# Imon Bera

+1 267-466-7408 | ib385@drexel.edu | linkedin.com/in/imonbera13/ | github.com/invcble

## About me

I am pursuing **Master's** in **Computer Science** at **Drexel University** (exp. Grad in June, 2025) with specialization in **Artificial Intelligence** and **Machine Learning (AI/ML)**. I have a sound understanding of various **AI frameworks, cloud-based solutions** and **deployment methods**. I have strong knowledge in **AI automation, Python, Databases (SQL, noSQL, vectorDB)** and various cloud platforms (**GCP, AWS** and **Azure**). I have hands-on-experience of implementing **Deep Learning models** with variety of Machine Learning libraries like **TensorFlow, Pytorch, and Scikit-learn**. I am a quick learner and avid problem solver. I am seeking a challenging role that allows me to utilize my skills and continuously enhance my expertise.

## Education

**Drexel University** **Sep. 2023 – Present**
*Master of Science in Computer Science, GPA: 3.85* *Philadelphia, PA*

**Bengal Institute of Technology** **Jul. 2019 – Jun. 2023**
*Bachelor of Technology in Computer Science and Engineering, CGPA: 9.14* *Kolkata, WB*

*Coursework: Machine Learning, Deep Learning, Computer Vision, Data Structures & Algorithms, Software Design (Java), Full Stack, Agile Methodology, Object Oriented Programming, Operating Systems, Database, DevOps, CI/CD.*

## Technical Skills

**Languages**: Python, Java, Scala, SQL, C, C++, HTML, CSS
**Dev Tools**: Git, Matlab, Google Colaboratory, Visual Studio Code, PyCharm, IntelliJ, Eclipse, LaTeX, Spark, Tableau, Jira, Confluence, Swagger/OpenAPI, Groq, Ollama (LLM Tool), LM Studio
**Cloud Tools**: Artifact Registry, Vertex AI, Firestore, Terraform, Pulumi, Docker
**Cloud Platforms**: Google Cloud Platform, AWS, Azure
**Databases**: MySQL, PostgreSQL, Firebase, Qdrant, Pinecone, Milvus
**Libraries**: TensorFlow, Pytorch, Keras, Scikit-learn, Flask, pandas, NumPy, Matplotlib, OpenCV, PIL, CV2, VADER, Word2Vec, GloVe, FAISS, PyMongo, pyautogui, Tkinter, Customtkinter, Win32GUI, PyQt

## Experience

**AI Engineer Intern** **Jun 2024 - Dec 2024**
*Tapistro, Inc.* *San Francisco, CA*

- Designed an end-to-end LLM Agentic RAG framework with a periodic web scraper for signal accounts and parallel error reduction, integrating Function Calling, Code Interpreter, and a JSON extractor, and storing embeddings in Firebase's vector DB enabling automation and business processes optimization.
- Developed a fallback search engine pipeline integrating Bing Search API v7 and Google's Custom Search API, reducing 'no-answer' error rates by 35% for queries unsupported (yet to populate/live data) by Firestore's Knowledge Vector.
- Collaborated with core team to create AI-Infused RESTful API systems for account engagement to deliver automated persona enrichment, reducing manual work by 80%. Fine-tuned machine learning models to improve decision making and client interactions.
- Leveraged scalable, on-demand GPU computing on Hugging Face models through vLLM and llama.cpp, optimizing resource usage with quantized formats (GGUF, INT8, BF16) and achieving a 90% cost reduction compared to GPT-4o API pricing at the time.
- Managed cloud infrastructure with GCP tools including Kubernetes, Docker, Terraform, and Pulumi scripts, and automated testing and deployments via CI/CD pipelines using GCP Cloud Run and Cloud Artifacts.

**Research Assistant** **Jan. 2024 – Mar. 2024**
*LeBow College of Business* *Philadelphia, PA*

- Developed and deployed Windows automation macro scripts in Python utilizing libraries such as pyautogui, pygetwindow, subprocess, SimpleXMLRPCServer, and ctypes, reducing manual data handling in research workflows.
- Designed and optimized an automation control panel and survey browser using Python libraries Tkinter and PyQt, enhancing accessibility and reliability.
- Engineered Artemis log extraction automation solutions for streamlined data analysis, alongside an RPC macro for remote execution of Python scripts to improve research operations.
- Migrated legacy Perl scripts to Python, achieving a 70% improvement in runtime and significantly enhancing automation reliability.

