Nebiyou Hailemariam

412-251-7954 | nebhailema@gmail.com | linkedin.com/in/nebhailemariam | nebhailemariam.github.io/

Summary

I'm a **software engineer** with **4 years of experience**, I've built **applications used by tens of thousands of users**, combining my expertise in **Software Engineering** and **Machine Learning**. I specialize in **Python**, **Flask**, **C#**, **ASP.NET Core**, and **React.js** to deliver robust solutions. With a strong foundation in **software engineering** and **machine learning**, I leverage **PyTorch** to create **AI-driven systems** and **fine-tune large language models (LLMs)** for customized user experiences. In my free time, I contribute to **open-source projects**, such as enhancing the **Gorse Recommender System**.

Education

Carnegie Mellon University — CGPA 3.93 *Pittsburgh, PA*
Masters in Information Technology, Applied Machine Learning Specialization *Aug. 2022 – May 2024*

Addis Ababa University — CGPA 3.8 *Addis Ababa, Ethiopia*
Bachelor of Science, Computer Science *Sep. 2016 – May. 2020*

Experience

Machine Learning Software Engineer *Aug. 2024 - Present;*
Bizu *US, Remote*

- Worked on **eezly**, a grocery price comparison application used by over 10,000 users, leveraging **ASP.NET Core Web API**, **Python (FastAPI)**, **PyTorch**, and cloud-based microservices to build scalable, AI-driven solutions.
- Built a **Recipe Recommendation System** using **LangChain**, **OpenAI**, and the **Recipe1M+** dataset, creating a **Retrieval-Augmented Generation (RAG)** system to suggest recipes based on the products users purchase. Incorporated the **Weaviate** vector database to enhance search and recommendation.
- Employed **PyTorch** and **Hugging Face** to train hierarchical machine-learning models for classifying retail products from various stores (e.g., Walmart) into **aisles, categories, and subcategories**.
- Integrated **Gorse**, a recommender system, and contributed to open-source recommender systems.
- Designed and implemented **RESTful APIs** for inventory management using n-tier architecture and developed a **single-page application** with **React.js**.
- Implemented **OAuth 2.0 client-credential flow** using **OpenIddict** for secure machine-to-machine communication, **Single Sign-On (Firebase, Cognito)**, and **ASP.NET Core Identity** for user management.
- Designed a messaging system using **Kafka** with **Golang** as the message producer.
- Employed **xUnit** and **pytest** to write unit and integration tests for microservices using a custom web application factory.

Machine Learning Software Engineer II *Jul. 2021 - Jan. 2024; May 2019 - Jul. 2020*
Matrix Information Technology *Remote*

- Developed scalable microservices for **eQub**, a platform used by **tens of thousands of users**, transforming a traditional financial communal saving system into a modern **digital solution**.
- Implemented key features like **transaction processing** and **account management**, leveraging **ASP.NET Core Identity**, **Entity Framework**, and **AWS RDS**.
- Developed microservices applications for the **no-code data and operations management platform WorkOnGrid**, a platform used by **thousands of users**, leveraging **AWS** (**EC2**, **RDS**, **EKS**, **SES**, **SQS**).
- Built and deployed **predictive models** for the **user dashboard** of the **no-code platform WorkOnGrid**, enhancing users' **decision-making processes**.
- Led the development of a **subscription management microservice** integrated with **Stripe**, simplifying subscription workflows and boosting **customer satisfaction**.
- Wrote comprehensive **unit and integration tests** for **ASP.NET Core Web API** using **xUnit** and for **Python** using **pytest**, significantly improving **code reliability** and reducing bugs across services.

Software Engineer I *Aug. 2020 – Sep. 2021*
Digital Medarbeider *Oslo, Norway*

- Developed a **Cloud-Based School Management System** for managing students, employees, and schedules, enhancing administrative efficiency and automation for **Norway municipalities**.
- Implemented a **microservices architecture** for student and employee management, and scheduling systems
- Employed **IdentityServer4** for securing microservices and **ASP.NET Core Identity** for user authentication.
- Leveraged **Azure App Service** for web hosting, **Azure DevOps** for CI/CD pipelines, and **Azure Key Vault**.
- Utilized various databases, including **Azure Cosmos DB** for scalable NoSQL data storage, and **Azure SQL Database** for relational data needs, ensuring **efficient** and **reliable data handling**.
- Wrote **unit and integration tests**, used **Bogus** for realistic test data, and documented APIs with **Swagger** to facilitate integration and maintenance.

