Shuvam Banerji Seal

Aspiring Computational Chemist | AI/ML Researcher | DeepTech & AI-Fintech Co-Founder

BS-MS Student (Chemistry Major, Computer Science Minor)

Indian Institute of Science Education and Research - Kolkata

sbs22ms076@iiserkol.ac.in | github.com/Shuvam-Banerji-Seal |

linkedin.com/in/mastersbs | shuvam-banerji-seal.github.io | orcid.org/0009-0000-0714-569X

Research Experience

A Hybrid RAG Architecture for Verifiable Answer Synthesis with Automated Source Citation *July 2025*

Under Dr. Dwaipayan Roy, Dept. of Computation and Data-Science, IISER-Kolkata

- Developed a comprehensive RAG pipeline for IISER-K intranet documents, including automated data acquisition via Selenium, OCR processing with PyMuPDF and EasyOCR, and text extraction Implemented context-aware text chunking with metadata enrichment and high-dimensional embeddings using Qwen3-4B model, stored in FAISS vector database Designed hybrid search strategy combining dense vector retrieval with BM25 keyword search, fused using Reciprocal Rank Fusion for optimal relevance Integrated query refinement with hypothetical answers and verifiable inline citations in a Streamlit application for transparent, grounded responses

Developing an Advanced Retrieval Model for TREC-Tip of Tongue Queries *2024 (TREC 2024 Proceedings)*

Guided by Dr. Dwaipayan Roy, Dept. of Computation and Data-Science, IISER-Kolkata

- Engineered multi-layer BM-25 filtering system (in Lucene, Java) with dynamic search-domain contraction based on extracted keywords and implemented transformer-based query expansion and semantic matching using Local LLMs through multi-shot and chain-of-thought prompting achieving results on par with specifically-modified DPR models

Industrial Experience

Research Intern, HistoXai (Astroloop Technologies Pvt. Ltd.) *May 2025 - July 2025*

Bangalore

- Conducted an extensive comparative analysis of digital histopathology slide quality assessment tools, reviewing 30+ open-source frameworks (e.g., HistoQC, PathProfiler, GrandQC, HistoROI, FASTPathology) to evaluate architectural designs, datasets, performance metrics, and limitations for future model integration
- **Technologies:** Python, OpenCV, PyTorch, Scikit-learn, Pandas, Literature Survey Tools

Publications

AgriIR: A Scalable Framework for Domain-Specific Knowledge Retrieval *Jan 2026 (Accepted at ECIR 2026)*

***Shuvam Banerji Seal**, Aheli Poddar, Alok Mishra, and Dr. Dwaipayan Roy*

- Developed a scalable information retrieval framework specifically designed for agricultural domain knowledge accessibility in developing regions, implementing advanced retrieval algorithms optimized for domain-specific terminology and contextual understanding to advance retrieval systems for social good.

Hierarchical Opinion Classification using Large Language Models *2025 (Accepted at FIRE 2025)*

***Shuvam Banerji Seal**, Alok Mishra, and Utkarsha Ghosh*

- Developed a novel parameter-efficient fine-tuning approach for the Gemma LLM using a custom two-layer classification head and class-weighted cross-entropy loss to address data imbalance. Reformulated three-level hierarchical opinion labels into an 8-class flat scheme, exploring dual methodologies like selective fine-tuning and instruction-tuning to advance LLM adaptation for hierarchical classification under computational constraints.

IISERK@ ToT_2024: Query Reformulation and Layered Retrieval for Tip-of-Tongue Items *2024 (TREC 2024 Proceedings)*

***Shuvam Banerji Seal** & **Subinay Adhikary**, Soumyadeep Sar, and Dr. Dwaipayan Roy*

- Explored various approaches for known-item retrieval in the TREC 2024 Tip-of-the-Tongue track, focusing on retrieving previously encountered items when searchers struggle to recall exact identifiers
- Implemented four-step query reformulation technique combined with two-layer retrieval using BM25 with optimized parameters and Large Language Models for query enhancement Achieved best performance with two-layer retrieval approach, obtaining Recall@1000 of 0.8067 and enhanced NDCG metrics through systematic query reformulation

Computational Modeling of [VO(SALIEP)(DTP)] as Water Reducing Catalyst *2024-25 (to be submitted)*

***Shuvam Banerji Seal**, Guided by Dr. Soumyajit Roy*

- Implemented DFT methods (B3LYP) for MO energy calculations and electronic structure analysis using Gaussian and conducted mechanistic studies using transition state theory and reaction pathway analysis catalyst performance prediction for water reduction

Entrepreneurial Ventures

Co-Founder & Lead Developer, Synapse *2025 – Present*

Incubated at RISE Foundation, IISER Kolkata

- Co-founded an AI-driven fintech startup funded by **MeitY Startup Hub (MSH)** through the **GENESIS** program.
- Developing a user-friendly platform with simple UI/UX to democratize access to essential ML and AI tools for automated smart trading alerts across stocks, crypto, and mutual funds.

