Sai Shailesh Nanisetty

3701- 82 Dalhousie St, Toronto, ON M5B 0C5

(437) 566-0979 linkedin github shailesh005@cs.toronto.edu

Education

University of Toronto Sep 2023 - April 2025
Master of Science in Applied Computing (Artificial Intelligence) *GPA: 4.0/4.0*
[Courses] Neural Networks & Deep Learning, Natural Language Processing, Visual and Mobile Computing, Computer Vision, Probabilistic Learning & Reasoning

Indian Institute of Technology, Kharagpur July 2018 - May 2023
Dual Degree (B.Tech+M.Tech) in Industrial Engineering, Micro Specialization in AI *GPA: 8.65/10*
Silver Awardee, Department Rank 2
[Courses] Optimization & Heuristics, Data Structures & Algorithms, Probability, Multivariate Statistical Modelling, Foundations of ML, Linear Algebra, Engineering Mathematics

Work Experience

External Research Collaborator @ServiceNow Research | *Python, LLMs, Prompt Engineering, Git* Oct 2024 - Present

- Creating AgentADA: an autonomous agent that learns & applies advanced data analytics tools from unstructured code sources
- Developed the Insight Toolkit by extracting, verifying & categorizing code blocks from various analytics notebooks.
- Automated metadata generation & skill filtering using few-shot & instruction-based LLM prompting techniques respectively.
- Conducting research to optimize RAG workflows, focusing on improving tool retrieval & adaptation for domain-specific analytics.

ML Developer Intern @Synapsis Medical Technologies Inc. | *Python, PyTorch, Tensorflow, React,Docker* Aug 2024 - Present

- Created react-based modules for real-time facial emotion, head pose & HeartRate prediction respectively with latency of 55ms/frame.
- Developed anomaly detection model for wearable health monitoring, improving accuracy by 15% through reduced false alarms.
- Optimizing ML models for real-time health insights on resource-constrained devices from asynchronous data streams.

Summer Associate @Balyasny Asset Management | *Python, PostgreSQL, Streamlit, Databricks* May 2024 - Aug 2024

- Developed an Apache Airflow DAG for BAM Elevate, automating data population processes & improving workflow efficiency by 30%
- Structured multi-level datasets from investor & deal interaction feedback, enhancing analysis speed by 25%
- Optimized SQL queries for faster geographical expansion insights, reducing extraction & aggregation time by 40%.
- Built dynamic Streamlit dashboards for various use cases providing real-time insights, supporting key investment decisions.

Generative ML Intern @16 Bit Inc. | *Python, PyTorch, HuggingFace, Transformers, Git* Nov 2023 - Feb 2024

- Developed PaddleOCR-ViT model to extract patient details from DXA sheets, achieving 92 % accuracy on test set of 1200 samples
- Performed zero-shot & few-shot prompting on Llama2 7B & Zephyr 7B, achieving ROUGE-L score of 45.2 in report summarization.
- Fine-tuned models using PeFT-LoRA techniques enhancing contextual accuracy of report summaries with a ROUGE-L score of 48.3.

Computer Vision Intern @Bharat Smart Services | *Python, Tensorflow, Android Studio, Docker, Git* May 2021 – Aug 2021

- Developed & deployed custom CNN models, achieving 91% accuracy on test set, to classify electric meters & extract readings using YOLOv5. Built an Android app for real-time electric meter reading & automated billing, streamlining the process for end-users.

Major Projects

Prompt-GS: Segment Anything in 3D Gaussians with Multi-View Text Prompting | link Feb 2024 - May 2024
Collaborators: Kai Zhu, Lakshya Gupta, Anannya Popat, University of Toronto

- Developed an enhanced 3D segmentation method by integrating Gaussian Splatting with prompt-based segmentation.
- Employed LangSAM for text-prompted object segmentation & DUSt3R for efficient point-cloud initialization.
- Developed multi-view mask generation & label voting for accurate 3D object segmentation from sparse views.
- Achieved higher accuracy & IoU scores with 50 percent fewer views alongside cutting down compute requirements by 23%.

TinyHR @University of Alberta | link Aug 2021 - Aug 2022
Collaborators: Preetam Anbukarasu, Ganesh Tata, Prof. Nilanjan Ray, University of Alberta

- Built a hybrid FFNN-CNN pipeline that extracts heart rate from pressure data acquired on low-power ESP32 device.
- Wrote C++ Scripts for suitable deployment of implemented PyTorch models onto ESP32 edge device.
- Proposed method cuts energy and time inference by 82 % & 28 % compared to state of the art methods.

Improvised sequential few-shot segmentation - UG Thesis @IIT KGP | link Dec 2021 - May 2022
Collaborators: Prof. KS Rao, IIT Kharagpur

- Designed a Few-Shot CNN algorithm for segmenting low-labelled images by reducing perceptual bias.
- Incorporated set of Difference of Gaussians and bi-directional ConvLSTM algorithm in the framework.
- Performance measured in mean IOU shot up by 6.26 % & 1.2 % for 1 & 5 shot cases respectively.

Technical Skills & Certifications

Languages: Python, C++, C, R, Matlab
Machine Learning: Transformers, PyTorch, JAX, Tensorflow, Keras, OpenCV, Hugging Face
Data Science: Pandas, Numpy, Scikit-Learn, Statsmodels, Scipy, Plotly, Seaborn, PySpark, Databricks, Microsoft Power BI
Database and Web Frameworks: SQL, PostgreSQL, MongoDB, React, NodeJS, FastAPI
Tools and OS: Docker, Heroku, Flask, Spark, Dask, Git, Linux, Ubuntu

Achievements

- Awarded the MITACS Globalink Graduate Fellowship worth 15000 CAD to aid further research in Canada.
- One among 1190 students across the globe selected for the prestigious MITACS Global Research Internship.
- Winner of Smart India Hackathon, Government of India, 2020
- Achieved top 0.3 percentile in IIT-JEE examination, 2018 among 1.25 million candidates.