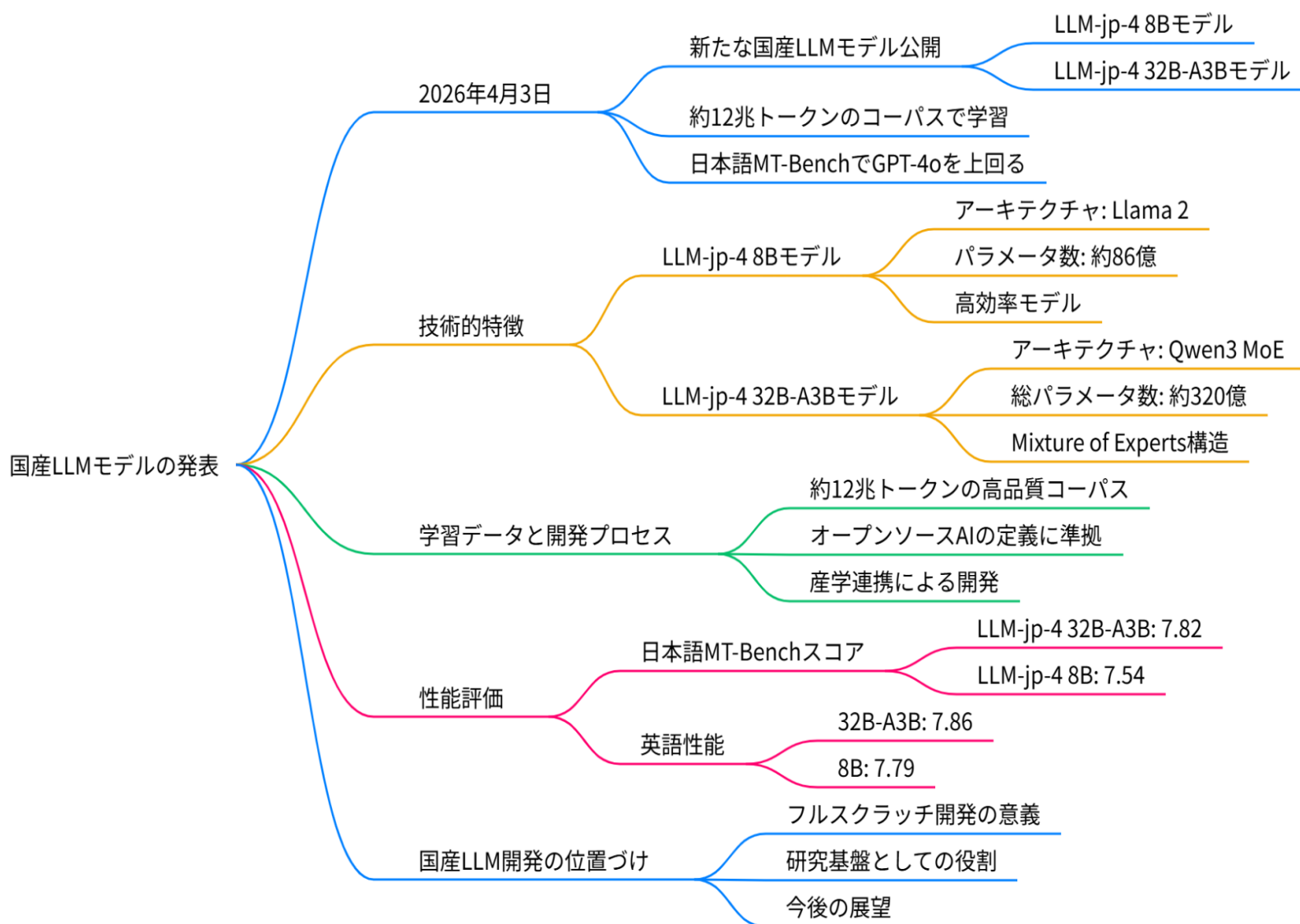


# 国立情報学研究所（NII）が約 12 兆トークンの良質なコーパスで学習した新たな国産 LLM 「LLM-jp-4 8B モデル」 「LLM-jp-4 32B-A3B モデル」 をオープンソースライセンスで公開

Felo AI



# 概要

2026年4月3日、国立情報学研究所（NII）の大規模言語モデル研究開発センター（LLMC）は、LLM 研究開発コミュニティ「LLM-jp」の活動を通じて、新たな国産大規模言語モデル「LLM-jp-4 8B モデル」および「LLM-jp-4 32B-A3B モデル」をオープンソースライセンスで公開した [19](#)。これらのモデルは、約 12 兆トークンという大規模かつ高品質な日本語中心のコーパスでゼロから学習（フルスクラッチ学習）されており、特に日本語の対話能力を測るベンチマーク「日本語 MT-Bench」において、OpenAI の「GPT-4o」や Alibaba Cloud の「Qwen3-8B」といった世界トップクラスのモデルを上回るスコアを記録した [111](#)。

