

Nishan Pantha

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

EDUCATION

The University of Alabama in Huntsville (UAH) <i>MS in Computer Science</i>	Huntsville, AL <i>Jan 2022 – Present (Expected Graduation Summer 23)</i>
Tribhuvan University, Institute of Engineering, IOE, Pulchowk Campus <i>Bachelor of Engineering in Computer Engineering</i>	Kathmandu, Nepal <i>2012 – 2016</i>

EXPERIENCE

Graduate Research Assistant <i>NASA Interagency Implementation and Advanced Concepts Team (IMPACT), UAH</i>	Jan 2022 – Present
<ul style="list-style-type: none">Built an extensible framework for benchmarking large-scale data transfer (5 PetaBytes) from ESA to NASA. framework RepositoryCollaborating with NASA's Bio-Physical Science team, I developed an ML-based gene ranking algorithm for gene expressions from space-flown rodents and identified a few important genes (out of 25k genes) that affect the physical attributes (<i>gender, age,...</i>) of the rodents. I presented this work as a poster for AGU 22 and waiting for the final paper to be published. Final Slide Deck, Poster LinkCurrently building an extensible evaluation framework for Large Language Models on downstream Earth-Science tasks.	
Senior Data/Research Scientist & ML Engineer <i>Docsumo</i>	Nov 2019 – Dec 2021 <i>Kathmandu, Nepal</i>
<ul style="list-style-type: none">Developed different internal tooling to scale up Docsumo's ML infrastructure.Built an end-to-end document table extraction pipeline (<i>ensemble of model such as CascadeTabNet, DBScan clustering, table header detection, bbox detection, etc.</i>) 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 pipeline.Worked on transformer-based Language Model (<i>modified-BERT with 2d position embedding</i>) for key-value extraction that significantly increased accuracy to 90%+ on documents like <i>Invoices, Receipts</i>, etc.	
Co-Founder, Directory of Technology <i>MPercept Technology</i>	May 2018 – April 2019 <i>Kathmandu, Nepal</i>
<ul style="list-style-type: none">I was responsible for all the technical decisions for all the projects.During this time, I also helped co-found an AI community in Nepal, <i>AIDevNepal</i>, where we organized 14 Saturdays workshops teaching and mentoring 400+ students on Data Science and Machine Learning.Went to Amman, Jordan as a big-data consultant to Umniah Telecommunication 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.	
Software Engineer <i>Fusemachines</i>	Nov 2016 – Feb 2018 <i>Kathmandu, Nepal</i>
<ul style="list-style-type: none">Worked on research projects related to chatbot and Intelligent Character Recognition (ICR). We were able to achieve ICR accuracy of 70% using RNNs.Built recommendation engine as a POV for ML systems.Built Salesforce integration API for a client (<i>Enhatch</i>)	
Data Science Intern <i>Phunka Technologies</i>	Jun 2016 – Aug 2016 <i>Kathmandu, Nepal</i>
<ul style="list-style-type: none">Worked on data crawling and analytics system using pandas, scrapy, Django, etc.	
Miscellaneous <i>Miscellaneous</i>	2016– 2021 <i>Nepal</i>
<ul style="list-style-type: none">From <i>May 2019-July 2019</i> I taught Artificial Intelligence course for the final year B.E. Computer Engineering students (20) as a part-time lecturer at <i>Janakpur Engineering College, Kathmandu</i>. Course slidesFrom <i>August 2018-Feb 2019</i>, I taught 2 batches of students (10 per each) as a course instructor for Data Science at <i>MPercept Academy</i>.	

PROJECTS

- Docsumo: Document AI** | *Python, Flask, Docker, Pytorch, Tensorflow, mlflow, spacy* Nov 2019 – Dec 2021
- Built most of the initial infrastructure for end-to-end Machine Learning Engineering. Eg: Table Extraction Framework, Graph-based key-value extraction frameworks, etc.
- Damage Segmentation Detection** | *Python, PyTorch, Tensorflow, scikit-learn*
- This is a project collaboration between MPercept Technology and a Germany-based automobile insurance company to help identify damage segments from Vehicle images which are then used for calculating insurance costs.
 - Worked on image processing pipeline and built Deep Learning models such as Mask-RCNN, and U-Net to detect damaged segments. We were able to achieve **mAP score of 74%**.
- humT: Query by humming system, BE final year project** | *Python, numpy*
- Worked on a temporal pattern-matching algorithm like Dynamic Time Warping(DTW). Using DTW and audio segmentation, we were able to **correctly identify 15/20 songs**.
 - Published at ICTAES, humT GitHub repo
- anuwadak: Statistical Nepali-English text translation, BE third-year project** | *Python, numpy, Markov Models*
- Worked on Markov and N-gram models for predictive text generation. GitHub repo
- playx: open-source music assistant for Linux** | *Python, numpy, beautifulsoup*
- This project has nurtured **more than 200 stars** where I worked on cores such as crawler for songs+lyrics, song-caching mechanism, and string matching algorithm.
 - Developed Markov-model-based song recommendation algorithm that uses user logs to auto-generate playlists.
 - Experimented on item2vec model to embed songs and auto-generate playlists.
 - GitHub repo

TECHNICAL SKILLS

Languages+Technologies: Python, C/C++, Java, SQL, postgres, mongoDB, Docker, kubernetes, Google Cloud, AWS, S3

Web Frameworks: Flask, Django, FastAPI,

ML+NLP+CV Frameworks: scikit-learn, PyTorch, pytorch-lightning, Tensorflow, scikit-optimize, mlflow, open cv, mmcv, spacy, transformers, Large Language Models

Data Science Tools: pandas, numpy, matplotlib, seaborn, plotly, Jupyter Notebooks