

# BÁO CÁO TÌNH HÌNH PHÁT TRIỂN LLM TIẾNG VIỆT

Dựa trên VMLU – Nền tảng đánh giá & xếp hạng năng lực tiếng Việt các mô hình LLM

# NỘI DUNG

Giới thiệu	03
Tình hình phát triển LLM năm 2024	11
Bảng xếp hạng From-scratch models	16
Bảng xếp hạng Fine-tuned models	22
Phụ lục	28

---

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.

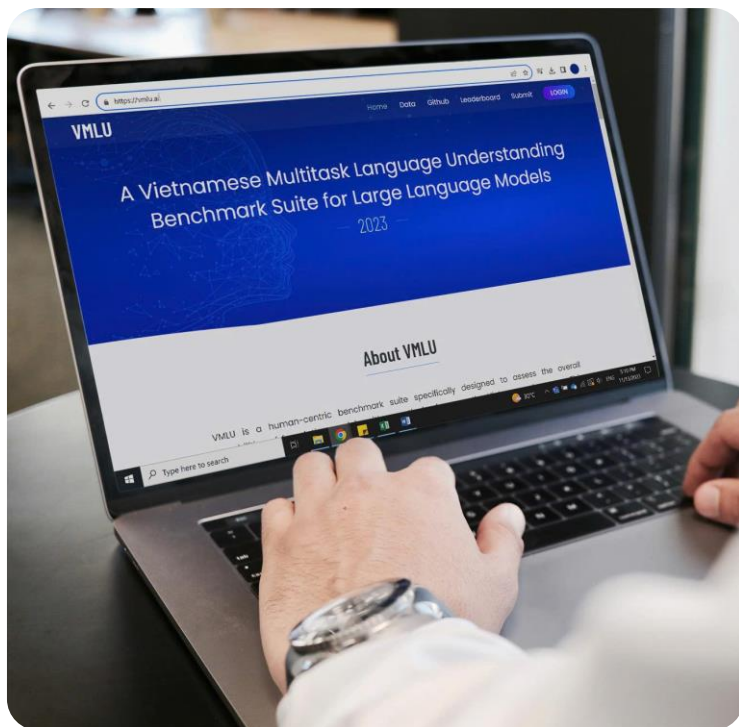


# GIỚI THIỆU

---

Về VMLU | VMLU Leaderboard | Đánh giá của các đơn vị





## VMLU – Vietnamese Multitask Language Understanding Benchmark Suite for Large Language Models

---

Là bộ tiêu chuẩn đánh giá năng lực tiếng Việt cho các mô hình ngôn ngữ lớn (LLM), do Zalo AI phối hợp với Viện Khoa học & Công nghệ Tiên tiến Nhật Bản (JAIST) xây dựng và cung cấp miễn phí từ tháng 11/2023.

Trong bối cảnh lĩnh vực AI vừa “chớm nở” tại Việt Nam, với VMLU, Zalo AI & JAIST mong muốn cung cấp một bộ tiêu chuẩn chung, giúp các đơn vị phát triển LLM có căn cứ đánh giá mô hình và chiến lược huấn luyện phù hợp.

## VMLU CUNG CẤP BỘ DỮ LIỆU & TIÊU CHUẨN ĐÁNH GIÁ TOÀN DIỆN, GỒM:

**10.880**

câu hỏi

**58**

chủ đề

**4**

lĩnh vực

STEM, Khoa học xã hội,  
Khoa học nhân văn & mở rộng

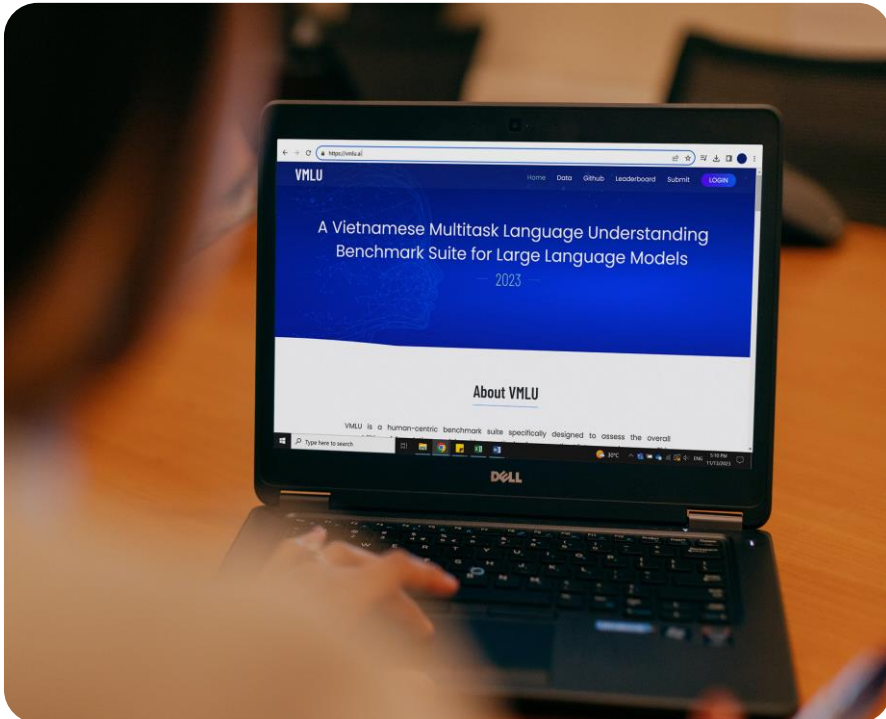
## ĐỘ KHÓ TĂNG DẦN

Tiểu học

Trung học cơ sở

Trung học phổ thông

Chuyên nghiệp (ĐH & sau ĐH)