- **Status:** Funded; DPIIT registration in progress (Q1 2026); Product shipping by 2027.

Co-Founder & CTO, UnderWater AI *2025 – Present*
Incubated at RISE Foundation, IISER Kolkata
- Co-founded a deeptech startup funded by **MeitY Startup Hub (MSH)** through the **GENESIS** program.
- Developing software solutions to enhance underwater image quality using deep neural networks and early fusion learning for marine species identification, serving researchers and industrial marine applications.
- **Status:** Funded; DPIIT registration in progress (Q1 2026); Product shipping by 2027.

Research Libraries and Frameworks Developed

Fernholz Stochastic Portfolio Theory Python Library *Ongoing*
Quantitative Finance Library
- Developing a Python library for Stochastic Portfolio Theory, implementing diversity-weighted portfolios and relative arbitrage strategies using stochastic calculus and continuous semimartingales.

LAMMPS Data Web Viewer *2025*
Active Web Service
- Developed a web-based 3D visualization tool for LAMMPS molecular dynamics data files using React, Three.js, and Flask, enabling interactive molecular dynamics analysis and real-time atom/bond rendering.

Tutorials Developed

Complete C Programming Course *2025*
Open Source Curriculum
- Created a comprehensive 20-module C programming curriculum covering fundamentals to advanced topics including Network Programming, Machine Learning, and GUI development with GTK4.

Python Course for Beginners *2025*
Open Source Curriculum
- Designed a progressive Python learning resource spanning core concepts to local LLM deployment, featuring interactive notebooks, real-world project implementations, and database integration.

Projects

TCP Activity Monitor for Arch Linux *2025*
System Utility
- Developed a system-wide TCP monitoring utility for Arch Linux integrating systemd services to log kernel statistics, socket states, and per-process connections, enhancing network diagnostic capabilities.

WeLearn Bot in C *2025*
Utility Application
- Engineered a multi-threaded educational resource automation tool in C featuring both CLI and GTK4 GUI interfaces with secure session management and encrypted credential storage for efficient bulk downloads.

IndicAgri: AI Platform for Indian Agriculture *Aug 2025*
Capital One Launchpad Hackathon 2025
- Architected a multi-modal RAG chatbot for Indian agriculture supporting 20+ languages using Gemma/DeepSeek LLMs and Indic-Conformer, backed by an autonomous 15k+ document curation pipeline.

Automated Event Coupon Management System *2025*
Event Automation System
- Built a secure Flask-based event management system featuring Google OAuth 2.0 authentication, automated bulk emailing, and real-time QR code verification with mobile scanner integration.

Legal Document Retrieval RAG Application *Feb 2025*
Legal Tech RAG System
- Developed a legal document retrieval RAG stack leveraging LLaMA-3.2, Nomic embeddings, and ChromaDB, with a Streamlit interface and Mistral-OCR for processing complex legal texts.

Full-Stack E-Commerce GUI in C *2024*
Under Dr. Kripabandhu Ghosh, IISER-Kolkata
- Implemented a full-stack E-Commerce GUI in C using GTK4 and SQLite3, incorporating a static context chatbot with optimized BM25 retrieval for intelligent customer interaction.

Wi-Fi Channel Optimizer *2024*
Network Optimization Tool
- Engineered a Bash-based Wi-Fi optimization script to automate network scanning, speed benchmarking, and channel selection, significantly enhancing wireless network performance on Linux systems.

SMC Canteen System Modernization *2024*
IISER-Kolkata
- Modernized the IISER-Kolkata Canteen System by porting legacy code to a modular Django-based architecture, improving maintainability and scalability for 2000+ users.

Coordination Chemistry Twists Simulator *Nov 2024*
Computational Chemistry Simulation
- Developed a Python molecular dynamics simulator for visualizing Ray-Dutt and Bailar twist mechanisms in coordination complexes, providing interactive 3D stereochemical analysis.

Agentic Database Builder *Aug 2024*
Autonomous Data Engineering
- Designed an AI-driven agentic system for autonomous database construction, utilizing LLM decision-making for automated schema generation, data validation, and scalable curation.

