Sanjay Suryanarayanan

✉ sanj.ai@outlook.com | in linkedin.com/in/sanjay73 | ⬛ github.com/sanjay-73

EDUCATION

Manipal Institute Of Technology, MAHE   **Manipal, Karnataka**
Bachelor of Technology in Data Science & Engineering   August 2020 – July 2024

EXPERIENCE

AI4Bhārat, IIT Madras   **Chennai, Tamilnadu · On-site**
Research Engineer   July 2024 – Present

- Advisors: Mitesh M. Khapra, Raj Dabre, Anoop Kunchukuttan
- Working on MULTILINGUAL LANGUAGE MODELING AND MACHINE TRANSLATION, with a focus on Indian languages.
- Investigating effectiveness of MULTIMODAL LLMs as evaluators for VITA tasks through targeted perturbations.
- Fine-tuned INDICTRANS2 by transitioning from absolute sinusoidal PEs to relative PEs like RoPE, and ALiBi for improved long-context handling.

Research Intern   January 2024 – June 2024

- Advisors: Mitesh M. Khapra, Raj Dabre, Anoop Kunchukuttan
- Introduced PRALEKHA, an Indic document alignment evaluation benchmark, featuring a novel margin-based greedy bipartite matching algorithm for mining large-scale parallel document pairs. Leveraged LaBSE, SONAR embeddings, and FAISS similarity search for efficient cross-lingual alignment.

Smart Collared Eye Lab, MIT Manipal   **Manipal, Karnataka · On-site**
Undergraduate Researcher   July 2022 – June 2024

- Advisors: Vidya Rao, Abhilash K Pai, Poornima Panduranga Kundapur
- Spearheaded the SCENE SEGMENTATION & PATH PLANNING division, focusing on developing wearable assistive technology for the visually impaired leveraging COMPUTER VISION and SENSOR FUSION.
- Optimized semantic segmentation (SAM, U-NET) and path planning algorithms (D*, RRT*) to improve indoor navigation for the visually impaired, ensuring real-time deployment on compact systems like RASPBERRY PI.

ThirdEye Data   **San Jose, California · Remote**
Data Science Intern   June 2023 – August 2023

- Contributed to a retail project utilizing the MICROSOFT AZURE SUITE to blend in-store and online shopping seamlessly.
- Utilized AZURE DATABRICKS for data analysis, uncovering insights into purchase patterns & customer behaviour, and developed personalized recommendation models in AZURE ML STUDIO.

Apache   **Chennai, Tamilnadu · On-site**
Data Analyst Intern   November 2022 – April 2023

- Developed and maintained an ETL pipeline using AZURE services (SQL Server, Power Automate, Power Query) to load SharePoint data into POWER BI for compliance analysis, and designed interactive dashboards providing actionable insights to NQ CRANES offices.

FunctionUp   **Bengaluru, Karnataka · Remote**
Data Science Instructor   July 2023 – June 2024

- Facilitated 500+ hours of hands-on workshops on data science concepts and problem-solving strategies for students with non-technical backgrounds, helping them transition into data science. Also led weekly masterclasses, doubt sessions, and curated assignments to provide a comprehensive learning experience.

PUBLICATIONS

[1] **Pralekha: An Indic Document Alignment Evaluation Benchmark** 📄 🗄 ‹/›
Sanjay Suryanarayanan, Haiyue Song, Mohammed Safi Ur Rahman Khan, Anoop Kunchukuttan, Mitesh Khapra, Raj Dabre
arXiv preprint arXiv:2411.19096

[2] T-Copilot: Copilot for All Things Translation
Sumanth Doddapaneni, Mohammed Safi Ur Rahman Khan, **Sanjay Suryanarayanan**, Haiyue Song,
Abhishek Ranjan, Sparsh Jain, Anoop Kunchukuttan, Mitesh M. Khapra, Raj Dabre
(Work in Progress)

[3] Finding Blindspots in Multimodal Language Model Evaluations with Interpretable Checklists
Mohammed Safi Ur Rahman Khan, **Sanjay Suryanarayanan**, Oikantik Nath, Sumanth Doddapaneni,
Mitesh M. Khapra
(Work in Progress)

[4] Beyond Sight: Leveraging Semantic Segmentation for Assisting the Visually Impaired
Sanjay Suryanarayanan, Tanay Gupta, Abhilash Pai, Poornima Kundapur, Cifha Crecil Dias, Vidya Rao
Proceedings of the 6th International Conference on Computational Methods in Engineering & Health Sciences

PROJECTS

Indic Parallel Corpora Scraper
- Developed the Indic Parallel Corpora Scraper, a robust tool for extracting parallel corpora from diverse Indian websites, including government portals, news platforms, and tourism sites. It efficiently collects aligned text segments across multiple Indian languages by leveraging **BeautifulSoup** for static pages and **Selenium** for dynamic content, ensuring comprehensive coverage of multilingual web data.

Music Genre Detector & Lyrics Generator
- **Music Genre Detector** - Developed a system that accurately categorises music tracks into distinct genres based on their audio features by leveraging various classification models and, additionally, utilising CNN for genre prediction. **XGBoost** achieved the best accuracy of **90%** and a prediction accuracy of **93.5%** for **CNN**.
- **Music Lyrics Generator** - A **LSTM**-based text generation model for generating lyrics with the help of **NLTK**.

Amazon Recommender System
- Developed a recommendation system for Amazon products employing both **Popular-Based** Method and **Collaborative Filtering** techniques, with a focus on the **Singular Value Decomposition (SVD)** method for **user-item recommendation**. The goal is to suggest popular items based on overall user preferences & purchase history and also utilize user-item interaction data, such as user ratings, to identify patterns & similarities between users.

Formula 1 Race Predictor
- Developed a Formula 1 Race Predictor using machine learning to forecast race outcomes by analyzing historical data, including driver performance, team dynamics, and track conditions. Achieved an **F1-Score of 93.6** with **RandomForest** and **90.1** with **Support Vector Machines (SVM)**, demonstrating the model's ability to accurately predict results under diverse scenarios.

POSITIONS OF RESPONSIBILITY

Project Mudra   **Manipal, Karnataka · On-site**
Advisory Board   August 2022 - July 2023
- Served as a senior advisor, providing guidance and support to the new board in critical decision-making and management processes.

Board Member (Head of Web Development)   August 2021 - July 2022
- Spearheaded a team of 27 students, revamped the website with enhanced back-end functionalities, and managed site analytics, SEO, and administrative tasks.

Web Developer   January 2021 - July 2021
- Defined the design direction, built the UI from scratch and developed the login page to ensure a seamless UX.