発明情報を扱う案件には、PFN が公表した社内文書限定の閉域調査という選択肢もある。評価の際は、接続している情報源を明記すると、モデルの性能と検索環境の差を見分けやすい。［12］

出典

参考資料

資料名のリンクから発表原文を参照できます。参照日 2026 年 11 月 20 日。

［1］ **OpenAI　2024.09.12**　Introducing OpenAI o1-preview
［2］ **Preferred Networks　2026.03.19**　PLaMo 3.0 Prime β 版のモニター企業募集
［3］ **Preferred Networks　2026.06.22**　国産生成 AI 基盤モデル PLaMo 3.0 Prime を正式リリース
［4］ **ITmedia AI＋　2025.01.27**　DeepSeek R1 の日本語追加学習モデルをサイバーエージェントが公開
［5］ **OpenAI　2023.06.13**　Function calling and other API updates
［6］ **カラクリ　2024.06.20**　KARAKURI LM 8x7B Instruct v0.1 を一般公開
［7］ **Salesforce　2024.10.29**　Agentforce general availability announcement
［8］ **NEC　2024.11.27**　高度な専門業務を自動化する AI エージェントを発表
［9］ **NEC　2025 年 3 月期**　第 187 期 有価証券報告書「研究開発活動」
［10］ **Google　2024.12.11**　Try Deep Research in Gemini Advanced
［11］ **OpenAI　2025.02.02**　Introducing deep research
［12］ **Preferred Networks　2026.10.27**　PLaMo Prime に Web 検索・Deep Research 搭載
［13］ **Preferred Networks　2025.10.07**　自動ツール連携機能を実装した PLaMo 2.1 Prime の提供開始
［14］ **富士通　2026.05.25**　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
［15］ **SB Intuitions　2026.06.30**　Cloud PF Type A で Sarashina3 シリーズを提供開始
［16］ **Noetra ほか　2026.07.16**　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
［17］ **NEC　2026.08.18**　Microsoft Marketplace で NEC cotomi Agent を販売開始
［18］ **Sakana AI　2026.09.17**　Sakana Chat を Fugu Max に刷新、メモリー機能を追加
［19］ **国立情報学研究所　2026.09.28**　大規模言語モデル「LLM-jp-4.1」シリーズを公開
［20］ **ELYZA　2026.10.02**　ELYZA RSI Research 設立、基盤モデル・評価・自律改善の成果公開
［21］ **SB Intuitions　2026.10.29**　国産 LLM Sarashina3.2 mini thinking の紹介
［22］ **Preferred Networks　2026.11.09**　金融特化型の小規模言語モデル PLaMo 3.0 Fin を公開
［23］ **NTT　2026.11.12**　業務知識を少量データから LLM に定着させるチューニング技術
［24］ **Preferred Networks　2026.10.27**　PLaMo 3.1 Prime 技術解説、学習・比較評価・評価条件
［25］ **LLM-jp　2026.09.28**　LLM-jp-4.1 モデルの公開とツール利用評価
［26］ **METR　2026.09.08 版**　Task-Completion Time Horizons of Frontier AI Models
［27］ **PwC Japan　2026 年 6 月**　生成 AI に関する実態調査 2026 春 6 カ国比較（p.136）
［28］ **デジタル庁　2026.08.21**　源内における国産クラウド上での国産基盤モデルの試用開始
［29］ **デジタル庁　2026.10.20 発表**　源内の AI エージェント利用環境と源内基準（11 月 6 日更新）
［30］ **ストックマーク　2026.07.15**　Aconnect の技術探索エージェントに根拠文献参照機能を追加
［31］ **ELYZA　2026.10.02**　音声対話 AI エージェント ELYZA Voice Agent を開発
［32］ **OpenAI　2026.11.17**　Introducing GPT-6.2 Astra and Sol（GDPval v2 等）
［33］ **カラクリ　モデルカード**　KARAKURI LM 8x7B Instruct v0.1 のベースモデル
［34］ **Swallow 開発チーム　モデルカード**　Swallow-MX-8x7b-NVE-v0.1 の事前学習元
［35］ **楽天グループ　2026 年 11 月**　2026 年度第 3 四半期決算説明資料（スクリプト付、特に p.58）