**VMLU LEADERBOARD** là bảng xếp hạng các mô hình ngôn ngữ lớn (LLM), dựa theo Bộ tiêu chuẩn đánh giá năng lực VMLU.

Các mô hình được đánh giá & xếp hạng trên VMLU Leaderboard bằng cách lấy API (đối với các LLM mã nguồn mở); hoặc các nhà phát triển LLM tự chạy thử nghiệm mô hình của mình trên bộ dataset câu hỏi VMLU, sau đó gửi kết quả cho đội ngũ VMLU để được trả bảng điểm chi tiết theo từng bộ môn và từng lĩnh vực.

VMLU Leaderboard công bố điểm và xếp hạng các LLM khi nhận được yêu cầu từ các đơn vị phát triển.

Với VMLU Leaderboard, các nhà phát triển có thể tự so sánh năng lực mô hình của mình với các LLM hiện có trên thị trường, qua đó có thêm động lực để nâng cao trình độ huấn luyện và thúc đẩy quá trình phát triển AI, phục vụ nhu cầu thực tiễn của người Việt.

Lưu ý:

\*Bảng xếp hạng VMLU Leaderboard có giá trị tham khảo trong một khoảng thời gian nhất định nhằm cung cấp góc nhìn về năng lực các LLM cũng như bối cảnh phát triển LLM tại Việt Nam, và cần được sử dụng một cách thận trọng.

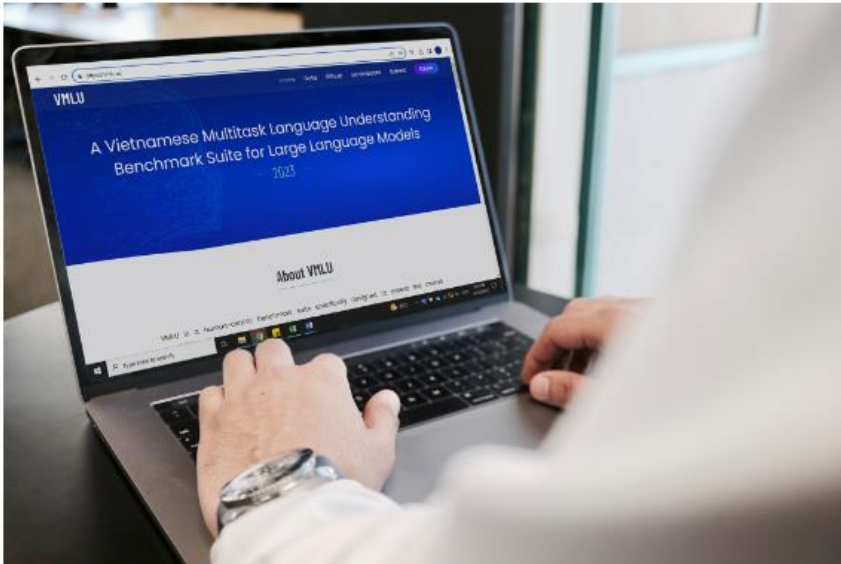
\*Việc đánh giá & xếp hạng các LLM có thể có sai lệch vì một chỉnh sửa nhỏ trong quá trình prompting cũng có thể dẫn đến kết quả không hoàn toàn chính xác. Điều này rất đáng lưu tâm vì một số LLM có mã nguồn đóng và không thể truy cập công khai.

THÔNG TIN VÀ TRUYỀN THÔNG > CHUYỂN ĐỔI SỐ

## Việt Nam đang làm tốt việc bản địa hóa các mô hình ngôn ngữ lớn

Trong bối cảnh AI tạo sinh ngày càng ứng dụng rộng khắp, lãnh đạo cấp cao của Intel cho rằng, Việt Nam đang làm tốt việc bản địa hóa các mô hình ngôn ngữ lớn.

Trí tuệ nhân tạo tạo sinh hay Generative AI (AI tạo sinh) đang là tâm điểm của giới công nghệ toàn cầu. Đây là loại AI có khả năng sinh ra các nội dung như văn bản, hình ảnh, âm thanh và video, thông qua việc dự đoán dựa trên lượng dữ liệu khổng lồ đã được đào tạo.



Bộ tiêu chuẩn VMLU giúp đánh giá khả năng của các mô hình ngôn ngữ lớn tiếng Việt. Ảnh: ZaloAI

Znews

## Thước đo đánh giá năng lực tiếng Việt cho cộng đồng AI

Hai chuyên gia đặt kỳ vọng bộ tiêu chuẩn đánh giá tiếng Việt sẽ trở thành một thước đo chung cho cộng đồng AI Việt. "Chúng tôi hy vọng VLMU sẽ..."



