

Khalid Mehtab Khan

AI Engineer | NLP Researcher

San Francisco, CA | 4042637813 | khalidmehtabk@gmail.com | linkedin.com/in/khalidm-khan
github.com/Kahl-d | https://khalidmk.vercel.app

EDUCATION

Master of Science, Data Science & Artificial Intelligence | GPA: 4.0, Aug 2023 – present | San Francisco
San Francisco State University

- **Coursework:** Statistical Modeling, Data Mining, Deep Learning, Data Visualization, Algorithm Analysis

SKILLS

Programming & Development: Python, C++, PostgreSQL, JavaScript, React, Next.js, Git
Data Engineering & Cloud Technologies: AWS, Flask, MongoDB, REST APIs
AI & Machine Learning: PyTorch, Hugging Face, Generative AI, Transformer Models, AI Agents, RAG
NLP Expertise: LLM Fine-Tuning, Prompt Engineering, Word Embeddings, ML Pipelines, Model Evaluation
Data Science: Statistical Modeling, Feature Engineering, Clustering, Decision Trees, Data Visualization

PROFESSIONAL EXPERIENCE

Graduate Research Assistant, AI & ML Research, SFSU Aug 2024 – present | San Francisco

- Fine-tuned DeBERTa and T5 models using prompt engineering and hyperparameter optimization, improving theme identification accuracy to 91% and reducing manual annotation time by 50%.
- Integrated Masked Language Modeling (MLM) and hybrid transformer architectures, achieving a 15% improvement in model generalization across diverse datasets.
- Evaluated data augmentation techniques (SMOTE) for structured data, improving data pipeline efficiency.

Healthcare Data Analyst, Innovaccer Jan 2021 – Sep 2022 | Noida, India

- Automated data pipelines using Python and SQL, increasing lead generation efficiency by 20%.
- Designed interactive dashboards and visual analytics, enabling real-time performance tracking and data-driven decision-making for healthcare networks.
- Led a cross-functional team of 3 to gather requirements and deliver customized data solutions, improving operational efficiency and reducing the lead generation time by 15%.

PROJECTS

Context-Aware Data Augmentation for Low-Data Scenarios Aug 2024 – Dec 2024

- Led a team of frontend, backend, and ML design leads to develop a context-aware data augmentation system that solved low-data training challenges.
- Built a generative AI-powered pipeline using structured perturbations aligned with dataset schemas, enhancing model training accuracy by 3–5%.
- Improved model decision reliability by 15% through implementation of local feature importance analysis.

Chain of Thoughts Analysis for Multi-Class Image Classification Jan 2024 – May 2024

- Implemented a framework to analyze decision-making processes and enhance classifier performance.
- Applied chain-of-thought analysis to identify and address misclassification patterns in multi-class classifiers.
- Leveraged visual analytics to interpret model behaviors and implement targeted accuracy improvements.

POSITION OF RESPONSIBILITY

Teaching Assistant – NLP, San Francisco State University Aug 2024 – present

- Conducted two classes for CSC 820 on logistic regression and fine-tuning LLMs using Hugging Face.
- Mentored peers in data science concepts via the AI Stars initiative and workshops on AI innovation.

Resident Assistant, ResLife, San Francisco State University Aug 2024 – present

- Fostered a supportive environment for 48+ residents through active programming and engagement initiatives.
- Curated and facilitated community events focused on Justice, Equity, Diversity, and Inclusion (JEDI) themes.