Emerald Henry

Fullstack A.I Engineer | LLM + GNNs

emeraldhenry3@gmail.com | Portfolio Page | Github | Linkedin

EDUCATION

Covenant University | Lagos, **B.S. in Mechanical Engineering** | September 2017 - July 2022 Background: Carried out Multiple Machine Learning and Deep Learning Projects.

Stanford University, Online Stanford: Machine Learning with graphs (CS224W)

EXPERIENCE

TechWave | Remote Fullstack AI Engineer (LLM) | March 2024 – July 2024

- Spearheaded the development of TechWave's groundbreaking Autonomous Course Generator, revolutionizing the process of skill acquisition.
- Developed an AI agent capable of autonomously curating comprehensive courses, encompassing beginner to expert levels, based on user inputs, up to 70% completion.
- Orchestrated the integration of multiple AI modules, including a course structure generator, learning video transcript generator, and sample assignment creator, streamlining the course development process.
- Developed a Django-based backend application to support LLM agents that are developed using Langchain, Llama-Index and CrewAI, and deployed as RESTful APIs for agent communication, managing data storage and retrieval, and ensuring system scalability and reliability.

LCA | Remote Data Science & ML Tutor | October 2023 – March 2024

- Created and executed end-to-end machine learning applications with a specialization in Natural Language Processing (NLP). Proficient in leveraging Language Models (LLMs) from HuggingFace, Google, and OpenAI and utilizing frameworks like Langchain, Llama-Index and Crewai developing diverse AI Agents (LLMOps).
- Utilized the OpenAI API and open source models from Ollama to run RAG applications, develop AI Agents and build chatbots with a chainlit interface, as well as a HTML and CSS initial interface page.
- · Developing comprehensive training materials, including notebooks and files, for Data Analysis using R, Database Management (MySQL), and MLOps practices (Docker, FastAPI, CLI scripts, CI/CD, and deployment with Azure and AWS technologies).

· Clinton Health | Hybrid

Data Scientist | January 2023 – October 2023

- Spearheaded the development and management of Healthcare Databases tailored for the Nigerian National Healthcare Sector.
- Conducted exploratory data analysis (EDA) to extract meaningful insights and designed Dashboards to facilitate data-driven decision-making within the Nigerian National Healthcare Sector.
- Innovatively crafted HTML-based tools to streamline the collection of healthcare data from both patients and healthcare facilities.

Covenant University | Lagos, Nigeria A.I Researcher | August 2022 - January 2023

- Investigated fine-tuning ResNet50, using ImageNet base weights, with domain-specific images for automated diagnosis of lesions and cancer across various medical imaging modalities, including MRI, CT, Ultrasound, and Whole Slide Imaging (WSI).
- Conducted a comparative analysis of pure CNN and Hybrid Vision Transformer methods for efficient classification, segmentation, and reconstruction of medical imaging, reviewing over **200** literature sources. Found that the hybrid method outperformed the pure CNN method by approximately 20%, this led to a publication.

TEE Research Group | **Lagos, Nigeria A.I Research Intern** | October 2021 - August 2022

- Developed a wind turbine power curve model that employs a data filtration technique based on experimentally determined quantile ranges (0.2–0.25) and a Radial Basis Function network, achieving an 11.2% improvement over SOTA methods. This research was published as a chapter in a book with Taylor and Francis.
- Developed a novel method combining the Kolmogorov-Smirnov non-parametric test, Mariano-Preve's test for predictive ability, and machine learning techniques (XGBoost, MLP, and RBF) to enhance wind turbine fault detection speed and accuracy, focusing on qualitative improvements given the stochastic nature of the study. This research was published in Energy Reports.

COMPUTATIONAL PROJECTS (more)

Train/Finetune Base LLM Model

- Built GPT-2 and LLaMA2-variant models from scratch with parameter sizes up to 124M, optimizing training through flash attention, mixed-precision training with TF32 and BF16, PyTorch compile, and Distributed Data Parallel (DDP). These methods achieved a **10x speedup and 40% reduction in GPU memory usage** compared to baseline training runs without optimizations, with DDP enhancing multi-GPU scaling efficiency on an 8-GPU box setup.
- Conducted fine-tuning and preference optimization using Hugging Face pipelines for Supervised Fine-Tuning (SFT), Direct Preference Optimization (DPO), and Low-Rank Adaptation (LoRA), achieving up to 25% improvement in BLEU score after SFT and DPO.

Comprehensive AI Agent/Chatbot/RAG

- Built a Robust AI Agent comprising multiple tools such as RAGs, web scraping and browsing tools, along with automatic tool routing based on user query using the Langchain framework.
- Built a Robust AI master agent for multiple sub agent orchestration, for solving complex reasoning problems such as system 2 thinking, using the Crewai framework.
- Build a comprehensive RAG application with multiple customized tools for specific retrieval operations such as summarizing part-or-whole text, itemizing facts etc using the Llama-Index framework.
- The entire project's page is built using HTML, CSS and Flask.
- Open Source LLM Framework Contribution

 A contributor to a few open source LLM frameworks such as CrewAI

DL & LLM with C/CUDA

• Ongoing, Attempting to replicate the previous LLM project and a few Deep Learning techniques in pure C and CUDA.

Graph Representation Learning Projects

• Completed a robust project comprising the implementation of numerous graph neural network research for graph generation, classification and traversal, with very useful applications in recommender systems.

PUBLICATIONS

Vision Transformers in Medical Imaging: A review. (link)

Conditional Monitoring and Fault Detection of Wind Turbines Based on Kolmogorov-Smirnov's Non-Parametric Test and Machine Learning. (<u>link</u>)

A Neural Network-Based Wind Turbine Power Curve Model Using Several Wind Farm's Influencing Parameters and Topography. (<u>link</u>)

SKILLS

Language: Python, C, C++, SQL, HTML, CSS Frameworks: Pytorch, Tensorflow, PyG, Django, Flask, FastAPI LLM Skills: Langchain, Llama-Index, Crewai, VectorDB, HuggingFace Other Skills: Linux, AWS, GCP, git Soft Skills: Leadership, Communication, Teamwork, Presentation