

2026年（令和8年）2月25日

商用利用可能な同時双方向日本語音声対話モデル 「LLM-jp-Moshi-v1」の公開

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（エヌアイアイNII、くろはし きたお所長：黒橋 禎夫、東京都千代田区）の大規模言語モデル研究開発センター（LLMC）は、主宰する LLM 勉強会（LLM-jp）の対話 WG の活動の成果として、日本語の音声対話に特化した AI モデル「LLM-jp-Moshi-v1」を開発し、商用利用可能なライセンス（Apache2.0）で一般公開しました。日本語で、商用利用可能な同時双方向（Full-duplex）音声対話モデルは世界初となります。本モデルは、音声を逐次入力として受け付けながら音声を出力し、人間同士のような自然な音声対話を実現します。今後の日本語音声対話研究ならびにコールセンターやカウンセリング、その他、リアルタイムな対話が期待される場面での応用が期待されます。

LLM-jp-Moshi-v1 は、日本語の雑談対話データ約 1,000 時間を含む大規模な音声・会話データを用いて学習されており、相槌や返答のタイミングなど、会話の「間合い」を含めたふるまいを学習しています。評価の結果、既存のモデル（J-Moshi, 研究利用のみ可能なモデル）に比べて、音声の聞き取りやすさや会話の自然さに関する指標において改善が確認され、また、人間による主観評価でも、自然性や意味性の向上が示されました。

なお、本モデルは、国立研究開発法人産業技術総合研究所および株式会社 AIST Solutions が提供する AI 橋渡しクラウド（ABCI）を計算環境として開発されました。また、対話 WG は、NII/LLMC 科学主幹の東中竜一郎教授が主担当を務め、早稲田大学の小川哲司教授、慶應義塾大学の高道慎之介准教授との緊密な連携のもと推進される研究グループです。

LLMC では、LLM-jp-Moshi-v1 の公開を通じて、音声会話 AI の透明性・再現性の向上を図るとともに、日本語における音声対話技術の研究開発と社会実装に向けた基盤整備を進めていきます。

1. 今回公開した日本語音声対話モデルの概要

(1) 利用計算資源

- AI 橋渡しクラウド(ABCI) 3.0 を利用して学習を実施。

(2) モデル学習用コーパス

- J-CHAT : ポッドキャストによる対話データ (約 69,000 時間)
 - <https://huggingface.co/datasets/sarulab-speech/J-CHAT>
- LLM-jp-Zoom1 : Zoom によるオンライン対話データ (約 1,000 時間)
 - LLM-jp にて独自に収集した Zoom での雑談対話データ (近日公開予定)

(3) モデル

- パラメータ数 : 約 70 億個 (7B)
- モデルアーキテクチャ : Moshi (<https://github.com/kyutai-labs/moshi>)

(4) チューニング

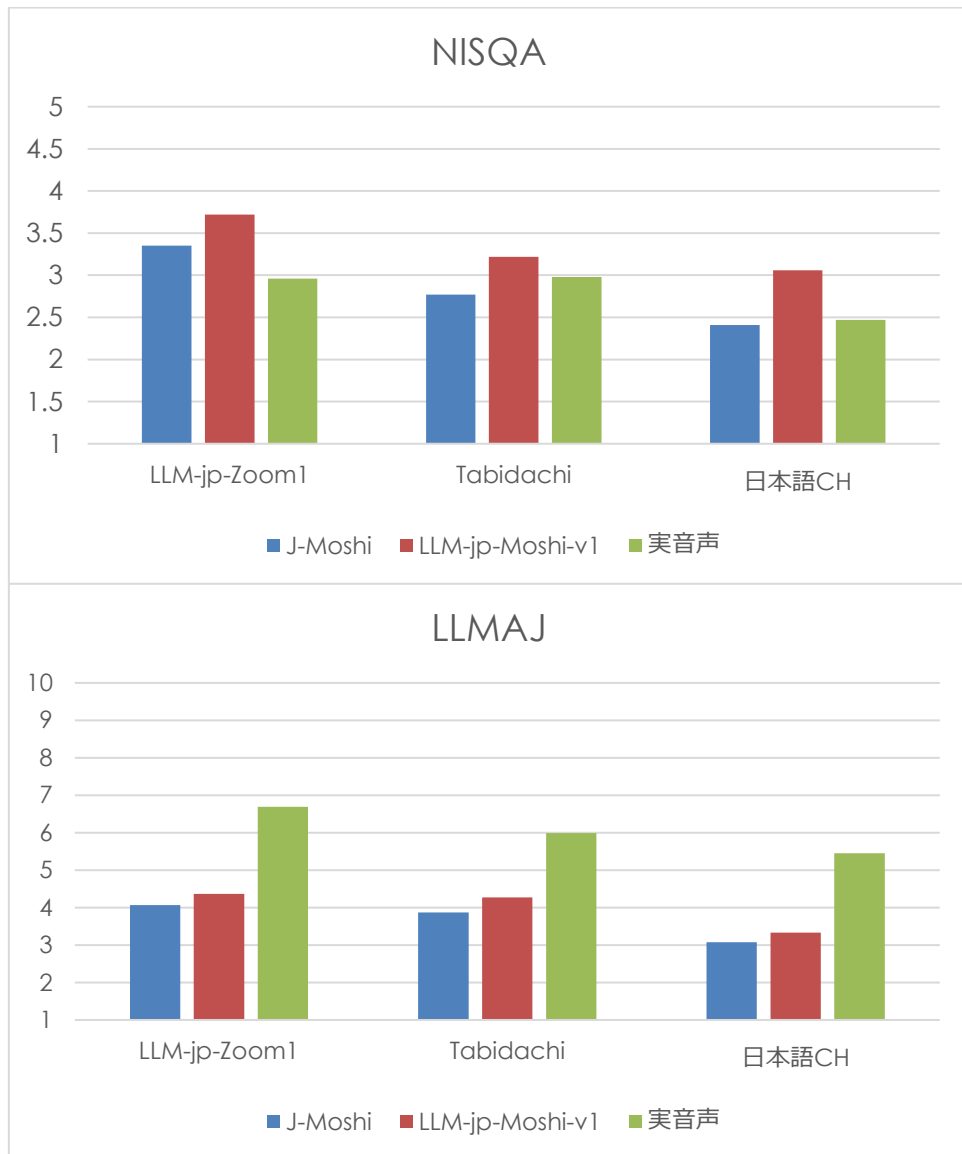
- J-CHAT を用いて事前学習を行い、LLM-jp-Zoom1 を用いてファインチューニングを実施しました。

