Phuc Van Phan

Ho Chi Minh city

+84 835349558 | phanphuc1100@gmail.com | github.com/pphuc25 | linkedin.com/in/pphuc

Work Experiences

ZenAI *VietNam*

AI Engineer *Apr 2023 - Current*

- **AI Companion: An social media product offering interactive interfaces for uploading and reacting to images, coupled with an engaging AI-driven role-play chat feature**:
  - Performing multi-label image classification on a dataset comprising 20,000 tags specifically associated with Anime-related imagery.
  - Accomplished classification of images into general, 15+, and 18+ categories for effective content filtering based on age appropriateness.
  - **Developed** data conversion, intent generate image request in chatting for character chatbot powered by GPT-4. **Engage** in train Bert-architecture model for intent generate image request, testing to see mistakes of model in role-play (Llama2). **Deployment** model to API, **design flow end-to-end** for data to user.
  - Optimized inference speed and resource usage in using CTranslate2, FlashAttention, etc. Successfully led initiatives that achieved a **fourfold increase** in inference speed while nearly maintaining the accuracy of results.

VietAI *VietNam*

AI Researcher/Engineer Collaborator & Teaching Assistant *Jan 2023 - Current*

- Identify performance improvement opportunities and integrate trained models.
- **Team project - Leader** (5 members and 1 mentor): **Fine-tuning** the **Whisper model for the Vietnamese language.**
  - Achieving **SOTA performance** in Vietnamese language by curated and processed Vietnamese language audio data over **500 hours**, trained model with migrate from GPU to TPU infrastructure.
- Provided in-depth responses to student inquiries, focusing on the intricacies of Deep Learning and Natural Language Processing. Excelled in clarifying complex concepts and theories, ensuring a comprehensive understanding in these advanced courses.

Gradients Technologies *Tan Binh, VietNam*

AI Researcher Intern *Jan 2023 - Apr 2023*

- Conducted in-depth research on relevant papers and authored comprehensive reports summarizing key findings.
- Analyzed the source code of machine learning models implemented in the JAX framework, providing documentation for code.

Side Projects

Open Sources Contributor *Personal*

Personal *Nov 2022 - Current*

- Contributed **over 300 lines of code** to prominent open source platforms including **Huggingface**, **Meta**, and **Microsoft**, with a primary focus on the example sections for PyTorch and Flax in **Transformers repositories**.
- Significantly enhanced open source projects by **debugging** and rectifying critical calculation errors in gradient accumulation and learning decay, **authoring comprehensive documentation**, **refining code** for optimal hardware performance, and **boosting processing speed** through minimized redundancy calls.

Generate IELTS Writing Task 2 | Third Prize *FPTU Research Festival contest*

Team Research FPT edu campus HCM *Dec 2022 - Jun 2023*

- An app to create the next sentences for IELTS writing task 2 by providing a specific topic or some sample phrases.
- Using GPT architecture to implement and data is crawled from some published sources about IELTS writing task 2.

Education

FPT University *Ho Chi Minh city, VietNam*

BSc in Artificial Intelligence *Sep 2021 - Current*

- **GPA:** 3.75/4

Skills

Programming	Python, R(ggplot2), C++, SQL.
Libraries/Frameworks	PyTorch, JAX, Tensorflow, Scikit-learn, Pandas, NumPy
Miscellaneous	Linux, Shell (Bash/Zsh), LaTeX (Overleaf/R Markdown), Git, Google Cloud, FastAPI, Docker, TensorRT.
Soft skills	Leadership, Time Management, Teamwork, Problem-solving, Documentation, Engaging Presentation, MC.

Achievements

2022	**Most Promising Group**, FPT Entrepreneurial Hackathon	*Ho Chi Minh city*