Rushil Chandrupatla

(408) 916-5410 | rushilcs@gmail.com | linkedin.com/in/rushil-c | github.com/rushilcs

Education

University of California San Diego San Diego, CA

B.S. Data Science | GPA: 3.9/4.0 (Major), 3.87/4.0 (Cumulative) *Sept 2022 - March 2026 [Expected]*

Courses: Data Structures, Algorithms, Deep Learning, ML: Learning Algorithms, Trustworthy Machine Learning, Data Management, Data Science Principles, Probabilistic Modeling and Machine Learning, Theoretical Foundations of Data Science, Linear Algebra

Skills

Languages & Tools: Python, Java, SQL, PostgreSQL, Bash, Git, Docker, Kubernetes, Amazon Web Services (EC2, ECS, S3, Lambda), Google Cloud Platform (Vertex AI, Vision API), Linux, PyCharm, VSCode, Jupyter, Agile/Scrum

Technical Skills: Deep Learning, LLMs, In-Context Learning (ICL), RAG Systems, Prompt Engineering, Computer Vision, Unsupervised Learning, Clustering Algorithms, Transformer Models, Model Evaluation & Benchmarking, MLOps, API Integration

Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, NumPy, Pandas, Matplotlib, OpenCV, LangChain, OpenAI API, Claude API, HuggingFace, Flask, FastAPI, NLTK, BERT, Google Vision AI

Experience

BILL | *Machine Learning Engineer Intern* June 2025 - September 2025 | *San Jose, CA*

- BILL is a leading fintech platform providing intelligent, automated payment solutions for millions of small and mid-sized businesses.
- Designed an AI-driven code-generation framework that reduced new ERP integration build time by **45-55% (approx. 6 months)**, enabling BILL to expand its Sync Platform to new accounting systems significantly faster.
- Reverse-engineered existing ERP handlers (NetSuite, QuickBooks, Xero, Intacct, Sage, etc.) to derive invariant sync patterns and construct generalized, model-ready templates for automatic generation.
- Built a multi-model prompt-engineering pipeline using **Claude Sonnet 4**, GitHub Copilot, and **Claude Code** to generate entity-level handler scaffolds with clear boundaries between reusable sync logic and ERP-specific implementation requirements.
- Developed CLI automation that leverages Claude Code to generate full integration skeletons in **under 2 hours**, enabling engineers to generate new ERP scaffolds with a single command. directly contributing to BILL completing the new **Acumatica integration in a record 5 months**.
- Created a **novel ML-driven code-evaluation metric** using GumTree AST-edit diffs, action-weighted structural scoring, and cosine-similarity boosts to assess functional and structural similarity between AI-generated and production code.
- Shipped a reusable generation + evaluation framework now adopted by the Sync team, accelerating BILL's ERP expansion roadmap and serving as a long-term ML assistive system for engineering velocity.

Data Science Student Society | *Consulting Director* October 2024 - June 2025 | *San Diego, CA*

- Secured 6 industry-facing projects across 3 San Diego startups through targeted outreach and relationship building
- Led selection process for **30+** students from a 200-applicant pool by designing applications and conducting interviews
- Oversaw project execution and provided technical + strategic support to ensure successful outcomes and student growth
- Directed final client presentations to company executives, with deliverables informing business decisions or product development

BILL | *Machine Learning Engineer Intern* July 2024 - September 2024 | *San Jose, CA*

- Built a custom clustering system to automatically group millions of historical ERP–BILL sync error messages, enabling standardized error codes and improving the end-to-end debugging experience for customers and support teams.
- Analyzed years of production error logs using SQL and benchmarked multiple unsupervised approaches (dendrogram-based clustering, GPT/LLM semantic grouping) before designing a bespoke, scalable clustering algorithm.
- Developed a boosted cosine similarity scoring method that incorporates token-level similarity with start/end-match boosts and domain-specific heuristics, resulting in a **30% higher clustering efficiency** compared to the prior implementation.
- Implemented a continuous-update pipeline where new, unseen errors are automatically flagged for PM review and seamlessly integrated into the global cluster set, ensuring long-term adaptability with no system downtime.
- Deployed the system as a real-time inference service on **AWS EC2/ECS**, achieving **99%+ classification accuracy** while cutting operational costs by **50%** versus the previous LLM-heavy approach.

