# Niyar R Barman

niyarrbarman.github.io/
+91 847-386-9166
barmanniyar@gmail.com

## Education

2021 - 2025 **National Institute of Technology Silchar, India**, *Bachelors of Technology in Electronics and Communication Engineering*, CGPA - 7.19/10

## Experience

May 2024 - July 2024, Jan 2025 - Mar 2025 **Research Intern**, *Institut de Recherche en Informatique de Toulouse*, France

- Conducted research at IRIT under Dr. Nicholas Asher, examining the influence of discourse structure and premise semantics on language model reasoning; investigated numerical reasoning failures stemming from tokenization methods and proposed a unified embedding strategy to improve models' comprehension of numerical values.
- Modified a GPT-style architecture to seamlessly integrate both text and numerical data processing. Trained this 125 M parameter model from scratch on a 800 M tokens dataset on a GPU cluster.

Jan 2023 - Present **Research Intern**, *Artificial Intelligence Institute of UofSC*, United States

- $CT^2$: Analyzed 100K data points with 15 language models to determine negative log curvature and perplexity values, performed a range of statistical tests, and developed task-specific pipelines.
- Visual Paraphrasing: Conducted an empirical analysis demonstrating the fragility and susceptibility of current image watermarking methods to visual paraphrasing techniques, highlighting their inability to withstand watermark removal. Investigated the effectiveness of visual paraphrasing in removing watermarks under two scenarios: with access to the original AI-generated prompt and when encountering potentially watermarked images in the wild

## Publications

EMNLP 2023 **Counter Turing Test $CT^2$: AI-Generated Text Detection is Not as Easy as You May Think – Introducing AI Detectability Index (2023)**, *Megha Chakraborty, S. M Towhidul Islam Tonmoy, S M Mehedi Zaman, Krish Sharma,* ***Niyar R Barman****, Chandan Gupta, Shreya Gautam, Tanay Kumar, Vinija Jain, Aman Chadha, Amit P. Sheth, & Amitava Das.*

**Honored with the Outstanding Paper Award in the Resources and Evaluation Track**

SIGDIAL 2025 **DIMSUM: Discourse in Mathematical Reasoning as a Supervision Module (2025)**, ***Niyar R Barman****, Krish Sharma, Nicholas Asher, Akshay Chaturvedi*

arXiV Preprint 2024 **The Brittleness of Image Watermarking Techniques: Investigating Visual Paraphrasing for De-Watermarking AI-Generated Images (2024)**, ***Niyar R Barman***, *Krish Sharma, Ashhar Aziz, Shashwat Bajpai, Shwetangshu Biswas, Aman Chadha, Vasu Sharma & Amitava Das.*

IEEE CICT 2023 **A Transformer-Based Approach to Automate Disease Prediction from Patient Descriptions (2023)**, ***Niyar R Barman***, *Krish Sharma & Ranjay Hazra.*

FIRE 2023 **Addressing Hate Speech: ATLANTIS for Efficient Hate Span Detection (2023)**, ***Niyar R Barman***, *Krish Sharma, Yashraj Poddar, Advaitha Vetagiri & Partha Pakray.*

## Open Source Contributions

**pytorch/examples**: Implemented Vision Transformer using only native PyTorch libraries, trained on the CIFAR-10 dataset

## Technical Skills

**Programming languages**: Python, JavaScript, C, C++, Bash

**Libraries & Frameworks**: PyTorch, TensorFlow, Keras, Transformers, Scikit, Django, FastAPI, ReactJS, NextJS, React Native

**Tools**: Git, Docker

**Miscellaneous**: AWS, LaTeX, Linux, Windows, Blender

## Key Projects

**swiftGPT**

- Developed SwiftGPT, a natural language generation model based on the GPT-2 architecture, specifically trained on an artist's songs dataset to emulate their writing style.
- Implemented a custom BigramLanguageModel using PyTorch, incorporating multi-head self-attention and feed-forward layers. Conducted 5000 training iterations with an AdamW optimizer and monitored train and validation loss to fine-tune the model for creative songwriting.

**Viper**

- Applied Reinforcement Learning and Deep Q Learning principles to develop an AI capable of playing the classic game Snake.
- Implemented the project from scratch using Pygame for the game environment and PyTorch for neural network-based decision-making.

**MapSnap**

- Developed MapSnap, a web application that leveraged the Segformer model for real-time semantic segmentation of satellite images. The application generated masked outputs that highlighted affected areas
- Successfully fine-tuned the Segformer model using the IARAI landslide4sense dataset, resulting in accurate and efficient identification of landslide-prone areas

**DiagnoAI**

- Fine-tuned the BERT transformer model to detect 24 common diseases and generate probability scores for the top 3 diseases.
- Curated and labeled a comprehensive dataset of 1200 symptom descriptions associated with 24 diseases, which was used for training and validating the model.

## Awards

Aug 2023 **2**$^{nd}$ in HateNorm'23, a component of HASOC 2023, by excelling in Task 3, which involved the identification of hateful spans within sentences already categorized as hateful.

Apr 2023 **1**$^{st}$ in Neurathon'23, the annual machine learning hackathon organized by the ML Club at NIT Silchar.

Jul 2022 **4**$^{th}$ in Un-Flood Assam'22 hackathon organized by the Ministry of Electronics & Information Technology.

## Positions of Responsibility

Dec 2023 **EMNLP Presenter**: Presented Counter Turing Test $CT^2$ (Winner of Outstanding Paper award in the Resources and Evaluation Track) at EMNLP Main Poster Presentation 2023.

Jun 2023 - Present **Teaching Assistant/Moderator**: Conducted introductory classes on Python and Machine Learning for freshmen and sophomores as part of the Machine Learning Club at NIT Silchar.