Research Assistant in Machine Learning *May 2021 – Jan 2022*

Empathic Computing Lab *Auckland, New Zealand*

- **Empathic Computing Laboratory (ECL)** is an academic research laboratory directed by **Prof. Mark Billinghurst** at the **University of South Australia** in Adelaide, Australia, and the **University of Auckland** in Auckland, New Zealand.
- Collaborated with **Ph.D. students** to refine methods for detecting **emotions from physiological signals**.
- Conducted extensive **literature reviews** and analyzed the performance of various **machine learning** and **deep learning models**, applying rigorous Hyperparameter tuning. Authored a **14-page paper (IUI - ACM)**.

Projects

XLM-R Based Extractive Amharic Question Answering with AmaSQuAD Sep 2023 – May 2024

- Completed my thesis project at Carnegie Mellon University in multilingual question-answering research (NLP).
- Developed a novel framework for translating the SQuAD 2.0 dataset into Amharic, resulting in the creation of the AmaSQuAD dataset. Implemented a translation-based data generation framework valuable for extractive Question Answering (QA) systems, contributing to the advancement of natural language processing (NLP) for low-resource languages. Leveraged XLM-R, a pre-trained language model, and fine-tuned it specifically for Amharic QA tasks, achieving 7% F1 improvements in baseline performance.

Machine Learning Based Rain Gauge Using Acoustic Data Sep 2022 – May 2023

- Addressed the challenges of conventional weather stations such as high setup costs, instrument fragility, and measurement errors by exploring alternative sound sources for making rainfall predictions.
- Used a Convolutional Neural Network (CNN) regression model using PyTorch and TensorFlow to estimate rainfall intensities from MFCCs extracted from acoustic recordings. Employed CNN model and achieved a Mean Absolute Percentage Error (MAPE) of 35.20% and a Mean Squared Error (MSE) of 0.66, outperforming a baseline Support Vector Regression (SVR) model with 152.55% MAPE and 1.73 MSE.

Technical Skills

Languages: Python, C#, C/C++, Python, Go, JavaScript, Java, HTML/CSS, R
Frameworks: ASP.NET Core Web API, Entity Framework, IdentityServer4, OpenIddict, Dependency Injection, Flask, FastAPI, Django, Gin, Express.js, Node.js, Angular, React, Node.js, PyTorch, TensorFlow, Scikit-learn, Pandas, Numpy, RabbitMQ, Kafka
Dev Tools: Git, Docker, Kubernetes, GCP, AWS, Azure, Postman, Jenkins, Grafana, Prometheus, Grafana-Loki, Jaeger
Web & Frontend Technologies: TypeScript, ReactJS, JavaScript, CSS, Bootstrap, Redux, Angular.js, full-stack
Cloud Services & Platforms: AWS (EC2, ECS, EKS, RDS, SES, SQS, MQTT), Google Cloud (Google Compute Engine, Google Cloud Storage, Google Colab), Firebase
Security: OAuth 2.0, JWT, IdentityServer4, ASP.NET Core Identity, Passport.js
Databases & Data Storage: Relational (MySQL, PostgreSQL, Postgres), NoSQL (MongoDB, DynamoDB, Cosmos DB, Redis), Search & Indexing (ElasticSearch), Vector (Weaviate), ORM (SQLAlchemy, Entity Framework)
Testing & Quality Assurance: Unit test, Integration test, xUnit, WebApplicationFactory, Bogus, pytest, Mocha, Jest
DevOps & CI/CD: Docker, Kubernetes, Jenkins, GitHub Actions, Containerization
Machine Learning & Deep Learning Skills: Linear Regression, Support Vector Machines, Bagging and Boosting, Neural Networks, CNNs, RNNs, LSTMs, NLP, Graph Neural Networks (GNNs), clustering, Graph Attention Networks (GATs), Hyperparameter Tuning, Transfer Learning, TensorFlow, PyTorch, keras, Scikit-learn, XGBoost, Hugging Face Transformers, Data Preprocessing, Weights & Biases, OpenCV, Computer Vision, predictive models, Derivatives, LangChain, data pipeline
Data Science: Data Warehousing (BigQuery, Redshift), Data Pipelines (RabbitMQ, MQTT), Data Engineering, Data Mining, Data Visualization, Data Analysis, Scikit-learn, NLP, Pandas, NumPy, PyTorch, NLTK, Anaconda, PySpark, Apache Spark, data analytics, Business Intelligence, Matplotlib, data structures, data modeling, Kafka
Architectures: Event-Driven, Microservices, Serverless, Monolithic Architecture, Test-Driven Development (TDD), Data-Driven, Design patterns, Agile, REST API