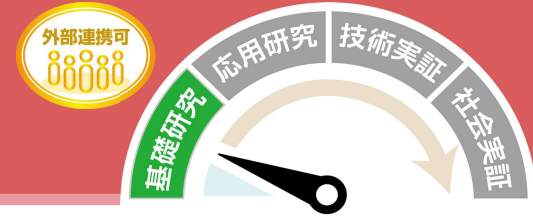


# NICTにおける大規模言語モデル(LLM)への取り組み



## 概要

NICT独自の大規模言語モデルNICT LLMとして、日本語特化版では国内最大規模(3,110億パラメータ)のモデルを始め17個のモデルを国内最速で開発。  
事前学習のみで、質問応答、議論などの基本的な応答が可能なことを確認済

### NICT LLMとユーザとの議論の例

NICT 大規模言語モデル (179B, b16, 議論モード)

ユーザが議論のテーマとなる主張を入力

議論のテーマ  
地球温暖化の解決のために増やすべき自動車のタイプは、電気自動車ではなく、ハイブリッド車である。

その理由は?

電気自動車は、電気を作るために、火力発電所で化石燃料を燃やすので、二酸化炭素を増やすからである。

ユーザが「その理由は？」と尋ねることで、議論のテーマとなった主張の根拠をNICT LLMに生成させる

しかし、ハイブリッド車も二酸化炭素をたくさん排出する。

NICT LLMの生成した根拠にユーザが反論

このため、結論は、「自動車を減らす」ということになる。

ユーザがさらに反論

しかし、自動車を減らすことによる経済的ダメージで人類の存続が危うくなるというのであれば、

ユーザがさらに反論

自動車を減らすことはできない。このため、自動車を減らすことによる経済的ダメージを減らすために、自動車の燃費を向上させる必要がある。自動車の燃費を向上させるためには、自動車の軽量化が必要である。自動車の軽量化のためには、自動車のボディを鉄から炭素繊維に変える必要がある。炭素繊維は、鉄の1.0倍の強度を持ち、重さは鉄の4分の1である。このため、炭素繊維を自動車のボディに使えば、自動車の軽量化が可能になる。

注：議論のプロセスや結論が正しいことを保証するものではありません

しかし、炭素繊維を作るには二酸化炭素をたくさん排出しなければいけない。

このため、結論は、「二酸化炭素を減らすためには、自動車を減らす」ということになる。

テキストを入力(議論)

NICT LLMが議論のテーマの根拠を生成

反論を受けて、NICT LLMが「全ての自動車を減らすべき」だと大胆に主張を変更

反論を受けて、NICT LLMがさらに主張を変更

## 特徴

- 130億から日本語特化型で国内最大の3,110億パラメータまで、パラメータ数や学習用データサイズの様々な組み合わせのLLMを独自開発。モデルサイズや学習データ等の大規模化による高精度化を確認
- LLMの学習によく使われるCommonCrawlの5倍程度の日本語データを含む日本語中心の661億Webページを収集済み。そこから書籍にあるような文章のみを抽出する手法を開発し高品質・大規模なLLM学習用日本語データを整備(5.56TB構築済み。今後さらに増強予定)、外部提供も準備中

## ユースケース

- 事前学習のみで一定の精度で質問への回答、要約、存在しない映画のあらすじの創作、議論等の基本的なやり取りが可能であり、様々な領域への応用可能
- 様々な尖った仮説や将来シナリオを生成してユーザに「気づき」を与えることで有用な意思決定に活用

## 今後の展開

- WISDOM XでLLMの生成内容の裏を取る機構や、複数のLLMが議論を行って偽情報等の生成を抑制、裏取りと合わせてより良い生成を行う機構を開発
- MICSUSなど、対話システムへの導入も計画中
- NICT LLMや関連データを活用した外部連携を実施