この成果は、単に高性能な国産モデルが登場しただけでなく、その開発プロセスと成果物が完全にオープンである点に大きな意義がある。学習に用いたデータは第三者が入手可能なものに限定され、オープンソース AI の定義（OSAID）に準拠しているため、高い透明性と信頼性を確保している [11](#)。特に、専門性の高い日本語データの不足が国産 LLM 開発の大きな課題とされる中で [16](#)、大規模で質の高い学習コーパスを構築・公開したことは、日本の AI 研究開発エコシステム全体に貢献する重要なマイルストーンと言える。

## 詳細レポート

### LLM-jp-4 シリーズの技術的特徴

今回公開されたモデルは、特性の異なる 2 種類である [1](#)。

- **LLM-jp-4 8B モデル:**
  - **アーキテクチャ:** Meta 社の Llama 2 アーキテクチャをベースに採用 [19](#)。
  - **パラメータ数:** 約 86 億 (8B) [1](#)。
  - **特徴:** 比較的小規模ながら、後述するベンチマークで GPT-4o を超える日本語性能を達成しており、コストパフォーマンスに優れた高効率モデルとなっている [11](#)。
- **LLM-jp-4 32B-A3B モデル:**
  - **アーキテクチャ:** Alibaba Cloud の Qwen3 MoE アーキテクチャをベースに採用 [19](#)。
  - **総パラメータ数:** 約 320 億 (32B) [1](#)。
  - **構造:** 複数の専門家 (Expert) ネットワークを持ち、入力に応じて一部のみを稼働させる Mixture of Experts (MoE) モデル [113](#)。
  - **アクティブパラメータ数:** 約 38 億 (3.8B) [1](#)。128 個のエキスパートのうち 8 個がアクティブになる [1](#)。
  - **特徴:** 総パラメータ数は大きいものの、推論時に実際に動作するパラメータは 3.8B と非常に軽量である。これにより、大規模モデルの表現力を持ちながら、推論コストを大幅に抑制できる [11 13](#)。この効率性は、近

年の高性能モデル（DeepSeek V3, Mistral など）でも採用されているトレンドである [13](#)。

両モデルともに最大で約 65,000 トークンのコンテキスト長に対応しており、長文の読解や要約、複雑な対話タスクにも対応可能である [111](#)。

## 学習データと開発プロセス

LLM-jp-4 の卓越した性能は、その学習データと厳格な開発プロセスに起因する。

**約 12 兆トークンの高品質コーパス:** 学習には、インターネット上の公開データ、政府・国会の公式文書、LLM-jp が独自に開発した合成データなどから構成される約 12 兆トークンの巨大なコーパスが使用された [111](#)。これは、過去の LLM-jp-3.1 シリーズの約 6 倍に相当する規模である [11](#)。特に、国産 LLM 開発における最大の課題の一つが、著作権をクリアした高品質な日本語データの不足であると指摘される中で [16](#)、これほど大規模なコーパスを整備したことは特筆に値する。

**透明性と再現性の確保:** 開発にあたってはオープンソース AI の定義（OSAID）が重視され、学習コーパスは第三者も入手可能なデータで構築されている [111](#)。これにより「モデルが何を学習したか」が明確になり、研究やビジネス利用における信頼性が向上する。このアプローチは、ベースモデルの出自が不明確であったり、ライセンス表記に不備があったりして問題となった他の事例とは一線を画す [13](#)。

**産学連携による開発体制:** このプロジェクトは、NII の LLMC が主宰するコミュニティ「LLM-jp」が中心となり、計算資源として産業技術総合研究所（AIST）の AI 橋渡しクラウド「ABCI 3.0」を利用するなど、日本のトップレベルの産学連携によって実現された [111](#)。

## 性能評価とベンチマーク分析

LLM-jp-4 の性能は、LLM-jp が独自に開発した評価フレームワーク「llm-jp-judge」を用いて、GPT-5.4 を評価者（Judge）とする LLM-as-a-Judge 方式で測定された [111](#)。

