Vraj Patel

(510) 579-9479 | vrajmp999@gmail.com | https://www.linkedin.com/in/vrajmp/ | Boulder, CO

EDUCATION

University of Colorado at Boulder, College of Engineering and Applied Sciences ⸻ Graduated May 2025
B.S. in Computer Science - *Business Minor* ⸻ *Boulder, CO*
- GPA 3.65/4.00 (Dean's List: 5 Semesters), awarded **1st Place** across all CS Capstone projects

Relevant Courses: Cybersecurity I: Computer Systems, Cybersecurity 2: Networks & Security, Cybersecurity 3: Ethical Hacking, Data Structures, Computer Systems, Software Development & Tools, Discrete Structures, Linear Algebra w/ CS Applications, Data Science + Probability/Statistics, Algorithms, Intro to Artificial Intelligence, Database Systems, Object-Oriented Analysis/Design, Human-Computer Interaction, Advanced Data Science, Computer Vision, Principles of Programming Languages, Machine Learning, Data Mining, Deep Learning & Neural Networks, Natural Language Processing

WORK EXPERIENCE

CU Boulder Institute of Behavioral Science ⸻ May 2025 – Present
Desktop Support Engineer (Automation & Internal Tools) ⸻ *Boulder, CO*
- Developing a **full-stack ticket classification system** using **FastAPI** and **PostgreSQL**, architecting a microservices based solution to automate incident tagging via a fine-tuned **BERT transformer** model
- Developed an interactive analytics dashboard using **React (Vite)** and **Recharts**, utilizing complex state management to visualize historical data and identify critical operational trends, such as pinpointing peak ticket volume (Tuesdays at 10 AM) to proactively optimize staffing schedules
- Engineered a PowerShell automation tool to recursively scan IBS OU under Colorado AD and purge group memberships, reducing per user offboarding time by **93% (15 mins to <1 min)**
- Enforced Secure Compute compliance standards across 50+ endpoints by implementing **Windows Autopilot** and **Jamf Pro** enrollment workflows

Princeton University – Studio Lab ⸻ March 2023 – Present
Research Assistant (Data Engineering & ML) ⸻ *Princeton, NJ*
- **Political Data Pipeline (ETL)**: Engineered a cloud-hybrid pipeline to ingest **40+ GB** of election speeches (Modi & Gandhi, 2014/2019) and archive **40+ years (1981–2024)** of Lok Sabha debates, creating the largest unified dataset for Indian political linguistic analysis
- **ML & Cloud Optimization:** Architected an automated **AWS** workflow (**S3, Transcribe, Translate**) using **Boto3**, implementing Custom Language Models (CLM) to recognize niche political entities and reduce WER
- **Web Scraping System**: Developed a resumable **Selenium** crawler with **SQLite** state management to index the Parliament Digital Library, implementing logic to handle dynamic pagination and sync with **OneDrive** storage
- **Unstructured Data Parsing**: Designed a text extraction engine using **PyMuPDF** and **FuzzyWuzzy** (string matching) to structure thousands of raw PDF statements, mapping OCR text to standardized Ministry entities

The Data Mine – Purdue University @ L3Harris ⸻ August 2024 – May 2025
Undergraduate Research Assistant - Satellite Telemetry Data Analysis ⸻ *West Lafayette, IN*
- Developed machine learning models to detect and identify cyberattack-based anomalies using satellite telemetry data from the NASA Simulator for Small Satellites (**NOS3**)
- Analyzed and modeled potential space-cyber threats using the **SPARTA Matrix** and **MITRE ATT&CK** frameworks, creating synthetic telemetry data to simulate various cyberattack scenarios
- Presented research findings at the **Data Mine of the Rockies Symposium** to stakeholders from the **US Space Force**, **Lockheed Martin**, **CrowdStrike**, and **L3Harris**

University of Colorado Boulder - Institute of Behavioral Science ⸻ March 2023 – May 2025
Lead Technical Research Assistant ⸻ *Boulder, CO*
- **Computer Vision Pipeline (PyTorch/ResNet50):** Collaborated with PhD researchers in weekly Agile sprints to architect a visual bias analysis pipeline for NYT COVID-19 imagery. Engineered a custom preprocessing workflow using **OpenCV** and **Scikit-image** to perform Z-score normalization and channel-wise intensity rescaling
- **Unsupervised Learning:** Leveraged **Transfer Learning** by deploying a truncated **ResNet50** model to extract high dimensional feature embeddings, which were fed into **K-Means clustering** algorithms to uncover latent patterns in media datasets without reliance on labeled data
- **ETL Architecture:** Engineered a resilient Data Engineering pipeline to aggregate **15+ years** of legislative data. Built a custom **Selenium** and **BeautifulSoup** scraper to navigate dynamic DOM elements, implementing JSON checkpointing to ensure data integrity during long running jobs

- **API Optimization:** Developed a Python wrapper for the LegiScan API with in-memory caching and rate-limit handling, reducing redundant network requests by **40%** during bulk data ingestion

