Sampras Manuel Dsouza

201-616-6095 | sd6701@nyu.edu | linkedin.com/in/sampras-dsouza-sd6701 | github.com/sampras-dsouza

Education

New York University, Courant Institute of Mathematical Sciences August 2025 - May 2027
Masters of Science in Computer Science

University of Mumbai August 2017 - June 2021
Bachelors in Information Technology CGPA - 3.87/4.0

Experience

Senior Software Engineer | Cimpress (Sr. Software Engineer, Since December 2023) August 2021 – August 2025
Tech Stack: ReactJS, NodeJS, AWS, SQL, Python, Java Mumbai, India

- Built an **mapping tool** for files migrated to different data storage systems and developed **custom ETL pipeline** to migrate **60 Million** records from Google drive to AWS Dynamodb, optimizing the process and improving data management. Also, Developed a User facing portal to handle **10K requests** daily.
- Built an entire **Internal developer portal** with **ReactJS** designing and developing **100+ reusable components** with React Bootstrap. implementing advanced state management using **Redux-Saga, Redux Toolkit** and creating a centralized component library.
- Developed and maintained **Attribute-based Authorization Service** while optimizing the existing RBAC system to support **3 million daily active users and billions** of transactions monthly.
- Build and managed infrastructure pipeline for systems on **AWS infrastructure** using **terraform, AWS CDK** to enhancing scalability and Availability.

Research Experience

Research Volunteer | Carnegie Mellon University, Pittsburgh, PA Feb 2024 – May 2025
Tech Stack: Pytorch, VideoMAE (Transformer-based Masked Autoencoder)

- Enhanced few-shot particle detection in cryo-ET ("SaSi: A Self-augmented and Self-interpreted Deep Learning Approach for Few-shot Cryo-ET Particle Detection") by adapting **VideoMAE**, a transformer-based masked autoencoder, for 3D tomogram reconstruction and segmentation.
- Utilizing self-augmented data and self-interpreted segmentation strategies to maximize data efficiency, reduce dependency on labeled data, and enhance localization accuracy on the **SHREC Cryo-ET 2021 Dataset: Classification in Cryo-Electron Tomograms**.

Research Intern | Tata Institute of Fundamental Research (TIFR), Mumbai, India June 2024 – Oct 2024
Tech Stack: Python, PyTorch, Autoformer, LSTM, Scikit-learn, Matplotlib, Pandas Under Prof. Jatin Batra

- Analyzed historical weather data (1900–2021) to model and interpret long-term rainfall variability across India using **Autoformer** and machine learning algorithms including **Linear Regression**, **Decision Trees**, and **Random Forests**.
- Designed **Autoformer-based time series forecasting pipelines** for sequence modeling and temporal rainfall prediction, leveraging its series decomposition blocks to capture periodic and trend components for better interpretability.
- Experimented with **LSTM networks** for short-term forecasting ($n+1$, $n+3$ days), comparing performance against transformer-based architectures to assess generalization under non-stationary climate conditions.
- Developed end-to-end workflows integrating data preprocessing, model training, evaluation, and visualization to support climate trend interpretation and decision-making for regional hydrological analysis.

Projects

Self-Supervised Learning Using VICReg

- Pretrained **VICReg-based self-supervised models** on a **700K-image custom dataset** and evaluated representations on downstream image classification tasks under ImageNet-style constraints.
- Trained **ResNet-50×2 architectures (<100M parameters)** using **large-batch training (batch size 1024)** with the **LARS optimizer** for stable non-contrastive SSL.
- Optimized **data augmentations, loss coefficients (variance, invariance, covariance), and hyperparameters** to prevent representation collapse and ensure convergence.
- Performed **linear probing and finetuning** on downstream tasks, analyzing the impact of architecture and augmentation choices on transfer performance.

Eyeris — Computer Vision, Raspberry Pi, Node.js, React.js, HTML

- Developed an **AI based assistive system** using a **DenseNet-based CNN** for **real-time object recognition and anomaly detection** on the NYU v2 dataset, integrating speech-to-text and text-to-speech for audio-based guidance.
- Implemented **YOLOv3** with the COCO dataset for efficient edge-device object detection and automated user feedback workflows.

Achievements

- Completed **Distributed Systems (IIT Kanpur, NPTEL)** in **Top 5%**, mastering concepts in **fault-tolerant architecture**, **multi-node synchronization**, and **consensus algorithms**.