(5) 評価

- 対話継続タスク (入力音声対話に対して、その続きとなる対話音声を生成させ、その妥当性について評価を行うタスク) において、クラウドワーカーによる客観的自動評価と人間による主観評価の双方を実施し、既存の公開モデルである J-Moshi と比較して、自然性および意味的適切性の両面で優位な性能を示した (下表参照)。
- 入力音声としては、LLM-jp-Zoom1 のヘルドアウトしたテストデータ、Tabidachi (旅行案内の音声対話)、日本語 CallHome (日本語 CH) を用いた。実音声は、入力音声対話の続きとなる実際の人間同士の音声を表す。

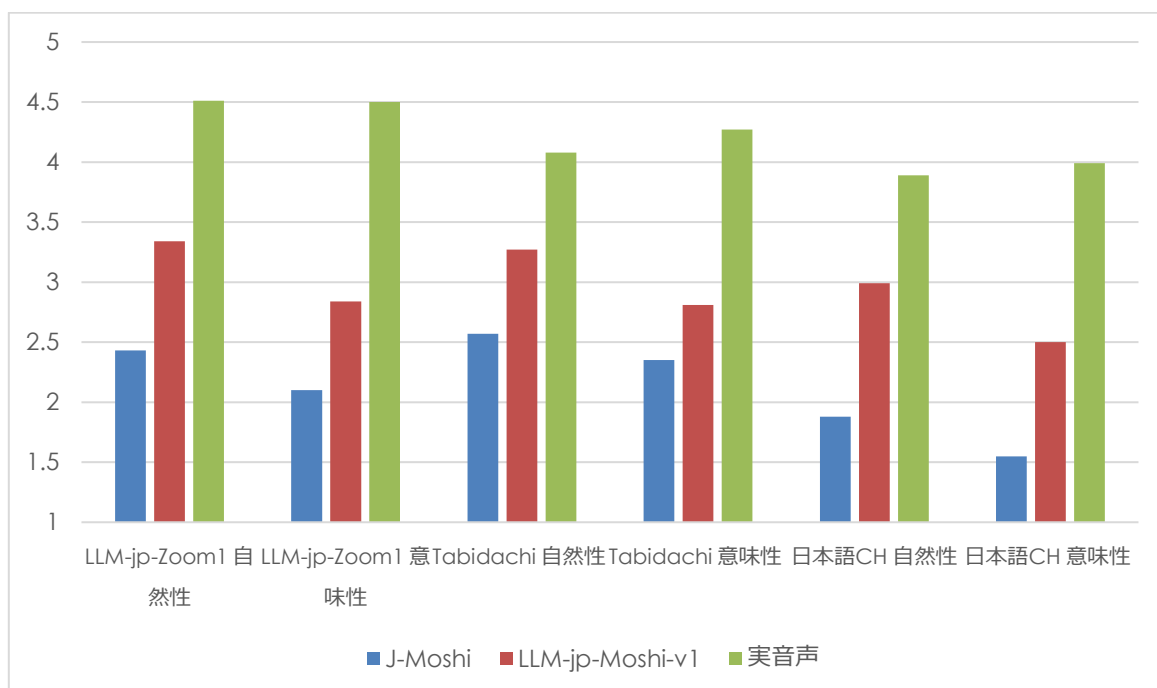
客観評価結果

NISQA は音声の自動評価尺度の値（5段階）を表す。LLMAJ は LLM-as-a-judge の枠組みを用い、大規模言語モデルに対話としての自然性や流暢性を評価させた値（10段階）を表す。



