

# Shuvam Banerji Seal

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

## Education

<b>Indian Institute of Science Education and Research - Kolkata</b> <i>BS-MS (Chemistry Major, Computer Science Minor)</i>	<i>CGPA: 8.2</i> 2022-2027 ( <i>expected</i> )
<b>Calcutta University</b> <i>B.Sc Honours in Physics (1st Year Only)</i>	<i>CGPA: 8.308</i> 2021-2022
<b>Jodhpur Park Boys' High School</b> <i>Higher Secondary Education in Physics, Mathematics, Chemistry, Computer Science (WBCHSE)</i>	<i>83%</i> 2019-2021
<b>The New Horizon High School</b> <i>Secondary Level Schooling(English Medium) under WBBSE</i>	<i>83.75%</i> 2009-2019

## 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 though 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

- Hierarchical Opinion Classification with Large Language Models** 2025 (*Submitted to FIRE 2025*)  
*Independent Research with Co-authors: Alok Mishra and Utkarsha Ghosh*
- Proposed a parameter-efficient fine-tuning approach for hierarchical opinion classification using the Gemma-1B model, employing a lightweight two-layer classification head, selective transformer block updates, and an 8-class flat reformulation with class-weighted loss to enhance minority-class performance and structural sentiment understanding.
- IISERK@ ToT 2024: Query Reformulation and Layered Retrieval for Tip-of-Tongue Items** 2024 (*TREC 2024 Proceedings*)  
*Co-authors: 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*)  
*Guided by Dr. Soumyajit Roy, Dept. of Chemical Sciences, IISER-Kolkata*
- 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

# Research Libraries and Frameworks Developed

## Fernholz Stochastic Portfolio Theory Python Library

Ongoing

Independent Project

- Developing modular components for diversity-weighted portfolio construction and relative arbitrage strategies using entropy measures, market capitalisation ranking algorithms and log rate optimisation
- Researching on stochastic differential equation applications for portfolio drift processes and growth rate to outperform benchmarks under volatility conditions
- Technology Used: Python, NumPy, SciPy, Pandas, Stochastic Calculus, Continuous Semimartingales

## Projects

### IndicAgri: AI Platform for Indian Agriculture

Aug 2025

Capital One Launchpad Hackathon 2025

- Developed a multi-modal (text, voice) RAG chatbot for 20+ Indian languages, delivering scientifically cited and hyper-localised outputs by integrating real-time data with OpenSource LLMs (Gemma, DeepSeek)
- Engineered an agentic data pipeline to autonomously build and curate a novel 15k+ document dataset for Indian agriculture, released on [Hugging Face](#)
- Technology Used: Python, RAG, LangChain, FAISS/ChromaDB, HuggingFace, Gemma, Indic-Conformer, BM25

### Automated WhatsApp Bulk Message Sender with Google Sheets Integration

2025

Independent Project

- Developed an automated messaging bot using Python and Selenium for bulk WhatsApp messaging with support for CSV files and Google Sheets API integration
- Implemented robust error handling for network issues, Unicode/emoji compatibility, and session persistence using Chrome user data for seamless re-authentication
- Technology Used: Python, Selenium, Google Sheets API, ChromeDriver, Unicode Processing

### Automated Event Coupon Management System with QR Verification

2025

Independent Project

- Built a Flask-based event management system with Google OAuth 2.0 authentication for secure bulk email distribution and unique QR code coupon generation
- Implemented real-time mobile QR scanner interface for instant coupon verification with encryption, status tracking, and automated thank-you emails post-verification
- Technology Used: Python, Flask, Google OAuth 2.0, Gmail API, QR Code Generation, Ngrok, HTML/JavaScript, Encryption

### Developed RAG Application for CCA-2015 based Legal Document Retrieval for INCI Entrepreneurial Summit

Feb 2025

Independent Project

- Created a robust RAG stack using Selenium for web scraping, Mistral-OCR for building the JSON document base, ChromaDB for vector storage, Nomic embeddings, and LLaMA-3.2:1B for retrieval, all wrapped in a Streamlit-based user interface.

### Developed a complete GUI full-stack application using GTK4 in C

2024

Under the supervision of [Dr. Kripabandhu Ghosh](#), Dept. of Computation and Data-Science, IISER-Kolkata

- Implemented a complete GUI application in C using GTK4 and Glade, with SQLite3 as the local Database Management System, along with a static Context ChatBot using optimized BM25-based retrieval for NLP.

