Sahana Hariharan

408-930-4973 | sahanahariharan@gmail.com | linkedin.com/in/sahana-hariharan/ | sahana-h.github.io

Education

University of Illinois at Urbana-Champaign May 2027
Bachelor of Science in Computer Science & Statistics *GPA: 3.95/4.0*
Relevant Coursework: Data Structures, Algorithms & Models of Computation, Computer Systems, Applied Statistics, Machine Learning, Statistical Methods, Statistics and Probability, Computational Linear Algebra

Technical Skills

Programming Languages: Python, Java, C/C++, Go, JavaScript, TypeScript, SQL, HTML/CSS, R, Arduino C
Frameworks: Flask, React.js, Node.js, REST APIs, Firebase, MongoDB, Neo4j, AWS, Docker, Databricks
Libraries: Pandas, NumPy, PyTorch, TensorFlow, OpenCV, Scikit-learn, Plotly, Hugging Face Transformers
Developer Tools: Git, GitHub, XCode, Eclipse, VSCode, Atlassian (Jira), Unit Testing, Figma

Experience

Data Engineering Intern | KPMG Jun 2025 – Aug 2025
- Engineered a low-discrepancy, message-passing Monte Carlo engine that streamlined monetary-unit sampling, improving sample coverage consistency and cutting simulation run-time by 40%
- Developed PySpark pipelines to validate synthetic audit datasets and implemented automated QA checks to ensure robust data integrity across various ERP formats for robust ETL testing
- Implemented a unified Databricks pipeline (within KPMG's proprietary workflow) to extract and format various ERP exports (ex. SAP S/4HANA, NetSuite), streamlining pre-processing and debugging

Full-Stack Project Manager | CS124 Honors Sept 2024 – Present
- Led a 10-person team in building a full-stack study room tracker for 40+ locations using React.js, MongoDB, and Google Maps API with real-time location updates
- Mentored developers in implementing sentiment tracking for products using Gemini and Vader
- Managed agile development lifecycle from feature scoping to deployment on Firebase, ensuring alignment with real student needs and learning objectives

Statistical Analyst Researcher | Illinois Risk Lab Aug 2024 – Jan 2025
- Developed Python-based financial modeling tools for RGA's Global Financial Services Valuation team to evaluate $10B+ in transactions, such as longevity swaps, asset-intensive transactions, and pension risk transfers
- Built and calibrated 15 proxy models using actuarial and statistical techniques, enabling faster performance monitoring, assumption testing, and risk profile assessment

Software Engineering Intern | Prognosis May 2024 – Aug 2024
- Leveraged Large Language Models (LLMs) to transform 3,000+ unstructured patient records into structured knowledge graphs, enabling AI-driven predictions of diseases and personalized treatment recommendations
- Conducted a comparative analysis of Node2Vec and OpenAI embedding methods within a Neo4j database to evaluate their effectiveness in generating meaningful patient knowledge graphs
- Developed a responsive web application using React.js, integrating a robust front-end with an intuitive user interface to display patient information in a dynamic, searchable table

Projects

Deepfake Detection Model | Disruption Lab Sept 2024 – Jan 2025
- Partnered with an NGO to build a digital content authenticity system supporting global trust and safety efforts
- Trained 15+ deepfake detection models (CNNs, ViTs) on a 20K+ media dataset using PyTorch and TensorFlow
- Developed model testing and flagging pipelines to reduce false positives by 25%, supporting real-time content moderation
- Integrated confidence-based thresholds and auto-flagging logic into backend services using Flask and REST APIs

Predictive Crime Analysis | altREU Teuscher Lab May 2024 – Aug 2024
- Processed features from 8.2M+ rows of historical Chicago crime data, incorporating time, location, and category-based attributes to support meaningful model creation and analysis
- Trained Random Forest and LSTM models to identify crime trends and forecast likelihoods, using cross-validation to compare performance
- Built and deployed an interactive Flask web app with RESTful APIs and Plotly visualizations, enabling dynamic geospatial filtering, time series analysis, and real-time updates for public use