

# Sourav Kumar Khan

Portfolio: [sourav-kumar-khan.github.io/portfolio/](https://sourav-kumar-khan.github.io/portfolio/)

Github: [github.com/Sourav-Kumar-Khan](https://github.com/Sourav-Kumar-Khan)

Email: [khansourav1999@gmail.com](mailto:khansourav1999@gmail.com)

Mobile: +91-7908341191

LinkedIn: [Sourav Kumar Khan](#)

## EDUCATION

- **Defense Institute of Advanced Technology (DIAT), DU, DRDO** Pune, India  
*Master of Technology - MTech, Data Science; GPA: 8.34* July 2021 - June 2023  
*Courses: Operating Systems, Data Structures, Statistics, Artificial Intelligence, Machine Learning, NLP, Computer Vision, BigData*

## TECHNICAL SKILLS:

- **Machine Learning** - DL, RL, RAG, NLP, PEFT, LoRA, CV, Neural Networks
- **Programming** - Python, C, C++, LATEX, (Familiar with JavaScript, MySQL)
- **Frameworks/Libraries** - PyTorch, TensorFlow, Transformers, Keras, Flask, FastAPI, Pandas, YoLo, Matplotlib, OpenCV, Spacy, HuggingFace
- **Cloud and MLOps** - AWS, Microsoft Azure, CI/CD, Jenkins, Deployment using AWS ECS and Azure container app
- **Other** - LLMs, LangChain, Text-To-Speech, Object-Oriented Programming, SQL, MongoDB, VectorDB, Docker, Git & Github

## EXPERIENCE

- **SproutsAI** United States (Remote)  
*Machine Learning Engineer* June 2023 - Present
  - **AI Model Development:** Developed task-specific AI models to automate business needs, including fine-tuning text embedding models on company-specific data.
  - **Model Deployment:** Deployed models in scalable, containerized environments with API services.
  - **Large Language Models:** Instruct tuned large language models (LLMs) like Mistral-7B and Zephyr-7B Alpha using AWS SageMaker.
  - **LLaMA3 Hosting:** Hosted Instruct-tuned LLaMA3 on Azure ML to reduce OpenAI API usage.
  - **ICL Integration:** Incorporated ICL with OpenAI API and privately hosted models to address specific business requirements.
- **SproutsAI** United States (Remote)  
*AI Engineer (NLP)* September 2022 - May 2023
  - **Autonomous Recruitment Process:** Contributed to the development of an end-to-end autonomous recruitment process.
  - **Resume Data Analysis:** Analyzed and parsed unstructured resume data into structured formats for candidate data analysis.
  - **AI-Driven Candidate Matching:** Implemented AI-driven solutions to identify optimal candidate profiles for job openings, incorporating feedback mechanisms and advanced filtering techniques.
- **Indian Institute of Tropical Meteorology Pune** Pune, Maharashtra, India  
*Project Intern* June 2022 - May 2023
  - **Geospatial Time Series Analysis:** Analyzed complex geospatial time series data for master's thesis using Python modules such as NetCDF, xarray, and dask.
  - **Data Visualization:** Utilized libraries like Seaborn and Matplotlib to extract insights into Mixed Layer Depth (MLD) over a 20-year period in the Bay of Bengal.
  - **Advanced Modeling:** Leveraged the Pratyush supercomputer to train over 100 complex ConvLSTM and ConvGRU models for studying seasonal variations.

## PROJECTS

- **Construction Site Safety Detection using YOLOv5:** Trained and validated a state-of-the-art YOLOv5 model on a custom construction dataset to detect helmets on workers. Implemented real-time object detection for automated monitoring of basic safety measures on construction sites using video footage.
- **WhatsApp Group Chat Analysis and Visualization:** Exported WhatsApp group messages and performed data cleaning and preprocessing using NLTK and spaCy. Generated visualizations including word clouds, highly active members, and most discussed topics using Matplotlib and Seaborn.

## CERTIFICATIONS

- Machine Learning A-Z: AI, Python & R + ChatGPT
- Learn C++ Programming - Beginner to Advance - Deep Dive in C++
- Ultimate AWS Certified Solutions Architect Associate SAA-C03
- The Complete Web Development Bootcamp

## HONORS AND AWARDS

- **First Position in Poster Presentation:** Annual Monsoon Workshop, IMSP Pune, 2023
- **Gold Badge:** NPTEL course (National Programme on Technology Enhanced Learning)