Shivashish Khanna

612-517-4924 | Madison, WI | skhanna32@wisc.edu | LinkedIn

Education

University of Wisconsin–Madison Madison, WI

Bachelor of Science in Computer Science, Mathematics, and Economics *Expected Graduation May 2029*

- **Related Coursework:** Data Structures and Algorithms, Programming III, Computer Architecture, Discrete Mathematics, Linear Algebra, Calculus I-III, Probability and Statistics, Statistical Modeling and Regression

Technical Skills

Developer Platforms: Git, Docker, Linux, VS Code, IntelliJ, PyCharm, Jupyter, Hugging Face, Codex, Claude Code
Frameworks and Libraries: React.js, Node.js, TensorFlow, PyTorch, Pandas, NumPy, Scikit-learn, PyMC, FreeRTOS
Languages: Python, Java, C/C++, SQL, JavaScript, TypeScript, HTML/CSS, R, MATLAB, X-64 Assembly
Technical Skills: Data Structures and Algorithms, Object-Oriented Programming, Software Development, AI/ML, LLMs, Data Modeling, REST APIs, CI/CD, Embedded Systems

Experience

Software Development Intern July 2026 – Present

Minnesota NLP Lab *Minneapolis, MN*

- Developed a unified Python framework for **OpenAI Codex and Claude Code**, standardizing agent execution, logging, and performance monitoring across **30+** artificial intelligence research workflows.
- Implemented a **FastAPI** backend supporting session management, turn execution, and **dynamic model switching**, integrating **Model Context Protocol (MCP)** tooling and modular adapters for extensibility.
- Established and tested a structured **SQL-based data layer** for tracking model interactions, token usage, execution metrics, and task state, supporting reliable experimentation and performance analysis to support **LLM routing**.

LLM Research Intern May 2025 – Oct. 2025

University of Minnesota - Twin Cities *Minneapolis, MN*

- Architected **SpireEval**, a modular Python framework integrating SpireComm, **custom communication interfaces, and JSON-based experiment pipelines** for evaluation of **LLM agents** in gaming environments.
- Implemented scalable JSON schemas along with **preprocessing pipelines** that converted live SpireComm game objects into **normalized datasets**, extracting **combat context, card metadata, player state, and derived strategic features** from **100,000+ agent decisions** to support large-scale automated evaluation.
- Created **Bayesian evaluation models** using **PyMC, PyTensor, SciPy, and NumPy**, quantifying long-horizon strategic performance of modern LLMs and contributing to an accepted IEEE ICDMW publication.

Firmware Developer September 2026 – Present

Formula SAE *Madison, WI*

- Upgrade **embedded systems firmware** in C for onboarding projects, implementing **GPIO control**, **interrupt handlers**, and **FreeRTOS tasks** and queues for **button-driven LED channels and per second logging**.
- Integrate **STM32 temperature monitoring firmware** integrating an **I2C sensor and display**, temp-based LED indicators, and a UART console for **battery management systems**.

Projects

Adaptive LLM Router | *Python, Codex, vLLM/SGLang, SQL, Git, Hugging Face, APIs* July 2026 – Present

- Deployed an **adaptive multi-LLM routing system**, ensuring precise agent state management while **dynamically switching models** at tool call boundaries across **500+** SWE-Bench Verified software engineering tasks.
- Created a **backend system** with modular model adapters, parallel task execution, human-in-the-loop routing, session management, persistent turn history, and real-time tracking of **token usage and tool calls**.
- Integrated **local Qwen models via vLLM/SGLang** along with **frontier OpenAI Codex models** through an adapter interface, enabling cross provider model switching and context preservation within active agent tasks.
- Refined automated routing signals from testing data including **runtime failures, token usage, dependency errors, and execution progress**, reducing token costs by **36%** while maintaining task success rate over single model baselines.

Samsung Solve for Tomorrow | *Python, XGBoost, Matplotlib, Pandas, NumPy* Nov 2025 – Jan 2026

- Built an **XGBoost machine learning pipeline** using environmental data, feature engineering, and model evaluation to forecast Lake Superior nearshore wind production.
- Compared and validated two predictive models, **linear regression and gradient-boosted models** to assess nearshore wind-farm feasibility, contributing to selection as a **Minnesota semifinalist**.