Vasu Sharma

https://vasusharma.github.io/
sharma.vasu55@gmail.com | +1(412)-616-6880

EDUCATION

SCHOOL OF COMPUTER SCIENCE, CARNEGIE MELLON UNIVERSITY

Masters in Language Technologies
4.19/4.33 (Dept. Rank: 1)

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

B.Tech. in Computer Science and Engineering
Cum. GPA: 9.9/10.0

ST. COLUMBA'S SCHOOL

AISSCE (Class XII, CBSE)
Percentage: 97%
IIT-JEE All India Rank 165

AREAS OF INTEREST

Multimodal Foundation Models
Computer Vision
Natural Language Processing
Speech and Music Processing
Agentic AI workflows

RELEVANT COURSEWORK

Deep Reinforcement Learning (A+)
Neural Networks for NLP (A)
Deep Learning (A+)
Advanced Machine Learning (A+)
Advanced Multi Modal Machine Learning (A+)
Recent Advances in Computer Vision (A)
Natural Language Processing (A)
Visual Recognition (A)
Machine Learning Techniques (A)
Human Centered Computing (A*)
Human Cognitive Processes (A*)
Advanced Algorithms (A)
Data Structures and Algorithms (A)
Digital Signal Processing (A)
Probability and Statistics (A*)
Linear Algebra (A*)
Data Science Specialization(Coursera)

WORK EXPERIENCE

APPLIED RESEARCH SCIENTIST LEAD

Meta AI (Facebook AI Research)
FAIR team | Aug 2022 - Present| Menlo Park, CA, USA

- Working on creating multimodal foundational LLM models like Llama 3/4 and Chameleon. Experimenting with effective pretraining and customized finetuning approaches, Mixture of Expert models, data curation, studying scaling laws at trillion tokens and 100B+ parameters scale, speeding up inference and training and designing evaluation and safety tuning benchmarks
- Leading the team working on creating foundation models for enabling Multimodal Interactive conversational agents with full body avatars and full duplex expressive conversational capacity
- Researching text conditioned audio-video generation models for enabling scalable large scale content creation with long form coherent video generation
- Created a large scale Audio-Video self supervised learning based representation learning model called MaVIL which leverages masking and inter and intra modal reconstruction objectives to achieve state of the art performance beating even state of the art supervised approaches
- Created a text quality aesthetic score pipeline to perform effective data pruning allowing large language models to obtain state of the art performance despite using only 50% of the original data

APPLIED SCIENTIST

Amazon Alexa
Alexa AI team | Aug 2021 - Aug 2022| Sunnyvale, CA, USA

- Created a new benchmark for dialog enabled visual-language navigation as a part of the Alexa Prize Simbot challenge leveraging the multimodal data sources to faithfully navigate a virtual environment based on user instruction and designed benchmark models for the same
- Designed efficient multimodal transformers to speed up their training and deployment by improving the computational complexity of the self attention mechanism
- Video processing applications like video action recognition, video question answering, video summarization, moment retrieval etc working directly with compressed video streams
- Created a benchmark for cooperative heterogenous multi agent reinforcement learning platform including open sourcing the collected dataset and it's benchmark models

QUANTITATIVE RESEARCH ANALYST

Citadel LLC
Global Quantitative Strategies (GQS)| Aug 2019 - Aug 2021| Chicago, USA

- Worked on cross asset class Alpha Construction using hybrid linear and non-linear fitting techniques to best exploit statistical arbitrage opportunities in financial markets
- Built Automated fitting pipeline for incorporating advanced Machine Learning techniques into trading strategies
- Optimized existing machine learning framework to scale up to large volumes of incoming data and improve performance and speed

INTERNSHIPS

CITADEL LLC

Summer Intern, Machine Learning Team

Global Quantitative Strategies | May 2018 – Aug 2018 | Chicago, USA

- Worked on "Deep Neural Networks for Time series modelling of financial markets" and "Effective Feature scalability for Machine Learning models". In this project I explored a variety of Deep Learning models and effective training techniques to perform time series analysis on the large scale and highly noisy financial markets data. I also ensured that the models scale to arbitrarily large dimension feature sets.

UNIVERSITY OF TORONTO

Summer Intern, Machine Learning Team

Raquel Urtasun, Sanja Fidler | May 2016 – Jul 2016 | Toronto, Canada

- "FlowSeg: A Deep Learning based approach for simultaneous semantic segmentation and flow estimation from videos"
- The project focused on building Deep Convolutional Neural Network architectures to study the problem of Instance and Semantic segmentation of videos. We experiment with fairly advanced and novel Deep CNN architectures to jointly estimate semantic segmentation and flow from videos. The approach shows promising results on various datasets.

LOCALITE INC

Head of AI

2017-2018 | Remote

- Founding team at Localite Inc - a tours and activities marketplace for connecting people with local tour agencies and local tour guides. Raised $200k in pre-seed funding. Implemented the core recommendation systems, feed ranking and search retrieval systems

ABZOOBA INC.