SỨC MẠNH SỐ

## GenAI "made in Vietnam" góp phần giúp doanh nghiệp chuyển đổi số toàn diện

Vào top 4 bảng xếp hạng "Năng lực tiếng Việt"

Sau khi ra mắt chỉ vài tháng, mô hình VIGPT 1.6 tỷ tham số (VIGPT-1.6B) do VinBigdata (thuộc Vingroup) phát triển đã nhanh chóng chứng minh chất lượng khi trở thành mô hình ngôn ngữ đứng trong top 4 bảng xếp hạng "Năng lực tiếng Việt". Mô hình GenAI "made in Vietnam" này chỉ đứng sau ChatGPT, GPT-4 và Gemini. Kết quả cho thấy khả năng nổi bật của VIGPT trong việc nắm bắt tri thức và tư duy ngôn ngữ tiếng Việt.

arXiv > cs > arXiv:2403.15470

Computer Science > Computation and Language

[Submitted on 20 Mar 2024]

## Vi-Mistral-X: Building a Vietnamese Language Model with Advanced Continual Pre-training

James Vo

The advancement of Large Language Models (LLMs) has significantly transformed the field of natural language processing, although the focus on English-centric models has created a noticeable research gap for specific languages, including Vietnamese. To address this issue, this paper presents vi-mistral-x, an innovative Large Language Model designed expressly for the Vietnamese language. It utilizes a unique method of continual pre-training based on the Mistral architecture, which incorporates grouped-query attention and sliding window attention techniques. This model, vi-Mistral-X, marks a significant step forward in improving the understanding and generation of the Vietnamese language. It introduces an additional phase of continual pre-training, specifically adapted for Vietnamese, enhancing the model's capability in understanding complex language nuances and generating accurate, context-aware Vietnamese text. Through comprehensive testing on various benchmarks, vi-mistral-x has shown to outperform existing Vietnamese LLMs in several key areas, including text classification, question answering, and text generation. Particularly, in the Vietnamese Multitask Language Understanding (VMLU) benchmark, vi-mistral-x sets a new standard, outperforming other available models significantly. This paper highlights the critical role of continual pre-training in advancing language-specific LLMs and opens new avenues for the development of multilingual models. We aim for vi-mistral-x to not just be an important asset for processing the Vietnamese language but also to encourage more advancements in creating large language models for languages that are less represented.

Mindmaid

Apr 19, 2024 • 4 min read



Phát triển sản phẩm từ LLM nguồn mở, đặc biệt LLM nguồn mở tiếng Việt đang là mối quan tâm chung của nhiều AI Developer hiện nay. Năm bắt nhu cầu đó, tôi 17/4 vừa qua, Mindmaid đã tổ chức thành công webinar "Ứng dụng LLM nguồn mở tiếng Việt vào các bài toán thực tế" với khách mời là Mr Quân Nguyễn - cựu kỹ sư của OpenAI và đồng tác giả của nhiều mô hình nguồn mở như VinaLlama, Vistral...

## VinaLLaMA: LLaMA-based Vietnamese Foundation Model

Anonymous ACL submission

### 4 Evaluation

In our research, we utilized three distinct evaluation benchmarks: VLSP, VMLU, and the Vicuna Benchmark, with the latter being translated into Vietnamese by VinAI Research<sup>3</sup>. Specifically, we conducted separate experiments for pre-trained and instructional fine-tuning models. This methodology allowed us to compare the baseline capabilities of the pre-trained models against their performance post-instructional fine-tuning.

VMLU đóng vai trò quan trọng trong quá trình phát triển LLMs cho tiếng Việt của chúng tôi vì cho phép đánh giá toàn diện khả năng xử lý tiếng Việt và hiểu biết văn hóa Việt Nam của mô hình.

Việc sử dụng VMLU đã giúp chúng tôi phát triển LLMs không chỉ thành thạo về mặt ngôn ngữ mà còn có hiểu biết sâu sắc về văn hóa Việt Nam, từ đó đáp ứng tốt hơn nhu cầu của người dùng Việt.

Dựa trên kết quả sử dụng và đánh giá mô hình, chúng tôi đã hoàn thiện và phát hành thành công Vistral-7B-Chat, một mô hình LLM mở dành riêng cho tiếng Việt, được cộng đồng người Việt đón nhận với hơn 100.000 lượt tải từ Hugging Face chỉ sau 7 tháng.

## **GS. Nguyễn Hữu Thiện**

Khoa Khoa học máy tính, ĐH Oregon, Hoa Kỳ

VMLU cho thấy sự hữu ích trong việc đánh giá hiệu suất của mô hình LLMs trên tiếng Việt, giúp chúng ta hiểu rõ hơn về khả năng của mô hình trên QA task. Qua đó cải thiện mô hình toàn diện hơn.

Từ góc nhìn tổng quan, tôi thấy bộ benchmark còn những tiềm năng phát triển sâu rộng. Trước tiên là khả năng nâng cao độ phức tạp của câu hỏi, độ đa dạng của dữ liệu và tính chính xác của label để đánh giá các mô hình chính xác hơn.

Bên cạnh đó, phát triển thêm những task khác như reasoning, text generation, code generation, text summarization, và text classification nhằm đánh giá toàn diện các mô hình. Cuối cùng, phát triển Responsible AI để đảm bảo các mô hình có thể hoạt động đúng cách.

## **TS. Bạch Hưng Nguyên**

Principal Engineer, Microsoft





Đây là một bộ benchmark có dữ liệu tương đối hoàn chỉnh và toàn diện để đánh giá năng lực về tri thức của LLM cho tiếng Việt. VMLU không chỉ hữu ích để đánh giá chất lượng LLM ở từng giai đoạn phát triển, mà còn là thước đo cho tính hiệu quả của các thử nghiệm của chúng tôi trong quá trình huấn luyện.