PROJECTS

Full-Stack Biometric Marathon Platform (SeeMyRace)   August 2024 - May 2025
- Architected a **full-stack web application** using **Flask (MVC)** and **Docker**, integrating a **PostgreSQL (pgvector)** backend to enable **sub-50ms** semantic facial recognition search of race participants
- Engineered an **asynchronous ingestion pipeline** that decouples heavy **TorchVision** and **DeepFace** inference from the main server thread, ensuring UI responsiveness and **<100ms latency** during bulk image uploads
- Implemented a hybrid identification algorithm combining **128-dim vector embeddings** with a custom **OpenCV adaptive thresholding OCR pipeline**, boosting athlete retrieval recall by **~30%** on high-variance images (glare, truncated/crumpled bibs)
- **Developed an interactive admin portal** using **Jinja2** and **Vanilla JS** to parse **GPX data** for dynamic map rendering, allowing race organizers to visualize athlete locations via extracted EXIF GPS metadata

Generative AI Mental Health Pipeline (BERT & Llama-3)   March 2025 - May 2025
- Engineered a **hierarchical NLP classification** pipeline by **fine-tuning BERT** with **PyTorch** and **Hugging Face Transformers**, achieving a **0.79 Macro-F1** score across 7 clinical mental health categories through rigorous hyperparameter tuning and Early Stopping
- Architected a resource-efficient LLM workflow utilizing **4-bit QLoRA quantization** to **fine-tune Llama-3.2-1B** with **Unsloth**, reducing VRAM usage by **50%+**
- Orchestrated a **Supervised Fine-Tuning (SFT) loop** using the **TRL (Transformer Reinforcement Learning)** library to align model outputs with **Cognitive Behavioral Therapy (CBT)** protocols, integrating a dynamic prompt injection system that conditions responses based on real-time sentiment analysis
- Developed a robust **ETL pipeline** processing **53,000+** clinical text samples using **Pandas** and **Regular Expressions**, implementing **Stratified Sampling** and custom **PyTorch Dataset** classes to mitigate class imbalance and optimize tokenization throughput

Real-Time NLP Detection Engine (HateBlocker)   January 2025 - May 2025
- Architected a real-time detection system coupling a Chrome Extension frontend with a Python FastAPI backend, achieving **<1s inference latency** on live Twitter feeds by efficiently processing dynamic DOM content
- Engineered a **metadata-augmented NLP** pipeline achieving **90.72% test accuracy (0.89 Weighted F1)** by stacking **TF-IDF vectors** with 16 custom linguistic features (e.g., capitalization ratios, punctuation density) to train a robust **XGBoost classifier**
- Optimized production inference by implementing a **Strategy design pattern** to dynamically hot-swap between Deep Learning (fine-tuned BERT) and Classical ML backends via environment variables, ensuring gradient-free execution using torch.no_grad()

End-to-End Disease Prediction Pipeline   October 2024 - December 2024
- Architected a dynamic Deep Learning classifier achieving **75.5% accuracy**, implementing **Neural Architecture Search** with **Keras Tuner** (Hyperband) to automatically optimize network topology (depth/width) and dropout rates
- Optimized training efficiency on the BRFSS 2015 dataset by engineering a multi-stage coarse-to-fine GridSearchCV pipeline for **SVMs** and **TensorFlow Callbacks** (EarlyStopping), reducing computational overhead by **~40%**
- **Engineered** model interpretability modules to demystify "black box" predictions, integrating **SHAP (DeepExplainer)** to translate raw tensor weights into risk factors (identifying BMI & GenHlth as primary drivers)

Automated Interview Scoring Engine   October 2024 - December 2024
- Engineered a **semantic alignment algorithm** using **SentenceTransformers (MiniLM)** and **Cosine Similarity**, converting unstructured multi-turn dialogue into structured datasets by mapping responses to a reference question bank (threshold > 0.40)
- Architected a multimodal predictive regression pipeline using **TensorFlow/Keras** and **Scikit-Learn**, achieving a **0.74 Pearson correlation** by fusing dense text embeddings with **100+ prosodic audio features** (pitch, jitter, pause duration) into a unified input tensor
- Optimized model interpretability and feature selection by deploying **SHAP** (SHapley Additive exPlanations) and **Mutual Information Regression**, isolating the top **15 critical behavioral indicators** (e.g., lexical diversity, sentiment consistency) to reduce dimensionality and prevent overfitting
- Deployed a robust validation framework using **K-Fold Cross-Validation** (k=4) and **Gradient Boosting Regressors**, reducing relative prediction error to **8%** by iteratively tuning hyperparameters (learning rate, tree depth) to minimize **Mean Absolute Error (MAE)**

Full-Stack Student Marketplace Application (Ralphie Bites)   Oct 2022 - Dec 2022

- Architected a containerized deployment using **Docker Compose**, orchestrating communication between the **Node.js/Express** application service and persistent **PostgreSQL** database to ensure environment consistency
- Engineered a secure **RESTful API** handling **10+ endpoints**, implementing custom **Express Middleware** for session management and **bcrypt** encryption to strictly validate university domain credentials
- Optimized server-side rendering performance by developing **asynchronous (async/await)** controller logic, aggregating persistent **SQL** user records with real time session states to dynamically update **EJS** dashboards

SKILLS

Languages: Python, C++, Java, JavaScript (NodeJS, EJS), SQL (PostgreSQL), Bash/Shell
ML & Data: PyTorch, TensorFlow, Keras, Scikit-Learn, Pandas, OpenCV, Hugging Face, NumPy
Tools & Platforms: Git, Docker, AWS (S3, Lambda), Linux, LaTeX, REST APIs