Technical Consultant

Labhesh Patel | Aug 2016 – Jul 2017 | California, USA (working remotely)

- Worked on building "A Smart E-commerce Virtual Assistant". Implemented features like cloth parsing from images, similar image retrieval from a huge fashion catalogue and a state of the art Deep Recommender system.
- Implemented a Multi Turn Conversational Voice Agent to facilitate user interaction. Involved the use of Memory Networks and a soft attention mechanism over previous queries and responses to figure out the best response to a given user query.
- Also worked on "Query based document retrieval" by learning rich semantic document embeddings using a deep LSTM pipeline and using these to find the match the queries to relevant documents
- "Abstractive summarization using Attention based encoder-decoder networks" : Worked on building a deep residual LSTM pipeline which used temporal attention over both encoder and decoder networks to generate an abstractive summary of documents.

CARNEGIE MELLON UNIVERSITY

Summer Intern, School of Computer Science

Bhiksha Raj, Rita Singh| May 2014 – Jul 2014 | Pittsburgh, USA

- "Deep Recurrent Gated Neural Networks for Dynamic Audio Denoising"
- The project focused on construction of a Deep Recurrent neural network to achieve signal reconstruction by denoising noise corrupted signals by dynamic spectral subtraction.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL)

Summer Intern, Machine Learning and Optimization Lab

Martin Jaggi | May 2017 – Jul 2017| Lausanne, Switzerland

- "Learning semantic sentence embeddings using Hierarchical Convolutional Neural Networks
- In this project I worked on creating Deep Hierarchical Convolutional Neural Networks to learn unsupervised semantic textual embeddings. The representations learnt capture both local and global textual information and hence perform competitively against major state of the art approaches on both supervised tasks like sentiment analysis and unsupervised ones like similarity matching.

XEROX RESEARCH LABS, EUROPE

Research Intern, Computer Vision Team

Diane Larlus, Albert Gordo | Sep 2015 – Dec 2015 | Grenoble, France

- Worked on "Large Scale Image Recognition using Deep Convolutional Neural Nets"
- The projects primarily focused on constructing Deep Learning frameworks for Image Recognition. Worked on designing some novel Deep Learning frameworks for the image recognition task on the ImageNet dataset. Also made extensive use of GPUs and the popular Caffe library for training Deep Convolutional Neural Nets.

SELECTED PUBLICATIONS

- "DINOv2: Learning Robust Visual Features without Supervision"
  Published at Transactions of Machine Learning Research, 2024

- "MAViL: Masked Audio-Video Learners"
  Published at NeurIPS, New Orleans, 2023

- "A Picture is Worth More Than 77 Text Tokens: Evaluating CLIP-Style Models on Dense Captions"
  Published at CVPR, Seattle, USA, 2024

- "Branch-Train-MiX: Mixing Expert LLMs into a Mixture-of-Experts LLM"
  Published at Conference on Langauge Modelling (COLM 2024), 2024

- "Demystifying CLIP data"
  Published at ICLR, Vienna, 2024

- "Chameleon: Mixed-Modal Early-Fusion Foundation Models"
  Arxiv, 2024

- "Alexa Arena: A User-Centric Interactive Platform for Embodied AI"
  Published at NeurIPS, New Orleans, 2023

- "Scaling Autoregressive Multi-Modal Models: Pretraining and Instruction Tuning"
  Under review at ICCV 2025

- "FLAP: Fast Language-Audio Pre-training"
  Accpeted at ASRU, Taiwan, 2024

- "Attend, Attribute & Attack Model: Multimodal Adversarial Attacks to Investigate Vulnerability in VQA Models"
  Under Review at ICCV 2025

- "Tweet Based Reach Aware Temporal Attention Network for NFT Valuation"
  Published at EMNLP, Abu Dhabi, 2022

- "E-ViLM: Efficient Video-Language Model via Masked Video Modeling with Semantic Vector-Quantized Tokenizer"
  Published at WACV, Hawaii, 2024

- "CHMARL: A Multimodal Benchmark for Cooperative, Heterogeneous Multi-Agent Reinforcement Learning"
  Published at Robotics Science and Systems (RSS), New York, USA, 2022

- "Attend and Attack: Attention Guided Adversarial Attacks on Visual Question Answering Models"
  Published at NeurIPS 2018 (Security in Machine Learning Track), Montreal, Canada

- "PISA: PoIncaré Saliency-Aware Interpolative Augmentation"
  Published at Interspeech, Korea, 2022

- "Community Regularization of Visually-Grounded Dialog"
  Published at International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, Canada, 2019

- "Multimodal Behavioral Markers Exploring Suicidal Intent in Social Media Videos "
  Published at 21st ACM International Conference on Multimodal Interaction (ICMI) 2019

- "BioAMA: Towards an End to End BioMedical Question Answering System"
  Published at Annual Meeting of the Association for Computational Linguistics(ACL), BioNLP track, Melbourne, Australia 2018

- "Mind Your Language: Learning Visually Grounded Dialog in a Multi-Agent Setting"
  Published at CVPR, VQA Challenge and Visual Dialog Workshop, Salt Lake City, USA, 2018

- "Induced Attention Invariance: Defending VQA Models against Adversarial Attacks"
  Published at NeurIPS 2019 (ViGIL), Vancouver, Canada