SARTHAK ARORA

sarthakvarora@gmail.com | P: +91-8081880257 | sarthakvarora.github.io

RESEARCH INTERESTS

Computer Vision, Domain Adaptation, Reinforcement Learning, Climate Science & Modelling, Remote Sensing

EDUCATION

PLAKSHA UNIVERSITY — Chandigarh, India
Post Graduate Diploma — 2021 – 2022
Major in Computer Science
Cumulative GPA: 3.88/4.0; Department Rank 2
Relevant Coursework: Machine Learning; Computer Vision; Reinforcement Learning; GraphML

RAMJAS COLLEGE, DELHI UNIVERSITY — New Delhi, India
Bachelor of Science — 2015 – 2018
Major in Statistics; Minors in Economics
Cumulative GPA: 7.176/10.0; First Division
Relevant Coursework: Actuarial Sciences; Time-Series Analysis; Statistical Modelling; Hypothesis Testing

PUBLICATIONS

Conferences & Journals

Queer In AI: A Case Study in Community-Led Participatory AI - FAccT'23 **Won best paper* 🏆
Organizers of Queer in AI
pdf | abstract

Queer in AI - ACM'22
Hetvi Jethwani, Arjun Subramonian, William Agnew, MaryLena Bleile, Sarthak Arora, Maria Ryskina, Jeffrey Xiong
pdf | abstract | cite

RESEARCH EXPERIENCE

WUUII — Remote
ML Engineer — Jun 2024 – Present
- Using Segment Anything on street view imagery and translated the identified features onto an aerial map for improved Wildfire Risk Assessment.
- Utilizing paired training on street view and aerial data to create an invertible model to identify features across modalities with enhanced geolocation using a Vision Transformer Model.
- *Keywords:* Computer Vision, Transformers, Deep Learning, Wildfires

COLLABORATIVE EARTH — Remote
ML Engineer — Jan 2023 – Present
- Using Machine Learning and GEE to identify and cluster different "Naalas" for the Ganges river from Satellite Images.
- Using GIS Tools and datasets to create a comprehensive site suitability map for Coastal Forests Wetlands Lab.
- Conducted fieldwork around the Ganges River which involved scoping the area of research and meeting stakeholders.
- Conducting research for publications around "Naala" Identification, Taxonomy and Remediation.
- Led fundraising efforts for the Ganges Lab, reached the final round for the Frontier CDR Prepurchase 2024.
- *Keywords:* GIS, Deep Learning, River Remediation

UNIVERSITY OF CALIFORNIA, BERKELEY Remote
Research Scientist Sep 2022 – Present
- Conducted a time-series Anomaly Detection analysis of Power Plant Emissions in San Diego to find areas of Climate policy improvements.
- Analyzing Regressors for a Wildfire Survey conducted by MWPA to cater to the needs of people in Marin County.
- *Keywords:* Time Series Analysis, Anomaly Detection, Power Plant Emissions, Wildfires

UNIVERSITY OF CALIFORNIA, BERKELEY Berkeley, CA
Research Scholar Apr 2022 – Jul 2022
- Conducting Research to develop a wildfire risk assessment tool under the mentorship of Dr. Thomas Azwell.
- Working on Remote Sensing and GEE datasets for the state of California to predict wildfires with an accuracy of 82%.
- Gaining valuable insights in applying Data Preparation, Machine Learning, Model Deployment and GIS mapping.
- *Keywords:* Wildfires, Remote Sensing, Machine Learning

AI FOR GOOD FOUNDATION Remote
ML Engineer Feb 2022 – Present
- Working on the Climate SDG project, which is backed by Microsoft and the UN Development Program.
- Creating a Graph based Visualisation for Climate Entities using Named Entity Recognition and Neo4J.
- Cleaning Scraped data using LDA Model for topic modeling, leading to a Recall value of 83.5% in noise detection.
- *Keywords:* NLP, GraphML, Climate