### Wi-Fi Channel Optimizer for Network Performance Enhancement

2024

Independent Project

- Engineered a Bash automation script for Wi-Fi channel optimization through automated network scanning, speed testing, and channel selection
- Implemented real-time speed benchmarking across multiple channels using speedtest-cli and network interface management via nmcli
- Technology Used: Bash, Shell Scripting, Network Programming, Linux System Administration

### Porting the SMC Canteen system for IISER-Kolkata to Python3 Django

2024

Student Monitored Canteen System of IISER-Kolkata

- Contributing to the SMC system that serves 2000+ students at IISER-K by modernizing it into a Django-based web application with improved maintainability and modularity.

### Ray-Dutt and Bailar Twists Simulator for Coordination Chemistry

Nov 2024

Independent Project

- Developed a Python-based molecular dynamics simulator for octahedral coordination complex isomerization mechanisms
- Implemented 3D visualization of stereochemical rearrangements and transition states for Ray-Dutt and Bailar twist mechanisms
- Technology Used: Python, Molecular Dynamics, 3D Visualization, Computational Chemistry

### Agentic Database Builder for Autonomous Data Curation

Aug 2024

Independent Project

- Developed an AI-powered agentic system for autonomous database construction and data curation using LLM-driven decision making
- Implemented automated data validation, schema generation, and quality control mechanisms for scalable database

management

- Technology Used: Python, LLMs, Autonomous Agents, Database Management

### Spherical Harmonics Visualizer with Interactive Web Interface

2023-24

#### Independent Project

- Built an interactive 3D visualization tool for spherical harmonics using Flask/Dash with Plotly for quantum mechanical orbital representation
- Implemented real-time parameter controls for quantum numbers (l, m) with multiple visualization modes (surface, wireframe, contour, points) and mathematical computation using SciPy
- Technology Used: Python, Flask, Dash, Plotly, SciPy, NumPy, Mathematical Visualization

## Technical Expertise

**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:

<b>Qualified JEE Mains and Advanced</b> <i>Ranked in Top 0.1 fraction of Candidates</i>	2022
<b>Qualified IAT(IISER-Aptitude Test)</b> <i>Ranked in top 0.06 fraction of candidates</i>	2022
<b>Qualified WBJEE</b> <i>Ranked in Top 0.05 fraction of Candidates</i>	2022

Scholarships and Honors:

<b>Reliance Foundation Undergraduate Scholar</b> <i>Qualified the RF-UG Aptitude test to be in the top 5000 students to be awarded this</i>	2023
<b>Best Young Scientist Speaker on NanoTechnology</b> <i>Successfully presented at the prestigious conference</i>	2019 World Science Conference, Jadavpur University

Professional Experience

<b>Web Development for Anicon 3.0</b>  <i>Inquivesta XI, IISER Kolkata</i> <ul style="list-style-type: none"><li>Developed and leaded the web development of Anicon 3.0 Event.</li></ul>	2024-2025
<b>Web Development for Material Science Laboratory</b>  <i>EFAML, IISER Kolkata</i> <ul style="list-style-type: none"><li>Developed and designed the researchers lab info page available under Dr. Soumyajit Roy’s homepage.</li></ul>	2025
<b>Private Educator &amp; Technical Trainer</b> <i>Advanced Computing &amp; Basic Sciences Instruction (Self-Employed), Kolkata</i> <ul style="list-style-type: none"><li>Developed and conducted courses in Computer Science, Physics, Chemistry and English for High School students (ICSE, CBSE, WB Board)</li><li>Mentored 50+ students for Board and competitive examinations</li></ul>	2018-Present
<b>Technical Consultant</b> <i>Self-Employed, Kolkata</i> <ul style="list-style-type: none"><li>Designed and implemented high-performance computing solutions</li><li>Expertise in system optimization, BIOS/UEFI configuration, and OS installation</li><li>Successfully completed 50+ custom build projects with 100% client satisfaction</li></ul>	2021-Present
<b>Published Author</b> <i>MindScapes (ISBN: 978-9389923209), Kolkata</i> <ul style="list-style-type: none"><li>Published creative anthology work focusing on metaphorical and philosophical themes</li><li>Conducted workshops on technical and creative writing</li></ul>	2020

Leadership & Community Impact

<b>Event Management:</b> Organized Anicon 3.0 (500+ participants) in 2025   Organized Supra-Molecular Discussions 2024   Organized GIAN Courses on Soft-Oxometalates and X-Ray Crystallography
<b>Social Impact:</b> COVID-19 relief coordinator   Educational mentor for Ek-Pehal program
<b>Additional Activities:</b> District Level Debate and Quiz finalist   Shotokan Karate practitioner   4th Year Art and Painting Student with major works in Stroke art and Portraits