

Tesfaye Alemayehu Erena

AI Engineer

CONTACT INFORMATION

Email: tesfayealemayehu27@gmail.com

LinkedIn: [linkedin.com/in/tesfaye-alemayehu](https://www.linkedin.com/in/tesfaye-alemayehu)

Github: <https://github.com/tesfayealex>

Website: <https://sites.google.com/view/tesfaye/home>

Blog: <https://hubpages.com/@tesfayealemayehu>

Phone number: +251964345364

PROFESSIONAL SUMMARY

Machine Learning and AI Engineer focused on Generative AI, NLP, and the development of end-to-end AI solutions. Experienced in delivering LLM-powered applications, including multi-agent systems (Autogen, LangGraph), RAG pipelines, and vector databases (Weaviate, Pinecone). Proficient in full-stack development, data engineering, feature engineering, and applying frameworks like PyTorch and TensorFlow for impactful solutions. Committed to algorithm development, model evaluation, and advancing AI with attention to detail and AI ethics.

SKILLS

Programming Languages: Python, JavaScript, Go

Frameworks & Libraries: Autogen, Langgraph, OpenAI, Huggingface, Transformers, PyTorch, TensorFlow, Langchain, n8n, Scikit-learn, Django, Flask, FastAPI, NLTK, SciPy, Matplotlib, Pandas, NumPy, Node.js, Kedro, dbt, Apache Superset, Spark, Kafka, MLFlow, Airflow, Airbyte, Pytest, Redash, OpenCV, Langfuse, Data Visualization, Data architecture, Data integration, Data quality assurance

Database Technologies: Vector Databases (Weaviate, Pinecone), PostgreSQL, MongoDB, MySQL (SQL and NoSQL), Redis, Big Data, Spark

DevOps Tools: Docker, Kubernetes, GitLab CI/CD pipeline, GitHub Actions, Linux, Git Cloud

Service providers: AWS, Azure (VMs, Blob Storage, Azure AI), DigitalOcean

ML Algorithms & Techniques: Linear and logistic regression, decision tree, XGBoost, random forest, Naive Bayes, K-means clustering, CNN, LSTM, RNN, SVM, Transformers, neural networks, hyperparameter tuning, ensemble learning, data preprocessing, statistical analysis, Model Deployment

Methodologies: AI Agents and Assistants, LLMs (OpenAI GPT series, Google Gemini, Huggingface models), Prompt engineering, Workflow automation, Semantic-Aware Text (SAT) algorithm, Function Calling, Microservices Architecture, API development, Model evaluation, MLOps

Soft Skills: Adaptability, time management, attention to detail

Languages: English and Amharic

PROFESSIONAL EXPERIENCE

Tenacious Intelligence Corp., California, USA

2024 – Present

GenAI and Data Engineer

- Directed a team of AI engineers in creating a legal AI assistant using multi-agent architecture (Autogen), projected to enhance legal research efficiency by up to 10x for an Italian law firm.
- Engineered a Python-based data pipeline for complex legal datasets, processing over 80GB into Weaviate and developing LLM-generated schemas to improve Retrieval-Augmented Generation (RAG) accessibility.
- Developed an AI-powered survey management system (Django, Next.js), automating analysis and reducing survey data processing time from days to seconds, significantly cutting associated costs.
- Instituted a data warehouse (PostgreSQL, Kedro, dbt) and Redash dashboards for a content creator, achieving an 80% satisfaction rate from senior management by enhancing data visibility for strategic decisions.
- Managed FastAPI and Go/MongoDB backend integrations for AI services, ensuring robust support for applications managing large-scale user interaction and data processing.

Machine learning engineer

- Enhanced AI talent search for a SaaS platform using Generative AI and RAG, incorporating an SVM CV classifier and semantic resume ranking, projected to save clients tens of recruitment hours.
- Designed an agentic AI chatbot (LangGraph, Autogen) with GPT-4 function calling and multi-channel integration, streamlining engineering team access to AI tools and boosting internal productivity.
- Implemented Langfuse for comprehensive AI interaction tracing and Redis caching, optimizing system performance, and contributing to a notable reduction in API operational costs.
- Advanced algorithm development for candidate matching using Python and embedding models, achieving a 90% accuracy rate and elevating average client satisfaction scores by 20 points.
- Led R&D for an autonomous video subtitling system (English to Amharic), delivering over 90% accuracy in automated subtitling and documenting key AI dubbing challenges.

Ethiopia Taxi Solutions, Addis Ababa, Ethiopia

2021– 2024

Research and Development Lead, Software Engineer

- Led a team of engineers in developing ML models (Logistic Regression, feature engineering) for ETA prediction on the Zmall platform, achieving a 10% reduction in operational costs.
- Headed backend refactoring (Node.js, Express.js) and API development for the “Zmall” delivery system, significantly lowering operator action latency and improving system scalability.
- Devised a personalized recommendation engine (K-Nearest Neighbors, Random Forest) using Python and Spark, resulting in a 10-fold increase in sales and a 30% lift in user engagement with recommendations.
- Pioneered an Airbyte-powered Data Warehouse (PostgreSQL) for “Zmall,” enhancing data preprocessing and statistical analysis capabilities, which improved decision-making accuracy by 30%.
- Spearheaded 15 Odoo ERP implementations; directed “Zmall’s” migration to Odoo and engineered “Zoorya” (SaaS ERP), reducing data migration project timelines by an average of 50%.

PERSONAL PROJECTS

- [Deep search Chat](#) (LLMs, Tool calls)
- [Amharic speech-to-text using deep learning](#) (neural networks)
- [MCP Agent Gateway](#) (MCP Proxy with RAG)

EDUCATION

- **10 Academy, Santa Clara, CA, USA (remote learning)** **2022**
Professional Certificate - Machine learning, data engineering, and Web3
- **Addis Ababa Science And Technology University, Addis Ababa** **2016–2021**
Bachelor of Science in Software Engineering

CERTIFICATION

- Certified Tensorflow Developer, From Tensorflow
- Azure Data Engineer Associate, From Microsoft
- AWS Cloud Practitioner, From Amazon AWS
- Deep Learning Specialization, From Coursera