

Suchandra Chakraborty

+1 (267) 266 2284 | schndrac@seas.upenn.edu | github.com/SuchandraC | linkedin.com/in/suchandrac | suchandrac.vercel.app

EDUCATION

University of Pennsylvania

Aug 2025 – May 2027

Master of Science in Computer and Information Sciences

Relevant Coursework: Machine Learning, Artificial Intelligence, Software Systems, Analysis of Algorithms, Theory of Computation

Maulana Abul Kalam Azad University of Technology

Aug 2019 – Jun 2023

B.Tech in Computer Science and Engineering, GPA: 9.4/10 | Top 3% (6th/220); Batch Topper, 3rd Semester (800+ students)

TECHNICAL SKILLS

Programming: Python, C/C++, SQL, JavaScript, R, HTML/CSS | **Cloud & Tools:** AWS, GCP, Docker, Git, Weights & Biases, MLflow, Jupyter

ML/NLP: PyTorch, TensorFlow, Hugging Face Transformers, HF Datasets, RoBERTa, BERT, SpaCy, NLTK, Scikit-Learn, NumPy

Frameworks: React, Node.js, Flask, FastAPI, Streamlit, LangChain, Pandas, Matplotlib, Seaborn | **Databases:** PostgreSQL, MongoDB

EXPERIENCE

Tata Consultancy Services Research

Kolkata, India

Researcher

Sep 2023 – Jul 2025

- Architected an AI-powered Interview Assistant being rolled out to TCS panels, using LangChain to parse job descriptions, extract the top 10 technical and behavioral skills, and auto-generate targeted questions per skill.
- Engineered a fine-tuned RoBERTa model for mental health evidence extraction across 343 Reddit posts from 30 user timelines, achieving second-best recall (0.602) at CLPsych 2025 (ACL); Random Forest outperformed DeepSeek-7B (MSE: 2.994 vs. 6.610).
- Constructed MuRFE, a 5,684-sample synthetic dataset (8 relation types) via Llama-3.3-70B with SO/SRO/document-level prompting, 81.7% clean rate; benchmarked 5 SOTA models, best F1 of 0.844; under review at CoNLL 2026.
- Optimized an end-to-end pipeline for document-level relation extraction across 3,008 DocRED documents, validating mention extraction (0.930 F1), clustering (0.804 F1), and triple extraction (0.405 F1) using BART-large.

Magic EdTech (collab. with ABP Learning Technologies, ABP Group)

Remote

NLP Intern

Aug 2022 – Oct 2022

- Designed a distractor generation system using Sense2Vec and MMR over 10K+ samples, achieving ~60% acceptance rate on human evaluation with contextually diverse outputs.

RESEARCH EXPERIENCE

Dr. S. Datta, Dr. S. Majumder, HIT Kolkata

Jan 2022 – May 2023

- Modeled temporal developer interactions across 4 networks (853 nodes, 17,019 edges) in OpenStack/Eclipse using BTERGM with 500 bootstrap replications; identified PageRank as the most significant predictor of tie formation; validated via GOF analysis.

Dr. D. Bhandari (HIT Kolkata), Dr. D.P. Mandal (ISI Kolkata)

Aug 2021 – Jun 2023

- Synthesized 4,000 retinal image-mask pairs using StyleGAN2 (FID: 9.55) and Pix2Pix, boosting DICE from 0.842 to 0.931, outperforming classical augmentation (0.887) with 99.2% accuracy on clinician-verified data from 6 benchmark datasets.

PUBLICATIONS

(full list: suchandrac.vercel.app/papers)

- Saini, P., Jain, B., **Chakraborty, S.**, et al. (2026). MuRFE: Multi-valued Relational Facts Extraction. *Under review, CoNLL 2026*
- **Chakraborty, S.**, Jana, S., Sinha, M., Dasgupta, T. (2025). Self-State Evidence Extraction. *CLPsych 2025 (ACL Workshop)*
- **Chakraborty, S.**, Basu, A., et al. (2024). Developer Communication over Time. *COMSYS-2024, Springer [SCOPUS]*
- Bhuiya, S., **Chakraborty, S.**, et al. (2023). Retinal Image Segmentation Data Generation. *PReMI 2023, Springer [SCOPUS]*

PROJECTS

PennCloud - Distributed Cloud Platform | C++, gRPC, SMTP, WebSockets

Fall 2025

- Engineered a full SMTP server with external email relay via MX record lookup and RFC 2822-compliant headers, supporting intra/inter-domain routing across 4 backend nodes with Paxos-based replication.
- Designed the email storage schema (row-based KV pattern with pipe-delimited serialization), enabling efficient per-user inbox retrieval via GetRow() queries; built an extra-credit sent folder and an address book with one-click compose.
- Implemented the Admin Console backend aggregating data across 8 gRPC ports with 1-second liveness probes, std::set deduplication, and a paginated KV data viewer with color-coded rows and soft-shutdown control plane.

NeurIPS - Ariel Data Challenge (Bronze Medal (top 9%)) | PyTorch, NumPy, Pandas, Scikit-Learn

- Constructed a time-series model with manual feature engineering and ridge regression to estimate exoplanet atmospheric properties from telescope observations.

BiDAF QA Model | PyTorch, GloVe, NLTK

- Built a QA model with bi-directional attention flow over 100K+ SQuAD samples, achieving 65.7 F1 / 52.5 EM with 2.6M parameters

LEADERSHIP & AWARDS

- ACM Secretary (Dec 2021 – Jul 2022): Organized inclusive hackathons with 200+ participants promoting women in tech.
- Won Silver Idea at **Tata eHackathon** for an innovative approach to streamlining the claims process.