

Building and Nurturing AI Development in Vietnam

BY HUNG BUI, MINH HOAI NGUYEN, DAT QUOC NGUYEN, LINH PHAM, AND DINH PHUNG

IS IT POSSIBLE for a developing country like Vietnam to be a competitive player on the world stage in cutting-edge artificial intelligence (AI) research and development? Will it be able to tap into the \$US15.7 trillion projected for the AI global economy by 2030? For Vietnam, these questions often went unchallenged; contemplating answers was daunting. VinAI Research, however, aims to embrace these challenges by laying the groundwork for AI innovation and growth for the region.

Founded in 2019, VinAI leapfrogged to the 20th ranking on Thundermark Capital's list of "Global AI Research Companies" by 2022, and was the only Southeast Asian (SEA) representative on the list.^a Today, its research has been shared in more than 100 publications from top AI venues spanning across three main areas: machine learning (ICML, NeurIPS, ICLR), computer vision (CVPR, ICCV, ECCV), natural language processing (ACL, EMNLP, InterSpeech), reflecting some of the highest concentrations of AI expertise worldwide. Coincidentally (or not), Vietnam, which had never appeared on global AI research rankings before, jumped to 26th worldwide the same year.

a <http://bit.ly/3YZ2EaZ>



AI recognizes facial expressions to promptly issue warnings when the driver is tired and sleepy.

Research at VinAI aims to push the frontiers of AI and disrupt the status quo. We question the theoretical foundations and practices in machine learning, and deep learning, and we investigate how they can enable new AI technology in natural language understanding and computer vision—two areas of applications fundamental to cognitive AI capabilities with unique challenges in our geographic locations. Our leading research exemplars in machine learning, include the young mathematics field of optimal transport for AI, self-supervised learning and domain adaptation, and in computer vision, including 3D vision, robust AI, few-shot problems, and image restora-

tion and enhancement. In natural language processing (NLP), we are the creator of the top Vietnamese machine translation system as well as many large language models for Vietnamese.

We are also naturally drawn toward problems of developing countries, which might otherwise be overlooked in the research community. Take low-resource (LR) language models, for example. At this writing, ChatGPT has stunned the world and completely dominates social media.^b There is no denying the success of large language models is easily one of the biggest highlights of AI research in recent years. But with close

b <https://openai.com/blog/chatgpt/>

to 200 billion parameters to be trained, trillions of words needed, and millions of dollars to train a model such as GPT-3, what are the chances for LR languages like Vietnamese with little, and underrepresented, data available? Our goal is to not only create new tools and knowledge agonistic to LR languages in the region, but also to create the best NLP technology for Vietnamese. As a result, we introduced our first public large-scale monolingual language model pretrained for Vietnamese in 2020.^c Named after the famed Vietnamese noodle soup, PhoBERT¹ was built using

c <https://github.com/VinAIR-research/PhoBERT>

a 20GB pretraining dataset of 145 million Vietnamese word-segmented sentences. Multiple downstream NLP tasks such as part-of-speech tagging, named entity recognition, dependency parsing, natural language inference, and text classification have achieved state-of-the-art performances based on PhoBERT. It is publicly available with 100K+ downloads per month.^d

Another highly challenging problem for the region is language translation, which includes not only the traditional machine translation (MT, foreign text to local text), but also speech translation (S2T, foreign language speech to local text). Vietnam has achieved rapid economic growth in the last two decades, becoming an attractive destination for tourism, trade, and investment. Understandably, one common hurdle identified by both foreign and domestic business is the communication barrier—as Vietnamese (or ngôn ngữ tiếng Việt) is also an incredibly complex language to learn.

Latin-based with six different tones, a sentence can take a totally different meaning in Vietnamese without correct tonal usage. Indeed, without proper use of tones, ‘muoi’, for example, can mean salt (muối), mosquito (muỗi), each (mỗi), lip (môi), or fishing bait (mồi). High-quality text and speech translation to the target language, such as Vietnamese, has become more important than ever. The current publicly available datasets for Vietnamese are nowhere near what is needed to achieve quality close to human-level translation. Once again, we are pushing the state of

the art and making them available to the community: For English-to-Vietnamese, PhoMT is the first high-quality, large-scale parallel dataset for MT, consisting of 3.02M sentence pairs;^e and PhoST consists of 508 audio hours, 331K triplets (audio, script, target language), for S2T.^f Combined, the VinAI Translate system² obtains a state-of-the-art performance for each translation direction and outperforms Google Translate in both automatic and human evaluations; it is now available for public use.^g

