Nguyen Le

lenguyen18072003@gmail.com +84-942-142-797 lenguyen.vercel.app Nguyen Le lenguyen1807

Experience

- **VinBigdata - Ha Noi, Viet Nam** *(July 2025 - Recent)*: *AI Engineer Trainee.*
- **Gameloft Vietnam - Ho Chi Minh, Viet Nam** *(Nov 2024 - March 2025)*: *C++ Game Programmer Intern.* Contributed to the development of in-game features (UI) for Asphalt 8 using C++.
- **Bosch Global Software Technologies Vietnam - Ho Chi Minh, Viet Nam** *(Aug 2024 - Nov 2024)*: *AI Engineer Intern.* Developed a full-stack RAG application, featuring a VSCode extension front-end and a FastAPI/Langchain back-end, to automate fuzz test generation for C source code. **Tech stack**: Typescript, Langchain, FastAPI. Details.

Education

- **VNUHCM - University of Science, Viet Nam** *(Oct 2021 - Oct 2025)*: Bachelor of Science. Major in Artificial Intelligence. **GPA**: 8.4/10.
- **Bachelor Thesis** (Grade: 9.6/10): *Mechanistic Analysis of Representation Misdirection Method (RMU) for Large Language Model Unlearning.* This work shows that RMU does not deeply "unlearn" knowledge but instead injects a "junk direction" into the residual stream for harmful prompts, causing incoherent outputs to simulate unlearning.

Technologies

- **Programming Languages**: C++(17/20), Rust, Python, Typescript.
- **System Programming**: OpenGL, Metal, CUDA.
- **Frameworks**: Pytorch, Jax, Numpy, Transformers, Langchain.
- **Tools**: Git, Unix Shell, Docker, LaTeX (*General*), CMake (*Specialized*).

Projects

- **Simple Neural Network in C++** *(July-24 → Jan-25)* (Details): Developed a simple neural network trainer from scratch with simple custom matrix operations in C++ to classify MNIST, CIFAR-10, etc. datasets. Architecture supports configurable layers (Convolution, Dropout, Pooling, Linear, etc.). Unit testing through Catch2 framework. **Tech stack**: C++17, BLAS (Apple's Accelerate), OpenCV, Catch2, CMake. [1]
- **Banhxeo** *(May-25 → July-30)* (Github): A simple NLP library built with Python and Jax/Flax. Implement NLP models (RNN, MLP, GPT-2, etc.) and Tokenizer system from scratch. **Tech stack**: Python, Jax, Flax.
- **Optimize GEMM (Metal)** *(August-25 → Recent)* (Github): Implemented and optimized General Matrix Multiplication (GEMM) for Apple (Metal) GPUs. Focused on learning about GPU (Nvidia and CUDA) through Apple's Metal and how to optimize Matrix Multiplication operation. **Tech stack**: C++17, Metal (Apple's GPU).
- **PRML for Viet** *(June-24 → Recent)* (Website): Attempt to provide a resource/reference for Vietnamse student to learn about Machine Learning in general and learn Pattern Recognition and Machine Learning book in specific.

Languages

- **English**: Professional working proficiency. (*TOEIC Reading & Listening*: 870/990, *TOEIC Speaking & Writing*: 320/400, Details).
- **Vietnamese**: Native.

[1] Each project will has template: "name (start date → finish date) (link): description". If finish date is "Recent", that means the project is still in active developement.