Mustafa Sameen

Colorado Springs, CO | s_sameen@coloradocollege.edu | (719)-201-2330 | LinkedIn: mustafasameen |
GitHub: mustafasameen | Portfolio: https://mustafasameen.github.io/

EDUCATION

COLORADO COLLEGE — **Colorado Springs, CO**
Bachelors of Arts in Computer Science, Mathematics — *Aug 2021 - May 2025*
- *Dean's List, Euclid's Scholar, Phi Beta Kappa, Thomas Post Rawles Mathematical Prize* — **GPA:** 4.00 /4.00

TUCK SCHOOL OF BUSINESS AT DARTMOUTH — **Hanover, NH**
Tuck Bridge Business Program — *Winter 2024*
- Selective 4-week conducting a Costco valuation analysis with DCF modeling and presenting findings to executives.

AQUINCUM INSTITUTE OF TECHNOLOGY — **Budapest, Hungary**
Semester Abroad Program in Computer Science — *Aug 2023 - Dec 2023*
- Focused on software engineering and mobile app development in an international setting. — **GPA:** 5.00 /5.00

RESEARCH EXPERIENCE

Colorado College Math and Computer Science Department — **Colorado Springs, CO**
Undergraduate Student Researcher — *Aug 2024 – Present*
- Designed the FORMAL system (Feedback-Oriented Retrieval Method for Automated Lean translation) combining Retrieval-Augmented Thinking (RAT) with an agentic feedback loop to translate natural language mathematics into Lean 4 code.
- Developed specialized vector stores for theorem statements and tactic examples, enabling dynamic retrieval during different proof stages to enhance formal verification.
- Created a robust agentic feedback loop that executes generated code in Lean 4, analyzes errors via a Reasoner LLM (DeepSeek-R1 8B), and autonomously refines the formalization process.
- Achieved 92% syntactic correctness and 83% semantic accuracy on the Lean-Workbook dataset, compared to only 58% and 42% for pure LLM approaches, with most problems solved in 2-3 iterations.
- Implemented a low-resource efficient system using local LLMs (Qwen2.5-14B and DeepSeek-R1-8B) that democratizes access to formal verification without requiring massive computational resources.

Colorado College Math and Computer Science Department — **Colorado Springs, CO**
Summer Research Assistant — *Summer 2024*
- **Advisor**: Professor Robert Rennie
- Investigated *LLM-assisted formal verification* using **Agda** and **UniMath**, with a focus on automating and verifying proofs of Sylow Theorems in **Homotopy Type Theory (HoTT)**.
- Contributed 10+ formal proof entries to open-source libraries (Agda/UniMath) addressing nuanced type dependencies and complex inferences.
- Developed iterative prompting strategies to reduce model hallucinations, resulting in a 15% decrease in formal proof errors during generative tests.

PROFESSIONAL EXPERIENCE

SocialTechLab Inc. — **New York CIty, NY**
Software Engineer Intern — *June 2024 – Aug 2024*
- Engineered and launched 5+ innovative features for the mental wellness app using **SwiftUI**, including real-time push notifications, and QR code-based friend addition, significantly boosting user engagement by 30%.
- Spearheaded AI image processing setup on **AWS**, resolving critical issues and increasing application stability by 25%, contributing to the app's readiness for a wider release, and optimized front-end and back-end connectivity with **Go**.

State of the Rockies — **Colorado Springs, CO**
Research Fellow — *May 2023 – Present*
- Developed NLP pipelines to transcribe and analyze Indigenous oral histories from the Bears Ears region.
- Linked linguistic themes from transcriptions with environmental data (e.g., light pollution levels).
- Created a GIS-integrated storytelling platform to contextualize Indigenous narratives within conservation efforts
- Conducted comprehensive analysis of dark sky preservation's cultural, economic, and ecological significance, informing policy recommendations for the Bears Ears National Monument Resource Management Plan.

Quad Innovation Alliance **Colorado Springs, CO**
Data Analyst *Aug 2022 – Dec 2022*

- Led a cross-functional team in analyzing K-12 education data for strategic decision-making in Colorado Springs school districts, identifying critical areas for resource allocation using Python, SQL, Excel.
- Presented findings to over 20 local stakeholders, including school administrators, nonprofit leaders, and donors, driving a 30% increase in funding interest for targeted educational initiatives and adoption of data-driven strategies.

PUBLICATIONS & PREPRINTS

- **Sameen, M.** "FORMAL: Democratizing Lean 4 Formalization Through Retrieval-Augmented Thinking and Agentic Feedback Loops." *Undergraduate Thesis*, *Colorado College*, 2025.
- **Sameen, M.** "Type-Theoretic Foundations for Enhancing RAG Efficiency and Correctness." *Manuscript Under Review, Submitted to TYPES* 2025.
- **Sameen, M**. "Preserving the Starry Skies: Light Pollution's Impact at Bears Ears National Monument." *State of the Rockies Project Research Report, Colorado College*, 2024.

PROJECTS

AFK: SCAC Esports Statistics Tracker
React, Flask, AWS, OCR, HTML, CSS, JS

- Automated **OCR-based** data extraction from esports match screenshots (Apex Legends, Valorant, Rocket League) for the **Southern Collegiate Athletic Conference (SCAC)**.Deployed a **predictive model** for match outcomes by synthesizing historical match data; achieved a 75% accuracy rate in pilot evaluations. Centralized **web dashboard** for integrated analytics, scheduling, and real-time Twitch streaming, facilitating streamlined data-driven coaching strategies.

Formal Proof Generation Tools
Lean, Python, Agda

- Built custom scripts to parse and label theorem statements, bridging *informal text* with *formal proof scripts*. Collaborated on plugin enhancements for Agda to improve type-checking speed in multi-file proof developments.

Type-Theoretic RAG Enhancement
Type Theory, Dependent Types, Linear Types, Retrieval-Augmented Generation

- Designed a formal model using dependent, linear, and affine types to enhance RAG correctness by enforcing constraints on information usage.Created a type-driven framework where retrieved facts are treated as typed resources, making hallucinations analogous to type errors.Formalized a generation process as proof derivation where every output must be supportable by retrieved evidence, ensuring outputs are type-safe.

SKILLS

Python, Java, Kotlin, C/C++, JavaScript, TypeScript, Ruby, R, Haskell, Stata, SQL, MATLAB, AWS, Docker, NLP, Git, OCR, Data Science (pandas, scikit-learn, NLTK, OpenAI API), HTML/CSS, Swift, React, Next.JS, Flask

RELEVANT COURSEWORKS

Computer Science: Computer Science I & II, Computer Organization, Software Design, Data Structures and Algorithms, Programming Language Implementation, Python for Science, Natural Language Processing, Theory of Computation, Mobile App Development, Computer Graphics, Data Science.
Mathematics: Calculus 3, Linear Algebra, Number Theory, Ordinary Differential Equations, Graph Theory, Abstract Algebra I, Abstract Algebra II, Real Analysis I, Partial Differential Equations.

ACTIVITIES AND LEADERSHIP

Learning Assistant, Colorado College Math & CS Dept. *Aug 2022 – Present*
Supported 100+ students in Linear Algebra, CS2, Computer Organization, and ODEs, improving pass rates by 10% through peer-led sessions.
Residential Advisor, Colorado College *Aug 2022 – Present*
Managed a hall of ~20 freshmen, organizing monthly panels on minority identities, mental health, and substance abuse.
Chair, South Asian Student Association *Dec 2022 – Present*
Led 30+ members, organized major cultural events (Diwali, Holi, Eid), and collaborated with CU Denver & CU Boulder.