Ananya Malik

LinkedIn: https://www.linkedin.com/in/ananyamalik/
Github: https://ananyamalik.github.io/

Email: ananyamalik1999@gmail.com
Mobile: +1 470 753 7363
Atlanta, GA

Education

- **Georgia Institute of Technology** — Atlanta, USA
  Masters of Science - Computer Science- ML specialisation; GPA: 4.0 — Aug 2021 - Dec 2022
  Courses: Machine Learning, Deep Learning, Web Search, Data Management with ML, Deep Learning for Text, Data Visualisation

- **Dwarkadas J Sanghvi College of Engineering** — Mumbai, IN
  Bachelors of Engineering - Computer Engineering; GPA: 9.79 — Aug 2017 - Jun 2021
  Courses: Analysis of Algorithms, Machine Learning, Artificial Intelligence Soft Computing, Big Data, Software Development, OS, HMI

Experience

- **Amazon** — Seattle, USA
  Software Development Engineer and Intern — Aug 2023 - Present; May 2022-Aug 2023
  - **Build Data Debugger**: To identify and highlight causally redundant attributes in the data pipeline
  - **Compare Text Based Data for similarity**: Developed an algorithm to compare the attributes in a JSON data by the text values using embedding that lead to a reduction in false positives by 17 %
  - **Created Tool**: Implemented the backend on AWS Lambda, pulling data from S3 buckets and frontend on React to visualise the lineage of the data by type of data loss and attribute

- **CLAWS Lab, Georgia Tech** — Atlanta, USA
  Research Engineer — Aug 2022 - Aug 2023
  - **Misinformation Platform**: Worked on building a platform to identify and interact with tweets flagged as misinformation, assisted data collection and creating the database
  - **Counter Misinformation**: Working on finding the motivation, classifying counter misinformation on Twitter using LLMs, HCI

- **Tata Communications Limited** — Pune, IN
  AI Research and Project Trainee — Jun 2020 - Aug 2020
  - **Research on Fraud Detection Models**: Researched on techniques to interpret the fraud detection model using Call Data Records.Experimented and Reported the findings on ExplainableAI and interpretability of the data
  - **Explain Fraudulent Entries in Data**: Implement LIME to understand the fraudulent entries, increasing efficiency by 30%
  - **Analysis**: Analysed the Call Data Records to detect patterns incall location, timing, duration using K-Means analysis, thresholding in python

Research and Publications

- **Correcting Misinformation Online**: Researched to understand the semantics such as demographics, and motivation behind countering misinformation on social media and suggest ways to increase the frequency of countering. Collected observational data from Prolific and analysed to see whether demographics and topical relevance played a role in countering of misinformation. Observed that people are likely to assume a public role while countering than assume a private responsibility.

  Paper: Working Paper Link

- **Evaluating Large Language Models through Gender and Racial Stereotypes**: Studied gender and racial bias in a professional context across language models: GPT, Flan, Claude. Built a classifier to classify the responses generated by the LLMs into a gender out of Male, Female and Neutral. Created a dataset classifying 99 professions into one gender, and evaluated the model's biases as a shift from the baseline. Recreated a similar set-up for evaluating racial bias, where we prompted the LLMs to generate descriptions of individuals with the same profession against different races and noted the existence of similarly worded qualities for the same race.

  Paper: Preprint Link

- **Syncphonic: Context-Aware Music Generation**: Developed a model to assign a song to a sequence of images. Trained a word2vec model to detect the scene of the images. Trained a CLIP model to generate captions from the images with context, to identify the mood and sentiment using a BiLSTM classifier. Depending upon the combination of the scene and the sentiment/mood, mapped the image sequence to a classified music category. Conducted qualitative analysis able to achieve 90% agreement on scene and music combination.Presentation

- **Generation of a visual storyline from a single sentence**: Developed a multi-modal model to generate a story depending upon a starting prompt using GPT2, and then trained a network of StackGANS to generate images to visualise the each sentence of the story. Used the Pororo SV dataset, to capture image sequences from the to train our model using the word description. To maintain the context we trained LSTMs on the word embeddings of each sentence. Presentation

  Paper: Successive Image Generation from a Single Sentence, Amogh Parab, Ananya Malik, Arish Damania, Arnav Parekhji, Pranit Bari, ITM Web Conf. 40 03017 (2021), DOI: 10.1051/itmconf/20214003017.

