

SAYANTAN DAS

Toronto, ON | M: +1 6137702006 | sdas.codes@gmail.com | www.ucalyptus.me |
www.github.com/ucalyptus | www.linkedin.com/in/ucalyptus/

SUMMARY

Machine Learning Research Engineer at Prem Labs with 2 years of industry experience. 4 years of academic and industrial research experience with Vector Institute, ETH Zurich, Ingenuity Labs and ISRO.

Skills: Pytorch, Docker, Kubernetes, Kubeflow, MLFlow, Airflow, CUDA, Terraform, Django, AWS, GCP

EDUCATION

QUEEN'S UNIVERSITY

Kingston, ON

MASc (Master of Applied Science) – Artificial Intelligence

September 2021 – March 2024

Cumulative GPA: 4/4.3. [Thesis: Fair and Accurate Deepfake Detection](#)

WORK EXPERIENCE

Machine Learning Engineer

Prem Labs Inc

United States(remote)

January 2024 – Present

- Developed [Prem-1B](#) and [Prem-1B-chat](#) small language models (LLMs), also its [WebGPU version](#) .
- Developed POC (proof of concept) for an evaluation suite for LLMs via LLM-as-a-judge techniques.
- Worked with the Platform (app.prem.ai.io) team for full-stack improvements and deployment.
- Developed a video trailer for a comic book character using custom-made Flux LoRA and Diffusion Transformer (DiT) models, as part of a verticals project.
- Led the technical content team till June 2024. Here are some of my [blogs](#). Also responsible for two of the Prem AI whitepapers.
- **Skills:** Django, PostgreSQL, REST, Terraform, Amazon Web Services, Comfy UI, HF Transformers, Pytorch, WebGPU, WASM, transformers.js, Selenium, Playwright

Machine Learning Associate

Vector Institute & Realsage

Canada(remote)

May 2024 – July 2024

- Developed a **Text-to-SQL** chatbot using [SQLCoder2](#), incorporating two steps: **Table Retrieval** and **Intent Classification** before inference on **private multi-family residential real estate data**.
- Prepared a dashboard using **react** and **node.js** for visualization and inference.
- Achieved a bump of **25%** on the **execution accuracy** compared to the existing solution at the company that relied on **GPT-4-turbo**.
- **Skills:** Text to SQL, Ollama, Docker, AWS Lambda, Node.js, Next.js

Software Developer (Contract)

Quantum Growth Advisors

Canada(remote)

December 2023 – January 2024

- Developed an LLM-based web scraper using GPT4 API and Pydantic. Replaced GPT4 with LLAMA2-7B to save costs. Deployments were done using **Docker** and **Google Cloud Run**.

- **Utilized** Instructor for **structured-response-generation** to scrape investment and startup metadata from their landing pages and news platforms. Reduced failure cases to 0.4 %
- **Developed** an async scraping automation tool for clutch.co using Playwright, selenium, and multiprocessing frameworks.
- **Skills:** Pydantic, REST, OpenAI API, Docker, Google Cloud Run, Selenium

Machine Learning Developer (Contract)
Ready Tensor Inc

United States(remote)
October 2023 – November 2023

- **Contributed** to expanding Ready Tensor's model-first portfolio by incorporating graph neural networks.
- **Trained** Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) on the Planetoid datasets (Cora, Cite Seer, PubMed) for node classification tasks.
- **Facilitated** multiple installation methods, including Docker and local setups.
- **Supported** real-time inference capabilities through the integration of Fast API using node id as input.
- **Skills:** Fast API, Pytorch Geometric, Docker, Kubeflow Pipelines

Graduate Research Assistant
Ingenuity Labs Research Institute

Kingston, ON
September 2021 – December 2023

- **Developed** self-supervised models, specifically masked autoencoders and vision transformers, to enhance facial forgery detection accuracy by 5%
- **Published** at **International Joint Conference on Biometrics (IJCB 2023)** [\[Read paper\]](#)
- **Implemented** a novel fairness promotion and evaluation protocol for deepfake detection (submitted to ECCV 2024).
- **Devised** mitigation strategies to improve the fairness of deepfake detection, reducing demographic bias by 10%.
- Previously worked on developing **controllable GANs (Generative Adversarial Networks)**, enabling targeted data generation and enhancing model interpretability.
- **Received** funding from the Vector Institute and Irdeto BV for my research and Alliance (Compute Canada) for research compute (NVIDIA A100 clusters).
- **Skills:** Distributed Training, SLURM, Anyscale Ray Train, Pytorch DDP, MSR fairlearn, DEI training