Sự ra đời của VMLU là cần thiết đối với cộng đồng nghiên cứu AI. Đây sẽ là “bàn đạp” để thúc đẩy sự phát triển của AI nói chung và LLM nói riêng, vì phải có những bộ benchmark tốt thì chúng ta mới có căn cứ huấn luyện những mô hình chất lượng cao.

## **TS. Đặng Trần Thái**

Trưởng phòng xử lý ngôn ngữ tự nhiên,  
Khối công nghệ trợ lý ảo VinBigData



Mô hình của chúng tôi được fine-tuned dựa trên Qwen1.5-14B-Chat, với tập dữ liệu 2 triệu sample (dữ liệu private do chúng tôi xây dựng). Trong thời gian tới, LLM là hướng phát triển quan trọng của VNPT AI để ứng dụng vào nhiều bài toán như chatbot, callbot, tóm tắt văn bản, tổng hợp thông tin,...

VMLU là một benchmark rất chất lượng để đánh giá và so sánh các mô hình. Chúng tôi cũng sử dụng các tập đánh giá khác, nhưng VMLU vẫn là một tập dữ liệu quan trọng giúp quá trình đánh giá chất lượng các LLM của chúng tôi nhanh chóng và chính xác hơn.

## **Đội ngũ phát triển LLM, VNPT AI**

VNPT AI



Là một dự án hợp tác giữa Zalo AI cùng Viện Khoa học & Công nghệ Tiên tiến Nhật Bản (JAIST), nhằm cung cấp bộ tiêu chuẩn đánh giá năng lực tiếng Việt cho các mô hình ngôn ngữ lớn, thông qua VMLU, Zalo AI mong muốn tạo động lực để các đơn vị nghiên cứu AI nâng cao trình độ huấn luyện LLM, từ đó đóng góp tích cực cho sự phát triển của cộng đồng AI Việt Nam. Zalo AI hy vọng rằng, sự phát triển của tầng nghiên cứu sẽ tạo tiền đề cho sự phát triển của tầng ứng dụng, và cho ra đời các sản phẩm AI đa dạng và hữu ích, phục vụ nhu cầu thực tiễn của người Việt.

## TS. Châu Thành Đức

Trưởng Bộ phận Nghiên cứu & phát triển tại Zalo AI



# TÌNH HÌNH PHÁT TRIỂN LLMs NĂM 2024

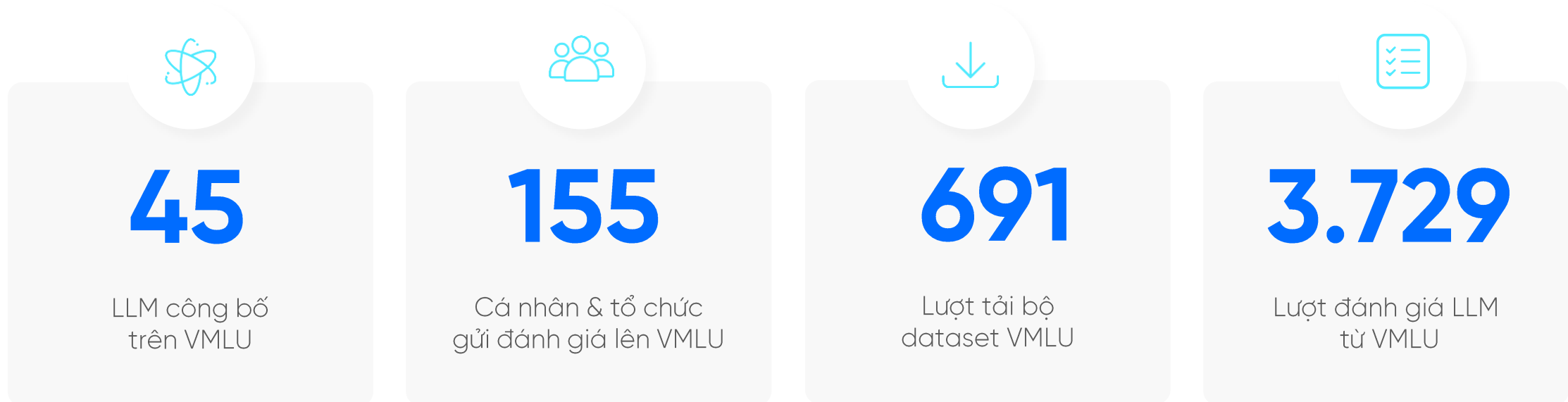
---

Hiện trạng & Xu hướng phát triển LLM cho người Việt | Nhận định của chuyên gia

# Sự nở rộ nhanh chóng của LLM hướng đến người dùng Việt

Chỉ trong vòng hơn 2 năm kể từ khi làn sóng AI tạo sinh (GenAI) lan tỏa mạnh mẽ, và trong bối cảnh lĩnh vực AI còn khá mới tại Việt Nam, đồng thời việc phát triển LLM còn nhiều trở ngại (thiếu dữ liệu do tiếng Việt thuộc nhóm low-resource, hạ tầng đắt đỏ và nguồn lực con người còn ít kinh nghiệm), VMLU đã ghi nhận số lượng LLM và số lượng đơn vị/cá nhân gửi đánh giá lên VMLU khá ấn tượng.