## Projects

**Climate Risk Analysis Platform, Chubb Challenge Winner** **Apr. 2024**
*Hackathon: Philly Codefest* *Philadelphia, PA*

- Developed a sophisticated climate risk analysis platform, awarded first place in the Chubb Challenge at Philly Codefest, employing Poisson's distribution for statistical modeling of disaster event frequency and custom deep learning model with an autoencoder for accurate financial damage estimation.
- Implemented a user-friendly Flask web application with an interactive map interface, enabling users to visualize and interact with risk data effectively. The frontend was crafted using HTML, CSS, and JavaScript to ensure a responsive and engaging user experience.
- Deployed the platform on AWS EC2, using Gunicorn as the WSGI HTTP server and Nginx as the reverse proxy, to ensure scalable and reliable access, handling varying loads with optimal performance.

## Projects

**NoMoreBorder: A Window Managing Utility** **Feb. 2024 - Mar. 2024**
*Open Source GitHub Project (Over 2500 downloads)* *Philadelphia, PA*

- Developed NoMoreBorder, a portable pyinstaller executable utility enabling any windowed application to run in fullscreen borderless mode, enhancing multitasking experience without changing screen resolution.
- Implemented the backend and UI using Python, leveraging libraries such as customtkinter, pywin32, ctypes to streamline system window manipulation and future developements.

**Voc-Notes : A Free-to-use AI Notebook Tool** **Apr. 2024 – Jun. 2024**
*Drexel University* *Philadelphia, PA*

- Developed a standalone application to transform live lecture audio into structured notes using Google Cloud Speech-to-Text for real-time, accurate transcription.
- Used Apache Spark for data preprocessing to clean and segment text, preparing it for analysis and summarization by Large Language Models (LLMs).
- Orchestrated a summarization and note generation process with Mixtral-8x7b and Llama3-70b, leveraging LLMs to create concise and accessible class notes. Mixtral-8x7b handles large context chunks, while Llama3-70b combines the chunked results into coherent summaries.
- Managed data storage with PyMongo and MongoDB, ensuring efficient data retrieval and scalability.

**Alzheimer's Classification using OASIS Dataset** **Nov. 2023 – Dec. 2023**
*Drexel University* *Philadelphia, PA*

- Leveraged the OASIS Alzheimer's Detection Dataset, containing 80,000 brain MRI images of 461 patients, for classifying Alzheimer's progression based on Clinical Dementia Rating (CDR) values.
- Implemented Logistic Regression and LDA machine learning models from scratch, achieving 81% and 73% accuracy, respectively, for binary classification.
- Executed Multiclass Analysis using KNN and Naive Bayes models with image preprocessing (Blur, Canny edge), achieving a peak accuracy of 73% with KNN using Canny edge.
- Developed custom one vs. one Logistic Regression and LDA models, attaining 84% accuracy for Logistic Regression and 78% for LDA; subsequently ensembled them to achieve a final accuracy of 86% for Multi-Class Classification.

**Sentiment Analysis of Indian Political Tweets 2023** **Feb. 2023 – Jun. 2023**
*Bengal Institute of Technology* *Kolkata, WB*

- Created a robust Long Short-Term Memory model for sentiment analysis, accurately categorizing tweets into positive, neutral, or negative sentiments.
- Efficiently gathered tweets using Tweepy and merged with *Labeled 2021 Indian Political Dataset* from Kaggle, expanding the data source for analysis.
- Utilized Natural Language Processing techniques and Global Vector word embeddings for data preprocessing, ensuring the dataset's readiness for analysis.
- Validated sentiment predictions with a 96% accuracy, by integrating the VADER sentiment analysis, offering a comprehensive model accuracy comparison.

**Mono, Multi and Cross Lingual Audio Emotion analysis** **Dec. 2022 – Mar. 2023**
*Bengal Institute of Technology* *Kolkata, WB*

- Developed and deployed LSTM, Bidirectional LSTM, and GRU models for precise emotion recognition in audio samples, tuning their proficiency across diverse languages.
- Utilized EMODB, EMOVO, RAVDESS, TESS, and Polish audio datasets, ensuring comprehensive and multilingual training and testing for emotion classification.
- Applied Sklearn to generate classification reports and confusion matrices, gaining an in-depth understanding of emotional cues across different linguistic and cultural contexts.

## Activities

**Course Assistant**, College of Computing and Informatics **Apr. 2024 – Present**
**CCI Dean's Fellowship**, College of Computing and Informatics **Jul. 2023 – Present**
**Active Member**, Drexel IEEE **Oct. 2023 – Present**
**Coordinator**, BITS-2-BYTES, 2K22 **Sep. 2022**

## Certificates

**Reinforcement Learning**, Coursera by Google **March 2023**
**Google Data Analytics Professional Certification**, Coursera by Google **August 2022**