ANSH MITTAL

Mountain View, CA | (213) 573-9188 | manshm1811@gmail.com | linkedin.com/in/mittalansh/ | github.com/AnshMittal1811 | medium.com/@anshm18111996

EXPERIENCE

Guidewire Software Inc

- **Machine Learning Engineer** | *San Mateo, CA* — **April 2024–Present**
  - Orchestrated Incident AI-data pipeline with DocLLMs for grounding Incident data in product Cyence
  - Implemented Apache Medallion architecture for time-travel and near-real time score and loss simulation using DBT Models
  - Implemented customized OCR-CNN workflow for Payment Processor for firmographic loss simulation using EMR Serverless
  - Proposed exposure signal for 3D Wifi Radio S/N to profile cyber attacks with Wireshark and CVE-DB dataset

- **ML Engineer Intern** | *San Mateo, CA* — **May 2022–August 2022**
  - Containerized Data Quality framework (PyDQ) for data artifacts (Datalake, PostgreSQL, S3) & deployed on ECS
  - Added statistical tests (K-S & χ-squared) to monitor data distribution drifts and exceptions
  - Designed Semantic Segmentation Pipeline to automate Fire-Hydrant detection reducing 25% cost & 70% time in HazardHub
  - Demonstrated use-case of Neural Radiance Fields team-wide for various geopatial and radiowave applications

Remix Inc

- **Senior Computer Vision Engineer (Founding)** | *Palo Alto, CA* — **September 2023–December 2023**
  - Innovated 3D Dynamic View Reconstruction Pipeline and reduced latency to 38 ms/27 FPS
  - Compiled Gstreamer-, DNN-& CUDA-based OpenCV on Nvidia architectures for ARM64, AMD64, and X86
  - Trained LeRF with vision COLMAP embeddings for custom scenes using NeRF-studio for 97.56% recognized scenes
  - Rendered real-time Multi-Sphere Images with Human Gaussian Splats and pano-object detection with YOLO-GPT

KNOW Corp

- **AI/ML Research Engineer** | *Los Angeles, CA* — **January 2023–May 2023**
  - Personalized GenerativeQA model using Prompt Engineering (GPT-NeoX) (METEOR: 22.34; STS: 0.87)
  - Optimized training load and time with PEFT LoRA and Rotary Position Embeddings for long context window
  - Prepared PoC for 3D Neural Human Avatar & synchronized Text-to-Speech (SpeechT5) for FastAPI deployment

USC Research Labs

- **AI/ML CV Research Engineer, USC Information Science Institute** | *Marina Del Rey, CA* — **August 2023–July 2024**
  - Surveyed NeRFs and their applications in Neural Rendering here: NeRFs (Neural Radiance Fields): Past, Present, and Future
  - Applied 3D Reconstruction techniques for Astronomy, Geospatial, Medical, Autonomous driving, and Radiowave signals

- **Graduate Research Assistant, USC ISI & USC IMSC labs** | *Los Angeles, CA* — **January 2022–December 2022**
  - Annotated $\tilde{2}.5$ hours of video for Heavy-Vehicle Detection for different lighting conditions in Pascal VOC format
  - Street object classification with fine-tuned vision models (MobileNet, ResNet, ViT) on edge devices (accuracy: 83% (23ms))
  - Engineered 3DMM features regression (62-D) baseline w/o explicit Face Landmark detection
  - Demonstrated I-SPLIT algorithm for Split-Computing on edge-devices using CUmulative Importance Map (CUI GradCAM)
  - Customized 3D-Face Alignment baseline while comparing it with previous research (Normalized Mean Error: **10.7**)

Sociometrik

- **Data Scientist (GIS, Computer Vision)** | *Delhi, India* — **September 2020–August 2021**
  - Led development of Metal Roof-Detection pipeline (Avg IoU: 0.698) on AWS EC2 using AWS Lambda
  - Orchestrated Super-Resolution pipeline with LandSAT GIS-image data using PyTorch UDF (Horovod)
  - Built feature extraction (IoU: 0.79) & car segmentation (F-score: 0.73) workflow for multiple classes (Prec.: 0.87, Recall: 0.79)

Indian Institutes of Technology Delhi

- **Research Scientist (Game, Reinforcement Learning)** | *Delhi, India* — **January 2020–October 2020**
  - Hosted Unity Web Applications for Cyber Security & Threat Prevention using GraphQL at IIT Research Lab
  - Engineered Infotainment-based game to teach Blockchain to amateurs and professionals w/ Sequence-to-Sequence Chatbot

TECHNICAL SKILLS

- **Languages & frameworks**: Python, SQL, C++, C#, Kotlin, HCL, Rust, Flask, LaTeX, Bash, R
- **Libraries**: Pandas, Matplotlib, PySpark, TensorFlow/Keras, PyTorch/ONNX/JAX, OpenCV, LangChain, HuggingFace
- **Developer Tools**: AWS (EC2, ECS, EKS, S3, EMR (Serverless), RDS, Redshift, Lambda, Glue, Athena, Sagemaker), Docker, Kubernetes, Terraform/Terragrunt, AirFlow, Jenkins, Cassandra, DynamoDB, DeepSpeed, TensorRT-LLM
- **Skills**: NLP (BERT, CLIP, RAG, RLHF, FLARE, CoT, LLM), Computer Vision (Diffusion, NeRF, GAN, YOLO, LVLM), Multimodal Learning, Computer Graphics, Unity AR/VR, CUDA, TensorRT, MSVC, GDB, LLVM, MLIR, GCC, CMake, JIT

PUBLICATIONS [Scholar]

Detecting pneumonia using convolutions & dynamic capsule routing... | *Sensors, MDPI 2020*   (cited: 151)(paper)
- Detected: Pneumonia in Chest X-rays; Used: Capsule Networks in conjunction with CNNs; Accuracy Obtained: **96.36% (2.8%↑)**