Tuy số lượng này còn khiêm tốn khi so sánh với các quốc gia phát triển trên thế giới, nhưng phản ánh sự tích cực của các đơn vị trong việc tiếp cận công nghệ mới, cũng như triển vọng của lĩnh vực AI tại Việt Nam.





# Các nhóm nghiên cứu nước ngoài cũng tối ưu LLM cho tiếng Việt

Với vai trò là tiêu chuẩn đánh giá năng lực LLM, VMLU ghi nhận nhiều mô hình LLM mới từ các đơn vị cả trong và ngoài nước. Việc gửi đánh giá năng lực LLM lên VMLU cho thấy các đơn vị có chiến lược phát triển LLM hướng đến người dùng Việt Nam.

Các trường đại học tại Việt Nam cũng đóng góp vào quá trình nghiên cứu & xây dựng LLM cho người Việt, thể hiện sự tiến bộ của lĩnh vực AI trong môi trường giáo dục chuyên nghiệp, hướng đến sự phát triển bền vững.

## NHÀ PHÁT TRIỂN LLM NƯỚC NGOÀI TIÊU BIỂU

**DAMO**  
ALIBABA DAMO ACADEMY

DAMO Academy  
Alibaba Group

**O** UNIVERSITY OF OREGON

UONLP X Ontocord  
ĐH Oregon

SAMSUNG SDS  
R&D CENTER  
(SDSRV)

SDSRV Teams  
Samsung

Zalo AI

Zalo AI

VINBIGDATA

VinBigData

VNPT AI  
DEFINE THE FUTURE

VNPT AI

Ngân hàng số  
**CAKE**  
by VPBank

BeFinancial

viettel  
solutions

VTS DASC

BK  
TRHCM

ML4U

# LLM do người Việt huấn luyện đạt vị trí cao trong bảng xếp hạng VMLU Leaderboard

Nhìn chung, VMLU Leaderboard ghi nhận nhiều mô hình do người Việt huấn luyện lọt vào Top 10 LLM có năng lực tiếng Việt vượt trội nhất, với cả 2 dạng From-scratch models và Fine-tuned models.

Các mô hình được cải tiến với điểm số cập nhật liên tục, cho thấy nỗ lực trong nghiên cứu AI để phát triển và bản địa hóa LLM cho người Việt.



## FROM-SCRATCH MODELS

KiLM-13b-v24.7.1

ViGPT-1.6B-v1



## FINE-TUNED MODELS

VNPTAI.IO-Large-v3

Llama3-ZAI

CakebyVPBank-Large

Llama3-ViettelSolutions-8B

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.

# Nhận định của chuyên gia về sự phát triển LLM tại Việt Nam



“

Số lượng các mô hình ngôn ngữ lớn LLMs tại Việt Nam gia tăng đã cho thấy sự quan tâm của đông đảo các tổ chức, cá nhân đối với việc thúc đẩy tính ứng dụng của GenAI. Tuy nhiên, việc huấn luyện một mô hình ngôn ngữ lớn cần đầu tư khá nhiều thời gian, công sức và nguồn lực chi phí. Trong tương lai, tôi nghĩ rằng xu hướng phát triển LLM tại Việt Nam sẽ thiên về tận dụng những mô hình LLM mở như Llama. Từ đó chuyển đổi phù hợp với các bài toán và dữ liệu chuyên ngành. Dù vậy vẫn sẽ có các nhóm nghiên cứu tiếp tục huấn luyện mô hình ngôn ngữ lớn của riêng mình.

## GS. Nguyễn Lê Minh

Giám đốc Trung tâm nghiên cứu Interpretable AI, Viện Khoa học và Công nghệ Tiên tiến Nhật Bản (JAIST)



# BẢNG XẾP HẠNG FROM-SCRATCH MODELS

---

BXH toàn diện | BXH năng lực STEM | BXH lĩnh vực Khoa học xã hội | BXH lĩnh vực  
Khoa học nhân văn

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.





Với dạng From-scratch model, các mô hình của quốc tế vẫn chiếm ưu thế về năng lực tiếng Việt. Ít đơn vị của Việt Nam chọn huấn luyện mô hình theo hình thức này. Tuy nhiên, các nhà phát triển LLM của Việt Nam đang cho thấy sự bứt phá để rút ngắn khoảng cách về năng lực huấn luyện From-scratch models với thế giới.



KiLM-13b-v24.7.1 (Zalo AI) là from-scratch model duy nhất của Việt Nam góp mặt ở vị trí Top 5 BXH, xếp ở vị trí thứ 2 chỉ sau mô hình Llama-3-70B của Meta. Mô hình của Zalo AI dẫn đầu ở lĩnh vực Khoa học nhân văn, một điểm sáng cho nỗ lực nghiên cứu AI của người Việt.

Một mô hình khác của Việt Nam nằm trong Top 10 là ViGPT-1.6B-v1 (Vin BigData), xếp ở vị trí thứ 8.



Llama-3-70B (Meta) là mô hình dẫn đầu BXH, thể hiện năng lực vượt trội ở lĩnh vực Khoa học xã hội và mở rộng. Tuy nhiên, mô hình này xếp sau GPT-4 (OpenAI) ở lĩnh vực STEM và KiLM-13b-v24.7.1 (Zalo AI) ở lĩnh vực Khoa học nhân văn.

