Mustafa Sameen

Gainesville, FL | mustafasameen@ufl.edu | (719)-201-2330 | LinkedIn: mustafasameen |
GitHub: mustafasameen | Portfolio: https://mustafasameen.com/

EDUCATION

UNIVERSITY OF FLORIDA — **Gainesville, FL**
PhD Student in Civil and Coastal Engineering — *August 2025 - Present*
- **Research Focus:** *Agentic AI Systems, Multi-Agent Simulation, Large Language Models (LLMs) in Transportation, Spatial Bias Mitigation.*
- **GPA:** *4.00 /4.00, Graduate Research Assistant.*
- **Relevant Coursework**: *EEL5840 Fundamentals of Machine Learning, EEE6512 Image Processing and Computer Vision, EGN5215 Machine Learning Applications in Civil Engineering.*

COLORADO COLLEGE — **Colorado Springs, CO**
Bachelors of Arts in Computer Science, Mathematics — *Aug 2021 - May 2025*
- **GPA:** *3.99 /4.00*
- **Awards:** *Magna Cum Laude, Euclid's Scholar, Phi Beta Kappa, Rawles Prize*
- **Study Abroad:** *Aquincum Institute of Technology, Budapest (Software Engineering, GPA 5.00/5.00).*

RESEARCH EXPERIENCE

University of Florida — **Gainesville, FL**
Graduate Research Assistant — *July 2025 – Present*
- **Advisor**: Dr Xilei Zhao
- Architecting a LangGraph-based Multi-Agent System (Planner, Coder, Critic) to automate the analysis of large-scale mobile location data.
- Integrating cAST (Chunking via Abstract Syntax Trees) to enable agents to retrieve and apply complex codebase documentation for transportation agencies.
- Developed the Synthesizing Attitudes, Predicting Actions (SAPA) framework, utilizing LLMs to synthesize latent psychometric variables (Theory of Planned Behavior) from raw survey data. Achieved a 75.9% improvement in PR-AUC for ridesourcing mode choice prediction.
- Co-developed the Grid Based Home detection via Stay Time (GHOST) Python package for bias-mitigated home detection. Implemented a Stay-Time hierarchical metric and adaptive weekend fallback, reducing spatial error to 1.84 meters on the Boston Walks dataset.
- Leading the development of a household travel simulation where LLM agents, initialized with specific personas, negotiate shared resources (vehicles) via a structured conversation protocol.

Massachusetts Institute of Technology JTL-Transit Lab — **Cambridge, MA**
Summer Research Intern — *June 2025 – July 2025*
- **Advisors**: Dr Haris Koutsopoulos, Dr Muhammad Usama
- Investigated multi-agent LLM architectures for the Chicago Transit Authority (CTA).
- Designed a Spatial-RAG pipeline blending dense retrieval (policy documents) with sparse spatial lookup (GTFS/GIS geometry) to enable agents to reason about delay propagation and rerouting and simulate station-level dispatchers.

Colorado College Math and Computer Science Department — **Colorado Springs, CO**
Undergraduate Student Researcher (Thesis) — *Aug 2024 – May 2025*
- **Advisor**: Professor Robert Rennie
- Designed a feedback loop combining Retrieval-Augmented Thinking (RAT) with agentic reasoning to translate natural language mathematics into Lean 4 code.
- Achieved 92% syntactic correctness and 83% semantic accuracy on the Lean-Workbook dataset using local LLMs.

PUBLICATIONS & PREPRINTS

- Sameen, M., Zhang, X., Zhao, X. (2025). "*Synthesizing Attitudes, Predicting Actions (SAPA): Behavioral Theory-Guided LLMs for Ridesourcing Mode Choice Modeling*." arXiv:2509.18181.
- Recalde, A., Sameen, M., Zhang, X., Zhao, X. (2026). "*Design and Validation of a Grid-Based Home Detection via Stay-Time (GHOST) Software for Mobile Location Data.*" Proceedings of Transportation Research Board 105th Annual Meeting. (Accepted)

- Recalde, A., Sameen, M., Zhang, X., Zhao, X. (2025). *GHOST: Grid-based Home detection via Stay-Time [Computer software]*. University of Florida. https://github.com/SERMOS-LAB/Grid-Based-Home-Detection
- Anonymous Authors (incl. Sameen, M.). (Under Review). "*PEMANT: Persona-Enriched Multi-Agent Negotiation for Travel.*" Submitted to ICML 2026.
- Sameen, M. (2025). "*FORMAL: Democratizing Lean 4 Formalization Through Retrieval-Augmented Thinking.*" Undergraduate Thesis.

PROFESSIONAL EXPERIENCE

SocialTechLab Inc.   **New York CIty, NY**
Software Engineer Intern   *June 2024 – Aug 2024*
- Engineered 5+ features for a mental wellness app using SwiftUI (Real-time notifications, QR friend add), boosting engagement by 30%.
- Spearheaded AWS AI image processing pipelines, improving system stability by 25%. Optimized Go-based backend connectivity.

State of the Rockies   **Colorado Springs, CO**
Research Fellow   *May 2023 – May 2025*
- Developed NLP pipelines to analyze Indigenous oral histories and created a GIS-integrated storytelling platform linking linguistic themes with environmental data (light pollution) for the Bears Ears National Monument.

Quad Innovation Alliance   **Colorado Springs, CO**
Data Analyst   *Aug 2022 – Dec 2022*
- Led a team analyzing K-12 education data via Python/SQL. Presented strategies to stakeholders, driving a 30% increase in funding interest.

SKILLS

- Languages: Python, Lean 4, Agda, Haskell, C/C++, Java, Kotlin, TypeScript, SQL, Go, R.
- AI & ML Frameworks: PyTorch, TensorFlow, Scikit-learn, LangGraph, Hugging Face Transformers, Llama-3, DeepSeek, OpenAI API.
- Geospatial & Simulation: GeoPandas, ArcGIS, QGIS, MATSim, SUMO.
- Systems & Tools: Docker, AWS (EC2/S3), Git, LaTeX, CI/CD Pipelines.
- Core Competencies: Agentic AI, Neuro-Symbolic Reasoning, Retrieval-Augmented Generation (RAG), Formal Verification, Spatial Data Mining, Multi-Agent Systems.