Anoushka Sengupta

sengu22a@mtholyoke.edu | +91 9830255573 | LinkedIn: anoushka-sengupta | website: ankasengupta.github.io

Software Engineer | **Data Scientist** | **Analyst** | Experience in **Clinical Trials**, **AI**, **Machine Learning**, **Computational Linguistics**, and **Cloud Technologies** | Proficient in **Python**, **Java**, **R** | Skilled in **Agile** & **Scrum**
Indian Citizen currently based in Kolkata, West Bengal, but **open to relocation**

EXPERIENCE

Application Dynamics Technology LLP, Kolkata, West Bengal, India - *Software Developer*
MARCH 2025 - PRESENT

- Leading the design and development of an AI-powered chatbot using **Python**, **FastAPI**, and OpenAI's **GPT-4**, with support for both text and voice inputs via a speech-to-text API at an early-stage startup
- Building a modular, multilingual backend system with **MongoDB**, supporting domain-specific functionality and future integration with external services
- Architecting and implementing a high-performance **Retrieval-Augmented Generation (RAG)** pipeline using **FAISS** and **Hugging Face** transformer-based embeddings, enabling semantic search over 2.7 million question-answer pairs
- Leading experimentation and planning for cloud-based GPU infrastructure, optimizing for performance, cost, and reliability in processing large-scale data (~5GB raw text, ~12GB embeddings)
- Developing tooling for efficient batching and memory-safe data processing to support scaling and growth
- Maintaining the technical roadmap and system architecture to align with evolving business goals
- Driving technical decision-making on frameworks, infrastructure, and long-term scalability strategies in close collaboration with company leadership

4G Clinical, Wellesley, MA, USA - *Study Configuration Engineer*
APRIL 2023 - DECEMBER 2023

- Configured clinical studies utilizing Prancer RTSM, 4G Clinical's proprietary randomization and trial management supply management technology solution for optimizing clinical trials
- Coded **Python** scripts to rectify data discrepancies and maintain data integrity in the **PostgreSQL** relational database management system, and updated change logs to ensure accountability and create an audit trail for data modifications
- Created two Data Change Tool scripts to automate repetitive client requests to replace the need for at least 30 individual data change scripts
- Developed customized reports by collaborating with Client Service Leads and Business Systems Analysts to ensure alignment with client requirements and expectations
- Optimized existing report logic by modifying **Django** serializers and caching data to increase loading speed and improve functionality
- Investigated and resolved production issues by reproducing and diagnosing errors, leveraging **Sentry SDK**, **Kibana**, and **Stern** error logs from **Kubernetes** pods for comprehensive debugging
- Took ownership of the Anonymizer project to anonymize clinical data, utilizing **AWS DynamoDB** tables for primary key and email mapping management across various study stacks, and integrating Prancer's Gateway API for streamlined information exchange of **JSON** data
- Trained and mentored new hires

Wayfair, Boston, MA, USA - *Software Engineer*
JULY 2022 - MARCH 2023

- Developed and maintained the **Java**, **Python**, and **PHP** clients for Scribe, Wayfair's event tracking data engine, as part of the Event Pipelines & Consumption cross-functional scrum team, within the larger Data and Machine Learning Platforms Engineering Team
- Enhanced the Java client with batch support functionality by leveraging **Java Spring Boot** framework, implementing **JUnit** and **Selenium** tests, and integrating the testing into the **Buildkite CI/CD** pipeline

- Updated the PHP client to align with Scribe 2.0 specifications, enabling seamless data transmission via the new private Gateway, while conducting **PHPUnit** tests, focusing mainly on new HTTP response code scenarios
- Released updated versions of the Scribe clients and clearly communicated changes to impacted consumers
- Developed comprehensive documentation for Scribe consumers, enabling independent navigation of data monitoring, latency issues, storage concerns, and error log analysis via platforms like **Datadog** and **Kibana**, resulting in a 30% reduction in inquiries to the Scribe team and enhancing overall team productivity
- Led foundation dataset deprecation analysis to eliminate duplicate engineering efforts, collaborating with Storefront Data Engineering stakeholders, utilizing **Apache Airflow** to assess Directed Acyclic Graphs of the datasets, aligning with the company's cost optimization strategy transitioning to **Google BigQuery** for data processing

SKILLS
Programming and Scripting: Python, Java, R, PHP, C, SQL, MATLAB, JavaScript, bash/zsh shell scripting
Tools & Technologies: Microsoft Office, Google Workspace, Jira, LATEX, Git, Google Cloud (GCP), Amazon Web Services (AWS), Buildkite CI/CD, Sentry SDK, Kibana, Jenkins, Docker, Kubernetes
Languages: Fluent - English, Hindi, Bengali; Classroom study - Japanese, Italian

PROJECTS
Gene Expression Profiles for Breast Cancer Survival Prediction
- Built a **multivariate binary logistic regression** machine learning model in **Python** using z-score normalized mRNA gene expression data
- Predicted the survival of breast cancer patients with around 61% accuracy

Dengue Compartmental Model Simulation
- Developed an SEIR compartmental model in **R** to simulate the spread of a single strain of Dengue
- Implemented both **deterministic** and **stochastic** infectious disease modeling simulations

Tracovid
- Managed a team working on a cross-platform mobile application using **Dart** and **Flutter** in **Android Studio** to simplify the lives of students at Mount Holyoke during the COVID-19 pandemic
- Utilized the **Python** library **BeautifulSoup** for web scraping, and used **JavaScript** and a **MySQL** database to store the scraped data

EDUCATION
Mount Holyoke College, South Hadley, MA, USA - *Bachelor of Arts in **Computer Science**, minor in **Statistics***
AUGUST 2018 - MAY 2022
- **Honours**: *Magna cum laude*, Mu Sigma Rho (National Statistics Honor Society)
- **Cumulative GPA**: 3.85 on a scale of 0-4 (equivalent to 96.25%)
- **Relevant Coursework**: Data Science, Computational Linguistics, Machine Learning, Infectious Disease Modeling, Nonparametric Statistics, Software Design & Development, Computing Systems, Data Structures and Algorithms
- **On-Campus Work**: Peer Mentor, Teaching Assistant, and Grader - Computer Science and Statistics
- **Extracurriculars**: GirlUp MHC Co-Founder and President, Choir Secretary, TechTogether Boston Organizer, Hackathon Director of Prospective Sponsors, Harvard WECode Technology Fellow

Loreto House School, Kolkata, India - *Science Stream*
COMPLETED APRIL 2018
- **ICSE (Class X Board)**: 93.4%
- **ISC (Class XII Board)**: 92.25%