QUEER IN AI Remote
Organizer Jan 2022 – Present
- Program Chair for the Queer in AI Workshop at NeurIPS for the years 2022-2024.
- Conducting and publishing research to raise awareness of queer issues in the general AI community, received the best paper award at FAccT, 2023.
- Working on the Graduate Application Aid program, an initiative aimed at supporting queer applicants through funding and mentorship, which has led to scholarships worth $130,000 awarded to 130+ applicants across 30 different countries.
- *Keywords:* Ethics, Bias, AI Harms

INDUSTRY EXPERIENCE

PREDLI AB Remote
ML Engineer Sep 2023 – Feb 2024, May - June 2024
- Using OpenAI, ChromaDB and Langchain to generate a retriever-based DBQA system for a Climate consultancy client.
- Creating a Quiz Generator using LLMs from Transcripts with Topic Clustering and Evaluation using SerpAPI.
- Fine-tuning LLM Performance on a Tweet generation system using a RLHF model based on Tweet engagement.
- Created a Machine Learning Model trained on Geophysical Layers such as Bedrock, Magnetism, Resistivity, Gravity and Satellite Imagery for Mineral Site Prediction in Sweden with an accuracy of 89%.

INDIFI TECHNOLOGIES Gurgaon, India
Data Scientist Aug 2022 – Jul 2023
- Created a model which predicts Default Risk with 96% precision by analyzing CIBIL Trend using LSTM Model.
- Used Fuzzy Logic to match addresses for verification, leading to a reduction in manual checks by ~4 hours daily.

DISNEY+ HOTSTAR   Mumbai, India
Senior Executive Analyst   Feb 2021 – Aug 2021

- Reported daily figures for projects such as IPL, Burbank for Star and Disney. Reducing TAT by ~3 hours in the process.
- Designed a Health Score metric to predict user churn by modeling variables, leading to a rise in conversion by 1.5 pp.
- Analyzed campaigns over engagement, affluence, affinity on Conversion and CTR%, leading to 0.5 pp rise in renewals.

PROJECTS

- Created a real-time face and facial-features detector on live videos using **HAAR** and **CNN.**
- Created a Machine Learning model for Drug Discovery using **GANs**.
- Created an Image Classification model using transfer learning on **VGG-16**.
- Co-developed a business plan that models physiological indicators and provides actionable Mental Health insights.
- Predicted future order dates and SKUs purchased to create a Recommendation Engine for an e-commerce platform.

TALKS AND COURSES

- Delivered a talk on "South Asian Feminism/s in the 21st Century" at **ICML 2022**.
- Completed the Hilary 2024 School of Climate Change course offered by **Oxford Climate Society**.
- Completed the **Climate Change AI** Summer School 2024.
- Co-wrote a Literature Survey paper on **Sarcasm Detection in NLP** and presented it at **NAACL 2024.**
- Moderated a Panel at **NeurIPS 2022** on **Immigration and Queerness** as a part of the Queer in AI Workshop.

TEACHING

- Taught subjects such as ML, Data Science, Research, and Leadership to students based out of Africa with **Belong.**
- **Teaching Assistant** for the Economics Courses at Ramjas College, Delhi University.

ACHIEVEMENTS

- Received a Best Paper award at FAccT 2023 for the paper [Queer In AI: A Case Study in Community-Led Participatory AI](#).
- Silver Medalist for academic excellence at Plaksha University, TLP Batch of 2022.
- Winning Team for the Plaksha Entrepreneurial Challenge Lab where we pitched a Mental Health startup.

ADDITIONAL INFORMATION

Programming Languages	Python, R, SQL, Shell Script, Cipher
Machine Learning	TensorFlow, PyTorch, Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn, NLP, CV, RL, Graphs
Others	A/B Testing, Algorithms, AWS, Beautiful Soup, CSS, Data Structures, ETL, Excel, HTML, GIS, Google Analytics, Google Earth Engine, Product Management, Prototyping, Selenium, Spark, Tableau
Actuarial Papers Cleared	CT1 (Financial Mathematics), CT3 (Probability and Mathematical Statistics), CT4 (Models), CT5 (Contingencies), CT6 (Statistical Methods)