Vedant Vivek Jumle
+91 8879480965 | vedantjumle@gmail.com | LinkedIn | Medium | GitHub

EDUCATION

University of Mumbai, Thadomal Shahani Engineering College, Mumbai June 2020 – June 2024
Bachelor of Engineering (BE) – Computer Engineering (CGPA: 9.09)

CURRENT PROFESSIONAL ENGAGEMENTS/INTERNSHIPS

AI & Automation Engineer – R.U.D.R.A Cybersecurity
Dedicated AI engineer | Full-time | Since June 2024

- Lead the development of AI-driven solutions to enhance static code analysis, reducing false positives and improving the accuracy of security assessments by leveraging Large Language Models (LLMs).
- Design and implement automated pipelines for fine-tuning LLMs on curated datasets, focusing on optimizing security tasks, penetration testing, and vulnerability detection. Establish infrastructure for scalable model training, evaluation, and deployment.
- Developed LLM agents to automate security report generation, analyzing security operations data and delivering concise, actionable insights, thus allowing security analysts to focus on critical tasks.
- Emphasize explainability and interpretability through advanced prompt engineering techniques, such as Chain of Thought and iterative improvement, ensuring predictable and reliable AI outputs for sensitive security operations.
- Address challenges of integrating LLMs into security workflows by utilizing knowledge graphs and reviewer agents, refining LLM outputs to be highly context-specific and relevant to client environments.

PAST PROFESSIONAL ENGAGEMENTS/INTERNSHIPS

NLP Researcher – King's College London
Social Media Networks and (Dis)information Research Group | Part-time | Feb 2023 – June 2023

- Curated and analyzed a specialized dataset from Twitter, focusing on trending hashtags in India to explore patterns in social media dynamics and (dis)information spread.
- Enhanced data quality through advanced preprocessing techniques, addressing challenges related to noise and irrelevant content to ensure robust analysis.
- Applied Natural Language Processing (NLP) techniques, including word embeddings and Latent Dirichlet Allocation (LDA), to analyze language patterns and thematic shifts in social media discourse.
- Contributed to the research group's objectives by providing a foundational dataset, enabling further investigation into social media influence and user engagement.

Data Science Intern – Unilever, Mumbai, India
Data Science Intern | Full-time | June 2022 – Sept 2022

- Collaborated with the core data science team to develop a flexible data analysis module for Unilever's internal business intelligence dashboards, designed to integrate with any time-series data stream and generate hierarchical insights across various parameter groupings.
- Developed an algorithm to identify underperforming parameter groups(like region-city-product) in sales data by analyzing conversion rates from different advertisement channels. Applied anomaly detection techniques, such as a standardized deviation-based method, to identify areas requiring targeted marketing interventions.
- Extended the system to support any grouped time-series data by building a generalized pipeline for anomaly detection and trend analysis, enabling the application of the tool beyond sales data.
- Implemented time-series extrapolation using multivariate linear regression, enhancing the forecasting capabilities of the analysis tool.
- Built an ensemble recommendation system for Lakme's direct-to-consumer platform, using **Singular Value Decomposition (SVD)** and **K-Nearest Neighbors (KNN)** for collaborative filtering, improving product recommendations and user engagement.
- Addressed challenges in normalizing diverse sales datasets by developing a comprehensive preprocessing approach, joining data from various sources and normalizing values to ensure consistency across product categories.

Research Internship – NEWS lab, Indraprastha Institute of Information Technology, Delhi
Research Associate | Part-time | Oct 2021 – Dec 2021

- Contributed to the research project "*Decoding the Star System: Twitter and its Impact on Journalism in the Global South,*" focusing on how social media shapes networking practices among journalists in India.
- Conducted data extraction from Twitter, gathering information on journalists' profiles, follower networks, and tweets. Utilized tools like **BeautifulSoup** and **Tweepy**, employing **Latent Dirichlet Allocation (LDA)** to analyze content themes and patterns in social media discourse.
- Collaborated with fellow researchers to generate insights into the digital behaviors of national and regional journalists, with a focus on analyzing the language and themes used in their online bios.
- Supported the creation of a curated dataset that facilitated further analysis of social media interactions, contributing to the broader understanding of media dynamics within the Indian context.
- Findings and methodology contributed to a collaborative publication, highlighting the influence of celebrity culture and media hierarchies on journalistic practices in the Global South.

Internship – Trivedi Center for Political Data, Ashoka University
Research Associate | Part-time | May 2021 – Dec 2021

- Collected, cleaned, and compiled a detailed dataset on Indian Administrative Services (IAS) officers, covering comprehensive histories of all officers appointed until May 2021, addressing the challenge of disparate data sources.
- Utilized web scraping tools like **BeautifulSoup** and **Selenium** to automate the data collection process, standardizing data from varied sources for seamless analysis.
- Developed a sociological analysis dashboard to enable interactive exploration of the dataset, offering insights through visualizations like bubble plots and trend analysis graphs.
- Strengthened skills in data preprocessing and dataset structuring, key competencies for machine learning model development, and data analysis in research environments.

Research Assistant – Centre for Social Sciences and Humanities
Research Assistant to Jean-Thomas Martelli | Part-time | Mar 2021 – Sept 2022

- Assisted Dr. Jean-Thomas Martelli in various computational tasks, including optimizing data processing pipelines using parallel processing, and reducing preprocessing time for large datasets from 12-13 hours to 1-2 hours.
- Led the development of a dataset analyzing audio content from 'Republic TV' debates, focusing on identifying language patterns that blurred the line between 'entertainment' and 'news.' Extracted audio from YouTube and transcribed it into text using speech-to-text tools for subsequent analysis
- Presented the initial findings at The Broken Mirror symposium, hosted by CSH Delhi and IIITD, showcasing methodologies and preliminary insights.
- Gained hands-on experience in data preprocessing, speech-to-text transcription, and the structuring of large datasets, laying the groundwork for further research applications in machine learning.