**日本語 MT-Bench における性能比較:** 日本語の対話理解能力を測る「日本語 MT-Bench」において、LLM-jp-4 シリーズは既存の強力なモデルを明確に上回る結果を示した [1](#)。

モデル	日本語 MT-Bench スコア
<b>LLM-jp-4 32B-A3B (MoE)</b>	<b>7.82</b>
<b>LLM-jp-4 8B</b>	<b>7.54</b>
gpt-oss-20b	7.33

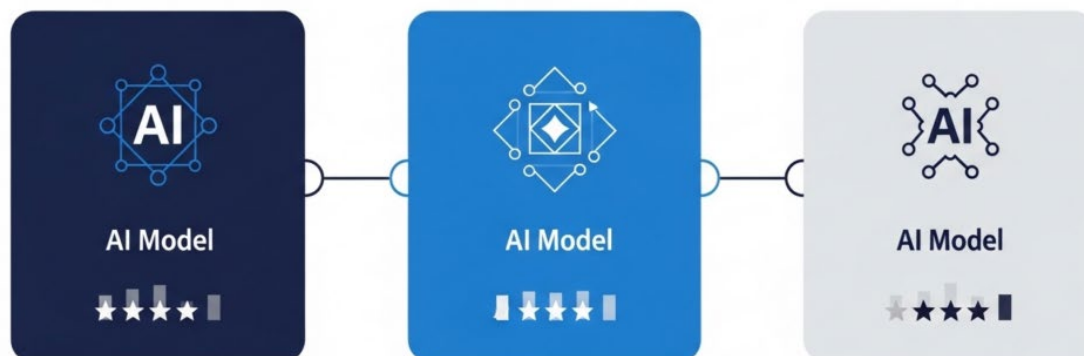
モデル	日本語 MT-Bench スコア
GPT-4o	7.29
Qwen3-8B	7.14
(出典: 国立情報学研究所の発表資料に基づく <a href="#">19</a> )	

この結果は、**8B** という比較的小規模なモデルでさえ、世界最高峰のクローズドモデルである **GPT-4o** や、同規模のオープンソースモデルとして評価の高い **Qwen3-8B** を日本語性能で凌駕したことを示している [330](#)。特に、**32B-A3B** モデルの **7.82** というスコアは、現行のオープンソースモデルの中で日本語対話能力がトップクラスであることを示唆している [11](#)。

**英語性能:** 英語の理解能力を測る「MT-Bench」においても、**32B-A3B** モデルが **7.86**、**8B** モデルが **7.79** を記録し、**GPT-4o** (**7.69**) と同等以上の性能を達成している [111](#)。これは、日本語に特化しつつも、グローバルなレベルでの汎用性を失っていないことを意味する。

## 国産 LLM 開発における位置づけと比較

LLM-jp-4 の登場は、日本の LLM 開発の潮流において重要な意味を持つ。



**開発アプローチ：「フルスクラッチ」の意義:** LLM-jp-4 は、既存の学習済みモデルに追加学習（ファインチューニング）を行うのではなく、データ収集から事前学習までを一貫して行う「フルスクラッチ」で開発された [1](#)。これは、海外製の強力なオープンモデル（例: DeepSeek V3）をベースにファインチューニングするアプローチとは根本的に異なる [13](#)。フルスクラッチ開発は膨大な計算資源とデータ収集コストを要するが、モデルの内部構造や学習データを完全に把握できるため、透明性、信頼性、そして特定の言語や文化への深い適応において優位性を持つ。楽天が発表した「Rakuten AI 3.0」が、当初「国産最大規模」とされながらも、その実態が中国の DeepSeek V3 ベースであったことが判明し批判を浴びた事例は、フルスクラッチ開発の価値を逆説的に示している [13](#)。

**目的とポジショニング：「研究基盤」としての役割:** 国内の他の LLM 開発プロジェクトと比較すると、LLM-jp-4 の立ち位置は明確である。

- **富士通「Takane」:** 官公庁や大企業向けに、セキュリティやガバナンス、既存 IT 基盤との親和性を重視した、ビジネスシステムへの組み込みを主眼に置いている [23](#)。
- **ソフトバンク「Sarashina」:** 通信インフラとの統合を視野に入れ、中長期的な AI 基盤の構築を目指している [23](#)。
- **LLM-jp-4:** これらに対し、LLM-jp-4 はオープンな研究開発基盤としての性格が強い。高性能なモデルと、その学習に用いた大規模コーパスをオープンソースで公開することにより、国内のスタートアップ、大学、研究機関が自由に利用・改変できる土台を提供する [16](#)。これは、特定企業による技術独占を防ぎ、日本の AI エコシステム全体の競争力を底上げする狙いがある [16](#) [23](#)。