- Delivered a production ML system that standardized error semantics across BILL's Sync Platform, giving teams visibility into top issues and enabling faster diagnosis of sync failures for tens of thousands of customers.

SEELab UCSD | *Research Intern* **|** https://arxiv.org/abs/2502.02883 **|** https://arxiv.org/abs/2501.04974 Feb - August 2024 | *San Diego, CA*

- Contributed to two research projects on wearable-sensor LLMs and QA dataset benchmarking
- Reproduced baseline benchmarks across multiple LLMs to validate the performance gains achieved by Compositional Attention models on temporal-reasoning and sensor-understanding tasks.
- Cleaned, organized, and standardized a large QA dataset collected via **Amazon Mechanical Turk**, preparing it for model training, profiling, and publication.
- Conducted extensive dataset analysis using **BERT embeddings** to identify latent structure across question types, sensor modalities, and user behaviors—discovering more reliable clusters than GPT-based alternatives.
- Leveraged **RAG with OpenAI GPT** to generate candidate answers and ground-truth references, enabling the use of exact-match accuracy metrics for rigorous model evaluation.

PromoDrone | *Data Science Consultant* April 2024 - July 2024 | *San Diego, CA*

- Promodrone is a drone-based ad-tech platform that displays digital ads from the air and uses onboard vision systems to analyze viewer engagement and demographics.
- Developed and trained a **Convolutional Neural Network** for gender and demographic data collection from drone video.
- Integrated emotion and demographic data collection into algorithm pipeline via **Google Cloud** using **Vision API and Vertex AI**.
- Deployed extraction model to PromoDrone's backend infrastructure, and is **commercially available** in the product as of 2025.

UCSD Research Group | *Undergraduate Student Researcher* May 2023 - September 2023 | *San Diego, CA*

- Developed an algorithm that recursively searches for open data lakes to augment unit tables with external data, giving insight into why two variables may be correlated.
- Used Postgres and SQL to work with data tables and altered similarly developed tools including JOSIE and MATE.

Publications

Yu, X., Hu, L., Reichman, B., Chu, D., **Chandrupatla, R**., Zhang, X., … & Rosing, T. S. (2025, Oct.). *S*ensorChat: Answering qualitative and quantitative questions during long-term multimodal sensor interactions. *ACM IMWUT*, 9(3), 1–35.

Reichman, B., Yu, X., Hu, L., Truxal, J., Jain, A., **Chandrupatla, R.**, ... & Heck, L. (2025, May). SensorQA: A question answering benchmark for daily-life monitoring. *ACM SenSys* (pp. 282-289).

Projects

In-Context Learning in Transformers Case Study | *UCSD Capstone Project* *September 2025 - Present*

- Built an ICL classification pipeline using a **single-layer transformer** from scratch in **PyTorch**, implementing custom forward passes, attention mechanisms, and task formatting to evaluate ICL behavior under controlled conditions.
- Analyzing how different transformer architectures (depth, width, attention variants) acquire in-context learning capabilities and identifying the regimes where these capabilities begin to fail.
- Investigating the conditions under which **benign overfitting** emerges in transformer training and when over-parameterization enhances in-context generalization.
- Evaluating the impact of task distribution shifts, context length, signal-to-noise ratio, and dimensionality on ICL performance to determine when ICL is beneficial versus when classical supervised learning is superior.

Seizure Detection Predictor | *UCSD Data Science Student Society* *February 2023 - June 2023*

- Conducted end-to-end data science project implementing a variety of ML models [**Naive Bayes, Decision Tree, SVM, k-Means**] on 5000 EKG samples to predict if a patient is prone to seizures
- Utilized **Python, Pandas, Matplotlib, Scikit-Learn, and Tensorflow** for data preparation/visualization and model implementation
- Found SVM model using a Radial Basis Function was most effective, performing at 98% accuracy, revolutionizing the field of medical diagnoses

Awards

Evergreen Valley HS Valedictorian, Eagle Scout, Provost Honors (All Quarters), Revelle College Honors (All Years), Silver Presidential Volunteer's Service Award, A-Z Hacks 2nd Place