Bemnet Beshah

Software Engineer

+1 (512) 210-1034 | bemnetbeshah@gmail.com | github.com/bemnetbeshah | linkedin.com/in/bemnetbeshah | renvane.com

Professional Summary

Computer Science & Mathematics student (Honors, 4.0 GPA) who designs and ships LLM/ML systems used by real clients, including compliance agents, RAG pipelines, MCP servers, and production APIs, through work with Renvane AI. Seeking an AI engineering internship to deepen skills on a strong engineering team.

Key Skills

Python | FastAPI | Docker | LLM Agents | Multi-Agent Systems | RAG | AWS | MCP Servers | PostgreSQL | REST APIs

Professional Experience

Undergraduate Researcher, IMICS Lab — Texas State University — Oct 2026 – Present
San Marcos, TX | Sign-language AI research

- Researching ways to create **realistic variations of sign-language movements** so AI models can learn from more diverse examples without changing the meaning of a sign.
- Built a process to compare original movements with computer-reconstructed versions across **20 previously unseen clips**, identifying errors in body and hand motion that must be addressed before the examples can be used for training.

Renvane AI — Founder & AI Engineer — Jan 2026 – Present
Remote / Texas — building LLM- and ML-powered automation for paying clients across infrastructure, government compliance, and food distribution

- **Afri Investment PLC** (paying client, large-scale mining & pumping engineering solutions): turned an opportunity-research process into a **full-stack app** with its own backend and database plus an **AI-enabled research pipeline**, surfacing **250+** infrastructure projects and contractors across **11** African countries and removing **~85 hrs/week** of manual research; produced **3** leads now in direct communication with Afri. View project
- **Ceda-Tex Services** (paying client, construction): built and shipped an **AI browser-automation agent** (Python) that searches the legacy Texas **CMBL/VetHUB** vendor databases to create the compliance packets required by law before subcontracting projects in Texas. Cuts an **8-hour** manual administrative process to **30–60 min** (~**$2,100/mo** in value). View project
- **Kare Baltina** (paying client, **7-figure-per-year** food distribution): a read-only **LangChain** agent interprets the delivery workbook's structure and proposes a column mapping, validated into normalized **DuckDB** tables; Python then applies every business rule and writes the **59-column** Zoho import CSV. A validator agent auto-fixes what is safe and flags inconsistencies (repeated order numbers, price changes) with a confidence score for **human-in-the-loop** review; decisions persist in **SQLite** so they auto-apply next run. View project
- Built internal tooling including a **Market Research Workspace** (Python stdlib HTTP/JSON API, **12-table** SQLite, SMTP outreach) for idea validation, evidence organization, and client relationship management.

AI Researcher (Diffusion Models / GNNs), Texas State University — CS Research Lab — Jan 2026 – Apr 2026
San Marcos, TX

- Explored methods to speed up convergence of **diffusion models** by learning transferable weight-initialization distributions across architectures to beat handcrafted schemes like **Kaiming**.
- Trained GNN-based denoisers across **4** architecture iterations on **8x A5000** GPUs with DDP, EMA, v-prediction, and per-slot normalization, progressing from divergent outputs to numerically stable generation.
- Designed a graph-conditioned **DDPM** pipeline that generates neural-network weight initializations from a zoo of **3,004** converged MNIST CNNs, using **PyTorch Geometric** heterogeneous graphs (row, slot, layer node types).

AI4ALL Ignite Fellow — Wind Energy Site Suitability Project — Aug 2025 – Jan 2026
Pandas, NumPy, Scikit-Learn, XGBoost, Streamlit

- Combined the NREL Wind Toolkit with the US Wind Turbine Database by physical proximity, creating a labeled geospatial dataset of **120K+** grid cells for turbine-site prediction.
- Engineered **56** features (geographic encodings, wind × capacity-factor interactions, spatial constraints), testing **12** feature combinations to optimize performance.
- Achieved **XGBoost ROC-AUC 0.9545**, Accuracy 0.8783, F1 0.8537 with **94.3%** turbine recall — a **30.4%** improvement over baseline; deployed to Streamlit and selected for the Ignite symposium.

