

Lance Nguyen

lancemnguyen@gmail.com | [Portfolio](#) | [LinkedIn](#) | [Github](#)

Summary

Highly driven and collaborative software engineer with a solid understanding of AI application architectures, seeking a challenging position to leverage expertise in generative AI technologies to deliver valuable innovation for the business.

Skills

Languages: Python, JavaScript, HTML, CSS

Frameworks: PyTorch, Hugging Face, LangChain, LlamaIndex, Django, Flask, asyncio

ML Techniques: Retrieval-augmented generation (RAG), LLM fine-tuning, Prompt engineering

Tools: LLMs (ChatGPT, Gemini, Copilot), Vector stores, Embeddings, Debugging, Git

Experience

Artificial Intelligence Engineer, Cyber Partner AI *March 2024 – Present*

- Fine-tune and evaluate the gemma-2b-it model to extract specific categories of information from user messages, resulting in better companionship experience for users
- Use GPT to create and annotate an inhouse dataset containing special prompt instruction, preprocessing the data with Python logic to prepare for model training
- Prompt engineer the initial model used for persona extraction with few-shot examples to instruct the model to return the response in JSON format

Artificial Intelligence Engineer Intern, Radical AI *April 2024 – Present*

- Develop Gemini Quizify, an application that dynamically generates quiz questions based on user-provided documents and offers instant feedback with comprehensive explanations
- Conduct thorough testing and performance logging, analyzing model outputs to ensure clarity and relevance in responses

Artificial Intelligence Engineer, Mentorship *January 2023 – Present*

- Engage in a private study with a senior AI researcher from Nvidia to discuss advanced AI application design, techniques and frameworks
- Develop expertise in interacting with LLMs to learn new skills and concepts, particularly in programming and technology

Professional Tennis Coach, Self-employed *April 2014 – Present*

- Empower 12 students by giving private lessons, resulting in improved ratings up to: Blue Chip, 13.5 UTR, and 5.5 NTRP. Previously lead training programs for 600+ players at an academy

AI Projects

RAG Optimization | [Website](#)

- Built a Streamlit application that allows users to upload documents and interact with their data
- Evaluated and upgraded the RAG system through an iterative process, improving the accuracy of responses for a 574-page document

Qdoc - Article assistant with memory | [Website](#)

- Identified user inefficiency in web content consumption and addressed it by creating an AI-powered article assistant that streamlines information retrieval and reduces screen time

Other Projects: [SpotOn](#) Analyzes reviews on Google Maps, [CooPA](#) Chatbot with Google Search API integration, [AI Chrome Extensions](#) Defines and transforms user-highlighted text on web pages

Education

Bachelor of Science in Kinesiology, San Jose State University

August 2019