Interactive Spherical Harmonics Visualizer *2023-24*
Scientific Visualization Engine
- Built an interactive 3D quantum orbital visualizer using Flask and Plotly, enabling real-time computation and rendering of spherical harmonics for educational visualization.

More projects can be found at my Portfolio Website or Github Profile.

Technical Skills

Core Competencies : Technical Leadership | Research & Development | Strategic Planning | Event Management | Mentoring | Public Speaking | Cross-functional Collaboration
Research Computing : Algorithm Development | Information Retrieval (Apache Lucene, BM25) | Bio-Informatics | Molecular Dynamics | DFT Computations | Reciprocal Rank Fusion | Information Theory
Programming Languages : Python | C/C++ | Java | Rust | QBASIC | GWBASIC | Fortran
Python Libraries - Core & Scientific : Numpy | Pandas | SciPy | Matplotlib | Plotly | Scikit-learn | OpenCV | PyTorch | Tensorflow
Python Libraries - NLP & ML : HuggingFace Transformers | LangChain | NLTK | Spacy | FAISS | ChromaDB | Selenium | BeautifulSoup
Python Libraries - Document Processing & Utilities : PyMuPDF | EasyOCR | Streamlit | Manim | fake_useragent
Large Language Models & Embeddings : Qwen2-4B | Qwen3-4B | Gemma (1B, 27B) | DeepSeek | LLaMA-3.2:1B | Nomic Embeddings | Mistral-OCR | Indic-Conformer
RAG & Retrieval Systems : Retrieval-Augmented Generation (RAG) | Dense & Sparse Retrieval | Vector Databases (FAISS, ChromaDB) | Hybrid Search (BM25 + Embeddings)
Digital Pathology Tools : HistoQC | PathProfiler | GrandQC | HistoROI | FASTPathology
Scientific Software : LAMMPS | VMD | Gaussian | Origin Pro | Scilab
Bio-Informatics : PyMol | ChimeraX | PyDock | AutoDock Vina
Development Stack : Full Stack (HTML/CSS/Javascript) | Databases (MongoDB, MySQL, SQLite3, ChromaDB) | Version Control (Git) | Containerization (Docker) | CMake | Make
Framework : Django
GUI Development : GTK4 in C | QT in C++ | Glade for Designing
Operating Systems & Server Management : Linux | Server Handling- SSH, OpenSSL | CUDA Programming (Python)
Laboratory Instrumentation : UV-Vis | ATR-FTIR | TGA | DSC | Optical Bench | XRD | GC-MS | Column Chromatography | Fluorimeter | SEM | TEM | AFM
Additional Tools : LaTeX| Shell Scripting (Bash) | UI/UX Design (Figma) | HTML E-mailing

Achievements & Honors

Hackathons:

Capital One Launchpad (Top 14 Team among 5000+ Teams) *2025*
Developed IndicAgri: Multi-modal RAG Platform for Indian Agriculture (Team Fibonacci) Capital One India

- Selected as one of the Top 14 teams from 5,073 registered teams across India in Capital One's flagship AI/ML innovation challenge
- Built an agentic AI-powered agricultural advisor supporting 20+ Indian languages with scientifically cited, hyper-localized outputs using RAG, LangChain, and open-source LLMs (Gemma, DeepSeek)
- Engineered autonomous data pipeline creating a novel 15k+ document dataset for Indian agriculture, released on Hugging Face

StatusCode1 (Awarded 1st Rank in GIAN Track) *2024*
Developed an AI-based Search Engine for GIAN's Database For Abandoned US Patents IIIT-Kalyani

- A searching algorithm based on Nomic Embeddings of the patent abstracts and similarity computations thus enabling the user to search for patents in natural language without the use of specific terminologies.
- Also created a web-scraping algorithm to get the patent data using Selenium, BeautifulSoup and fake_useragent for anti-scraping measures by the website

StatusCode0 (Awarded 1st Rank in MATLAB Track) *2023*
Developed a Domestic Waste Type Data analysis tool for a proposed Start-Up Solution IIIT-Kalyani

National Level Basic Sciences Competitions:

ChemEnigma (1st Rank) *2025*
Emerged champions at 72 hours Chemistry contest of theoretical, experimental and concept-presentation IISc-Bangalore

All Bengal Chemistry Quiz (2nd Runners Up) *2025*
Solved chemistry based questionnaires under time constraints Presidency University

Mimansa (Zonal Topper) *2024*
Contributed to the team in Mathematical problem solving IISER-Pune

NAEST-National Anveshika Experimental Skill Test (Zonal Runners Up) *2023*
Create extensive experimental setup using homely items- see papers NANI, IIT Kanpur and Shiksha Sopan

