# VITTHAL BHANDARI

+91 7529009938   •   vitthalbhandari98@gmail.com   •   vitthal-bhandari.github.io   •   linkedin.com/in/vitthal-bhandari

## EDUCATION

**Birla Institute of Technology & Science, Pilani Campus**  Rajasthan, India
Bachelor of Engineering in Computer Science  *July 2017 – July 2021*
*Minor in Data Science*

## WORK EXPERIENCE

**American Express**  Gurugram, India
*Software Engineer II*  Mar 2023 – Present
- Developed a language translation feature using Hugging Face's MarianMT framework for translating chats between English and Spanish, improving communication for 50+ Ask-Finance enterprise users
- Redesigned the Ask Finance chatbot UI using React and Redux to support graphics (images and tabular data) for 5K+ monthly user sessions, improving accessibility and enabling zoom-in capabilities
- Created an automatic dependency manager (Next.js, Sanic, TypeScript) to highlight and resolve package conflicts and vulnerabilities in Node and Python projects. Presented at Amex's internal hackathons: GrowthHack 2023, DevCon 2024

**Standard Chartered Bank**  Bengaluru, India
*Software Engineer I*  July 2021 – Feb 2023
- Led team of 4 in developing an API management framework (TypeScript, React, Flask) and creating configurable APIs (1.3K+ APIs) to fetch, add, and update data from Oracle SQL and Dremio DB
- Built reusable UI components (JavaScript, React, Redux) and REST APIs (Flask) for in-house chatbot Maya, improving its look and flow, and enhancing its capability to analyze data with speed and scale

**PayPal**  Bengaluru, India
*Software Engineer Intern*  Jan 2021 – July 2021
- Created PayPal's DropZone platform's 1st real-time metrics dashboard (JavaScript, React, Google Charts) and created REST APIs (Spring Boot, Elasticsearch), indexing > 2M file transfers/hour
- The dashboard is being used in production to analyze file transfer patterns, user demographics, and user behavior across geographies and time zones

## PUBLICATIONS  [Google Scholar Link]

[1] **On the Challenges of Building Datasets for Hate Speech Detection**  [arXiv preprint]

   Vitthal Bhandari

[2] **bitsa_nlp@LT-EDI-ACL2022: Leveraging Pretrained Language Models for Detecting Homophobia and Transphobia in Social Media Comments**  [ACL 2022]

   Vitthal Bhandari and Poonam Goyal

   *The Second Workshop on Language Technology for Equality, Diversity and Inclusion*

[3] **Image Processing in Retinal Imaging**  [Springer Nature]

   TMA Rehana Khan, Vitthal Bhandari, Sundaresan Raman, Abhishek Vyas, Akshay Raman, Maitreyee Roy, and Rajiv Raman

   *Digital Eye Care and Teleophthalmology: A Practical Guide to Applications*

## Research Experience

**Analyzing hate speech in popular political podcasts**                     Aug 2023 – Mar 2024
*Advisor: Dr. Sukrit Venkatagiri*                                                GitHub Repo

- Transcribed over 3k podcast episodes using Whisper based on a set of carefully curated 90 hateful keywords related to gender and gender identity
- Manually annotated randomized 30 transcripts for hate on the basis of a modular definition of hate
- Finetuned pre-trained language models (DistilBERT, DeBERTa, HateBERT) on an implicit hate speech dataset, and evaluated their performance on the podcast transcriptions

**Examining the challenges of building generalizable hate speech datasets**     Mar 2023 – May 2023
*Independent research*                                                              Paper

- Authored a comprehensive research paper examining the complexities of hate speech detection in NLP, focusing on data-centric challenges and ethical implications
- Analyzed 22 publicly available hate speech datasets, evaluating their methodologies and highlighting the impact of data source selection, annotation schema, and annotator demographics on dataset generalizability
- Developed a novel seven-point framework to guide reliable hate speech dataset creation, encompassing considerations for defining hate, choosing data sources, formulating annotation guidelines, and aggregating labels

**LGBQTweet: Building A large scale dataset for gendered hate speech**          July 2022 – Present
*Advisor: Dr. Poonam Goyal*

- Collected 150k+ publicly available tweets via extensive keyword search to build a dataset aimed at combating sexual and gender orientation discrimination, stratified for class imbalance using scores from Perspective API
- Crowdsourced 18k+ labels through LGBTQ+ annotators for over 6000 tweets
- Established high-performance classification and regression benchmarks by fine-tuning LLMs (Llama 2, Mistral, Gemma 2) and pre-trained language models (BERT, RoBERTa) using focal loss and LoRA

**Reviewing the collaborative role of image processing in retinal imaging**      March 2021 – July 2021
*Advisor: Dr. Sundaresan Raman*                                                    Paper

- Authored a comprehensive literature review on fundus image processing techniques, encompassing 60+ research papers and analyzing various image processing, machine learning, and deep learning methodologies
- Collaborated with a team of doctors and medical practitioners to investigate and document advancements in the diagnosis of Age-related Macular Degeneration and Diabetic Retinopathy using fundus imaging

## Projects

**Detecting Homophobia and Transphobia in YouTube Comments**              GitHub Repo / Paper

- Fine-tuned pre-trained language models (BERT, mBERT, XLM-RoBERTa, IndicBERT, HateBERT) to detect homophobia and transphobia in social media comments written in English and Tamil
- Developed a system for multi-class classification, employing techniques such as data augmentation (EDA) to handle class imbalance and preprocessing to remove punctuation and emojis
- Achieved a macro-averaged F1-score of 0.42, 0.64, and 0.58 in the English, Tamil, and Tamil-English code-mixed subtasks respectively, ranking 9th, 6th, and 3rd in the respective tracks of the LT-EDI Workshop shared task at ACL 2022

**TypeScript Tile-Matching Game** [GitHub Repo](#) / [Demo](#)

- Implemented a responsive game involving simple grid-based tile-matching in a Vercel project using Next.js and TypeScript
- Used Material UI for components and Redux for state management

**Analyzing IMDB Movie Reviews with Self-Trained Word Embeddings** [GitHub Repo](#)

- Developed custom word representations (vectors) from news articles using Python by training word2vec models with stochastic gradient descent, without using any external libraries
- Built RNN from scratch (Python, NumPy) to classify reviews in the IMDB dataset using the vectors trained above

## SKILLS

**Languages**: Python, TypeScript, JavaScript, C++, SQL

**Technologies**: Hugging Face, PyTorch, React, Next.js, Redux, Flask, Oracle SQL, Elasticsearch, Jenkins

## VOLUNTEER POSITIONS

**Reviewer**: ACL 2022

**Student Volunteer**: ACL 2022, AISTATS 2022, SIGIR 2021

**Academic Mentor**: Part of the Peer Mentorship Program (PMP) at BITS Pilani for helping freshmen with academic, co-curricular and personal doubts

## ACHIEVEMENTS

1 of 50 attendees selected for the Advanced Language Processing Winter School (ALPS 2024) in Grenoble, France