Vietnam, and SEA countries, are also (in)famously known for their traffic and mobility pain points. It is estimated that in 2019, 7,600 fatalities occurred on Vietnam roads,^h with 60% a direct result of driver distractionsⁱ and 20% of driver drowsiness.^j This is uniquely challenging due to mixed mode of travel, traffic conditions, local conditions, and unexpected behaviors of commuters. VinAI’s research aims to create products that, quite simply, save lives. Translated from research in embedded computer vision, Driver and Occupant Monitoring (DMS)^k technology ensures a safe driving experience with our in-cabin solutions, using low-cost cameras and highly efficient AI to analyze driver behaviors to warn against and prevent driving errors. We introduced the world’s first patented Auto Mirror Adjustment (AMA) technology at CES 2023—an AI-based feature to auto-

e <https://github.com/VinAIRsearch/PhoMT>

f <https://github.com/VinAIRsearch/PhoST>

g <https://vinai-translate.vinai.io/>

h <http://bit.ly/3JduYR0>

i <http://bit.ly/3LjaUz4>

j <https://bit.ly/3yyWMDp>


k <https://www.youtube.com/watch?v=qY7c1e3lo7s>

atically adjust the mirror to an optimal position with a single press of a button. A standard approach would need at least two cameras to triangulate, but our technology can predict eye position precisely with just one infrared camera. Beyond this, a full range of DMS includes highly accurate facial recognition for theft prevention, driver drowsiness and attention warning, advanced driver distraction warning, and dangerous behavior detection. DMS can run on multiple SoCs such as Nvidia, Qualcomm, Renesas, Ambarella, and others.^l Another cutting-edge product provided a real-time, 360-degree view around the vehicle, thereby completely removing all blind spots not only around but underneath the vehicle. Combined with AI that can detect the presence of pedestrians, cars, trucks, motorbikes, and small children, it is a product that significantly improves safety, especially for big SUVs, trucks, and commercial buses.

Vietnam has seen significant progress in AI in recent years, as the country has made efforts to advance its technology industry and invest in R&D in the field. VinAI has also established partnerships with leading technology companies and organizations to support its AI development efforts. Additionally, the government has implemented policies to support the growth of the technology sector and attract top talent in the AI field.

In conclusion, the use of AI in Vietnam has the potential to greatly benefit the country’s economy and improve various industries such as communication and

transportation. However, there are also challenges that must be addressed, such as the need for adequate regulations and the need to upskill the workforce to ensure the implementation of AI is done effectively and ethically. By addressing these challenges and fully embracing the potential of AI, we are confident Vietnam can become a leader in the field and reap the rewards of this technology.

Acknowledgments. Our work at VinAI is attributed to the dedication and expertise of all the VinAI members, including, but not limited to: Anh Tran, Tung Pham, Toan Tran, Son Hua, Rang Nguyen, Toan Bui, Khoi Nguyen, Dung Nguyen, Chan Vu, Vuong Cap, Duoc Nguyen, Tuan Le, Tri Pham, Viet-Anh Nguyen (The Chinese University of Hong Kong), Cuong Pham, Thien Huu Nguyen (University of Oregon, Eugene, USA), Van Anh Dam, and Ngoc Tran. 

References

1. Nguyen, D.Q and Nguyen, A.T. PhoBERT: Pre-trained language models for Vietnamese. In *Proceedings of the 2020 Findings of the Assoc. Computational Linguistics*, 1037–1042.
2. Nguyen, T.H. et al. A Vietnamese-English neural machine translation system. In *Proceedings of the 23rd Annual Conf. Intern. Speech Communication Assoc. Show and Tell*, 2022, 5543–5544.

Hung Bui is the Founder and CEO at VinAI in Hanoi, Vietnam. He is also an adjunct professor at Monash University, Clayton, Australia.

Minh Hoai Nguyen is the head of Computer Vision Group at VinAI in Hanoi, Vietnam. He is also an associate professor at Stony Brook University, NY, USA.

Dat Quoc Nguyen is Senior Research Scientist and the head of the NLP Group at VinAI, Hanoi, Vietnam.

Linh Pham is AI Residency Program Manager at VinAI, Hanoi, Vietnam.

Dinh Phung is a professor at Monash University, Clayton, Australia and has served as a senior technical consultant to VinAI, Hanoi, Vietnam, since 2021.

Copyright held by authors/owners. Publication rights licensed to ACM.