- **Analysing Dog Whistles in Social Media**: Researched to identify patterns and conducted a temporal analysis related to the usage of Dog Whistles in Social Media such as Twitter and Reddit. We analysed the usage of the dog whistles over time, analysed thematic patterns as clusters, analysed the sentiment, toxicity and political content associated with the dog whistles. We annotated data to create a predictor model using BERT to identify whether a social media post contained a dog whistle with the intent to spread implicit hate or not, achieving an accuracy of 0.92 Poster

  Paper: Final Paper for CS 6471

- **Self Supervised Learning for MRI Reconstruction**: Aim of this project was to leverage undersampled MRI scans and reconstruct them to produce high-quality scans, with limited or self-supervision. Developed a model agnostic framework to resolve this task. We used different masking strategies to created double undersampled images and a reconstruction network containing a 4 layer U Net network, to produce images that match the quality of fully supervised methods. Our combination of using a medical mask and U-Net model, with only 0.20 data, was able to achieve comparable performance with the state of the art. Presentation

- **Impact Analysis of COVID-19 News on the Economy**: Conducted an analysis to identify and predict the correlation between news articles and the global economy and markets. Conducted textual analysis of news headlines and temporally followed their correlation to the stock market. Created a model to predict and visualise the drop or rise in the stock market depending upon the news headline.

  Chapter: A.Malik, Y. Javeri, M. Shah, R. Mangrulkar, 'Impact Analysis of Covid 19 News Headlines on Global Economy', Cyber-Physical Systems for COVID-19, Elsevier

Publications

- **Paper**: Ananya Malik. (2023). Evaluating Large Language Models through Gender and Racial Stereotypes.
- **Article**: Intent to Hate: An study on the intent behind hate generated on Twitter against the Asian Community during the COVID-19 pandemic.
- **Article**: Of Multimodal Comprehension and Ignorance.
- **Paper**: Successive Image Generation from a Single Sentence, Amogh Parab, Ananya Malik, Arish Damania, Arnav Parekhji, Pranit Bari, ITM Web Conf. 40 03017 (2021), DOI: 10.1051/itmconf/20214003017.
- **Chapter**: A.Malik, Y. Javeri, M. Shah, R. Mangrulkar, 'Impact Analysis of Covid 19 News Headlines on Global Economy', Cyber-Physical Systems for COVID-19, Elsevier.
- **Paper**: Malik A. Survey paper on applications of generative adversarial networks in the field of social media. Int J Comput Appl (IJCA). 2020;175(20):13–18. doi:10.5120/ijca2020920728

TEACHING AND MENTORSHIP

- **Graduate TA: CS 3600 Intro to AI** — Atlanta, USA
  Prof James Rehg, Prof Mark Riedl; Held office hours, recitations, review sessions,graded papers Jan 2022 - Dec 2022

- **DJ Unicode** — Mumbai, IN
  Leading a 130+ member development team working on full-stack projects for colleges,non-profitsJul 2018 - May 2021

- **TA: Machine Learning Summer School** — Mumbai, IN
  TA for Machine Learning School, held recitations and office hours — Jul 2018 - May 2021

Skills Summary

- **Languages**: Python, C, C++, JavaScript, SQL, JAVA
- **Frameworks**: Scikit, NLTK, SpaCy, TensorFlow, Keras, Django, Flask, NodeJS, REST API
- **Tools and Areas**: Flask, Django, ReactJS, GIT, PostgreSQL, MySQL, SQLite, AWS Lambda, AWS Athena, Linux, Web, Windows, Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, GANs