

KHUSHANG ZAVERI

kzaveri1@jh.edu | +1 (443) 563-4154 | [Linkedin.com/in/khushang-zaveri/](https://www.linkedin.com/in/khushang-zaveri/) | [Github.com/Khushangz](https://github.com/Khushangz) | [Portfolio/Khushang](#)

EDUCATION

JOHNS HOPKINS UNIVERSITY | Baltimore, MD

Master of Science in Engineering in Data Science | Expected: May 2026

Relevant courses: Research and Design in Applied Mathematics, Data Mining, Algorithms in Data Science, Machine Translation

MAHINDRA UNIVERSITY | Hyderabad, India | CGPA: 3.7/4

Bachelor of Technology in Computation & Mathematics | August 2020- June 2024

Relevant courses: Deep Learning, Mathematics in Machine Learning, Design & Analysis of Algorithms, Probability and Statistics

Graduated 3rd position in class and awarded 2 merit-based scholarships for academic excellence

SKILLS

- **Programming Languages:** Python, SQL, R, C, HTML, MATLAB
- **Data Handling:** NumPy, Pandas, OS, Scikit-learn, SciPy, NLTK, Matplotlib, Seaborn, BeautifulSoup, Requests, ggplot2
- **Model Building and Training:** PyTorch, Transformers, Wandb, SageMaker, SpaCy, Flair, TensorFlow, OpenCV
- **Competencies:** Machine Learning, Generative AI, Feature Engineering, Deep Learning, Database System Management
- **Software:** Git, Docker, Jira, LIWC, Apache Hadoop, Apache Hive, Apache Spark, AutoCAD, Microsoft Excel
- **Certification:** Machine Learning Specialization, Generative AI with Large Language Model, Deep Neural Network with PyTorch
- **Soft Skills:** Communication, Leadership, Time Management, Problem-solving, Innovation

EXPERIENCE

Graduate Research Assistant | Center for Language and Speech Processing, Johns Hopkins University | November 2024 - Present

- Researched political polarization, scraped tweet data from multiple events, reviewed literature on detecting polarized content
- Developed custom classification algorithms, fine-tuned RoBERTa for polarized content detection

AI Intern | Veltris | Hyderabad, India | January 2024 - July 2024

- Engineered Biomedical NER model, fine-tuned BERT and RoBERTa, achieved **80% accuracy**, enhanced biomedical research
- Leveraged PyTorch to build CNN model from scratch to classify images into streets, forests, and mountains with **90% accuracy**
- Gained expertise in Ensemble Models, Feature Extraction, Transformer Model Architecture, PEFT through training program

AI Intern | KPMG | Mumbai, India | June 2023 - August 2023

- **Project 1: Devised application with Language Models** such as Flan-T5, GEC-T5_small, Falcon-7B-Instruct
- Helped auditors get started with relevant questions, developed GEC model, increased grammatical accuracy
- **Project 2: Researched and evaluated impact of BASEL IV norms on Indian banks**, evaluated climate risk for flood-prone areas
- Devised safe investment solutions for Indian banks, considered climate risk and mandates to optimize strategy

Research Assistant | University of Pennsylvania | Finding Social Norms | Research_Paper | January 2023 - January 2024

- **Led creation of largest cross-cultural dataset** from 5,435 movie transcripts to identify harmful social norms
- Engineered an automated Python pipeline (BeautifulSoup, Requests, OS) to extract and preprocess massive web data efficiently
- Applied advanced NLP techniques (LDA, LIWC, prompt engineering) to uncover 10,000 unique social norms
- Presented abstract of research paper in field of Natural Language Processing at **IC2S2**, archived in **ACL**

KEY PROJECTS

Medical Audio and SOAP Note Generation | OpenAI Whisper, pydub, librosa, blaze999/Medical-NER | [Github.com/SOAP](https://github.com/SOAP)

- Engineered advanced Whisper workflow for audio denoising, segmentation, and high-fidelity transcript generation
- Devised custom classification pipeline integrating NER modules, fine-tuned Flan-T5 for SOAP Detection

Automated Contextual News Summarization Framework | BeautifulSoup, Requests, Gradio, Transformers | [Github.com/News](https://github.com/News)

- Automated pipeline to scrape, filter, summarize latest news on any topic, deployed model using Gradio

AgentIntellect | Python, Streamlit, Pytest | Streamlit/Deployed_Game | [GitHub.com/AI_Agent_game](https://github.com/AI_Agent_game)

- Deployed game using Streamlit, used Minimax algorithm, Utility based algorithm, Goal based algorithm to develop agents

LEADERSHIP EXPERIENCES

- Led and hosted treks for 200+ students as Adventure Club Head of Events, Mahindra University, 2022
- Organized and coordinated events for Math Club for 1000+ students, Mahindra University, 2022
- Optimized and streamlined budget plan for AERO Club, ensured efficient resource allocation, Mahindra University, 2021
- Volunteered at TATA Memorial Hospital, streamlined patient registration process to enhance patient experience, 2018