# Top 10 From-scratch models



## Theo năng lực tiếng Việt tổng quát

**1** Llama-3-70B  
(Meta)  
66.44

**2** KiLM-13b-v24.7.1  
(Zalo AI)  
66.07

**3** GPT-4  
(OpenAI)  
65.53

4	Gpt-4o-mini	OpenAI	62.87
5	gemma-2-9b-it	Google	59.04
6	gemini	Google	51.03
7	ChatGPT	OpenAI	46.33
8	ViGPT-1.6B-v1	Vin BigData	42.34
9	gemma-7b-it	Google	41.9
10	microsoft/Phi-3-small-128k-instruct	Microsoft	41.24

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.

# Top 10 From-scratch models



Theo năng lực STEM (Khoa học, Công nghệ, Kỹ thuật và Toán)

**1** GPT-4  
(OpenAI)  
63.84

**2** Llama-3-70B  
(Meta)  
61.7

**3** KiLM-13b-v24.7.1  
(Zalo AI)  
60.29

4	Gpt-4o-mini	OpenAI	58.00
5	gemma-2-9b-it	Google	56.31
6	ChatGPT	OpenAI	43.24
7	gemini	Google	42.8
8	gemma-7b-it	Google	39.95
9	microsoft/Phi-3-small-128k-instruct	Microsoft	39.31
10	microsoft/Phi-3-small-8k-instruct	Microsoft	38.72

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.

# Top 10 From-scratch models



Theo năng lực Social Science (Khoa học Xã hội)

**1** Llama-3-70B  
(Meta)  
74.91

**2** KiLM-13b-v24.7.1  
(Zalo AI)  
73.15

**3** GPT-4  
(OpenAI)  
71.78

4	Gpt-4o-mini	OpenAI	70.95
5	gemma-2-9b-it	Google	65.79
6	gemini	Google	60.31
7	ChatGPT	OpenAI	51.67
8	ViGPT-1.6B-v1	Vin BigData	48.72
9	gemma-7b-it	Google	44.93
10	microsoft/Phi-3-small-128k-instruct	Microsoft	44.82

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.





## Theo năng lực Humanities (Khoa học Nhân văn)

**1** KiLM-13b-v24.7.1  
(Zalo AI)  
71.85

**2** Llama-3-70B  
(Meta)  
68.74

**3** GPT-4  
(OpenAI)  
66.14

4	Gpt-4o-mini	OpenAI	65.03
5	gemma-2-9b-it	Google	60.80
6	gemini	Google	55.35
7	ViGPT-1.6B-v1	Vin BigData	47.20
8	ChatGPT	OpenAI	46.96
9	gemma-7b-it	Google	43.39
10	microsoft/Phi-3-small-8k-instruct	Microsoft	42.32

\* From-scratch models: Kỹ thuật huấn luyện LLM từ đầu, bao gồm việc khởi tạo tham số, quyết định kiến trúc mô hình và thuật toán huấn luyện, trên một tập dữ liệu nhất định, nhằm mục đích tạo ra một LLM mới hoàn toàn theo mục tiêu huấn luyện.



# BẢNG XẾP HẠNG FINE-TUNED MODELS

---

BXH toàn diện | BXH năng lực STEM | BXH lĩnh vực Khoa học xã hội | BXH lĩnh vực  
Khoa học nhân văn

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.



Theo VMLU Leaderboard, **9/10** mô hình nằm trong **Top 10 Fine-tuned models** là do các đơn vị AI trong nước phát triển, gồm các mô hình của VNPT AI, Cake by VPBank, Zalo AI, Viettel Solutions, Trường ĐH Bách Khoa - ĐHQG TP.HCM. Điều này cho thấy Fine-tuned models là hướng đi phổ biến hơn cả khi các đơn vị nghiên cứu AI của Việt Nam phát triển mô hình ngôn ngữ lớn.

**VNPT AI** thể hiện năng lực vượt trội khi có mô hình **VNPTAI.IO-Large-v3** đang dẫn đầu **BXH Fine-tuned models**, xét trên năng lực tổng quát, năng lực ở lĩnh vực Khoa học xã hội, Khoa học nhân văn và mở rộng. Kết quả này vượt trội so với các mô hình từ đơn vị quốc tế như: **SeaLLM-7B-v2.5** (DAMO Academy – Alibaba Group), **Vistral-7B-Chat** (UONLP x Ontocord – thuộc ĐH Oregon, Hoa Kỳ).

Bên cạnh các mô hình AI của doanh nghiệp, **LLM** do nhóm nghiên cứu thuộc các trường **Đại học của Việt Nam huấn luyện cũng xuất hiện trong BXH Top 16 Fine-tuned models**, tiêu biểu như: **ML4uLLM-7B-Chat** (Nhóm ML4U, Trường ĐH Bách Khoa, ĐHQG TP.HCM), **T-Llama-7b** (ĐH FPT TP.HCM). Điều này cho thấy sự tiến bộ của lĩnh vực AI trong môi trường giáo dục Đại học của Việt Nam, hướng đến sự phát triển bền vững.



## Theo năng lực tiếng Việt tổng quát

**1** VNPTAI.IO-Large-v3  
(VNPT AI)  
75.4

**2** CakebyVPBank-Large  
(BeFinancial)  
73.99

**3** VNPTAI.IO-Large-v2  
(VNPT AI)  
72.65

4	CakebyVPBank-Small	BeFinancial	65.82
5	Llama3-ZAI	Zalo AI	65.34
6	Llama3-ViettelSolutions-8B	VTS DASC	56.20
7	VNPTAI.IO-14B	VNPT AI	55.83
8	Vintern-3B-beta	5CD-AI	54.81
9	SeaLLM-7B-v2.5	DAMO Academy	53.30
10	MI4uLLM-7B-Chat	ML4U	52.08

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.