Competitive Examinations:

Qualified JEE Mains and Advanced *2022*
Ranked in Top 0.1 fraction of Candidates

Qualified IAT(IISER-Aptitude Test) *2022*
Ranked in top 0.06 fraction of candidates

Qualified WBJEE *2022*
Ranked in Top 0.05 fraction of Candidates

Scholarships and Honors:

Reliance Foundation Undergraduate Scholar *2023*
Qualified the RF-UG Aptitude test to be in the top 5000 students to be awarded this

Best Young Scientist Speaker on Nanotechnology *2019*
Successfully presented at the prestigious conference World Science Conference, Jadavpur University

Professional Experience

Web Development for Anicon 3.0 *2024-2025*
Inquivesta XI, IISER Kolkata
- Developed and led the web development of Anicon 3.0 Event.

Web Development for Material Science Laboratory *2025*
EFAML, IISER Kolkata
- Developed and designed the researchers lab info page available under Dr. Soumyajit Roy's homepage.

Private Educator & Technical Trainer *2018-Present*
Advanced Computing & Basic Sciences Instruction (Self-Employed), Kolkata
- Developed and conducted courses in Computer Science, Physics, Chemistry and English for High School students (ICSE, CBSE, WB Board)
- Mentored 50+ students for Board and competitive examinations

Technical Consultant *2021-Present*
Self-Employed, Kolkata
- Designed and implemented high-performance computing solutions
- Expertise in system optimization, BIOS/UEFI configuration, and OS installation
- Successfully completed 50+ custom build projects with 100% client satisfaction

Published Author *2020*
MindScapes (ISBN: 978-9389923209), Kolkata
- Published creative anthology work focusing on metaphorical and philosophical themes
- Conducted workshops on technical and creative writing

Leadership & Community Impact

Office Bearer, Slashdot - The Programming and Design Club *Aug 2025 – Present*
IISER Kolkata
- Leading the official programming and design club of IISER Kolkata, organizing workshops, hackathons, and technical sessions to foster a coding culture on campus.

Office Bearer, Valence - The Chemistry Club *Aug 2025 – Present*
IISER Kolkata
- Leading the official chemistry club of IISER Kolkata, organizing seminars, student talks, and departmental events to foster a vibrant chemical science community.

Organizer, Qiskit Fallfest 2025 *Oct 2025*
Sponsored by IBM Quantum
- Organized the local chapter of Qiskit Fallfest (one of 150 selected institutes globally sponsored by IBM); conducted hands-on workshops and tutorials on Qiskit SDK and quantum computing fundamentals.

Event Management & Logistics *2023 – 2025*
IISER Kolkata
- **Anicon 3.0 & 2.0:** Led organization of two editions of the sci-fi convention with 500+ participants.
- **Academic Events:** Organized Supra-Molecular Discussions 2024 and GIAN Courses on Soft-Oxometalates and X-Ray Crystallography.

Social Impact & Community Service *2020 – 2021*
Volunteering
- Served as a COVID-19 relief coordinator and currently mentoring students under the Ek-Pehal education initiative.

Additional Activities: District Level Debate and Quiz finalist | Shotokan Karate practitioner | 4th Year Art and Painting Student (Stroke Art & Portraits)

Invited Talks and Positions of Jury

Technical Talk – Running and Optimizing Local LLMs *Aug 2025*
Slashdot Student Chapter, IISER Kolkata
- Delivered a hands-on session covering hardware setups, quantization methods, embedding pipelines, vector databases, and practical optimization techniques for local LLM deployment.

Invited Talk – Sustainability, AI, and Emerging Technologies *Aug 2025*

Valence – Chemistry Society, IISER Kolkata

- Presented an interdisciplinary talk bridging sustainability science with AI-driven modelling, retrieval systems, and computational chemistry tools.

Domain Judge – Computer Science Events *July 2025*

La Martiniere for Boys Annual Science Fest, Kolkata

- Invited to serve as an external judge for multiple Computer Science events involving project demonstrations, live coding, and problem-solving rounds.

Education

Indian Institute of Science Education and Research - Kolkata *CGPA: 8.2*
BS-MS (Chemistry Major, Computer Science Minor) 2022-2027 *(expected)*

Calcutta University *CGPA: 8.308*
B.Sc Honours in Physics (1st Year Only) 2021-2022

Jodhpur Park Boys' High School *83%*
Higher Secondary Education in Physics, Mathematics, Chemistry, Computer Science (WBCHSE) 2019-2021

The New Horizon High School *83.75%*
Secondary Level Schooling(English Medium) under WBBSE 2009-2019

References available upon request