

Aatman Vaidya


 [aatmanvaidya.github.io](https://github.com/aatmanvaidya)  aatmanvaidya@gmail.com  github.com/aatmanvaidya  [Google Scholar](#)

Education

Ahmedabad University, India
B.Tech in Computer Science & Engineering


Aug 2019 – Jul 2023

Experience

Tattle Civic Tech, India 







Aug 2023 – Present

Data Engineer and Researcher







- > I help build citizen centric **open-source** tools and datasets to understand and respond to **online harms and abuse**.
- > I lead the development of **Feluda**, a configurable engine for analyzing multilingual and multimodal content. Feluda helps fact-checkers and researchers in combating misinformation. 
- > I build data pipelines to cluster large amount's of Audio and Video data into social media thematic labels like politics, humor, memes, devotional content etc. This work helps fact-checker's better understand, vizualize and analyze the content received on their helpline.
- > Worked with **MLCommons** to develop an **AI safety benchmark dataset** for hate and sex related crimes. As a part of this project, I also conducted a **survey** on evaluating Indic LLM's for natural language tasks and online harms.

Publications

S=In Submission, C=Conference, W=Workshop, P=Poster/Demo, J=Journal, R=Report

- S.2 **Strategies to Mitigate Spread of Toxicity on a Social Network**
Aatman Vaidya, Harsh Bhagat, Seema Nagar, Amit Nanavati
[In preparation]
- S.1 **Analysis of Indic Language Capabilities in LLMs** 
Aatman Vaidya, Tarunima Prabhakar, Denny George, Swair Shah
[In Submission] [ArXiv]
- R.1 **AI Safety Benchmark Dataset for Hate and Sex Related Crimes in Hindi** 
Report Submitted to **MLCommons** and work included in the **AILuminate** Safety Benchmark Dataset
- W.1 **The Uli Dataset: An Exercise in Experience Led Annotation of oGBV** 
Arnav Arora, Maha Jinadoss, Cheshta Arora, Denny George, Brindaalakshmi...Aatman Vaidya, Tarunima Prabhakar
Workshop on Online Abuse and Harms at North American Chapter of the Association for Computational Linguistics
★ **Outstanding Paper Award** [NAACL'24]
- C.3 **Analysing the Spread of Toxicity on Twitter** 
Aatman Vaidya, Seema Nagar, Amit Nanavati
ACM International Conference on Data Science and Management of Data [CODS-COMAD'24]
- C.2 **Forecasting the Spread of Toxicity on Twitter** 
Aatman Vaidya, Seema Nagar, Amit Nanavati
IEEE International Conference on Cognitive Machine Intelligence [IEEE CogMI'23]
- C.1 **Overview of the 2023 ICON Shared Task on Gendered Abuse Detection in Indic Languages** 
Aatman Vaidya, Arnav Arora, Aditya Joshi, Tarunima Prabhakar
International Conference on Natural Language Processing [ICON'23]

Work Mentions and Media Coverage

- Center for Democracy & Technology** on “Moderating Tamil Content on Social Media”  June 2025
- FOSS United** on “Meet The Maintainers, 31 days 31 FOSS Maintainers from India”  May 2025
- World Economic Forum** on “Year of elections: Lessons from India’s fight against AI-generated misinformation”  Aug 2024
- NY Times** on “A Small Army Combating a Flood of Deepfakes in India’s Election”  Jun 2024
- The Nieman Lab** on “Indian journalists are on the frontline in the fight against election deepfakes”  May 2024
- Digital Public Goods** on “Safeguarding Information Integrity during Elections with Digital Public Goods”  Apr 2024

Talks

Tech4Dev AI Cohort Program on “Safety by Design in AI for Social Good Applications” 🎧	Remote, Jul 2025
MisinfoCon India’25 on “A Look at Open-Source Deepfake Detection” 🎧🎥	Bengaluru, Mar 2025
AI for Global Development by Agency Fund on “Evaluating Indic Language Performance in LLMs” 🎧	Bengaluru, Mar 2025
Ahmedabad University SNA Guest Lecture on “Modelling Hate Speech on a Social Network” 🎧	Ahmedabad, Apr 2024
ACM Winter School on Network Science on “Building NLP classifiers to detect Hate Speech”	Ahmedabad, Dec 2023
DEF India Digital Citizen Summit on “Tools to respond to Online Gender Based Violence” 🎥	Remote, Apr 2024

Featured Academic Projects and Collaborations

Modelling Spread of Hate Speech w/ <i>Amit Nanavati, Seema Nagar</i>	Aug’22 - Present
<ul style="list-style-type: none">> Worked on a novel model to capture the spread of hate speech on Twitter. Our model is based on user behaviour and captures two important factors: a) toxicity exists as a spectrum, i.e. hate is not binary, and b) toxicity is not conserved in a network.> An in-depth empirical analysis led us to find users change behaviour with time and is impacted by its neighborhood. Developing a model that captures this finer phenomenon gives insights into creating interventions to mitigate hate speech. [In Submission]	
Tracking Online Gender Based Abuse at <i>Tattle Civic Tech</i>	Jul’23 - Present
<ul style="list-style-type: none">> Created a YouTube Dataset to track gender-based violence using a crowd-sourced slur list by <u>Gender Right</u> Researchers and Activists. String matching algorithms and NLP models were used to analyse text in Indic languages. [🎧]> Help build Uli, a browser extension to redact slurs and abusive content.> Improved NLP systems to understand coded language (dog-whistle or double-meaning words).	
Financial Scams on Social Media w/ <i>Kiran Garimella</i>	Jun’25 - Present
<ul style="list-style-type: none">> Working on mapping financial scams in India. How do they move cross-platform (from Instagram to Telegram etc), what type of content is getting pushed?	

Awards and Honors

Outstanding Paper Award @ Workshop of Online Harms and Abuse NAACL 2024
Bosch Future Mobility Challenge Semi-Finalist [🏆] Programmed and Engineered an autonomous driving car to navigate a miniature city. One of the 24 teams out of 118, and only team to represent India for the finals in Cluj, Romania. [🎧][🎥]

Teaching and Academic Service

Reviewer: COLM’25, WOAHL ACL’25, ACL Rolling Review’24	
Discrete Mathematics (MAT 101) , Ahmedabad University. <i>Teaching Assistant</i>	Winter 2022
Center for Learning and Empowerment (NGO) , Volunteer Teacher. 🎧	Jul 2023 - Apr 2024
> Taught 20+ secondary school kids mathematics for an entire academic year, at a tribal village in Jharkhand, India.	
Programming Club , Ahmedabad University. <i>Content Head and Event Organiser</i> 🎧	July 2020 - May 2022
Reading Club , Ahmedabad University. <i>Organiser</i> 🎧	Dec 2021 - May 2022

Skills

Research Areas	Online Safety, Evaluation, Language Models & their Societal Impact, Content Moderation
Languages	Python, JavaScript, Elixir C/C++, Kotlin, GraphQL
Frameworks	PyTorch, Tensorflow, Nodejs, OpenCV, networkx, Elasticsearch, Phoenix
Technologies	MySQL, PostgreSQL, Docker, AWS (S3, EKS, EC2), Kubernetes, CI/CD, Android Studio