Team Lead / Backend Developer, HackTX 2025 — Computer Use Prize Winner — Oct 2025
Austin, TX

- Led a **4**-person team to win the Computer Use Prize among **1,000+** participants and **200+** projects — one of only **20** winning teams.
- Built an AI developer tool using **Playwright**, **Gemini 2.5 Computer Use**, and **FastAPI** enabling coding agents to visually probe, analyze, and interact with live web apps.
- Engineered the orchestrator backend and feedback loop that collects UI data (screenshots, DOM metrics, error states) and reprompts the agent, enabling autonomous UI debugging.

Software Engineering Intern, ETTA Solutions PLC Dec 2022 – Jan 2023
Addis Ababa, Ethiopia

- Automated data-processing workflows using **Python (pandas, NumPy, JSON)**, reducing manual entry time by over **50%**.
- Developed RESTful APIs with **FastAPI** and integrated **PostgreSQL**, standardizing the backend pipeline for reliable analytics reporting.
- Collaborated with engineers using **Git/GitHub** for version control and a clean, versioned codebase.

Projects

Interview Prep MCP Server — Shared Spaced-Repetition Study Backend for LLMs Python, MCP, OAuth/OIDC, Postgres

- Built a hosted **MCP server** that lets Claude and ChatGPT share one study-progress database, exposing **15+** MCP tools over a Study → Topic → Subtopic hierarchy.
- Implemented server-side **SM-2** spaced-repetition scheduling and **OAuth/OIDC + JWT** multi-user isolation; Postgres/SQLite, rate limiting, **22** test modules, Docker, Railway, and an OpenAI submission packet.

RAG Model Evaluation Tool — Side-by-Side LLM Comparison on Research Papers LangChain, Ollama, Anthropic

- Built a **LangChain RAG** pipeline (PDF → recursive chunking → Ollama nomic-embed-text embeddings → in-memory vector store) that streams side-by-side answers comparing local Ollama models against an Anthropic model.
- Added an optional **LLM-as-judge** that scores the two candidate answers, enabling head-to-head model quality assessment for RAG.

Quantitative Trading Research — ETH/USDT Momentum Strategies Python, Pandas, NumPy

- Built an **8-year** walk-forward ETH momentum backtest (365d train / 90d test) with a 1-bar execution delay and 0.13%/side cost model; returned **29.5%** annualized after costs at a **0.72** Sharpe, with rigorous risk-control analysis.

ToneTrace — Intelligent Writing Analytics for Educators FastAPI, HuggingFace, PostgreSQL, Docker

- Built a **FastAPI + HuggingFace Transformers** backend for multi-label tone/sentiment/style analysis, containerized with Docker and backed by PostgreSQL for per-user writing history.

Education

Texas State University — B.S. in Computer Science & Mathematics, Honors Expected May 2028
San Marcos, TX | GPA: 4.00, Honors College
Relevant Coursework: Linear Algebra, Data Structures & Algorithms, Probability & Statistics, Real Analysis, Numerical Analysis, Abstract Algebra, Calculus 3, Topology, Differential Equations, Discrete Math, Proof-Based Math.

Technical Proficiencies

Python | FastAPI | Docker | LangChain | Ollama | Scikit-Learn | NumPy | Pandas | SQL | PostgreSQL | LLM Agents | Multi-Agent Systems | RAG / Vector Embeddings | Diffusion Models (DDPM) | MCP (FastMCP) | OAuth / OIDC / JWT | Railway | Git / GitHub | GitHub Actions | Linux | Network Fundamentals | C++

Awards

- Computer Use Prize, HackTX 2025 — one of only 20 winning teams among 1,000+ participants.
- AI4ALL Ignite Fellow — wind energy project selected for the AI4ALL Ignite Fellowship student symposium.