Maxell G. Milay

AI & Machine Learning Engineer

Mandaue City, Cebu, Philippines | milaymaxell@gmail.com | +63 966 469 5096 | LinkedIn | Github | Portfolio

SUMMARY

I'm currently an AI Engineer, specializing in integrating AI solutions through LLMs and MLOps. I have been recently involved in designing agentic systems through prompt engineering and RAG for industry-level software solutions. I also create custom ML models and build pipelines for Continuous Integration (CI), Deployment (CD), and Training (CT) to prepare myself for a GCP Professional Machine Learning Engineer Certificate. Other than that, I particularly enjoy automating workflows, ensuring smooth transition from development and model training to production.

EXPERIENCE

ECLARO — Quezon City, Metro Manila

AI Engineer — *April 2025 - Present*

- Spearheaded the development of an enterprise-scale Generative AI framework, enabling secure and scalable internal adoption of LLM-based solutions across business units.
- Redesigned and refactored the Intelligent Document Processing (IDP) pipeline to enhance accuracy and performance for high-volume insurance documents, improving field-level extraction precision and system maintainability.
- Designed and evaluated AI-driven information extraction logic for First Notice of Loss (FNOL) fields, leveraging NLP techniques to process and structure unstructured claims data.
- Translated insurance business requirements into actionable model-level objectives, collaborating with product owners to define extraction targets, key entities, and validation rules.

BitWork Solutions — Florida, Miami

Lead Software Engineer — *December 2024 - April 2025*

- Led a pioneer team of 5 people in developing a image processing-centered mobile application company product
- Managed and planned epics and product backlogs across multiple sprints, ensuring 90% of tasks accomplished and delivered per development cycle iteration
- Spearheaded a major backend architecture migration from Directus headless CMS and AWS Lambda to Django backend and PostgreSQL database deployed using AWS App Runner, Elastic Container Registry, and RDS
- Ensured codebase integrity by reviewing code changes hands-on and made sure to mitigate technical debt early-on

Part-time DevOps Engineer — *July 2024 - December 2024*

- Setup CI linting and testing pipelines for across 3 projects within the company
- Integrated AWS S3 and DynamoDB into a mobile application that saves raw captured image data for preprocessing
- Migrated code from a custom built backend into AWS Lambda for better server scaling

BPOSeats — Mandaue City, Cebu

AI Engineer — *June 2024 - January 2025*

- Oversaw the initial planning and development phases of a machine learning productivity model projected to increase operational efficiency by 55%.
- Optimized LLM API calls through prompt engineering, leading to a 40% decrease in service costs
- Co-led a novel approach in implementing task estimation through engineering with OpenAI LLMs after integrating large volumes of company historical productivity data
- Spearheaded research in software effort estimation spanning the fields of NSGA genetic algorithms, LLMs, and simulations to improve current effort estimation implementations
- Integrated LLM API cost tracking in the entire application ecosystem by monitoring token usage per feature, and creating impactful visualizations and saved 35% of operational costs

Full Stack Web Developer — *September 2023 - May 2024*

- Pioneered the implementation of the company's first vector database in PostgreSQL, enabling the development of advanced AI retrieval algorithms
- Integrated Vue and Django unit tests into automated pipelines, reducing deployment failures to production by 80%.
- Built a help center chat bot that utilized Retrieval Augmented Generation, which was projected to reduce complaints to CX by 75%
- Created a pipeline migrating company documentation data to a vector database, which enabled vector operations for the chatbot
- Debugged code, resulting in a significant decrease in bugs and contributing to a 65% increase in system availability.

Symph — Cebu City, Cebu

Web Developer Intern — *December 2022 - March 2023*

- Deployed and implemented changes to client websites in collaboration with backend and DevOps teams, ensuring minimal disruption.
- Addressed issues and implemented recommendations over 2 projects on existing client web pages, improving usability and customer satisfaction.

