

SAYAR GHOSH ROY

+91-7980704678 @ sayar.1997@gmail.com sayarghoshroy Sayar's Work in sayarghoshroy 53/24 R.N Guha Road, Kolkata 700028

EDUCATION

International Institute of Information Technology, Hyderabad

Jun 2017 – Apr 2022

- Bachelor of Technology (Honors) in Computer Science
- Master of Science (by Research) in Computational Linguistics

CGPA: 9.76/10

HONORS AND AWARDS

- Winners, [Constraint Shared Task](#) at [AAAI 2021](#): Secured First Position for Adverse Content Detection in Hindi among 45 participant teams
- Winners, [LongSumm & LaySumm](#) at [EMNLP 2020](#): Achieved SOTA results on two text summarization tasks with ROUGE-1 scores of 0.50 and 0.46
- Recipient of Dean's List Academic Award, IIIT-H, 2017 - 2021
- 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
- Secured 1st School Rank, 8th State Rank in the [Indian School Certificate \(ISC\)](#) with an aggregate of 97.25% & a perfect score in Computer Science
- Award for General Proficiency, 2015, [Salt Lake Point School](#), Kolkata
- Student of the Year, 2014, [St. Mary's School](#), Kolkata

PUBLICATIONS

Summaformers @ LaySumm 20, LongSumm 20

S Ghosh Roy, N Pinnaparaju, R Jain, M Gupta and V Varma

[Proceedings of the First Workshop on Scholarly Document Processing at EMNLP '20](#), November 2020, Online | [url](#) • [slides](#) • [poster](#) • [code](#)

Leveraging Multilingual Transformers for Hate Speech Detection

S Ghosh Roy, U Narayan, T Raha, Z Abid and V Varma

[HASOC Identification in Indo-European Languages at Forum of Information Retrieval Evaluation \(FIRE\) '20](#), December 2020, Online | [url](#) • [code](#)

Task Adaptive Pretraining of Transformers for Hostility Detection

T Raha, S Ghosh Roy, U Narayan, Z Abid and V Varma

[First Workshop on Combating Online Hostile Posts in Regional Languages at AAAI '21](#), February 2021, Online | [url](#) • [code](#)

RELEVANT RESPONSIBILITIES

- Instructor at 'Intro to Deep Learning Workshop': Conducted various sessions for students joining [LTRC](#), IIIT-H | Summer '20, '21 | [created material](#)
- Project mentor: Mentored 5 teams for their 'Intro to Natural Language Processing' course project | Spring 2021
- Research mentor: Mentored 4 senior students in their research work at the Information Retrieval and Extraction Lab ([iREL](#)), IIIT-H | Jun 2020 - Present
- System Administrator: Managing resources shared by Centre for Innovation & Entrepreneurship ([CIE](#)) & [iREL](#), IIIT-H | Jan 2020 - Present

SKILLS AND COURSEWORK

python c++ c java javascript bash racket haskell matlab

pytorch scikit-learn tensorflow numpy spacy nltk huggingface

scipy cvxpy matplotlib pandas corenlp opencv opengl weblgl

Artificial Intelligence Statistical Methods in AI Optimization Methods

Deep Learning: Theory & Practices Probabilistic Graphical Models

Natural Language Processing · Advanced Information Retrieval & Extraction

Social Computing Neural & Cognitive Modelling Computational Linguistics

Linguistics Intro Social & Historical Typology & Universals

Linear Algebra Probability & Statistics Analysis Real Complex Multivariate

Programming Language Theory Formal Methods Database Systems

Algorithms Data Structures Operating Systems Graphics Discrete Math

EXPERIENCE

Applied Research Engineer Intern

Apple

Sep 2020 – Present

Remote

- Developed a system for automated procurement of pertinent webpages, & the extraction of appropriate information — in a multilingual setting
- Engineered a 7x+ speed gain in the Apple Maps' Web-Discovery pipeline
- Boosted existing document relevance-classifier precision by over 20%
- Oversaw the creation of multiple labelled datasets, built & deployed several models for specific Natural Language Processing tasks

Undergraduate Researcher

Information Retrieval & Extraction Lab

Aug 2019 – Present

Hyderabad, India

- Advised by Prof. Vasudeva Varma and Dr. Manish Gupta
- Designed an acclaimed architecture for generating summarized blog-like posts from Scientific Research Papers
- Formulated and built adverse content identification, categorization, and intervention systems for multiple languages

Software Engineering Intern

Digitalee

May – Jul 2019

Hyderabad, India

- Built an end-to-end system that produces text summaries of executive briefing documents as part of a three member team
- Implemented an information extractor that adheres to the structure of large executive briefing PDFs with all of their sections and subsections

Software Engineering Intern

Theodor

Sep – Nov 2018

Hyderabad, India

- Built a system facilitating the digitalization of hand-filled forms
- Utilized CTPNs for text detection and CRNNs for text recognition
- Created a web app using Django for users to upload scanned forms, receive digitized information, and have a dashboard of all activities

SELECTED PROJECTS

[Optimization and Learning](#) | python • scikit-learn

- From scratch implementations of various optimization techniques such as Adam Optimizer, Nesterov's Accelerated GD, Polyak's Learn Rate, Newton's Method, Rprop, and Quickprop: Tested out on classification & regression tasks

[Augment4Gains](#) | python • pytorch • tensorflow

- Enhancing Hate Speech Detection through various data augmentation methods

[Acronym Sense Disambiguator](#) | python • nltk

- Identifies acronyms in a text file and disambiguates possible expansions, where a guided disambiguator module runs queries on an indexed Wikipedia dump

[Image classification formulations](#) | python • scikit-learn • tensorflow

- Performance of various deep learning architectures for face recognition and miscellaneous image classification tasks: insights and analysis

[COVID-19 modelling for India](#) | python • scipy

- Formulation of a training algorithm for Susceptible-Infected-Recovered Models

[Manifold Learning Mechanisms](#) | python • scikit-learn

- Illustration of various Spectral Clustering Techniques + Manifold Visualization schemes such as MDS, LLE, & Isomap; with applications in Computer Vision

[Extractive summarization unit](#) | python • scikit-learn • spacy

- A lightweight easy-to-train extractive summarization formulation

Machine Translation systems | python • pytorch

- [Statistical MT system](#) for English → Hindi utilizing IBM models and an HMM-based formulation for relative token alignment
- [Supervised neural MT system](#) for English → Bengali using a bi-GRU encoder-decoder architecture with an attention mechanism

C shell | c

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

→ Do check out my work [here](#).