Harshil Jain

hjain@ucsd.edu | https://harshiljain.in
Harshil Jain | jain-harshil | jainharshil | jain_harshil

Experience

Hyperlinks at appropriate places

- **Senior Machine Learning Engineer at Meta, Menlo Park, California** *Feb 2024 - Present*

  - Working on various machine learning use cases including recommendations, ranking within Facebook.

- **Applied AI/ML Associate at JPMorgan Chase & Co, Palo Alto, California** *Jun 2023 - Sept 2023*

  - Worked on building a news title-only clustering production grade ML model on large-scale news articles datasets.

  - Used large scale data processing frameworks such as Spark, AWS EMR for feature engineering. Built a clustered dataset of size 15000 news article using GPT-4 with LangChain and prompt engineering.

  - Accounted for system-design constraints by ensuring efficiency of model inference and model consistency with the ML pipeline.

- **Member of Technical Staff - 2 at Nutanix Inc.** *Jan 2021 - Sept 2022*

  - Worked on building VM-centric storage policies for managing VMs at scale. Developed APIs on the control plane side which included application of storage policies, their enforcement and reporting of compliance of the policies on the VMs.

  - Designed storage policy RBAC requirements and implemented the backend changes to support RBAC for VM-centric policies.

  - Developed a debugging tool, Doogle by deploying the Loki, Logstash and Grafana stack for efficiently parsing service log data. Ensured that the parsing of data was efficient by load-balancing it across multiple Logstash instances.

Education

University of California San Diego *Sept 2022 - Dec 2023*

Masters in Computer Science, *Specialization - Artificial Intelligence* | GPA: 4.0/4.0

Courses: Probabilistic Reasoning and Learning for AI, Recommender Systems and Web Mining, Learning Algorithms in Machine Learning, Statistical NLP, Convex Optimization Algorithms, Graduate Networked Systems, Numerical Linear Algebra

Indian Institute of Technology, Gandhinagar *July 2017 - Jan 2021*

Bachelor of Technology (B.Tech) with Honours in Computer Science and Engineering | GPA: 9.46/10.0

Internships and Research Projects

- **Injecting Common Sense Knowledge into Unsupervised Multilingual Pretraining Models** *Summer 2020*

  ***Guide:* Prof. Dr. Goran Glavaš, Data and Web Science (DWS) Group, Universität Mannheim, Germany**

  DAAD WISE Scholarship 2020 [**Letter of Award**] [**Scholarship Certificate**]

  - Trained adapter-based multilingual BERT and XLM-R on ConceptNet relations and thus injected knowledge in transformers.

  - Employed the transformers to train on multi-choice datasets in English and evaluated on the target language from ConceptNet.

  - Inferred that intermediate ConceptNet training improves the zero-shot transfer from English to the target language.

- **End to End Binarized Neural Networks for Text Classification** *Summer 2020*

  Researcher at NeuralSpace, London, United Kingdom

  - Modelled end to end binarized networks on 4 intent classification chatbot datasets.

  - Implemented binarized hyper-dimensional vectors for embeddings and achieved state-of-the-art results.

  - Concluded that binarized networks are time efficient and reduce memory footprint compared to non binarized counterparts.

- **Incremental Training for Image Classification of Unseen Objects** *Summer 2019*

  ***Guide:* Prof. Soumitra Nandy, Indian Institute of Science (IISc), Bangalore, India**

  Indian Academy of Sciences Fellowship 2019

  - Implemented the YOLO object detection algorithm in Tensorflow and curated a smaller dataset for various experiments.

  - Trained the YOLO algorithm for image classification of a subset of classes and incrementally trained the remaining classes.

  - Verified that incremental training can be used for the classification of unseen objects on a video surveillance camera.

Publications

- **End to End Binarized Neural Networks for Text Classification**

  Kumar Shridhar, **Harshil Jain***, *Akshat Agarwal*, Denis Kleyko*

  SustaiNLP 2020 Workshop on Simple and Efficient Natural Language Processing, ***EMNLP 2020***

- **Blind Motion Deblurring through SinGAN Architecture**

  Harshil Jain*, Rohit Patil, Indra Deep Mastan, Shanmuganathan Raman*

  Deep Internal Learning Workshop, European Conference of Computer Vision 2020 ***(ECCV 2020) (Oral)***

- **Generative Fashion for Indian Clothing**

  Harshil Jain*, Rohit Patil, Utsav Jethva, Ronak Kaoshik, Shaurya Agarawal, Ritik Dutta, Nipun Batra*

  8th ACM IKDD CoDS and 26th COMAD Young Researchers' Symposium Track ***(CoDS COMAD 2021)***

*Equal Contribution

Projects and Presentations

- **Exploring Quantum Machine Learning Algorithms** | *Guide: Prof.Anirban Dasgupta* *Aug '20 - Nov '20*
  - Reviewed papers on quantum recommendation systems by Kerenidis and Prakash and its classical analogue by Ewin Tang.
  - Implemented the classical recommendation system proposed by Ewin Tang in Python.
  - Deployed quantum generative adversarial networks in TensorFlow Quantum and trained QGAN on the MNIST dataset.
- **Generative Fashion for Indian Clothing** | *Machine Learning Course Project* *Apr '20 - Jun '20*
  - Collated a dataset of 12000 clothing images pertaining to Indian context with descriptions from Myntra and Amazon.
  - Proposed a two-step GAN model for generation of the segmentation map and for texture rendering based on descriptions.
  - Experimented with BERT, TF-IDF, Doc2vec embeddings for textual descriptions.
- **Blind Motion Deblurring through SinGAN Architecture** | *Guide: Prof. Shanmuganathan Raman* *Aug '19 - Nov '19*
  - Introduced an architecture for blind motion deblurring involving passing image through finer scale GANs.
  - Demonstrated how the proposed model captures the internal distribution and helps in deblurring the image.
  - Showed that the training independent solution to motion deblurring could learn features from a single image like SinGAN.
- **Neural Title Generation for Scientific Papers** | *NLP Course Project* *Aug '19 - Nov '19*
  - Built a model involving encoders and attention layers to predict titles given a scientific paper/article.
  - Strategized heuristics to make the predicted title catchy and defined a mathematical metric to measure title's catchiness.
  - Concluded that the proposed pipeline outperforms baselines and demonstrates summarization of documents by neural approach.
- **Email Parser** | *Hackathon at IIT Gandhinagar* *Jan '19*
  - Devised an email parser to extract data of seminars, talks and workshops using NLP.
  - Employed the Gmail API, Stanford NLP API and regular expressions in Python to extract relevant information.
  - Presented a live demo of the same to the Computer Science community at IIT Gandhinagar.

Technical Skills

Programming Languages	C, C++, Python, Golang, LaTeX, HTML, XML, MySQL, Bash
Machine Learning Libraries	PySpark,TensorFlow, PyTorch, NumPy, Keras, Pandas, Scikit-Learn, OpenCV
Machine Learning Frameworks	AWS EMR, Amazon EC2, Amazon SageMaker, AWS S3

Academic Achievements & Awards

- Awarded the **Best Video Award** for the research presented on end-to-end binarized neural networks at the Undergraduate Research Conclave 2020 at IIT Gandhinagar. *2020*
- Appeared on the **Dean's List** for **6** semesters for exemplary performance at IIT Gandhinagar. *2017-2021*
- Recipient of the prestigious **DAAD WISE Scholarship** for pursuing research in Germany. *2020*
- Bagged the most coveted **NTSE** (National Talent Search Examination) Scholarship and **KVPY** (Kishore Vaigyanik Protsahan Yojana) Fellowship by the Government of India. *2015*