Mohak A. Prakash

✉ mprakash@ucsd.edu • 🔗 linkedin.com/in/mohakap • 🌐 tc960.github.io

Education

University of California, San Diego   Expected March 2027
- **B.S. in Data Science**, Minor in Neuroscience **GPA: 3.8**/4.00, Provost Honors (Winter 2024 - Present)
- **Relevant Coursework:** Probabilistic Modeling & Machine Learning, Representation Learning, Data Mining, Signal Processing for Data Analysis, Systems for Scalable Analytics, Data Management, Reinforcement Learning, IoT Sensors & Embedded Systems

Skills

Python, JavaScript, TypeScript, React, Flutter, FastAPI, Flask, Pandas, TensorFlow, PyTorch, Scikit-learn, Cvxpy, OpenCV, MediaPipe, LangChain, LangGraph, LlamaIndex, SQL, MongoDB, Firebase, AWS, GCP, Docker, Kubernetes, Spark, Dask, PCA, FFT/Signal Processing, Clustering (K-means, DBSCAN, GMM), SVM, Random Forests, Q-Learning, Transformers, OCR, LLMs & VLMs (OpenAI, Claude), RAG, Knowledge Graphs, NER, RelExtract, MCP, Twilio, Ngrok

Work Experience

Research Assistant — *Proto Lab, UC San Diego Design Lab*   Sep 2025 - Jan 2026
- Built **AI-powered meeting simulator** enabling multi-round agent discussions with configurable aggressivity; reduced latency from **12-14s to 0.1-0.5s** using custom **sliding-window optimization algorithm**.
- Developed **speaker diarization and persona generation pipeline** using **Pyannote, Rev AI, NVIDIA NeMo/Riva** to extract real speech patterns for human-like AI persona creation.

Data Analytics Intern — *EXL Services*   Jun 2025 - Sep 2025
- Contributed to **RiskXpert**, an agentic AI tool for automated insurance risk analysis; built and benchmarked multiple **OCR pipelines** including **AWS Textract, DocAI, PaddleOCR+LayoutLMv3**, and custom **CNN+RNN+CRF** architecture across diverse document structures
- Designed weekly **Power BI dashboards** visualizing risk metrics and insurance data for client deliverables and stakeholder decision-making; ensured **SOC 2 and GDPR compliance** across all solutions

Research Assistant — *Knight Lab, University of California San Diego*   Jan 2025 - Present
- Scraped and cleaned **2000+ research articles** from the MicrobiologyMetadataCrisis Project (**250+ members**) achieving **90% success rate across 50+ journals** for biomedical knowledge graph construction
- Fine-tuned **BioBERT and PubMedBERT** on MicroBioRel dataset surpassing **70% F1-score benchmark**; currently testing **BioMistral, Kimi-K2, etc** to identify optimal LLM for relation extraction for KnowledgeGraph creation.

Projects

Few Word Do Trick — Project Link   Winner @ CalHacks '25
- Built assistive communication platform for speech impediments combining **real-time EEG emotion classification** (**Fourier Transforms, PSD analysis, ICA, PCA, Random Forest** achieving **90% accuracy**) with **MBTI personality modeling** to generate context-aware, emotionally nuanced speech from simple keyword input via **GPT-5 and ElevenLabs**; implemented **FastAPI backend** with **React/Vite frontend** using **Ngrok** for real-time signal processing.

Clueless — Project Link   Winner @ HackMIT '25
- Developed **AI-powered web navigation assistant** that automatically maps web applications into **Neo4j knowledge graph** and provides voice/text-based guidance; built **multi-agent backend** with **Claude** for intent classification, **Pinecone + Sentence Transformers** for semantic RAG, and **Cerebras** for accelerated reasoning, plus **React/TypeScript Plasmo** extension with **BFS traversal** for real-time step-by-step UI guidance

TaxDaddy — Project Link   Winner @ Hacklytics '25
- Built **multi-agentic real-time voice tax assistant** using **WebSockets and Twilio** integration; implemented **RAG-based system** with **Pinecone VectorDB** for tax guidance, **OCR document processing** with **PyPDF and MongoDB** for automated form completion, **fraud detection using Benford's Law** analysis, and real-time analytics/anomaly detection across unstructured tax documents

Achievements and Activities

5x Hackathon Winner: CalHacks '25 (3rd - Lava Track) • HealthLink '25 (Winner) • HackMIT '25 (2nd - Wispr Track) • Hacklytics '25 (3rd Overall) • CodeDay '23 (1st Overall)
HDSI Department Tutor (DSC 40A, DSC 20) • **Triton NeuroTech Project Lead** • **DS3 DataHacks Organizer**