

# Nishan Pantha

[linkedin.com/in/nishparadox](https://www.linkedin.com/in/nishparadox) | [github.com/NISH1001](https://github.com/NISH1001)

## EXPERIENCE

---

### Computer Scientist II

NASA IMPACT, UAH | Sept 2023-Present

- At the intersection of Machine Learning research and engineering for Earth Science. Currently working with Large Language Models (LLMs) for hybrid search systems for NASA.

### Graduate Research Assistant

NASA IMPACT, UAH | Jan 2022 – July 2023

- Built an extensible framework for benchmarking large-scale data transfer (**5 PetaBytes**) from European Space Agency (ESA) to National Aeronautics and Space Administration (NASA). [GitHub Repository](#)
- Collaborated with NASA's Bio-Physical Science team to develop an ML-based algorithm for ranking gene expressions from space-flown rodents. Presented [poster for AGU 22 Final Slide Deck](#), IEEE SoutheastCon paper pending

### Senior Research Scientist/Engineer

Docsumo, Nepal | Nov 2019 – Dec 2021

- Built an end-to-end document table extraction pipeline (*ensemble of model such as CascadeTabNet, DBScan clustering, table header detection, pattern-matching, etc.*) with **row extraction accuracy of 85+**.
- Built a model-caching framework that helped **reduce the extraction latency by half (30s on average)** on any downstream ML inference pipeline.
- Worked on transformer-based Language Model (*modified-BERT with 2d position embedding*) for key-value extraction that significantly **increased accuracy to 90%+** on documents like *Invoices, Receipts, etc.*

### Co-Founder, Directory of Technology

MPercept Technology, Nepal | May 2018 – April 2019

- Technical decision-maker for all the projects such as *real-time face recognition, vehicle damage segmentation, etc.*
- Big-data consultant to **Umniah Telecommunication, Amman, Jordan** to help migrate their RDBMS data to HDFS in **2 months**. During this time, I also worked on real-time CDR file migration using streamsets.
- Co-founded an AI community, *AIDevNepal*, organizing weekend Data Science workshops for **400+ students**.

### Software Engineer

Fusemachines, Nepal | Nov 2016 – Feb 2018

- Built Salesforce Integration API for a client (*Enhatch*) as well as worked on research projects related to chatbot and Intelligent Character Recognition (ICR). We were able to achieve ICR accuracy of **70%** using RNNs.

## TECHNICAL SKILLS

---

**Languages:** Python(8+ years), C/C++(3+ years), Java, Bash, scala

**Database + Tools:** SQL, postgres, mongoDB, Docker, kubernetes

**Data Science + ML Libraries:** numpy, scikit-learn, scipy, scikit-optimize, mlflow, pandas, matplotlib, spacy

**Deep Learning Frameworks:** PyTorch, Tensorflow, transformers, sentence-transformers, mmcv, Large Language Models (LLMs), LangChain, Prompt Engineering

**Web Technologies:** Flask, Django, FastAPI, REST API, GCP, AWS

## EDUCATION

---

**The University of Alabama in Huntsville (UAH)**

MS CS | Jan 2022 - July 2023

**Tribhuvan University, Pulchowk Campus**

B.E. in Computer Engineering, Nepal | 2012 - 2016

## PROJECTS

---

**Damage Segmentation Detection** | *Python, PyTorch, Tensorflow, CNN*

- Project for a Germany-based automobile insurance company to identify damage segments using computer vision from vehicle images to estimate insurance costs. Worked on image processing pipeline and built Deep Learning models such as Mask-RCNN, and U-Net to detect damaged segments. Achieved **mAP score of 74%**.

**humT: Query by humming system, BE final year project** | *Python, numpy, signal processing* | [GitHub repo](#)

- Worked on a temporal pattern-matching algorithm like Dynamic Time Warping(DTW) and audio segmentation to **correctly identify 30/35 songs with top-5 accuracy of 90%**. ([ICTAES publication](#))

**anuwadak: Statistical Nepali-English text translation** | *Python, numpy, Markov Models* | [GitHub repo](#)

- Worked on Markov and N-gram models for predictive text generation for Nepali-English text translation.

**playx: open-source music assistant for Linux** | *Python, numpy, beautifulsoup* | [GitHub repo](#)

- Project nurturing **more than 200 stars** where I worked on cores such as crawler for songs+lyrics, song-caching mechanism, string matching algorithm, [playlist generation](#), and [song2vec embeddings](#).

**panim: open-source mathematical animation tool** | *Python, numpy, matplotlib*

- An open-source tool to implement (from scratch) various mathematical animations such as L-Systems, fractals, simulations, etc. [GitHub Repo](#)