

Sayar Ghosh Roy

☎ +1 (217) 979-4121

✉ sayagr3@gmail.com

🌐 [sayarghoshroy](#)

🔗 [Sayar's Work](#)

in [sayarghoshroy](#)

🎓 [scholar](#)

EDUCATION

University of Illinois at Urbana-Champaign

Aug 2022 – May 2024

- Master of Science in Computer Science

(Fully Funded)

GPA: 4.0, *Thesis*: Interpretable Local Citation Recommendation

Focus: Data Mining, Machine Learning, Algorithmic Trading & Market Microstructure

International Institute of Information Technology, Hyderabad

Jul 2017 – Jun 2022

- Bachelor of Technology (with Honors) in Computer Science

- Master of Science in Computational Linguistics

(Fully Funded)

CGPA: 9.76/10, **1st Rank in Class**

Focus: Information Retrieval & Extraction, Machine Learning, Natural Language Processing

Thesis: Proactively Forecasting Sentence-Specific Information Popularity [🔗₁](#)

EXPERIENCE

DMG, DAIS Lab, UIUC

Aug 2022 – Dec 2023

Graduate Researcher

Urbana, Illinois

- Advised by Prof. Jiawei Han
- Worked on Evidence-grounded distantly-supervised Interpretable Local Citation Recommendation for scientific claims

Apple

Sep 2020 – Aug 2022

Applied Research Engineer

Hyderabad, India

Part-time, Contract

- Built explainable Machine Learning models to score actionability of user reported issues within Maps
- Designed scalable unsupervised algorithms to cluster & derive definite business insights from geospatial data dumps
- Developed a focused web crawler for periodic retrieval of pertinent multilingual pages
- Engineered a 7x⁺ speed gain within Maps' web discovery pipeline while improving page-relevance categorization precision by over 20% by formulating an ensembling-based Deep Learning method

iREL, LTRC, IIIT Hyderabad

Jul 2019 – Jun 2022

Student Researcher

Hyderabad, India

- Advised by Prof. Vasudeva Varma and Dr. Manish Gupta
- Worked on proactive sentence-specific information popularity forecasting for online news documents, investigated connections between information salience & popularity (Joint work with Microsoft Bing AI) [🔗₂](#) [🔗₃](#) [🔗₄](#)
- Designed effective architectures for generating layperson-focused & extended summaries of research papers [🔗₅](#) [🔗₆](#)
- Built adverse content identification, categorization, and intervention systems for multiple languages [🔗₇](#) [🔗₈](#)

Digitalee

May – Jul 2019

Software Engineer Intern

Hyderabad, India

- Created an end-to-end pipeline for generating text summaries of lengthy executive briefing PDFs

Theodor

Sep – Nov 2018

Software Engineer Intern

Hyderabad, India

- Worked on digitization of hand-filled forms utilizing [CTPNs](#), [CRNNs](#) for text detection & recognition
- Built a platform with a central dashboard for users to upload & convert scanned images to digitized data

PUBLICATIONS

1. **S Ghosh Roy**, J Han. ILCiteR: Evidence-grounded Interpretable Local Citation Recommendation. Under review. | [code](#)
2. **S Ghosh Roy**, A Padhi, R Jain, M Gupta, V Varma. Towards Proactively Forecasting Sentence-Specific Information Popularity in Online News Documents. In [33rd ACM Conf. on Hypertext & Social Media \(HT '22\)](#) (Main Track), Jun 2022. | [url](#) • [pdf](#) • [code](#)
3. **S Ghosh Roy**, N Pinnaparaju, R Jain, M Gupta, V Varma. Summaformers at LaySumm 20, LongSumm 20. In [SDP Workshop at EMNLP '20](#), Nov 2020. | [url](#) • [pdf](#) • [video](#) • [poster](#) • [code](#)
4. **S Ghosh Roy**, U Narayan, T Raha, Z Abid, V Varma. Leveraging Multilingual Transformers for Hate Speech Detection. In [Forum for Information Retrieval Evaluation \(FIRE '20\)](#), Dec 2020. | [url](#) • [pdf](#) • [code](#)
5. T Raha, **S Ghosh Roy**, U Narayan, Z Abid, V Varma. Task Adaptive Pretraining of Transformers for Hostility Detection. In [Constraint Workshop at AACL '21](#), Feb 2021. | [url](#) • [pdf](#)