Cybersecurity Enhancement through Blockchain Training (CEBT)... | *IJIM Data Insights, Elsevier 2021*   (cited: 66)(paper)
- Developed NAF-based (Actor-Critic Variant) Cont. NPC Adaptiveness Algorithm to augment Game Design & Feedback Mechanics

Data Augmentation Based Morphological Classification of Galaxies... | *ESIN, Springer 2019*   (cited: 38)(paper)
- Classified: Galaxy Morphologies; Used: Machine Learning, CNNs with L1 and L2 Regularization; Accuracy: 97.92% (0.4%↑)

NeRFs (Neural Radiance Fields): Past, Present, and Future | *Arxiv Preprint 2023*   (cited: 14)(paper)
- Classified various 3D Computer Vision & Generative AI with History of model-based and image-based Novel View Synthesis
- Surveying various MPI & NeRF architectures, models, and extensions in terms of objective functions, datasets & evaluation metrics

AiCNNs (Artificially-integrated Convolutional Neural Networks) for Brain Tumor Prediction | *EAI 2019*   (cited: 11)(paper)
- Classified: 3 types of brain tumors; Used: ML Models, Ensemble CNNs (Regularization); Accuracy: 99.49% (3.72%↑).

Guess who?-A serious game for cybersecurity professionals | *9th International Conference, GALA 2020*   (cited: 7)(paper)
- Developed a cyber security threat game with clustering based on user feedback (Firebase DB)

SAVCHOI: Detecting Suspicious Activities using Dense Video Caption... | *Arxiv Preprint 2022*   (cited: 4)(paper)
- Detected: Suspicious activities (Videos); Used: BMT and DETR (Human-Object Interactions) (ResNet50 bb) + Genetic Algo & **BERT**; BLEU@1: 14.78, BLEU@2: 12.73, BLEU@3: 10.91, BLEU@4: 7.11, METEOR: 16.27, Detection Accuracy: 96.84%

FuNet-40: fundus disease/abnormality classification using ensemble of fine-tuned pretrained convolutions | *T&F'24*   (paper)
- Trained 56 models for 15-ensemble models for 40 different fundus diseases/abnormalities; Classification Accuracy: 99.6 %;
- Visualizations created in 3-D and 2-D for each image using T-SNE and PCA to depict distinct decision boundaries between diseases

PROJECTS [GitHub]

Research Paper Implementations (NLP) | *PyTorch, TensorFlow, DeepSpeed, HuggingFace*   (Gist) **May 2023–March 2024**
- Implemented LoRA (QLoRA, LongQLoRA), Attention, CoT (& ToT), RAG, HNSW for various Transformer architectures
- Leveraged HF-transformers, bitsandbytes, and Peft for LongNet, LLAMA, RetNet, RWKY

Organ Segmentation Using MonAI | *MonAI, 3D U-Net, CT, Pytorch, Segmentation*   (project) **July 2023-August 2023**
- Developed end-to-end Whole Body Organ Segmentation Pipeline using Monai and Whole Body-CT Segmentation
- FineTuned 3D-UNet for Liver Semantic Segmentation & Detection with refactored Monai-Torch Pipeline (**Loss: 0.29, DICE: 0.87**)

Real-Time Data Streaming with Object Detection | *PySpark, Kafka, StreamLit*   (project) **May 2022–June 2022**
- Deployed real-time data pipeline using Kafka to capture and stream data, enabling immediate analysis of incoming information
- Configured Object Detection model for streamed data using PySpark, integrated, and tested with StreamLit frontend

Few Shot Learning Approach to Dynamic Intent Satisfaction | *GPT3, GPT-Neo, NLP*   (video) **January 2022–May 2022**
- Proposed LLM-based problem–Intent Satisfaction (fulfill complex intents in slot filling systems) (BLEU (GPT_{dv}): 0.73)
- Curated function definition and docstrings benchmark dataset of 500,000 tokens for Intent Satisfaction

Computational Drug Discovery for Alzheimer's | *ChEMBL, SMILES, RDkit, PadelPy*   (project)(App) **June 2021–August 2021**
- Classified Acetylcholinesterase (AChE) activity for Alzheimer's Drug Discovery (with Lipinski and SMILE notations) using pIC50 Regression Analysis with LGBM Regressor (R^2 : 0.6) (LazyRegressor package) and StreamLit UI

EDUCATION

University of Southern California | *Los Angeles, CA*   **August 2021-May 2023**
- ***MS in Computer Science****; Courses: Artificial Intelligence, Algorithms, Deep Learning, Applied NLP, Advanced Computer Vision, Autonomous Cyber Physical System, Information Retrieval & Web Search Engines*

Guru Gobind Singh Indraprastha University | *Delhi, India*   **August 2015-May 2019**
- ***Bachelors of Technology in Computer Science and Engineering****; Rank: 5; Courses: Operating Systems, Compiler Design, Computer Networks, Artificial Intelligence, Data Mining, Machine Learning (Python), Discrete Mathematics, C++*

Extra Courses and Summer Schools
- Qiskit Global Summer School (QGSS'21)   (certificate) **May 2021–June 2021**
- Introduction to Quantum Computing (QxQ By Coding School)   **October 2020–May 2021**

HONORS & AWARDS

- Viterbi Grad Hack Winners (DeepFake Detection (w/wo LipSync) Chrome Extension)   (project) (video) **April 2023**
- Smart India Hackathon'19 3rd Runner-Up (INSUrance Recommendations with Integrated Mixed-reality)   **April 2019**
- Smart India Hackathon'18 2nd Runner-Up (LOgical Graph Augmented Virtual Indian Map)   (project) **April 2018**