Khalid Mehtab Khan

+1-404-263-7813 | khalidmehtabk@gmail.com | linkedin.com/in/khalidm-khan | github.com/Kahl-d

EDUCATION

Master of Science in Data Science and Artificial Intelligence Aug 2023 - May 2026
San Francisco State University *San Francisco, CA*

- GPA: 4.0/4.0
- **Thesis**: *AWARE: Beyond Sentence Boundaries: A Contextual Transformer Framework for Identifying Cultural Capital in STEM Narratives*
- Relevant Coursework: Deep Learning, Natural Language Technologies, Statistical Modeling, Analysis of Algorithms

Bachelor of Technology in Computer Science and Engineering Aug 2017 - Sep 2021
The LNM Institute of Information Technology *Jaipur, India*

- **Capstone Project**: *Textual Analysis of Aspects of Divinity Using Sentiment Analysis*
- Relevant Coursework: Data Structures & Algorithms, Operating Systems, DBMS, Discrete Mathematics

SKILLS

Languages & Tools: Python, SQL, C++, PyTorch, Hugging Face, Git

ML & Modeling: Supervised Learning, Unsupervised Learning, Classification, Reinforcement Learning, Transformers (BERT, Mistral, GPT), Embeddings, Fine-tuning (PEFT, LoRA), Model Optimization

LLMs & AI Systems: Generative AI, Large Language Models, Retrieval-Augmented Generation (RAG), Semantic Search, AI Agents, Voice Agents, Real-time Inference, Vector Databases, WebRTC

Evaluation & Deployment: Model Evaluation, Error Analysis, Metrics, Azure (Azure ML, Azure OpenAI), High-Performance Computing (HPC), On-device Inference, vLLM

PAPERS AND PREPRINTS

Khan, K. M. & Kulkarni, A. (2025). *AWARE: Beyond Sentence Boundaries: A Contextual Transformer Framework for Identifying Cultural Capital in STEM Narratives*. Submitted to ICMLA 2025. arXiv:2510.04983

RESEARCH EXPERIENCE

Graduate Research Assistant Aug 2024 - Present
NLP Lab (Advisor: Dr. Anagha Kulkarni) *San Francisco, CA*
Department of Computer Science - San Francisco State University

- **Project: ALMA: Cultivating Cultural Capitals in STEM through Reflective Journaling**
- Studying whether incorporating narrative-level awareness enables BERT-based language models to identify cultural capital in STEM student reflections, where sentence-level processing often loses semantic meaning
- Investigated failure modes of sentence-level DeBERTa-v3-large classifiers, showing that independent sentence processing obscures narrative cues needed to identify cultural capital themes
- Introduced domain, context, and cross-theme awareness through domain-adaptive masked language model pretraining (MLM), essay-level context aggregation with BiLSTM attention pooling, and multi-label learning for overlapping themes
- Improved downstream multi-label classification under severe class imbalance, achieving a +2.1% Macro-F1 gain and enabling an annotation-support pipeline that reduced manual labeling time by ~75% (SF BUILD Agents of Change Fellowship)

Graduate Research Assistant Jan 2024 - Aug 2024
Data Visualization & HCI Lab (Advisor: Dr. Shahrukh Humayoun) *San Francisco, CA*
Department of Computer Science - San Francisco State University

- Investigated hierarchical image classification models using an ImageNet-derived dataset of animal species (~1,000 classes), organizing classes according to biological taxonomies to support fine-grained species recognition
- Built a hierarchical classification pipeline to trace inference paths and identify where predictions diverged from ground-truth labels; applied feature-attribution methods to analyze visual evidence driving model decisions
- Observed that model predictions were often influenced by background and environmental correlations rather than animal morphology, revealing limitations of hierarchical structure for reliable visual reasoning

Undergraduate Researcher   Aug 2020 - May 2021

Advisors: Dr. Payel Pal & Dr. Vikas Bajpai   *Jaipur, India*
The LNM Institute of Information Technology

- Examined semantic and emotional shifts across theological and scientific texts, comparing how meaning and sentiment evolve in religious narratives versus evolutionary prose
- Analyzed large-scale textual corpora to quantify linguistic and affective patterns across distinct narrative structures, enabling cross-domain comparison of discourse-level properties
- Quantified statistically significant differences in the semantic framing of religion, gender, and societal norms using lexicon-based analysis, revealing systematic variation in expressive style across domains