**今後の展望:** LLMC は、今回公開したモデルを活用して LLM の透明性・信頼性の確保に向けた研究をさらに進めるとともに、現在、より大規模なパラメータを持つモデルの開発も進めている [12](#)。これらの次世代モデルは 2026 年度中に順次公開される予定であり、国産 LLM の性能をさらに引き上げることが期待される [19](#)。

- [1. 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 ...](#)
- [2. ニュースリリース 約 12 兆トークンの良質なコーパスで学習 ...](#)
- [3. 比較対象が古くないか。？ / 約 12 兆トークンの良質なコーパス ...](#)
- [4. 『pnpm vs npm vs yarn 2026 年版 公式ベンチマーク引用+ ...](#)
- [5. ローカル LLM のおすすめモデルと導入の全貌！スペック・商用 ...](#)
- [6. 日本語 LLM おすすめ 3 選 | 国内特化の大規模言語モデルを ...](#)
- [7. 5060/5090 日本語ローカル LLM 要約ベンチマーク \(2026/4/1\)](#)
- [8. おすすめの日本語対応ローカル大規模言語モデル](#)
- [9. 国立情報学研究所「gpt-oss-20b」超えの日本語性能うたう](#)
- [10. Compare GPT-4o \(March 2025, chatgpt-4o-latest\) vs Qwen3 8B ...](#)
- [11. 日本語に強いオープンソース LLM が欲しかった人に朗報！ NII ...](#)
- [12. DGX Spark で色々なローカル LLM を動かした比較結果](#)

13. [国産 LLM は作れるのか？ - RakutenAI 3.0 の炎上から考える](#)
14. [Nejumi Leaderboard4 からの考察 - 日本製モデル vs 海外 ...](#)
15. [Qwen モデル比較 : r/ollama](#)
16. [日本語 LLM の進化と課題 : 国産 AI 開発の現在地 | 0h-n0](#)
17. [ノート PC で動くローカル LLM 完全ガイド【2025 年 12 月版】](#)
18. [日本語コーディング性能が最も高いローカル LLM モデルは何か ...](#)
19. [リリース - LLM-jp](#)
20. [OpenAI GPT-4o と Hugging Face Qwen3-Coder-480B-A35B-Instruct ...](#)
21. [RP 約 12 兆トークンの良質なコーパスで学習した新た ...](#)
22. [日本語対応 LLM ランキング 2025 ~ベンチマーク分析レポート](#)
23. [いま国産 LLM はどうなっているのか](#)
24. [第 2 の DeepSeek ショック？ オープンな中国 LLM「Qwen3」 ...](#)
25. [各採択企業の性能評価結果詳細](#)
26. [リコー、強化学習で多段推論を獲得したビジネス文書向け LLM ...](#)
27. [Qwen3-Swallow & GPT-OSS-Swallow](#)
28. [日本語対応 LLM ランキング 2026 ~ベンチマーク分析レポート ...](#)
29. [GPT-4o や o1 より高性能な推論モデル「Qwen3」を Alibaba が発表](#)
30. [極東の島国でしか使われていない謎の言語が上手なモデル](#)
31. [まつにい \(@yugen\\_matuni\) / Posts / X](#)
32. [【賢い選択】LLM のパラメータ数比較ガイド | 2026 年最新 ...](#)
33. [AI が翻訳作業をどれだけ効率化するのか、実際に試してみた](#)
34. [「GENIAC」第 3 期においてリーズニング性能を備えたマルチ ...](#)
35. [llm-jp/llm-jp-4-8b-thinking](#)
36. [Qwen3 はローカル LLM の世界を変えたかも - きしだの Hatena](#)
37. [約 12 兆トークンのコーパスで学習した「LLM-jp-4 8B モデル」 ...](#)
38. [PLaMo 2.2 Prime をリリースしました](#)
39. [GPT-4o vs Qwen3 32B Comparison](#)
40. [shi3z/japanese-llm-benchmark](#)
41. [Qwen3.5 ローカル LLM 実機検証 9B vs 35B-A3B \(MoE\) 7 問 ...](#)
42. [日本語 LLM まとめ](#)
43. [タイトル「Qwen」を検索](#)
44. [Unsloth で MoE モデルを 12 倍高速にファインチューニング](#)

45. [2026年のローカル LLM 事情を整理してみた - DevelopersIO](#)