Mainak Chatterjee

Junior Project Engineer, **IIT Kharagpur**
+91-7980045963 | chatterjee01m@gmail.com | Mainak-Chatterjee

Education

Maulana Abul Kalam Institute of Technology — *West Bengal, India*
B.Tech (Hons.) in Electronics and Communication Engineering — *Nov 2020 – June 2024*

- **CGPA :** 8.72/10
- **Affiliations :** Core Member of Literary Club TINT(LITWIS), Core Member of Coding Club TINT

Experience

Indian Institute of Technology Kharagpur — 02/25 – ongoing
Junior Project Engineer with Saptarshi Ghosh — *West Bengal, India*

Indian Institute of Technology Kharagpur — 07/24 – 01/25
Project Assistant with Saptarshi Ghosh — *West Bengal, India*

Indian Institute of Technology Kharagpur — 03/24 – 06/24
Research Intern with Debarati Sen — *West Bengal, India*

Indian Institute of Technology (Indian School of Mines) Dhanbad — 06/23 – 08/23
Summer Intern with Samrat Mukhopadhyay — *Jharkhand, India*

Suvidha Foundation — 10/21 – 11/21
Machine Learning Intern — *Remote | Hyderabad, India*

Publications

[1] **CrILDSum: A Novel Benchmark Dataset for Cross-lingual Summarization of Long Indian Legal Judgments**
Debtanu Dutta, Mainak Chatterjee, Rohit Sharma, Adrijit Goswami, Saptarshi Ghosh
[Under Review]

Featured Academic Projects and Collaborations

Novel Question-Answering Dataset for Legal Queries — 07/24 – Present
w/ Saptarshi Ghosh — *IIT Kharagpur*

- Designed a novel **Legal Question-Answering (QA) dataset** comprising **244** questions based on **Indian legal contexts**, **97** questions on **U.S. Copyright Law**, and **53** questions on **U.S. Constitutional Law**, formatted in JSON with Yes/No answers and detailed explanations.
- Implemented **LLM-based workflows** using the **Groq API**, **GPT API**, and **Ollama framework**, to inference 10 state-of-the-art proprietary and open-source models, including **GPT-3.5-Turbo**, **GPT-4o-mini**, **LLama-3**, **Gemma3**, **Mistral** and reasoning-optimized models like **DeepSeek-R1-Llama-distilled**
- Currently conducting comprehensive evaluations by **benchmarking LLM-generated outputs against Gold Standard references**

Cross-lingual Summarization of Long Indian Legal Judgments — 02/25 – 10/25
w/ Debtanu Dutta, Saptarshi Ghosh — *IIT Kharagpur*

- Developed **CrILDsum**: a novel dataset comprising of **50 Indian Supreme Court judgments** paired with their summaries in **four scheduled Indian languages**: **Bengali**, **Hindi**, **Marathi** and **Malayalam**
- Generated both **extractive** and **abstractive** summaries using a variety of **off-the-shelf** and **fine-tuned encoder-decoder** and **decoder-only** models, including **BART**, **T5**, **LexRank**, **Luhn**, **LLaMA-3.2**, **Gemma2**, **Mistral**, and **Saul**

* Employed **IndicTrans2** Neural Machine Translation (NMT) model to translate English summaries into the four target languages
* Conducted cross-lingual summarization using the **mT5** model to generate summaries directly in multiple languages from the original judgment texts
* Evaluated summarization quality using advanced metrics such **ROUGE-2**, **ROUGE-L** and **BERTScore**, providing in-depth insights into the effectiveness of different summarization approaches

Modern Digital Communication with Feed Forward Neural Networks 03/24 – 06/24
w/ Swastik Chakraborty, Debarati Sen *IIT Kharagpur*

* Developed a comprehensive digital communication system using neural networks to enhance BPSK and QPSK modulation techniques
* Implemented a Two Tap Rayleigh Fading Channel with Normal, ZF, and MMSE precoders to simulate real-world signal degradation
* Developed a deep neural network model for BPSK signal reception, achieving upto **99** percent accuracy

Advanced Greedy Algorithms for Sparsity-Aided Generalized Online Caching 06/23 – 08/23
w/ Samrat Mukhopadhyay *IIT (ISM) Dhanbad*

* Developed and implemented several online learning algorithms for maximizing the **Reward** and minimizing the **Regret**, contributing to the development of several feasible solutions of the problem
* Optimized algorithm code by meticulously **fine-tuning input parameters**, resulting in highly efficient solutions that demonstrated a deep understanding of algorithmic optimization techniques
* Plotted and visualized algorithmic outputs, specifically Reward and Regret, providing clear insights into their behavior and effectiveness

Automatic Detection of Plant Species using Deep Convolutional Networks 10/21 – 11/21
Suvidha Foundation

* Trained Machine Learning model from the data extracted from several Deep Learning Algorithms for **Image Recognition** and **Classification**
* Used VGG16 **CNN** model for feature extraction and MobileNet **CNN** models Image Classification and compared their testing as well as validation accuracy
* Trained Random Forest Classifier machine learning model using the data extracted from VGG16 for making the **best possible decision** regarding correct classification

Achievements

- **Finalist of Smart India Hackathon 2023 organized by the Government of India**

Featured Coursework

- **Mathematics:** MIT RES-6-012: Introduction to Probability, MIT OCW; MIT 18.06: Linear Algebra, MIT OCW; Numerical Methods (4th Sem., MAKAUT)
- **Machine Learning:** Machine Learning Specialization, Stanford University; Applied Optimization for Wireless, Machine Learning, Big-Data IIT-K (NPTEL); Deep Learning Fundamentals - Introduction to Neural Networks, Deeplizard; Neural Networks, 3Blue1Brown; Artificial Intelligence (8th Sem., MAKAUT)
- **Natural Language Processing:** CS224N: Natural Language Processing with Deep Learning, Stanford University; Finetuning Large Language Models, OpenAI; ChatGPT Promt Engineering for Developers, OpenAI
- **Computer Vision:** CS294-158 SP24: Deep Unsupervised Learning, UC Berkeley

Skills and Interests

Programming Languages: Python, Java, C/C++, MATLAB, LaTex, bash, Git
Developer Tools: Google Collab, Huggingface, Ollama, Kaggle, Spider, Jupyter Notebook, Visual Studio Code, Google Translate API, Groq API, Git
Libraries: PyTorch, TensorFlow, keras, Scikit-learn, NumPy, pandas, Matplotlib, seaborn transformers, SpaCy, indicnlp
Research Interests: Large Language Models, Vision Language Models, Alignment, Reasoning, Interpretability, LLM as Agents, Safety in LLMs