VOLUNTARY WORK EXPERIENCE

AI Pilipinas Cebu Chapter   **Cebu City, Cebu**
AI Infrastructure and Deployment Advocate   *July 2024 - Present*
- Researched relevant techniques in MLOps, including ML workflows, from data ingestion and preprocessing, feature engineering, model development and training, model evaluation, deployment to production, and performance monitoring
- Created Medium articles revolving around Google Cloud's AI tools, including Vertex AI and other cloud services for MLOps and ML workflows

University of the Philippines Computer Science Guild   **Cebu City, Cebu**
Lead Software Engineer   *November 2023 - Present*
- Led a team of 7 designers and 10 developers in creating multiple versions of the official UP Computer Science Guild Website
- Established the entire workflow between designers and developers, rooting from relevant agile principles
- Deployed the website with CI/CD and GitHub Actions through Docker on university servers
- Managed and reviewed the UX of designers and code of each developer, ensuring code quality and performance

LEADERSHIP, PROJECTS, & RECOGNITION

Beacon
- Created a career guidance application that utilizes generative AI to provide users with personalized visual roadmaps towards their viable career option based on their user profile information.
- Led a team of 5 in building a working prototype of the application within the span of 6 hours
- Implemented a storage system and authentication through Google Firebase and leveraged OpenAI LLM models
- Garnered the 1st Place of the first UP Computer Science Guild Komsai Week Hackathon in 2024

Finite Automaton Visualizer
- Created a web application that generates DFA graphs based on regular expressions, and also verifies string inclusion by visually simulating the algorithm through a user-friendly and interactive interface.
- Led a team of 4 in creating a working prototype of our application through agile development
- Implemented the core algorithm which involved generating an abstract syntax tree from a regular expression in consideration of precedence

Lakbai
- Created a GIS application that uses a dynamic, data-driven pedestrian accessibility index powered by a fuzzy logic system for promoting sustainable and equitable urban mobility in the Philippines within the span of 2 weeks
- Awarded as the champion of the 2024 Philippine Junior Data Science Challenge hosted by the UP Data Science Society
- Led the integration of the fuzzy inference system and multilayer perceptron into the application
- Deployed the frontend using Vercel, the backend using Google Cloud Run, and the postgresql database using Google Cloud SQL
- Setup CI/CD pipelines using Github Actions that tied together GCP services for deployment and integration

EDUCATION

University of the Philippines Cebu   **Cebu City, Cebu**
Bachelor of Science in Computer Science
- Led 3 teams through small and large scale competitions and projects, achieving the best competition places and highest ratings

Philippine Science High School - Central Visayas Campus   **Argao, Cebu**
STEM Strand   *High Honors & Excellency in Physics*
- **Project DALOY** - Drone-based Depth and Atmospheric Level Open-Source monitoring device for the Filipino Youth
- **Thesis** - Simulated Biomimicry of Photovoltaic Tree Architecture Based on the Phyllotaxy and General Tree Crown Shape of *Pinus Strobus* (Eastern White Pine)

CERTIFICATIONS

Machine Learning Specialization   **DeepLearning.AI & Stanford**
- Completed 3 hands-on courses involving supervised, unsupervised, and reinforcement learning, along with neural networks using TensorFlow and Scikit-learn frameworks
- Replicated the Lunar lander reinforcement learning problem and solved through Deep Q Learning using the Proximal Policy Optimization (PPO)
- Created a movie recommender system implemented using vectorization and similarity search of metadata
- Built an unsupervised customer segmentation model using K-means clustering using each customer's spending score and yearly income
- Built and compared different prediction models for house pricing such as Linear, Random Forest, XGBoost, CatBoost regressors, and more

Deep Learning Specialization   **DeepLearning.AI & Stanford**
- Completed 5 hands-on courses around neural networks, and learned various optimization techniques, along with ML development iteration best practices
- Built industry grade CNNs and RNNs from scratch like VGG-16/19 and variations of the MobileNet, achieving 77% and 78% validation accuracies respectively both using the CIFAR-10 dataset