Technical Writer
Weights and Biases (wandb)

United States(remote)
March 2021 – Jan 2022

- **Authored** comprehensive guides and tutorials on machine learning experiment tracking, versioning, and hyperparameter tuning. My works are available at <https://www.wandb.ai/ucalyptus>
- **Skills:** wandb, Pytorch, technical writing

Research Intern
ETH Zurich

Basel, Switzerland
September 2020 – February 2021

- **Investigated** machine learning through topological data analysis, advancing feature extraction techniques.
- **Implemented** manifold learning algorithms for high-dimensional data visualization.
- **Co-authored** NeurIPS 2020 spotlight paper on optimizing GANs via topological constraints. [\[Read paper\]](#)
- **Skills:** Pytorch, TDA and Persistent Homology for ML research

Research Intern
Indian Statistical Institute Kolkata

Calcutta, India
March 2020 – July 2020

- **Explored** hyperspectral satellite imagery applications across agriculture, mineralogy, and environmental monitoring.
- **Developed** image processing algorithms with a specialized team, elevating hyperspectral classification accuracy by 7%.
- **Co-authored an IEEE GRSL** journal paper on channel selection for HSI classification. [\[Read paper\]](#)
- **Skills:** Pytorch

Research Intern

Indian Space Research Organization (ISRO)

Ahmedabad, India
June 2019 – August 2019

- **Analyzed** Synthetic Aperture Radar (SAR) microwave images from Sentinel-2 and Indian satellites to develop algorithms for Ground Moving Target Indication (GMTI).
- **Collaborated** with subject matter experts to adapt existing image processing methods for optimized GMTI applications.
- **Conducted** comparative studies of GMTI techniques, assisting in the selection of the most effective methods for operational use.
- **Skills:** ArcGIS Pro, Pytorch, Gradio UI, C, Docker, Microwave Image pre-processing techniques.

PROJECTS

BongoVaad

Queen's University

Toronto, ON

June 2023

- **Developed** a Python framework aimed at transcribing Bengali YouTube videos using Whisper Large V2.
- **Fine-tuned** the model via Low Rank Adaptation (LoRA) on Mozilla Common Voice 13 dataset to improve transcription capabilities.
- **Employed** rigorous testing methodologies to evaluate and enhance the accuracy and efficiency of the model.
- I am currently looking to **research** the efficacy of this product using ASR benchmarks available in the audioDL community.
- Currently available as a Chrome Extension that works when a YouTube tab is active. We will work to make it a paid service once we finalize the best-fit deployment strategy for this product.
- **Technologies/Framework Used:** Python, Whisper Large V2, Huggingface, PEFT, LoRA

N-BEATS Forecasting

Upwork

Remote

April 2023

- **Engaged** on a multivariate time series forecasting hourly electricity consumption for 321 customers with data of over 2 years using a state-of-the-art N-BEATS algorithm.
- **Evaluated** the model's performance on a comparative basis with a baseline model on SMAPE metric.
- **Generated** decomposition plots to interpret the predictions possible due to the N-BEATS' unique architecture.
- **Performed** a separate time series modeling task on ECG (non-seasonal) and site traffic (seasonal) data using S-ARIMA, GARCH modeling in R for the same client on Upwork.
- **Technologies/Framework Used:** PyTorch Forecasting, DeepAR, Prophet, PyTorch Lightning, R

AskTRS

Queen's University

Toronto, ON

May 2023

- **Created** a Gradio web app for vector-based search and response in NLP (Natural Language Processing), hosted on Huggingface Spaces.
- **Integrated** Langchain, FAISS, and OpenAI Embeddings(ada) to facilitate Retrieval Augmented Generation capabilities.
- **Designed** the app to function seamlessly across multiple platforms, ensuring broad accessibility and usability.
- **Technologies/Framework Used:** Gradio, Langchain, FAISS, OpenAI Embeddings, Huggingface Spaces