

# Mustafa Sameen

Colorado Springs, CO | s\_sameen@coloradocollege.edu | (719)-201-2330 | LinkedIn: [mustafasameen](#) |  
GitHub: [mustafasameen](#) | Portfolio: <https://mustafasameen.github.io/>

## EDUCATION

### COLORADO COLLEGE

Bachelors of Arts in Computer Science, Mathematics

- Dean's List, Euclid's Scholar, Phi Beta Kappa, Thomas Post Rawles Mathematical Prize

Colorado Springs, CO

Aug 2021 - May 2025

GPA: 4.00 /4.00

### TUCK SCHOOL OF BUSINESS AT DARTMOUTH

Tuck Bridge Business Program

Hanover, NH

Winter 2024

- Selective 4-week conducting a Costco valuation analysis with DCF modeling and presenting findings to executives.

### AQUINCUM INSTITUTE OF TECHNOLOGY

Semester Abroad Program in Computer Science

Budapest, Hungary

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

Undergraduate Student Researcher

Colorado Springs, CO

Aug 2024 – Present

- Designed a pipeline using **T5** (small variant) for auto-formalization, incorporating structure-aware representations to handle mathematical dependencies.
- Applied **distilled back-translation** and curated dataset augmentation (from MMA dataset) to boost model performance in resource-constrained scenarios.
- Deployed multi-task learning (formal proof generation + semantic annotation) to improve interpretability; evaluated using BLEU, ROUGE, F1 scores, and human-centric error analysis.
- Integrated the final outputs into **Lean**, achieving a 20% improvement in proof acceptance rate compared to baseline models.

### Colorado College Math and Computer Science Department

Summer Research Assistant

Colorado Springs, CO

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.

### Colorado College

*Independent Research: Cross-Lingual Mathematical Reasoning in Bengali*

Colorado Springs, CO

In Progress

- Developing a novel Bengali problem-solving benchmark by combining MGSM and **BenQA** datasets; bridging high-resource (Hindi, English) and low-resource (Bengali) language models.
- Implementing **retrieval-augmented generation** and **chain-of-thought prompting** to preserve semantic rigor while accommodating linguistic nuances.
- Submitting a manuscript based on these findings to **ICML 2025**, targeting improved cross-lingual, low-resource NLP methodologies.

## PROFESSIONAL EXPERIENCE

### SocialTechLab Inc.

Software Engineer Intern

New York City, NY

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

Research Fellow

Colorado Springs, CO

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.

## 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.** "Low-Resource Auto-Formalization in Lean: A Back-Translation Approach." *Undergraduate Thesis*, Colorado College, 2025.
- **Sameen, M.** "Cross-Lingual Reasoning for Mathematical Problem Solving in Bengali." *Manuscript Under Work*, 2025.

## 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.

### Minimal LLM Fine-Tuning for Mathematical Text

Python, PyTorch, TensorFlow, Hugging Face Transformers, TensorBoard, NLTK

- Prototyped a lightweight fine-tuning strategy using T5-based models, leveraging both paraphrasing and **structure-aware tokenization**. Achieved an 18% increase in exact match for mathematical statement formalization tasks compared to baseline generative LLMs.

## 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.