# Vageesha Datta Ganapaneni

vageeshadattag@gmail.com | linkedin | Github | +1(469) 805-1906 | Dublin,California

## Education

- **Master of Science in Computer Science** — Aug 2022 - May 2025
  The University of Texas at Dallas — *CGPA: 3.6/4.0*
  *Relevant Coursework: Machine Learning, Artificial Intelligence, Computer Vision, Natural Language Processing, Statistics in AI and ML*
- **Bachelor of Technology in Computer Science and Engineering** — Aug 2018 - June 2022
  SRM University, AP — *CGPA: 9.06/10.00*

## Experience

- **Allen Institute for AI (AI2)** — Dallas
  AI Researcher — *Jan 2025 - May 2025*
  - Implemented Optimization by Prompting (OPRO) for test-time LLM steering, boosting hypothesis generation accuracy by 12% and enabling more reliable few-shot inference in active reasoning benchmarks.
  - Built a belief-tracking and uncertainty quantification framework using AutoGen, binomial modeling, entropy, and KL divergence to analyze LLM confidence updates and evaluate hypothesis consistency.
  - Developed an interactive D3.js experiment tree to visualize belief evolution in LLMs, accelerating hypothesis debugging and collaborative insight generation, and iterative model tuning.
- **Rocktop Technologies** — Dallas
  Software Engineer Intern — *Sep 2023 - July 2024*
  - Developed and optimized microservices using Python and PyTorch to implement generative AI agents and fine-tuned LLMs, automating fixed-income data ingestion and risk analysis, reducing manual processing time by 30%.
  - Designed and deployed RESTful APIs with Flask to serve AI-generated insights, enabling seamless integration with Rocktop's fixed-income platform and accelerating portfolio risk evaluation workflows for faster decision-making.
  - Integrated NLP-based query modules to build conversational interfaces allowing non-technical users to interact with complex financial data via natural language, increasing platform accessibility and user efficiency.
  - Conducted rigorous unit and integration testing, debugging, and deployment of AI-driven anomaly detection pipelines, enhancing yield leakage identification accuracy and improving overall asset liquidity management processes.
- **Fiserv** — India
  Software Engineer — *Dec 2020 - May 2022*
  - Orchestrated the deployment of Kafka-based microservices using Jenkins CI/CD and CockroachDB, reducing latency for 166,000+ users and enhancing system scalability across distributed environments.
  - Implemented Prometheus and Grafana for real-time observability of microservices, enabling faster anomaly detection, improving uptime, and minimizing critical production incidents.
  - Developed a React.js and Spring Boot security dashboard with real-time production data and five integrated charts, saving 10+ hours weekly through streamlined monitoring and faster issue resolution.
  - Refactored legacy UI using Material UI and modular React components, containerized the frontend with Docker, and deployed to AWS EC2, improving performance and reducing support tickets by 25%.

## Projects

- **InsightBridge:An LLM-powered document analysis tool**:
  - Designed and implemented a recursive text chunking pipeline with LangChain's RecursiveCharacterTextSplitter, enabling efficient vectorization and semantic retrieval from long-form documents.
  - Integrated FAISS-based vector store for low-latency dense retrieval and constructed a Retrieval-Augmented Generation (RAG) chain with ChatOpenAI to produce grounded, context-aware responses.
- **MediQuery:An Instruction-Tuned Healthcare Chatbot**:
  - Developed a modular React frontend with real-time chat interface, integrating complex state management and optimized GPU batch inference for sub-second response times.
  - Implemented backend Flask APIs using FAISS and SentenceTransformers for dense retrieval, enhancing contextual accuracy by 19% via RAG-based instruction tuning.
- **StreamStyle:Scalable Real-time Video Style Transfer Platform**:
  - Engineered a React frontend featuring efficient video upload, processing status updates, and live previews, using CSS modules for responsive design and smooth UX.
  - Created scalable Flask REST APIs deployed on AWS EC2 with Docker, leveraging AdaIN and mixed-precision training to reduce inference flickering by 50% and computation overhead by 15%.

## Technical Skills

- **Languages:** Python, JavaScript, TypeScript, C++, SQL, Bash, HTML, CSS
- **Frameworks:** PyTorch, TensorFlow, Flask, FastAPI, React, Angular, Spring Boot, Material UI, AutoGen, D3.js
- **Tools:** Docker, AWS, Terraform, Jenkins, Kafka, CockroachDB, Git, Prometheus, Grafana, Unix/Linux, REST APIs