PUBLICATIONS

Replication Data: Twitter and the Projection of Political Personalities in India
Decoding the star system: Twitter and its impact on journalism in the global South
Bureaucrats of India - TCPD–IAS is a dataset on the officers of the Indian Administrative Service (IAS)

CERTIFICATIONS AND WORKSHOPS

- Google Developer's Machine Learning Bootcamp.
- Certified TensorFlow developer from Google (related to the badge at the top)
- Deep Learning Specialization

PROJECTS

Diffusion in TensorFlow | Class Conditioned Models: Medium-1 Medium-2
- **Objective**: Developed a novel class conditioning mechanism for diffusion models, enabling targeted generation of images based on specific class labels.
- **Key Contributions**: Implemented custom class conditioning methods using **TensorFlow** and **Keras**, extending the generative capabilities of diffusion models to control output generation.
- **Methodologies**: Combined concepts from **DDPM** and **DDIM** with class embeddings, integrating these into the diffusion model architecture.
- **Impact**: Provided one of the first publicly available guides on this method, contributing to the community's understanding of class-conditioned generative models. This approach enabled more precise control over the diffusion process, expanding its potential applications in multimodal AI.

Visual Audio Summarization: Github

- **Objective**: Created a tool to convert lengthy audio recordings into concise, user-friendly visual summaries, making audio content more accessible.
- **Key Contributions**: Utilized **Whisper** for accurate speech-to-text conversion and **BERT** for extractive summarization of transcribed text, translating spoken content into key points.
- **Methodologies**: Employed **NLP** techniques for summarization and visualization, integrating results into a user interface that displayed visual representations of summarized content.
- **Impact**: Enhanced accessibility to spoken content by providing concise visual summaries, which are useful for applications like meeting transcription analysis and podcast summaries.

Approximating Region of Interest in Images Using Attention Weights: GitHub
- **Objective**: Developed a method to identify key regions in images using **VisionTransformer (ViT)** models, focusing on attention-based analysis.
- **Key Contributions:** Extracted attention weights from ViT models to approximate regions of interest, enabling a better understanding of the model's focus areas in image classification.
- **Methodologies**: Applied advanced deep learning techniques and attention mechanisms to analyze images, contributing to research in **explainability** and **hierarchical object detection**.
- **Impact**: Demonstrated the potential of attention weights for improving interpretability in computer vision models, providing a foundation for future work in explainable AI and object detection.

Buzztrends: buzztrends.in
- **Objective**: Built an AI Copilot for moment marketing, automating content generation for brands based on trending topics and social media analysis.
- **Key Contributions**: Designed and deployed a system that uses **LLMs** like **GPT** and **LLAMA** to generate social media content, integrating **ChromaDB** for managing vectorized data from web scraping.
- **Methodologies**: Leveraged **Azure services** for deployment with a focus on **CI/CD** processes, utilizing **CosmosDB** for scalable data storage and **Azure OpenAI** for real-time content generation.
- **Impact**: Enabled brands to automate their social media content strategy, improving engagement through timely and relevant posts tailored to their audience, reducing manual effort in content curation.

Twitter Disaster Watcher: GitHub

- **Objective:** Developed an AI-driven system to monitor tweets for disaster-related information, providing early warnings and insights for effective disaster management.
- **Key Contributions**: Engineered a solution using transfer learning with BERT for context-aware disaster classification and implemented a vector similarity-based classifier to enhance detection accuracy. Addressed challenges posed by Twitter's data access restrictions through strategic data preprocessing and filtering.
- **Methodologies**: Utilized transfer learning techniques to fine-tune BERT, optimized data handling processes to mitigate noise and data sparsity, and employed semantic similarity measures for precise tweet classification.
- **Impact:** Demonstrated the feasibility of using NLP for real-time disaster monitoring, offering a proof of concept that could be extended to support emergency response efforts. Provided a simple and effective platform for everyone, including disaster response, to

OTHERS

Technical Skills: Python, TensorFlow, Torch, Numpy, Pandas, Polars, NLTK, OpenCV, PIL, Huggingface
Areas of Proficiency: Deep Learning, Computer Vision, Data Science, Machine Learning & AI, NLP
Languages Spoken: English *(Proficient),* Hindi *(Proficient),* Marathi *(Native)*

RELEVANT URLS

- https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ATGGJR
- https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-5899.13252
- https://tcpd.ashoka.edu.in/bureaucrats-of-india/
- https://medium.com/@vedantjumle
- https://github.com/vedant-jumle/MPR_sem_6
- https://github.com/vedant-jumle/VIT_region_approx
- https://github.com/vedant-jumle/TwitterWatcher
- https://buzztrends.in
- https://github.com/vedant-jumle/TwitterWatcher
- https://drive.google.com/file/d/1efQXUBBQ1qVoP5TYtxvf942QqFlpfi3c/view?usp=sharing
- https://www.credential.net/129561b4-60ef-49d7-8d99-a75b0af05d07#gs.f8nra1
- https://drive.google.com/file/d/1bVsgcoGRiQmKZfvz1CuDoz7GX5ctpIC_/view