

Sinjoy Saha

State College, PA 16803 • sinjoy.saha@psu.edu
sinjoysaha.github.io • linkedin.com/in/sinjoysaha • github.com/sinjoysaha • [Google Scholar](https://scholar.google.com/citations?user=...)

EDUCATION

The Pennsylvania State University, University Park, PA **Aug 2024 - May 2026**
Master of Science, Computer Science and Engineering GPA - 3.56/4.0
Relevant Courses: Algorithms, Machine Learning, Computer Architecture, Deep Learning, Computer Vision, DL for NLP
Graduate Assistantship: Grader for CMPSC 465 - Data Structures and Algorithms

University of Calcutta, Kolkata, WB **Oct 2016 - Sep 2020**
Bachelor of Technology, Electronics and Communication Engineering GPA - 3.9/4.0

TECHNICAL SKILLS

Languages: Python, C, C++, SQL, Java, MATLAB, JavaScript
Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, LangChain, spaCy, pySpark, Numpy, Pandas, Django, Flask, fastAPI
Databases: MySQL, MS SQL Server, PostgreSQL, MongoDB, Elasticsearch, Milvus Vector DB
Tools: Linux, Jenkins, Git, AWS, Docker, Power BI, Tableau, Excel, Jira, Confluence, REST, Unix, Shell

WORK EXPERIENCE

Siemens Healthineers **Bangalore, KA**
Senior Engineer - Research and Technology Mar 2023 - Aug 2024

- Fine-tuned a **Siamese BERT** network on query-result pairs for re-ranking search results and obtained NDCG at 10 of 0.71.
- Led the deployment of a search API using **fastAPI** on AWS cloud, leveraging on-premise LLM and **Milvus VectorDB** to serve **RAG** pipelines and low-code platforms like Power App and Teams chatbot, saving **\$20,000** per annum in license costs.
- Designed semi-supervised system for identifying failure modes from service tickets using **Named Entity Recognition (NER)**, HDBSCAN clustering and **generative labelling** by **Orca-2**, leading to an invention disclosure.
- Created a **Power BI** dashboard for visualizing failure modes across installed base and monitor trends in KPIs over time such as service costs, parts replaced and downtime, leading to the redesign of a SPECT subcomponent.
- Developed histopathology WSI segmentation model utilizing **DeepLabV3+** and **ResNet50** backbone (Dice: 0.85, IoU: 0.63). Optimized using **ONNX**, pruning, quantization, normalization and background separation, achieving sub-15 minute inference.

Siemens Healthineers **Bangalore, KA**
Engineer - Research and Technology Dec 2021 - Feb 2023

- Implemented end-to-end **Cause and Action** phrase extraction from service tickets leveraging **roBERTa** and **NER**, increasing average ticket coverage in search results from **14%** to **52%** compared to earlier POS tagging model.
- Developed an unsupervised method leveraging **Word2Vec** embeddings and **DBSCAN** to cluster domain-specific words into a thesaurus, increasing clusters from **15** to **346** and enhancing full-text SQL search for PET/SPECT service tickets.
- Engineered a novel image feature set for **Random Forest** classifier and an ROI-agnostic artifact segmentation model for **automated grading** of SPECT QC images, improving F1-score from **0.46** to **0.75**, resulting in a patent filing.

Cognizant Technology Solutions **Kolkata, WB**
Programmer Analyst Trainee Nov 2020 - Sep 2021

- Developed and executed automated testing scripts using Java Selenium, TestNG, and JUnit to ensure quality of API and web apps.
- Performed requirement analysis and test design for web/mobile apps and worked with developers to resolve defects.

RESEARCH EXPERIENCE

Human Language Technologies Lab, Penn State **University Park, PA**
Graduate Researcher *under Prof. Shomir Wilson* Aug 2024 - Present

- Optimized visualization rendering latency for **PrivaSeer**, a privacy policy search engine containing ~ 3.1 million policies.
- Leveraged on-premise **Llama 3.1-8B-it** to extract contradictory and inconsistent statements from privacy policies in JSON format.
- Designed output parsing and self-verification systems to filter incorrectly formatted outputs and reduce hallucinations.

Plant Vision Lab, University of Nebraska-Lincoln **Remote**
Research Intern *under Dr. Sruti Das Choudhury, Prof. Ashok Samal* Jul 2021 – Dec 2021

- Developed two methods to predict onset of plant drought stress using **DTW** algorithm on computed phenotypes and **1D-CNN** for temporal stress propagation on hyperspectral images (F1: 0.98, avg. SWC corr.: -0.85), leading to a journal publication.

Sensordrops Networks Pvt. Ltd., IIT Kharagpur **Remote**
Research Intern *under Dr. Rituparna Saha, Prof. Sudip Misra* Jul 2021 – Dec 2021

- Introduced a novel algorithm for **Federated Learning** in **non-IID** data by clustering on client data statistics, surpassing FedAvg accuracy by **2%** and reducing aggregation time by **67%** compared to clustering on local weights.

PATENT & PUBLICATIONS

- Daga S., **Saha S.**, Crawford T. E., Khan K., Morris B. Artifact Segmentation and/or Uniformity Assessment of a Gamma Camera. Patent filed in April 2023 (USPTO). File Ref. 18/297,062. Under review.
- Das Choudhury S. *, **Saha S.** *, Samal A., Mazis A. and Awada T. (2023) Drought stress prediction and propagation using time series modeling on multimodal plant image sequences. *Frontiers in Plant Science*. [\[link\]](#)
- Basu S., **Saha S.**, Pandit S., Barman (Mandal) S. (2020) Smart Health Monitoring System for Temperature, Blood Oxygen Saturation, and Heart Rate Sensing with Embedded Processing and Transmission Using IoT Platform. CIPR, 2019. [\[link\]](#)

PROJECTS

Tiktokenizer-js | *HTML, CSS, JavaScript, Bootstrap* | [\[github/tiktokenizer-js\]](#)

- Developed a static site for visualizing the **GPT-2 Byte-Pair Encoding (BPE)** tokenization process, replicating official OpenAI API.

Brain Tumor Segmentation from MRI using PSPNet | *Python, CNN, TensorFlow/Keras* | [\[kaggle/pspnet-brain-mri\]](#)

- Demonstrated efficacy of PSPNet for segmentation of brain tumors from MRI data, achieving Dice: 0.66 and IoU: 0.552.

AWARDS & ACHIEVEMENTS

- Nominated as **Finalist** in **DC Excellence Awards** for automated flood grading, Siemens Healthineers, Mar 2024
- **Top 5** in the **AI Innovation Sprint** for Predictive Maintenance and Incident Management, Siemens Healthineers, Feb 2024
- **Runners-up** in **Supply Chain Management Data Science Hackathon** among 60+ teams, Siemens Healthineers, Feb 2023
- **JBSTNS Senior Scholar**, 2016, Received a monthly scholarship and annual book grant throughout undergraduate studies

LEADERSHIP & OTHER ACTIVITIES

- As a Penn State ACM Co-Secretary (2024), managed email correspondence to deliver weekly updates to members and coordinated and scheduled weekly meetings to ensure effective communication and planning.
- Conducted hands-on session on Segmentation at [AI for Precision Medicine](#) workshop at IISc, Bangalore (Jan 2024)
- Invited talk on Introduction to AI and Data Science for undergraduate students organized by the AI & Robotics Club and IEEE Student Branch, University of Calcutta (May 2023)
- Mentored a team on NLP topics during the Data Engineering Re-Skill Initiative at Siemens Healthineers (Nov 2022)