# Mohid Tanveer

mohid.tanveer011@gmail.com | (731) 803-9164 | LinkedIn | Portfolio

## EDUCATION

**University of California, San Diego** **September 2025 - December 2026**
*M.S. Computer Science & Engineering, Overall GPA 3.9/4.0* *San Diego, CA*

- **Relevant courses**: Recommender Systems and Web Mining, Probabilistic Reasoning & Learning, ML: Learning Algorithms, Algorithm Design and Analysis, Unsupervised Learning, Computer Security

**Rhodes College** **August 2021 - May 2025**
*B.S. Computer Science, Overall GPA 3.91/4.0* *Memphis, TN*

- **Relevant courses**: Machine Learning, Artificial Intelligence, Mathematical Statistics, Statistical Analysis, Multivariable Calculus, Advanced Algorithms, Theory of Computation, Systems Programming and Computer Organization

## SKILLS

**Programming Languages:** Python, C, C++, Java, JavaScript, TypeScript, SQL, R, Bash, HTML/CSS
**Tools and Platforms:** Git, Node.js, React, REST APIs, Docker, Kubernetes, Linux (RHEL), Virtual Machines, PostgreSQL
**ML and Data Science:** Power BI, TensorFlow, PyTorch, LangChain, PySpark, NumPy, SciPy, Pandas, Matplotlib, Scikit-learn
**Languages:** Fluent in English, Urdu, and Punjabi; Conversational Proficiency in Spanish, Hindi | **Honors:** Eagle Scout

## EXPERIENCE

**St. Jude Children's Research Hospital** **September 2024 - May 2025**
*Student Machine Learning Engineer, Research Systems* *Memphis, TN*

- Reduced ServiceDesk ticket volume by 80% by analyzing support ticket data, deploying a production LLM-powered response system, and evaluating retrieval and response quality across real user queries.
- Designed a Retrieval-Augmented Generation inference pipeline using LangChain chains, custom retrievers, and embedding-based vector stores, combining NLP preprocessing, semantic retrieval, local LLM inference, and knowledge graphs to reduce hallucinations and improve answer precision on internal benchmarks.
- Enabled active research projects by designing and integrating machine learning pipelines, conducting data preprocessing, feature engineering, model training, evaluation, and end-to-end pipeline integration in collaboration with domain researchers.

*Applied Data Science Intern* **May 2024 - August 2024**

- Collected and analyzed large-scale storage and network telemetry data to identify performance degradations, building automated monitoring, anomaly detection, and data visualization pipelines.
- Streamlined researcher workflows by automating and scaling the preprocessing of large bioimaging datasets by building a backend microservice integrated with an internal imaging platform and Slurm-based batch processing.
- Improved resource allocation decisions by implementing logging, monitoring, and observability for Windows and Linux systems, analyzing workstation utilization, and generating actionable system metrics and data visualizations.

**Rhodes College Department of Computer Science** **August 2023 - May 2025**
*Peer Tutor* *Memphis, TN*

- Tutored and mentored ~200 introductory and intermediate Computer Science students through twice-weekly sessions covering data structures, algorithms, object-oriented programming, memory management, pointers, and system-level design.

## PROJECTS

**Tubify** **February 2025 - April 2025**
*Senior Capstone Project* *GitHub Repository*

- Designed and developed a full-stack web application enabling social music discovery by integrating Spotify and YouTube APIs to analyze user and friend listening histories, support shared playlists, and surface personalized recommendations.
- Engineered a scalable backend and data pipeline to batch-process, store, and query audio features and user interaction data using a PostgreSQL-backed architecture, supporting efficient retrieval and long-term persistence.
- Built a personalized music recommender system using collaborative and content-based filtering, taste clustering, and MMR reranking, leveraging audio feature extraction and CUDA-accelerated FFT analysis to improve recommendation quality.
- Evaluated recommendation quality using offline ranking metrics and user-level preference similarity.

**ScreenSense** **December 2025**
*Course Project* *GitHub Repository*

- Built a hybrid computer vision detection system utilizing Pytorch and signal processing to identify screen re-photography at verification time by combining EXIF metadata analysis, wavelet-based CNN models, and sub-pixel feature extraction.
- Achieved state-of-the-art detection performance by fusing a metadata-driven random forest prior with learned pixel-level signals, demonstrating robustness across diverse image conditions.