Salman Said

Saint Paul, MN

612-433-1186 linkedin.com/in/salman-said salmansaid.personal@gmail.com
https://github.com/currentlycrafting

Education

University of Minnesota Expected May 2027
Bachelor of Arts in Computer Science - Minor: Mathematics *GPA: 3.6/4.0*

Skills

Languages & ML: Python, Java, C/C++, JavaScript, HTML5, CSS, PyTorch, Hugging Face, scikit-learn, Pandas, NumPy

Backend & Tools: Flask, NodeJS, AngularJS, TailwindCSS, React, Spark, SQL/NoSQL, Docker, Git, CI/CD, Agile/Scrum, JIRA, TCP/IP

Experience

University of Minnesota — Human-Centered Computing Researcher Jun. 2025 – Present

- Optimized a scalable data ingestion and processing pipeline by integrating a **TikTok Research API** with **Apache Spark/MySQL**, handling over **12TB** of content. Leveraged `openai-whisper-large-v2` to transcribe audio content for input into **PyTorch/Hugging Face** models.
- Contributed to the design of **CLIP** and **LLaVA models** using **LoRA/QLoRA** on **PyTorch/Hugging Face Transformers** to interpret Algospeak and veiled intent in speech, achieving a **15% reduction in false positives** for classifications.
- Collaborated on architecting a **Human-in-the-Loop (HITL) Reinforcement Learning pipeline** that integrates expert feedback. Contributed to research slated for **official publication**.

Headstarter AI — Software Engineer Intern (Remote) Jun. 2024 – Sept. 2024

- Developed a `/risk-audit` API endpoint for anti-fraud, using an **XGBoost model** and **Redis caching** to cut data retrieval time by **15%** and support **10,000+ concurrent requests**. Applied **OOP** and **RESTful API** principles to create clean and maintainable code. Leveraged **TCP** protocols for high-throughput connections to **NVIDIA A100 GPUs**. Designed with CAP theorem in mind, balancing data consistency and partition tolerance.
- Engineered a risk management and reporting system using **GPT-4-Turbo** for SQL generation and **Microsoft 365 Copilot** for comprehensive reports.
- Implemented robust input-validation, ensuring **99%+ data integrity**, boosting risk coverage by **15%** and anomaly detection by **20%**. Achieved **25% faster feature delivery** within an Agile team.

Projects

Reddit Sentiment & Keyword Clustering Analysis

- Developed an **NLP-based sentiment analysis tool** for **10,000+ Reddit posts**, extracting sentiment and keywords using **TextBlob**, **scikit-learn**, and **KMeans**, with interactive **Plotly** visualizations.
- Achieved high-performance data ingestion (**1.2s** for large CSVs, **5MB RAM**) and implemented a **binary search keyword system** for rapid lookups (**10,000+ in 10.3ms**). Validated system performance with **multi-threaded tests** (**10 concurrent requests in 51ms**), maintaining efficient memory usage (**82MB**).

Aqoonta: Interactive CS Learning Platform

- Prototyped and developed a **desktop application** for interactive CS education using **Electron**, **NodeJS** for boilerplate, **Python/Flask**, and **PostgreSQL**. Engineered the **macOS executable (.app)** by leveraging **PyInstaller** for the Python backend and **Electron's native packaging**.
- Designed and implemented a **responsive UI** (**HTML**, **JavaScript**, and **Tailwind CSS**). Packaged into a **standalone macOS application** and containerized for streamlined development with **Docker**.