

# AINDRIYA BARUA

MACHINE LEARNING ENGINEER | ARTIST

They/Them

 [barua.aindriya@gmail.com](mailto:barua.aindriya@gmail.com)  
 [www.shhorai.com](http://www.shhorai.com)  
 [aindriya-barua](https://www.linkedin.com/company/aindriya-barua)  
 [AindriyaBarua](https://medium.com/@AindriyaBarua)  
 [barua.aindriya](https://github.com/barua.aindriya)  
 8787768696

## EXPERIENCE

---

### ShhorAI

Founder, Machine Learning Engineer

Oct, 2021, Jul 2022 – Present

[Website](#) | [GitHub Repo](#)

- Developed a Reddit bot that detects Hindi-English code-mixed hate speech in **text**/comments, **images**/memes, **videos**/reel-type content in real-time, replies with X warnings, followed by permanently banning the abuser. This solution can be extended to other social media platforms.
- Created **comment scrapers** for Instagram, Twitter, and Reddit: [GitHub Repo](#).
- Assembled and led a team of 45 activists and volunteers from marginalized communities for **data tagging**.
- Created a dataset with hate, and non-hate tags. The hate speech identified is further labeled as queerphobic, communal, sexist, political, casteist, and others.
- Conducted research on the dataset to identify **technology-facilitated polarization trends in India** and the intersectionality of marginalization.
- Developed a novel statistical method to extract a repository of abusive Hinglish vocabulary and stop-words.
- Trained a **classification model using BERT** with emoji analysis. Integrated it with **Reddit API**.
- Added technique for **doxxing detection**.

### AWARDS AND ACCOLADES FOR SHHORAI

Aug, 2022 – Dec 2022

- **Second runner-up position:** [International Hackathon organized by United Nations Population Fund](#)
  - › On "Addressing technology-facilitated gender-based violence and improving GBV prevention and response for persons with disabilities".
  - › Competed with 8500 teams. Other winners included start-ups from Canada and Nigeria.
- **Honored by USA Consulate General, Judith Ravin, and Political/Economic Chief, USA, Virsa Perkins**
  - › Invited to a detailed discussion on AI and social justice for my experience with grassroots-level activism through technology and art.

### MEDIA CITATIONS FOR SHHORAI

- **UNFPA Press Release: Global youth innovators crack the code on gender-based violence: UNFPA hackathon: The Print, ANI News, TOI, PTI News, Techgig, TechPanda, Dainik Sambad, NorthEast Colors**  
Links: [The Print](#), [ANI](#), [Times of India](#), [Press Trust of India](#), [Techgig](#), [TechPanda](#), [Dainik Sambad](#), [NorthEast Colors](#)
- **Interview: Times of India (All India Printed Media, Sunday Special)**  
Link: [Hinglish and Toxic Typos – How hate spreads on Reddit](#)
- **Interview: Analytics India Mag**  
Link: [This 24-Year-Old Built An Anti-Hate-India Reddit Bot](#)

---

### Bachchao Project

Technical Fellow: Without Fear Techno-feminist Fellowship

Dec 2022 – Jan 2023

[GitHub Repo](#)

- Collaborated with a lawyer to create a feature for ShhorAI, which provides Legal, Technical, and Mental Health Resources for the victims of online hate.
  - Designed a **Flask web app** for victims to submit a received comment, **classify** it as racist, communal, sexist, etc., display the **IPC laws** it violates, and further help them take legal action.
-

## Needl.ai

Machine Learning Engineer

Feb 2022 – Jul 2022

- Enhanced **large PDF document scrapping** techniques for **custom entity recognition** such as the names of sectors mentioned, company names, different forms of domain-specific tabled data, etc.
  - The extracted information was used for **feature engineering** of a **multi-class classifier** for 5 types of finance-related documents – annual, company, sector, general reports, and conference call transcripts.
  - Based on the classification, further extracted specific information such as the broker's name, broker's recommendation for stocks, people mentioned, rival companies mentioned, stock selling price, target price, etc.
  - If a document was classified as a conference call transcript, found the speakers in the conference and did a **sentiment analysis** of each speech with **FinBERT**.
  - Developed stable methods to regularly update to newer versions of models, choosing documents to analyze and check the performance of the model on real-world data.
- 

## Oracle (Cerner)

Software Engineer | Intern

Jan 2020 – Jan 2022

- Project Touchstone: Enhanced features on Cerner's auto-test **Java RCP** app to build, auto-execute, store and reuse workflows on any desktop or web app using **OCR** with multi-lingual support.
  - Auto Test Selection System: Created a system with Python to detect code changes on the entire Cerner codebase, which includes scripts in **Java, CPP and Python**, and find the exact tests impacted, so that they can be automatically regression-tested using Touchstone.
- 

