Rohan Siva

rohansiva@utexas.edu | LinkedIn | GitHub | Google Scholar | Website

Education

University of Texas at Austin — Austin, TX

B.S. in Electrical and Computer Engineering Honors (ECE Honors); ***GPA: 3.97/4.00*** — *Expected May 2027*

Minor in Statistics and Data Science

Relevant Coursework: Data Structures & Algorithms, Object Oriented Design, Data Science Principles, Data Science Lab, Probability & Statistics, Intro to Computing, Embedded Systems, Circuit Theory, Matrices, Linear Systems, Linear Algebra, Differential Equations, Discrete Math, Calculus I–III

Skills

Languages: Python, Java, C, C++, JavaScript, SQL, TypeScript, Swift, HTML, CSS, R, Assembly,

Technologies: Git, ROS, React.js, Node.js, PostgreSQL, SpringBoot, Flask, FastAPI, CI/CD, AWS, Elasticsearch, Docker, Kubernetes, Postman, Selenium, GCP, Kafka, Carla, ROS, SLAM, nuScenes, Figma, Gradio

AI/ML: PyTorch, TensorFlow, Jax, Keras, Sklearn, Transformers, LangChain, LlamaIndex, Unsloth, DeepSpeed, VLA, SFT(LoRA/QLoRA), RL(GRPO/PPO), WandB, Diffusion, Spark, MapReduce, RL Gym, OpenCV

Publications

Bhatt, N. P., Yang, Y., **Siva, R.**, Milan, D., Topcu, U., & Wang, Z. "Know Where You're Uncertain When Planning with Multimodal Foundation Models: A Formal Framework," In Proceedings, **Conference on Machine Learning and Systems (MLSys)**, 2025. Oral Presentation. Available: arXiv:2411.01639

Bhatt, N. P., Yang, Y., **Siva, R.**, Samineni, P., Milan, D., Wang, Z., & Topcu, U. "VLN-Zero: Rapid Exploration and Cache-Enabled Neurosymbolic Vision-Language Planning for Zero-Shot Transfer in Robot Navigation," Under Review, **IEEE International Conference on Robotics and Automation (ICRA 2026)**, 2026. Available: arXiv:2509.18592

Bhatt, N. P., Li, P., Gupta, K., **Siva, R.**, Milan, D., Hogue, A. T., Chinchali, S. P., Fridovich-Keil, D., Wang, Z., & Topcu, U. "UNCAP: Uncertainty-Guided Planning Using Natural Language Communication for Cooperative Autonomous Vehicles," Under Review, **International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)**, 2026. Available: arXiv:2510.12992

Yang, Y., Bhatt, N. P., Samineni, P., **Siva, R.**, Wang, Z., & Topcu, U. "RepV: Safety-Separable Latent Spaces for Scalable Neurosymbolic Plan Verification," Under Review, **Conference on Machine Learning and Systems (MLSys 2026)**. Available: arXiv:2510.26935

Experience

VITA Lab @ UT Austin — Austin, TX

AI Researcher — *Aug 2024 – Present*

- Research in Embodied AI, Formal Methods & Planning/Perception under Prof. Atlas Wang and Prof. Ufuk Topcu
- 1. Fine-tuned LLaVA multimodal models on low-uncertainty data using LoRA quantization with DeepSpeed for distributed training, resulting in improved specification compliance by **5%** & reduced decision variability by 40**%**
- 2. Designed a two-phase vision–language navigation framework combining SLAM based scene-graph generation & trajectory caching, achieving 2× success rate over SOTA zero-shot models & Sim2Real transfer via ROS
- 3. Developed an uncertainty-based vision–language framework enabling cooperative autonomous vehicles to exchange structured natural language messages, reducing bandwidth by 65% & decision uncertainty by 30%

PKU Lab @ Peking University — Beijing, China (Remote)

AI Researcher — *May 2025 – Present*

- Leading research under Prof. Hao Tang on integrating block-by-block diffusion into Chain-of-Thought reasoning for uncertainty-aware multimodal planning with real-time posterior variance feedback
- Designed a parallelizable diffusion-based reasoning architecture replacing autoregressive CoT, reducing latency by up to 34% while improving reliability for VLM-driven autonomous driving

Statistical Learning & AI Group Lab @ UT Austin — Austin, Texas (Remote)
AI Researcher — *November 2025 – Present*

- Exploring custom GRPO implementations for NanoChat w/ pass@1024 evaluation under Prof. Qiang Liu

Cisco — San Jose, CA
Machine Learning Intern — *May 2025 – Nov 2025*

- Developed kRAIG, an AI agent converting natural language into executable Kubeflow Pipelines for ETL workflows
- Deployed a RAG-based pipeline generator using Elasticsearch and safety guardrails for database operations
- Integrated MCP Server w/ PostgreSQL & AWS tooling, automating direct ETL into enterprise data storages
- Implemented CI/CD pipeline for Docker container deployment & automated unit tests w/ 90%+ code coverage

Mercor AI — Remote
Machine Learning Engineer (Contract) — *Aug 2025 – Nov 2025*

- Developed plan–code pairs for ML engineering tasks derived from Kaggle competitions with verifiable outputs
- Automated data generation and preprocessing pipelines, creating structured supervision signals for model training
- Improved data generation & training speed by 50% through scalable dataset construction and validation tooling

IX (Information eXperience) Lab @ UT Austin — Austin, TX
AI Researcher — *Aug 2024 – May 2025*

- Led LLM integration under Prof. Jacek Gwizdka into interactive information retrieval interfaces using Python, Node.js, and the OpenAI API, enabling real-time query refinement through dynamic user feedback loops
- Prototyped & tested multi-pane user interaction flows in Figma to support sequential querying w/ user studies
- Evaluated AI-driven search efficiency & models through user-centered interface testing and optimization

Canyon Technologies LLC — Austin, TX
Software Engineering Backend Intern — *May 2024 – Oct 2024*

- Built a cloud-based nzESL display management dashboard w/ SpringBoot, Docker & Postman for 1000+ users
- Integrated display tags and BLOZI base stations to enable real-time updates, QR-codes, and label management
- Designed and deployed a secure customer mailing & password system with Java and React.js using OWASP

Keitt Lab @ UT Austin — Austin, TX
Software Engineering Researcher — *Nov 2023 – Jul 2024*

- Built low-power sensor mesh using XBee radios & time-synced Raspberry Pis for remote wildlife monitoring
- Developed PyTorch acoustic detection system to identify bird species & location from chirp data w/ live analysis
- Implemented distributed computing architecture for remote data collection, reducing latency by 46%

Projects

MoodScribe – Social Sidekick | *GitHub*

- Fine-tuned Llama 3.1 model for sentiment analysis with custom synthetically generated dataset using LoRA adapters, optimizing training with Unsloth for memory reduction and FlashAttention for a 44% speedup

Minerva – Study App | *GitHub*

- Built a full-stack AI study app using Flask + React, integrating Google Classroom API and an OpenAI & LlamaIndex RAG pipeline for personalized study features, including flashcards, quizzes and chatbot

Medicina.ai – Health Risk Predictor | *GitHub*

- Designed full-stack health-risk prediction app w/ Python, HTML/CSS/JS & Flask using Random Forest & Logistic Regression to estimate diabetes & heart risk from user data, including data ingestion, preprocessing & inference UI

Poker-RL – Deep RL Blackjack/Poker Agent | *GitHub*

- Developed a DQN-based RL agent using PyTorch and RLCard, w/ experience replay, ϵ-greedy exploration, target networks, and reward-shaping to learn optimal strategies in Blackjack/poker environments, with training pipeline