Aman Sinha

iamansinha.github.io | aman.sinha@utah.edu
linkedin.com/in/ai-aman-sinha | github.com/iamansinha

Education

University of Utah — Salt Lake City, Utah, USA
PhD in Computing | Artificial Intelligence track — *Aug 2023 – Current*

- On Medical Leave of Absence from Spring 2024. Resuming in Fall 2025.

Indian Institute of Science Education and Research (IISER) Kolkata — Kolkata, India
5-Year BS-MS Dual Degree | Physics (Major), Computational & Data Science (Minor) — *2018 – 2023*

- CGPA: 8.18/10.0
- 4-point GPA (Scholaro): 3.46/4.0

Publications

[1] **Aman Sinha**, Priyanshu Raj Mall, Dwaipayan Roy. *Exploring the Nexus Between Retrievability and Query Generation Strategies*. In: ECIR '24: The 46th European Conference on Information Retrieval, Glasgow, Scotland: March 24 - 28, 2024.
[paper] [code]

[2] **Aman Sinha**, Priyanshu Raj Mall, Dwaipayan Roy, Kripabandhu Ghosh. *A Comparative Analysis of Retrievability and PageRank Measures*. In: FIRE '23: The 15th meeting of Forum for Information Retrieval Evaluation 2023. Goa University, Panjim, India: December 15-18, 2023.
[paper] [arXiv] [slides] [code]

[3] **Aman Sinha**, Priyanshu Raj Mall, Dwaipayan Roy. *Findability: A Novel Measure of Information Accessibility*. In: CIKM '23: The 32nd ACM International Conference on Information and Knowledge Management. Birmingham, United Kingdom: October 21 - 25, 2023.
[paper] [arXiv] [poster] [code]

Research Experience

RL Alignment of LLMs for Query Expansion — Oct 2024 – Present
Advisor: Dr. Dwaipayan Roy, IISER Kolkata

- Leading independent research on aligning small LLMs (1B-8B params) for query expansion, a previously unexplored fine-tuning objective.
- Designing an iterative training loop that optimizes nDCG@10 as a reward using modified PPO/DPO frameworks, with early stopping to prevent overfitting and enhance generalization from limited relevance-judged queries.
- Targeting production-ready LLMs for low-latency web search, with potential scalability to large search engines like Google.

Bayesian Optimization of UWBG HEMT Device Structure — Oct – Dec 2023
Advisor: Prof. Mike Kirby and Dr. Shandian Zhe, University of Utah

- RA project for US Army Research Office (ARO), focusing on ultra-wide-bandgap (UWBG) high-electron-mobility transistor (HEMT) device optimization.
- Worked on integrating device simulation software into an adaptive Bayesian Optimization (BayesOpt) framework for automating the optimization loop.
- Had to step back from the project due to health reasons.

Retrievability Bias in Information Retrieval Aug 2022 – Jun 2023

Advisor: Dr. Dwaipayan Roy, IISER Kolkata

- In my Master's Thesis, I researched algorithmic bias in retrieval models towards certain documents and their impact on retrieval performance.
- Evaluated Retrievability bias of baseline IR models, identifying BM25 as the retrieval model inducing the least bias.
- Discovered bias in the TREC relevance judgement document pool, revealing systemic skew in document pooling.
- Found Pseudo Relevance Feedback (PRF) to cause greater bias in search results.
- Found the *Fairness Hypothesis*—predicting an inverse correlation between model effectiveness and retrievability bias—to not hold for BM25+RM3.
- Investigated correlation between PageRank and Retrievability scores of all Wikipedia articles by preprocessing 100GB+ wiki dump, and reported the findings in our FIRE'23 paper.
- Revealed the lack of reproducibility and standardization in retrievability analysis literature due to varying query generation methods, in our ECIR'24 reproducibility track paper.
- Developed an improved statistical method for generating realistic artificial search queries.
- Coined *Findability* in our CIKM'23 paper, a measure of a document's findability in an IR system, and established *Findability* as an independent measure under the umbrella of Information Accessibility measures in IR.

Retrievability Analysis of Neural IR Models Aug – Dec 2022

Guide: Dr. Kripabandhu Ghosh, IISER Kolkata

- In this Independent study, I investigated biases in Doc2vec and SBERT by performing retrievability analysis, finding huge biases in comparison to classical retrieval models.
- Worked on TREC 678 news collection along with 580k artificial queries
- Performed a large scale retrieval simulation for SBERT on the institute's Supercomputer

Query Expansion and Relevance Feedback Jun – July 2022

Guide: Dr. Dwaipayan Roy, IISER Kolkata

- In this Summer project, I tested True Relevance Feedback (TRF) and Pseudo Relevance Feedback (PRF) for Rocchio and RM3 query expansion.
- Implemented Rocchio and RM3 query expansion algorithms in Python.
- Studied the impact of negative feedback for true relevance and pseudo relevance feedback.
- Integrated negative feedback part from Rocchio into RM3 to boost performance.