## CEN Labs, Amrita University

Research Intern

Jun 2018 – Dec 2019

- **Advisor:** [Dr. Soman K.P.](#)
  - **Contextual vs Non-contextual Embedding for Hindi Named Entity Recognition:**  
[GitHub Repo](#)
    - › Non-contextual: **Word2Vec, FastText**
    - › Contextual: **BERT, RoBERTa, ELECTRA, CamemBERT, Distil-BERT, XLM-RoBERTa, Distil-BERT, XLM-RoBERTa**
  - › Analyzed embedding models, used the best one to build a **Flask** web app for Hindi NER and **data collection from user feedback**, deployed on **AWS**.
  - **Resource-Poor Bangla POS tagging**  
[GitHub Repo](#)  
Compared performances of **RNN, CNN, LSTM, GRU, and Bi-directional counterparts**
- 

## Amrita Center for Wireless Networks & Applications

Research Intern

Feb 2019 – Jul 2019

- **Landslide detection from tweets**  
[GitHub Repo](#)
  - › Applied RNN to detect Landslide-related tweets.
  - › Government of India funded, used by Kerala and Sikkim State Disaster Management Authority.

## PERSONAL PROJECTS

---

- **Games for monitoring and early detection of sexual abuse in children (WIP)** Sept 2022 – Present
  - › A child-friendly chatbot that uses games to educate children on recognizing and addressing sexual, physical, and emotional abuse, while also supporting their daily mental health needs.
  - › Short-listed as 1 of the Top-8 finalists of the International United Nations Population Fund Hackathon.
  - › Designed with guidance from child psychiatrists, trauma-informed child therapists, and disability rights activists from Ummeed NGO, ChildLine India, Nemours KidsHealth, RisingFlame NGO, and UNFPA.
  - › To be merged with the start-up [TottoLearning](#).

- **Restaurant Chatbot**

Oct 2021

[GitHub Repo](#) | [Blog](#)

- › Chatbot with Natural Language Understanding, intent classification, and response generation
- › Classifies and responds to 18 intents: Greet, inquiries about the menu, offers available, vegetarian options, vegan options available, recipes of dishes, COVID protocols and hygiene, opening hours, order suggestions, contact information, address of the restaurant, availability of tables, table booking, some other general messages, identifies positive and negative feedback, responds accordingly.

- **NewsInShort-Text Summarization**

Aug 2021

[GitHub Repo](#)

- › Automated document merging and extractive summarization to take in a search query and provide a crisp version of the news article from over 5 reputed sources. Supported in various languages with FastText.

## PUBLICATIONS

---

- Barua, Aindriya, et al. "[Analysis of Contextual and Non-contextual Word Embedding Models for Hindi NER with Web Application for Data Collection.](#)" International Advanced Computing Conference. Springer, Singapore, 2020.

## SKILLS

---

- **Programming Skills:** Python (scikit-learn, numpy, pandas, matplotlib, tsne, tika, PyMuPDF etc), Java (Maven, Java RCP)
- **Classical Machine Learning:** Supervised, Unsupervised Learning, Gaussian NB, Adaboost, Random Forest, DecisionTree Classifier, SVN.
- **Deep Learning:** Keras, TensorFlow, Scikit-learn, RNN, LSTM, CNN.
- **Natural Language Processing:** NLTK, Gensim Word2Vec, FastText, Transformers, BERT, RoBERTa, ELECTRA, CamemBERT, Distil-BERT, XLM-RoBERTa, FinBERT.
- **Database Management:** Oracle SQL, MongoDB.

## EDUCATION

---

### Amrita Vishwa Vidyapeetham, Kerala

2016 - 2020

- Bachelor of Computer Science and Engineering.
- **CGPA: 9.05/10** | First Class with Distinction (Top 5%)
- **First Rank** in academics in First Year.

## ADDITIONAL INFORMATION

---

- **Featured Artist in Adivasi Lives Matter (Indigenous Media) Yearly Calendar** 2022
  - **Interviewed and featured for making socially aware art in national media outlets** 2018 - 2022
    - › Buzzfeed India: [27 Indian Artists Who Are Smashing Stereotypes With Their Fantastic Illustrations](#)
    - › The Quint: [Art Movement against Patriarchy and Colourism: Women Voices Interview](#)
    - › The Indian Express, News9Live, etc.
  - **Art Cited in Journal Paper Ecofeminism and Climate Change** Nov 2021  
By Art Historian Dr. Roma S [Link: Our Identity Remains Marked](#)
  - **Creative Head of Period Society NGO, Tripura Chapter:** Spreading menstrual hygiene awareness to underprivileged women in remote areas of Tripura. 2020 - Present
  - **Guest of Honour, Bhagat Singh Memorial Lecture, Tata Institute of Social Sciences, Mumbai** Sep 2020
  - **Art Exhibitions:** 2019 - 2023
    - › Bahujan Arts Festival, TISS Mumbai
    - › Ambar Art Exhibition, IIT Kharagpur
    - › Unmute Art, Vidyut National Multifest
    - › Woman Up Exhibition, Amrita
  - **VDM: INDIA ON THE MOVE:** Youngest to be awarded for contribution to society through art. Apr 2019
  - **All India Cultural Youth Festival, MHRD, India:** Silver medal, South Zone, won National scholarship. 2016
  - **Creator of socially aware art brand: [HuesOnMyCanvas](#)** (10k+ followers on Instagram) 2016 - Present
-