HONORS AND AWARDS

- ★ **Program Gold Medal**, IIIT Hyderabad, 2022
- ★ **Research Award**, IIIT Hyderabad, 2021
- ★ **Dean's List 1**, *every* academic semester at IIIT-H, Monsoon '17 – Spring '21
- ★ **Winners, Constraint Shared Task** for multilingual hostile content categorization, **AAAI '21**: 1st/45 teams globally
- ★ **Winners, LongSumm** (1st) & **LaySumm** (3rd) at **EMNLP '20**: state-of-the-art on 2 scientific summarization tasks
- ★ Endorsed by IIIT-H for the **S.N Bose Scholars Program**, 2018
- ★ Selected for **INSPIRE Scholarship**, 2016, by **Dept. of Science & Technology**, Govt. of India
- ★ 1st Institute Rank, 8th State Rank in the **Indian School Certificate** (ISC), 2016, aggregate: **97.25% + perfect score** in **CS**
- ★ Award for **General Proficiency**, 2015, **Salt Lake Point School**, Kolkata
- ★ **Student of the Year**, 2014, **St. Mary's School**, Kolkata

TEACHING + MENTORING

- Teaching Assistant, Introduction to **Data Mining**, UIUC | Spring '23, Fall '23, Spring '24
- Teaching Associate, Adv. Management Program in Business Analytics (**AMPBA**), Indian School of Business (**ISB**)
 - **Data Collection and Preprocessing** (2 offerings), Aug '21 – Jun '22 [[created material](#)]
 - **Natural Language Processing**, Jan '22 – Jun '22 [[created material](#)]
- Instructor, Introduction to Deep Learning Workshop, IIIT-H | Summer '20, '21 [[created material](#)]
- Teaching Associate, Basics of Ethics, IIIT-H | Monsoon '21
- Course Project Mentor, IIIT-H: Intro to **NLP** (Spring '21), **Information Retrieval & Extraction** (Monsoon '21)

TECHNICAL POSITIONS

- **System Administrator**: Managed compute servers shared by **CIE** & **iREL**, IIIT-H | Jan '20 – Jun '22
- **Web Administrator**: Developed + maintained lab website, **iREL**, IIIT-H | Nov '19 – Jun '22

ADVANCED COURSEWORK

ML	Statistical Methods in AI , Optimization Methods, Deep Learning Theory, Transfer Learning, Applied ML
info⁺	Data Mining, Information Retrieval & Extraction, Text Mining, Deep Learning for Healthcare
NLP	Computational Linguistics, NLP Applications , Social Computing, Language Typology & Universals
math	Multivariate Analysis, Probabilistic Graphical Models
misc.	Algorithmic Trading, Neural & Cognitive Modeling, Computer Security, Programming Language Theory

RELEVANT CERTIFICATIONS

- **Machine Learning for Trading**, Coursera Specialization (3 courses) by New York Institute of Finance & Google Cloud [↗₉](#)
- **Bloomberg Market Concepts** certificate [↗₁₀](#) & **Bloomberg Finance Fundamentals** certificate [↗₁₁](#), by Bloomberg for Education

TECHNICAL SKILLS

PLs	Python, C++, C, Java, SQL, Matlab, Octave, Bash, HTML, CSS, JavaScript, Racket, Haskell
Tools	PyTorch, TensorFlow, scikit-learn, NumPy, SciPy, spaCy, nltk, HuggingFace, pandas, RCM-X, OpenCV, \LaTeX , ...

SELECTED COURSE PROJECTS

... For more, go [here](#)

Arbitragedy | python • pytorch • C++ • RCM-X | Market Microstructure analysis and forecasting, Arbitrage-based High Frequency Trading strategies, Study on effect of latency on performance, Spectral Correlated Ticker clustering

Augment4Gains | python • pytorch • tensorflow | Enhancing hate speech detection through various data augmentation methods

Acronym Sense Disambiguator | python • nltk • xml-sax | Disambiguates possible expansions of acronyms within an article: disambiguator module runs queries on Wikipedia via **Wikisearch**: a built-from-scratch Search Engine for Wikipedia dumps

OptiLearn | python • scikit-learn | From scratch pedagogic implementations of optimization algorithms: Adam, Nesterov's accelerated GD, Polyak's LR, Newton's method, Rprop, Quickprop — classification & regression tasks

COVID Trends | python • scipy | Gradient Descent formulation for training Susceptible-Infected-Recovered models

Manifold Learning Mechs | python • scikit-learn | Implementations of various spectral clustering techniques & manifold visualization schemes such as MDS, LLE & Isomap: applications in Computer Vision tasks

LPProb | python • cvxpy | Formulations for miscellaneous Optimization Problems using Linear Programming Paradigms

LoClus | python • scikit-learn | Location clustering based on mutual geodesic distances, Rental pricing prediction in **place2crash**

C shell | C | User-defined interactive shell program capable of creating & managing processes

Func-Interpreter | haskell | Interpreter for a Functional Arithmetic Language

More with LDA | python • scipy | LDA for dimensionality reduction & classification, applications in sentiment analysis

Error from repair | python • spacy | Detecting errors in dependency parsing of speech transcripts due to active linguistic repair

PicIdentifier | python • tensorflow | Performance of various DL architectures for various image classification tasks: insights and analysis

Games | C++ • opengl • python • pygame | **Angry Paper Jet**, **Space Invaders Retro**, **Funride**, **Mario on Terminal**