## Theo năng lực STEM (Khoa học, Công nghệ, Kỹ thuật và Toán)

**1** CakebyVPBank-Large  
(BeFinancial)  
77.75

**2** VNPTAI.IO-Large-v3  
(VNPT AI)  
74.09

**3** VNPTAI.IO-Large-v2  
(VNPT AI)  
70.07

4	CakebyVPBank-Small	BeFinancial	63.95
5	Llama3-ZAI	Zalo AI	59.17
6	Vintern-3B-beta	5CD-AI	51.70
7	VNPTAI.IO-14B	VNPT AI	51.64
8	Llama3-ViettelSolutions-8B	VTS DASC	51.52
9	SeaLLM-7B-v2.5	DAMO Academy	49.35
10	Arcanic Cono 1.5	Arcanic AI	45.11

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.



## Theo năng lực Social Science (Khoa học Xã hội)

**1** VNPTAI.IO-Large-v3  
(VNPT AI)  
80.34

**2** VNPTAI.IO-Large-v2  
(VNPT AI)  
79.50

**3** CakebyVPBank-Large  
(BeFinancial)  
78.11

4	Llama3-ZAI	Zalo AI	71.73
5	CakebyVPBank-Small	BeFinancial	70.68
6	Llama3-ViettelSolutions-8B	VTS DASC	62.42
7	VNPTAI.IO-14B	VNPT AI	61.75
8	Vintern-3B-beta	5CD-AI	61.01
9	SeaLLM-7B-v2.5	DAMO Academy	60.66
10	SDSRV-7B-chat	SDSRV teams	60.55

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.





## Theo năng lực Humanities (Khoa học Nhân văn)

**1** VNPTAI.IO-Large-v3  
(VNPT AI)  
76.16

**2** VNPTAI.IO-Large-v2  
(VNPT AI)  
73.77

**3** Llama3-ZAI  
(Zalo AI)  
70.98

4	CakebyVPBank-Large	BeFinancial	70.38
5	CakebyVPBank-Small	BeFinancial	67.63
6	Llama3-ViettelSolutions-8B	VTS DASC	60.12
7	Vintern-3B-beta	5CD-AI	58.41
8	VNPTAI.IO-14B	VNPT AI	58.09
9	MI4uLLM-7B-Chat	ML4U	56.86
10	SeaLLM-7B-v2.5	DAMO Academy	55.95

\*\* Fine-tuned models: Kỹ thuật tinh chỉnh LLM đã được huấn luyện trước trên một tác vụ cụ thể hoặc một tập dữ liệu nhất định, để có được LLM mới với hiệu suất tối ưu và phù hợp hơn với mục đích sử dụng chuyên biệt.



# PHỤ LỤC

---

Về JAIST & Zalo AI | Các hoạt động thúc đẩy cộng đồng AI Việt Nam | Website VMLU

# Về Viện Khoa học & Công nghệ Tiên tiến Nhật Bản (JAIST)



Viện Khoa học và Công nghệ Tiên tiến Nhật Bản (JAIST) là viện nghiên cứu hàng đầu tại Nhật Bản được thành lập tháng 10/1990.

JAIST là viện sau đại học quốc lập đầu tiên tại Nhật Bản với mục tiêu tạo ra những đóng góp quan trọng cho sự phát triển xã hội trên toàn cầu, mang lại môi trường nghiên cứu và giáo dục cho các nhà lãnh đạo tương lai, những người làm chủ khoa học và công nghệ cho xã hội tương lai.

Tìm hiểu thêm về JAIST: <https://www.jaist.ac.jp/>



## Zalo AI

Zalo AI là đơn vị tiên phong về AI tại Việt Nam, từ năm 2017. Zalo hiện sở hữu 4 AI Lab, với 80 nhà nghiên cứu làm chủ các công nghệ nền tảng như: Xử lý ngôn ngữ tự nhiên (Natural Language Processing), Xử lý giọng nói (Speech processing), Thị giác máy tính (Computer vision) và Trí tuệ nhân tạo tạo sinh (Generative AI).

Zalo AI sở hữu hạ tầng AI mạnh mẽ, trong đó có hệ thống máy chủ gồm 8 DGX H100 có năng lực xử lý mạnh mẽ với hiệu suất lên đến 256 petaFLOPS (Floating-point Operations Per Second).

Các sản phẩm của Zalo AI có thể kể đến: trợ lý Kiki, công nghệ nhận dạng tiếng nói (dictation và voice-to-text), công nghệ tổng hợp tiếng nói (text-to-speech), công nghệ nhận dạng khuôn mặt (FaceID), công nghệ định danh điện tử (eKYC), công nghệ AI tạo sinh (AI Avatar, AI Sticker),...