Projects

Hopfield Neural Networks May 2022

Guide: Dr. Bhavtosh Bansal, IISER Kolkata

- Associative memory and efficient memory retrieval are often overlooked ingredients to machine intelligence. In this project, I studied Ising spin glass model inspired Hopfield Recurrent Neural Network.
- Studied the emergence of content addressable memory from Hebbian-learning-based update-rule and Hamiltonian of the network.
- Experimented with Hopfield net to perform image reconstruction from partial cues and dealt with mixed states problem

Pattern Classification and Network analysis on Colonoscopy data Apr 2022

Guide: Dr. Koel Das, IISER Kolkata

- In this project, I worked with Gastrointestinal Lesions in Regular Colonoscopy Dataset and LSVT Voice Rehabilitation Dataset.

- Applied Extra Trees Classifier to classify different types of lesions which outperformed the baseline by 11.5%.
- Performed network analysis approach to cluster the classes using gini feature importance scores from the classifier.

Nonlinear Dynamical Analysis of Rumor Spreading Nov 2021

Guide: Prof. Pradeep Kumar Mohanty, IISER Kolkata

- Motivated by the similarities between the spread of rumors and the spread of infectious diseases, I studied and simulated a rumor spreading model by Tingting Li & Youming Guo (2021).
- Simulated and studied the influence of different control parameters on the rumor propagation process.
- Compared the outcomes with models with no control parameters.

Quantum Circuit for solving Linear Differential Equations Dec 2018 – Jan 2019

Guide: Prof. Prasanta K. Panigrahi, IISER Kolkata

- Studied a novel Quantum algorithm for solving a system of LDEs by Tao Xin, et al.
- Developed a quantum circuit based on the algorithm to solve 4 x 4 LDE on IBM Quantum

Relevant Coursework

- Information Retrieval and Web Search
- Natural Language Processing
- Machine Learning and Network Analysis
- Statistics I
- Probability I
- Mathematical Methods of Physics
- Computational Physics
- Nonlinear Dynamics
- Programming and Data Structures I and II
- Introduction to Computation
- Introduction to Computer Programming

Skills

Programming Languages: Python (advanced), Java, Shell script (intermediate), C (basic), HTML,CSS,JS (basic).

Libraries: PyTorch, Transformers, BoTorch, Tensorflow, Keras, PyLucene, Pyserini, PyTerrier, Gensim, NumPy, SciPy, Pandas, Matplotlib

Software: MATLAB, Origin, GnuPlot, ImageJ

Misc.: LaTeX, Git, GitHub, Supercomputing, Cluster MPI

Mathematics: Probability and Statistics, Tensor algebra, Nonlinear dynamics, PDE and ODE, Linear Algebra, Network analysis

Foreign Language: English | TOEFL iBT score: 110/120 (R:30 L:30 W:27 S:23)

Teaching Assistantships

CS4201: Information Retrieval and Web Search Spring 2022

Instructor: Dr. Dwaipayan Roy *IISER Kolkata*

- TA to a class of more than 40 master's students

CS1101: Introduction to Computer Programming Autumn 2022

Instructor: Prof. Supratim Sengupta, Prof. Swastika Chatterjee *IISER Kolkata*

- Cleared doubts in lab sessions, Graded assignments, invigilated exams

CS1101: Introduction to Computer Programming Autumn 2021
Instructor: Prof. Supratim Sengupta, Dr. Dwaipayan Roy *IISER Kolkata*

- Held one-to-one tutorial sessions, Cleared doubts in lab sessions, Graded assignments

Training & Conferences

IndoML 2024 | The Fifth Indian Symposium on Machine Learning Dec 2024

- Winner of the Datathon@IndoML 2024 - national-level ML competition, to solve a industry-relevant problem, sponsored by NielsenIQ
- Received the Student Travel Grant to attend IndoML 2024

IBM Qiskit Global Summer School 2021 on Quantum Machine Learning Jul 2021

- Classical ML, Quantum Approximate Optimization Algorithm, Variational Algorithms, Quantum Kernels and Quantum SVMs, Quantum Boltzmann Machines, Quantum GANs

IBM Qiskit Global Summer School 2020 Quantum Computing and Quantum Hardware Jul 2020

- Quantum Circuits, Quantum measurements, Quantum Fourier Transform, Shor's Algorithm, Quantum Error Correction using Repetition Codes, Superconducting Qubits

International Conference on Quantum Information and Quantum Technology Jun – Jul 2019

- QIQT-2019 Conference at IISER Kolkata

Asia Pacific Conference and Workshop on Quantum Information Science Dec 2018

- APCWQIS-2018 Conference at IISER Kolkata

References

Dr. Dwaipayan Roy
Assistant Professor, Department of Computational and Data Sciences
Indian Institute of Science Education and Research Kolkata, India
Email: dwaipayan.roy[at]iiserkol.ac.in

Dr. Kripabandhu Ghosh
Assistant Professor and HoD, Department of Computational and Data Sciences
Indian Institute of Science Education and Research Kolkata, India
Email: kripaghosh[at]iiserkol.ac.in