PROFESSIONAL EXPERIENCE

Founding AI Engineer   Aug 2025 - Dec 2025

Xuman AI   *San Francisco, CA*

- Led 0-to-1 development as founding AI engineer, working closely with the CEO to identify use cases, shape product direction, and ship agent-based systems
- Built real-time voice agents using WebRTC for low-latency streaming, enabling interruptible, human-like conversations
- Developed AI agents combining retrieval-based search, structured prompting, and tool use to generate grounded responses. Added tools to perform actions like navigation and booking within the application
- Designed evaluation pipelines assessing factual correctness, response quality, cross-turn consistency, and behavioral alignment using LLM-as-a-judge and manual validation

Data Analyst   Jan 2021 - Sep 2022

Innovaccer   *Noida, India*

- Developed components of scalable ETL pipelines processing 10M+ patient records; mapped raw clinical data (HL7/C-CDA) to FHIR-based Unified Data Models
- Optimized clinical data ingestion latency by 30% through automated Python validation scripts for HEDIS and Risk Adjustment workflows

TEACHING EXPERIENCE

Instructor - Data Science & Machine Learning   Aug 2025 - Present

Genentech & College of Global and Professional Education (CPaGE), SFSU   *San Francisco, CA*

- Instructed a structured 5-course Data Science & Machine Learning certificate program for Genentech professionals, delivering hands-on Python, machine learning, and neural network training tailored to biomedical use cases

Teaching Assistant - CSC 620/820 (Natural Language Technologies)   Aug 2024 - Aug 2025

San Francisco State University   *San Francisco, CA*

- Served as Teaching Assistant for the Natural Language Technologies course, supporting a mixed cohort of graduate and undergraduate students
- Delivered full lectures on n-gram models and logistic regression, helping students build intuition for probabilistic language modeling and classification

SELECTED PROJECTS

Secure Sense   **Jan 2025 - Mar 2025**
Emerging AI Innovation Winner, SF Hacks | Incubated at Lam Family College of Business   *San Francisco, CA*

- Built Secure Sense, a privacy-focused AI system exploring on-device and local-only inference architectures for handling sensitive user data in real-world applications
- Recognized by OpenMind (title sponsor/co-host), SF Hacks organizers, and the Lam Family College of Business, with the project featured in the university's annual report following the Emerging AI Innovation win

Context Aware Data Augmentation   **Aug 2024 - Dec 2024**
Graduate Seminar - AI Entrepreneurship (Advisor: Dr. Isabel Song)   *San Francisco, CA*

- Built a context-aware data augmentation framework to address data scarcity and performance degradation in biomedical text classification caused by irrelevant or noisy training samples
- Implemented schema-preserving, constraint-based perturbations conditioned on tabular metadata to generate synthetic data while maintaining semantic consistency and feature relationships
- Validated the framework through downstream classification, achieving a 3.5% accuracy improvement and improved robustness under class imbalance compared to baseline training data

AWARDS & FELLOWSHIPS

Winner: Emerging AI Innovation Track   **2025**
SF Hacks   *San Francisco, CA*

- Awarded for "Secure Sense", a privacy-first AI tool leveraging on-device LLMs. Selected for incubation at Lam Family College of Business Innovation Center

SF BUILD Scholar (Agents of Change Fellowship)   **2024**
San Francisco State University / NIH

- Awarded $22,000+ in cumulative research funding. A highly selective initiative funded by the National Institutes of Health (NIH) in partnership with Purdue University to enhance diversity in AI and biomedical research

AI STARS Fellow   **2024**
San Francisco State University

- Selected as a high-potential research fellow in this competitive initiative designed to advance training in Artificial Intelligence and social impact. Received funding to pursue innovative interdisciplinary research

POSTER PRESENTATIONS

Khan, K. M. (2025). *Defining the Trajectory of Narrative Modeling in NLP*. College of Science & Engineering (CoSE) Student Project Showcase, SFSU

Khan, K. M. (2025). *AWARE: Essay-Aware Representation Learning*. Graduate Research and Creative Works Showcase, SFSU

Khan, K. M. (2024). *Visualizing Model Reasoning: An Audit of Hierarchical Classification*. College of Science & Engineering (CoSE) Student Project Showcase, SFSU

LEADERSHIP

Student Leader / Resident Assistant   **Aug 2023 - Aug 2024**
Residential Life, San Francisco State University   *San Francisco, CA*

- Received $30,000 in housing compensation based on demonstrated leadership excellence and commitment to fostering inclusive campus communities
- Fostered an inclusive living-learning community by organizing diversity-focused events and facilitating conflict resolution mediation