Swetha Saseendran

✉ ssaseendran@umass.edu — in LinkedIn — ⌂ GitHub — ⊕ Portfolio Website — 🎓 Google Scholar

EDUCATION

University of Massachusetts, Amherst *Expected Dec 2026*
MS Computer Science **GPA: 3.95/4.0**
SSN College of Engineering, India *May 2022*
BE Computer Science Engineering, Graduated *First Class with Distinction* **GPA: 8.97/10**

TECHNICAL SKILLS

Programming: Python, Java, C, TypeScript, Shell, OOPS, DSA
ML/AI: Transformers, Vision-Language Models, RAG, LangChain, TensorFlow, PyTorch, OpenCV, MediaPipe
Backend: FastAPI, Flask, NodeJS, Spring Boot, gRPC, GraphQL
Cloud/DevOps: AWS, Docker, Jenkins, Git, CI/CD, Terraform
Data/Databases: SQL, MongoDB, ElasticSearch, Kafka, Redis, Tableau
Web/App: Angular, React, NextJS, Android Java, Flutter, Tailwind, Material UI, Streamlit, Figma

LEADERSHIP & ACHIEVEMENTS

Committee Member, Junior Analyst Council, Citi *(Mar 2023 – Dec 2024)* — Selected as one of 20 analysts nationwide; led peer learning initiatives and coordinated tech sessions.

Alumni Relations Head, ACM SSN *(Jun 2021 – Apr 2022)* — Drove alumni engagement and coordinated coding and tech events.

Professional Awards:. A-Star Applause Award, Citi (2022) • Citi Developer GOLD (2023), SILVER (2024) • Winner, ICG Debate League 2.0 • Chennai Branch winner ICG Global Analyst Hackathon 2023

EXPERIENCE

AI Research Extern, UMass Industry Mentorship Program, Bloomberg Spring 2026
- Building a **multi-agent pipeline** where: data agents ingest and enrich court documents; extraction agents run LegalBERT and LLM-based fact/event extraction on document pairs; outputs are structured for **knowledge-graph** population and **RAG**-ready retrieval.

Research Assistant, Advanced Human and Health Analytics Lab, UMass Amherst Sept 2025 – Present
- Collaborating with Harvard Medical School and Mass General Brigham on post-stroke motor/cognitive impairment research under Prof. Ivan Lee's guidance. Building scalable models on GPU to automatically annotate linear movement in egocentric video using advanced visual language models (**V-LLaMA**) and RLHF.

Software Engineering Intern, Center for Data Science, UMass Amherst Mar 2025 – Present
- Built and led a production LLM platform called KeyMaker (LiteLLM-based) integrating multiple AI providers and custom services in UMass; deployed on AWS via Terraform (ECS/Fargate, Aurora Serverless), private VPC, internal ALB, VPN-only access; scaled to 2,000+ users with multi-provider orchestration, usage tracking, and benchmarking.
- Engineered an Android solution for Bluetooth-based sensor touch data acquisition from Shimmer devices for stroke patients, featuring a doze-resistant custom SPP protocol and dual-mode file persistence (local and **AWS S3**), real-time monitoring with **Crashlytics**. Created cloud sync APIs using **FastAPI** deployed on **AWS Lambda**. Medical trials for this app in progress in Northwestern University Feinberg School of Medicine.
- Deployed an **R-based API** for avian flu analytics leveraging MDP, Dockerized on **AWS EC2** with CI/CD. Built a React dashboard for real-time visualization. Bootstrapped a data scraping pipeline and orchestrating scheduled runs with **ECS Fargate** and **EventBridge** for daily automation. Optimized API by slashing CPU I/O wait time by 70% through in-memory caching.

Technology Analyst, Citi, Chennai Aug 2022 – Dec 2024
- Achieved a **60%** reduction in API response time by creating an API connector service with **asynchronous data streams** and reduced the development cycle by an entire sprint through designing a **proxy bridge service** for whitelisted APIs. Enhanced configuration management by developing a tool with the **DFS algorithm** to compare multiple **YML** config files, integrated into the **DevOps pipeline** improving deployment time by **22%**.
- Designed and developed a customizable process improvement project that creates real-time mock APIs from **OpenAPI** spec files, significantly reducing development time by eliminating API dependencies.
- Improved user engagement by creating an API to monitor services, onboarding two critical services, and achieving a **90%** code quality rating. Developed and maintained statistical APIs for financial markets, with data analysis using **Tableau** for algorithmic trading and risk modeling.

Data Science Intern, First Insight, Chennai Jul 2021 – Dec 2021
- Developed an aspect-based sentiment analysis system using LDA and BERT Transformers, improving topic coherence by 20%. Customized for user-defined aspects and deployed as a REST API, integrating it into a machine-learning pipeline for seamless access and efficiency.

Computer Vision Research Intern, SRIC, IIT Madras May 2021 – Nov 2021
- Developed a motion analysis system for athlete biomechanics using OpenCV, Mediapipe, and YOLO, deployed via Flask API, with a custom basketball dataset achieving 82% accuracy. Engineered a computer vision pipeline to monitor KPIs like shooting hand detection, knee angle analysis, and shot count for performance insights from video feed. Led and mentored a team of 5+ RAs through technical sessions and guidance.

PROJECTS & RESEARCH

AirgapAgentLite - Privacy-Preserving LLM Framework, UMass Amherst More Info Project Poster Github Repo
Built a RL based two-LLM AirGapAgent pipeline using Mistral-7B on GPU and enforced contextual privacy via a hybrid, lightweight data minimizer (rule-based + small LLM/transformer) optimized with **GRPO, Grouped PPO, RL**, targeting privacy retention under adversarial prompt attacks while preserving utility. Achieved **11–14× faster** inference than the LLM baseline, **+48% utility and +10% privacy**, and near-perfect utility and privacy with deterministic outputs versus high-variance LLM decisions. Evaluated baseline LLM minimizers (Qwen2.5, Mistral, Llama) on 2K+ samples.

Poker AI Agent, UMass Amherst More Info Github Repo
Developed AI agents using Expectiminimax with statistical opponent modeling, Q-learning with state abstractions and eligibility traces, and a hybrid MCTS–Minimax framework with dynamic opponent profiling; reduced effective state space from trillions to 1K per decision, trained across 10K+ simulations, achieving 80% win rate.

Analysis of Player Tracking Data from Football Match Feed, RRIA Journal, SSN Github Repo DOI:10.33436/v33i2y202307 Thesis
Developed a cost-cutting player tracking and decision-making system using GAN/Siamese networks for perspective transformation to extract 2D coordinates of all 22 players per frame from 3D view. Built a decision-making model combining pitch control and expected threat (xT) to quantify optimal passing choices.

Classification of Hate Speech Using DistilBERT, FIRE-WN 2021 CEUR Proceedings
Built an NLP pipeline using DistilBERT, LSTM, and ensemble ML models to detect hate, offensive, and profane content; achieved 77.7% accuracy for binary classification and 65% for multi-class categorization on social media datasets, leveraging TF-IDF, profanity filters, and hierarchical modeling, ranked 24th globally.