主観評価結果

クラウドソーシングにおける意味性と自然性の人間による主観評価結果の値（5段階）



(6) 開発モデル・ツール・コーパスの公開 URL

- <https://llm-jp.github.io/llm-jp-moshi/>
(上記ページにて、デモ音声をご試聴いただけます)
- <https://huggingface.co/llm-jp/llm-jp-moshi-v1>

2. 今後の予定

- 今回公開した llm-jp-moshi-v1 を起点に、学習データの拡充と学習手法の改善を進め、音声の聞き取りやすさ、応答の自然さ、会話の一貫性、ターンテイキング（相づち・割り込み・間合い）などの品質向上を図ります。
- オンラインで収録した対話データ（LLM-jp-Zoom1）に加え、より多様な会話場面や方言を含むデータを活用し、社会で必要とされる会話能力の向上を目指します。
- 本研究は以下の学会において発表予定です。
 - Yuto Abe, Mao Saeki, Atsumoto Ohashi, Shinnosuke Takamichi, Shiyna Fujie, Tetsunori Kobayashi, Tetsuji Ogawa, Ryuichiro Higashinaka, "Effects of dialogue corpora properties on fine-tuning a Moshi-based spoken

dialogue model,” Proc. International Workshop on Spoken Dialogue Systems (IWSDS), Feb. 2026.

- 阿部雄斗, 佐伯真於, 大橋厚元, 高道慎之介, 藤江真也, 小林哲則, 小川哲司, 東中竜一郎, “Moshi 音声対話モデルの日本語ファインチューニングにおける対話データ特性の影響,” 日本音響学会研究発表会講演論文集, March 2026.

(参考 1) LLM 勉強会 (LLM-jp) の概要

1. NII が主宰する LLM-jp では、自然言語処理及び計算機システムの研究者を中心として、大学・企業等から 2,600 名以上 (2026 年 2 月 25 日現在) が集まり、ハイブリッド会議、オンライン会議、Slack 等を活用して LLM の研究開発について情報共有を行うとともに、共同で LLM 構築等の研究開発を行っています。具体的には、以下の目的で活動しています。
 - オープンかつ日本語に強い LLM の構築とそれに関連する研究開発の推進
 - 上記に関心のある自然言語処理および関連分野の研究者によるモデル構築の知見や最近の研究の発展についての定期的な情報交換
 - データ・計算資源等の共有を前提とした組織横断的な研究者間の連携の促進
 - モデル・ツール・技術資料等の成果物の公開
2. 「コーパス構築 WG」「モデル構築 WG」「チューニング・評価 WG」「マルチモーダル WG」「実環境インタラクション WG」「対話 WG」「学術ドメイン WG」「安全性 WG」「原理解明 WG」を設置し、それぞれ、早稲田大学 河原大輔教授、東北大学 鈴木潤教授、東京大学 宮尾祐介教授、東京科学大学 岡崎直観教授、早稲田大学 尾形哲也教授、名古屋大学 東中竜一郎教授、国立情報学研究所 相澤彰子教授、国立情報学研究所 関根聡特任教授、東京大学 大関洋平准教授を中心に研究開発活動に取り組んでいます。このほか、東京大学 田浦健次郎教授、東京科学大学 横田理央教授 (並列計算手法等) を筆頭に、多数の方々の貢献により活動を進めています。
3. 詳細については、ホームページ <https://llm-jp.nii.ac.jp/> をご参照ください。

(参考 2)

本成果は文部科学省補助金事業「生成 AI モデルの透明性・信頼性の確保に向けた研究開発拠点形成」の結果得られたものです。

(謝辞)

本研究では、国立研究開発法人産業技術総合研究所および株式会社 AIST Solutions が提供する AI 橋渡しクラウド (ABCI) 3.0 を、「ABCI 3.0 開発加速利用」の支援を受けて実施しました。

〈メディアの皆様からのお問い合わせ先〉

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所

総務部 総務企画課 企画・広報チーム

TEL:03-4212-2164 E-mail : media@nii.ac.jp