

# SAJAG SWAMI

[sajagswami@gmail.com](mailto:sajagswami@gmail.com) ◇ [LinkedIn](#) ◇ [GitHub](#) ◇ [Google Scholar](#)

---

## EDUCATION

Indian Institute of Technology Roorkee 2019 – 2023  
B.Tech. in Production and Industrial Engineering Roorkee, India  
CGPA: 8.6/10.0, First Division with Distinction  
*Selected coursework:* MAN-001 (Matrix Algebra & Multivariable Calculus), Probability and Statistics, Data Mining for Business Intelligence, Operations Research, Programming and Data Structures, Product and Process Optimization

CFA Institute Mar 2026  
CFA Program Level I

---

## INDUSTRY EXPERIENCE

Associate, Mizuho Financial Group (Risk Management Department), Tokyo Jun 2026 – Present

- ▷ Conducting model-risk assessments for enterprise AI and machine learning systems as a member of the Model Governance Team
- ▷ Contributing to AI governance frameworks, including assessment standards and internal policies for AI/ML deployments

Data Scientist, Accenture (Strategy & Consulting), Tokyo Nov 2023 – May 2026

- ▷ Built and deployed production RAG pipelines and agentic workflows for enterprise clients, reducing selected data-analysis tasks from 30–40 minutes to 3–5 minutes (~90% reduction)
- ▷ Owned requirements, architecture, and end-to-end delivery across technical implementation and client-facing work

---

## RESEARCH EXPERIENCE

Independent Researcher, SpendSight / ExpenseSense (personal project) Jan 2026 – Sep 2026

- ▷ Designed a 115-query financial test suite spanning 5 analytical tools and benchmarked on-device tool calling across 7 quantized small language models (0.8B–3B parameters), comparing single vs. dual-agent (router–specialist) architectures.
- ▷ Showed that dual-agent decomposition shifted failures from silent parameter corruption (PARAM\_EXTRA, 22.6% → 1.9% for Qwen3.5-2B) to transparent omissions (PARAM\_MISSING), improving strict executable Correct Ratio to 77.6% (+17.4 pp).
- ▷ Demonstrated private on-device viability: Qwen3.5-0.8B achieved 89.1% Task Accuracy and 61.7% Correct Ratio in 3.0 s using 1.2 GB RAM, with inference performed entirely locally.

Contributor, Cohere Labs Open Science Community Mar 2026 – Sep 2026

- ▷ *Cultural Norm Robustness:* Evaluated 12 LLMs from 4 provider families on the NORMAD dataset across 3 countries and 5 language conditions, generating 55,000 LLM outputs to measure sensitivity to context granularity, translation, and adversarial perturbations.
- ▷ *MultiCulturalRiddle:* Contributed 24 English riddles and curated human annotations for 231 model outputs.

Research Intern, EADAL, IIT Roorkee Aug 2022 – Apr 2023  
Advisors: Prof. Dhish Kumar Saxena, Dr. Sukrit Mittal

- ▷ Investigated the sensitivity of Innovized Progress operators (IP2 for convergence, IP3 for diversity) in NSGA-III to the underlying machine learning algorithm, benchmarking 8 ML models across 39 multi- and many-objective problems.

- ▷ Evaluated 21 independent runs per method using Wilcoxon rank-sum tests on Hypervolume and runtime, finding that kNN provided relatively strong Hypervolume–runtime trade-off among the evaluated methods.

MITACS Globalink Research Intern, McGill University

May 2022 – Aug 2022

Advisors: Prof. Xiangfei Meng, Dr. Yingying Su

- ▷ Investigated the association of lifetime stress with alcohol dependence across 1,400 adults (ZEPSOM dataset) using Latent Profile Analysis (LPA).
- ▷ Clustered subjects into 3 distinct cohorts; the high-stress cohort had more than twice the adjusted odds of alcohol dependence relative to the low-stress cohort.

Research Intern, Virginia Commonwealth University

Feb 2021 – Oct 2021

Advisor: Prof. Ekaterina Smirnova

- ▷ Developed survey-weighted multivariable logistic regression models with forward selection using 10-fold cross-validated AUC on 2,500+ subjects (NHANES dataset) to study associations between physical activity and NAFLD.
- ▷ Found that moderate-to-vigorous physical activity (MVPA) was associated with 43% lower odds of NAFLD.

---

## OPEN SOURCE EXPERIENCE

Student Developer, Python Software Foundation (GSoC)

May 2021 – Aug 2021

Built open-source 3D molecular visualization tools for FURY using VTK-based rendering.

---

## PUBLICATIONS

*SpendSight: Privacy-Preserving Personal Finance Q&A with Sub-2B LLMs*

Sajag Swami, Yash Pandey

FinNLP Workshop at EMNLP 2026; extended from *ExpenseSense: Privacy-Preserving On-Device Personal Finance Tool Calling with Sub-2B LLMs*, HAIPS Workshop at COLM 2026

*Robustness of Cultural Norm Reasoning Under Language and Context Perturbations*

Ankita Maity, Sajag Swami, Van Ngo, Akhil Arora, Nikita Moghe

Culture × AI Workshop at ICML 2026; StereACuLT Workshop at ACL 2026

*MultiCulturalRiddle: A Multicultural Benchmark of Riddles*

Tianyi Hu, Henry Gagnier, ..., Sajag Swami, et al.

Multilingual Representation Learning (MRL) Workshop at EMNLP 2026

*Socioeconomic status and health disparities drive differences in accelerometer-derived physical activity in fatty liver disease and significant fibrosis*

Lucia Tabacu, Sajag Swami, Mark Ledbetter, Mohamad S. Siddiqui, Ekaterina Smirnova

PLOS ONE, 2024

*Investigating Innovized Progress Operators with Different Machine Learning Methods*

Drishti Bhasin, Sajag Swami, Sarthak Sharma, Saumya Sah, Dhish Kumar Saxena, Kalyanmoy Deb

Evolutionary Multi-Criterion Optimization (EMO), 2023

---

## HONORS AND AWARDS

- ▷ JEE Advanced Rank 4,627 (Top 3% of 160,000+ candidates) 2019
  - ▷ JEE Main Rank 6,246 (Top 0.6% of 1,150,000+ candidates) 2019
  - ▷ NTSE Scholar, National scholarship by NCERT, India 2017
-

## SERVICE, TEACHING & MENTORING

- |                                                                |             |
|----------------------------------------------------------------|-------------|
| ▷ Reviewer, FinNLP 2026 (co-located with EMNLP 2026)           | 2026        |
| ▷ Student Mentor, Student Mentorship Program, IIT Roorkee      | 2021 – 2022 |
| ▷ Undergraduate Teaching Assistant, Mathematics-I, IIT Roorkee | 2020 – 2021 |
- 

## SKILLS

- ▷ Programming: Python, R, C/C++, SQL
  - ▷ Engineering: Git, Docker, Azure DevOps, FastAPI
  - ▷ Data: PostgreSQL, Milvus, Azure Blob Storage
  - ▷ Languages: English (Native), Hindi (Native), Japanese (JLPT N3, CEFR B1)
- 

## EXTRACURRICULARS

- ▷ Athletics: Avid runner, 4,300+ km and counting (2024 – present); Participant, CBSE National Football Championship (2018); CBSE Intrastate Football Winners (2018)
- ▷ Volunteering: Tutored underprivileged students via National Service Scheme (2019 – 2020); Raised INR 30,000+ for annual treatment of a DMD patient (2018)