Tìm hiểu thêm về Zalo AI: <https://zalo.ai/>

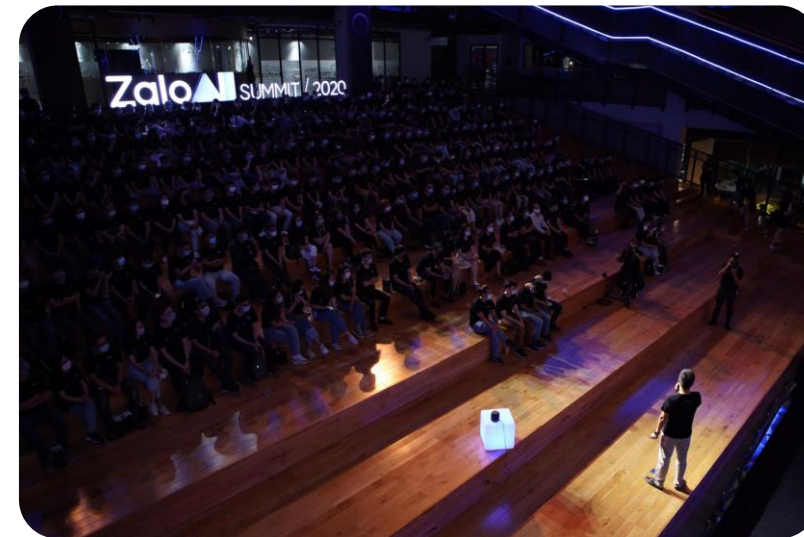


# Các hoạt động khác của Zalo AI nhằm thúc đẩy cộng đồng nghiên cứu AI Việt Nam



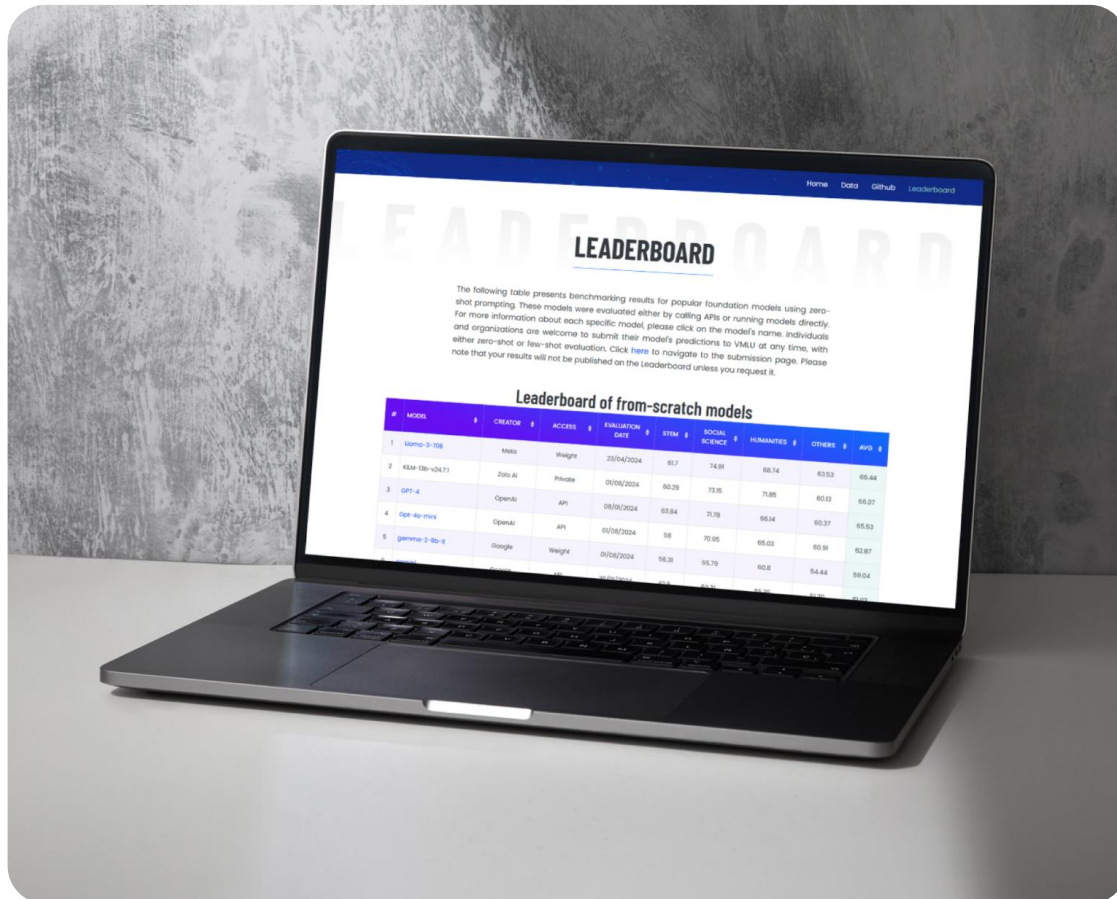
## Zalo AI Challenge

Cuộc thi AI thường niên, từ năm 2017 diễn ra vào Quý IV hàng năm



## Zalo AI Summit

Diễn đàn AI thường niên, từ năm 2017 diễn ra vào Quý IV hàng năm



## Tìm hiểu thêm

Tìm hiểu thêm để sử dụng VMLU, hoặc theo dõi VMLU Leaderboard:

- Website: <https://vmlu.ai/>
- Github: <https://github.com/ZaloAI-Jaist/VMLU/>

## Liên hệ

Với bất kỳ thắc mắc nào về VMLU, vui lòng liên hệ:

- Email: [contact@vmlu.ai](mailto:contact@vmlu.ai)

Để hợp tác, vui lòng gửi email về địa chỉ:

- Email: [developer@vmlu.ai](mailto:developer@vmlu.ai)





Thank you

