HIMANSHU GOYAL

+1 773-231–3457, goyalhimanshu414@gmail.com, Google Scholar

Work Experience (7+ years)

JCI (Acquired FogHorn.io) **Data-Scientist** *Oct'20 - Aug'24*

- **OpenBlue-Building-AMA (OBAMA):** Actionable insights generation and NLQ tool for technical dashboards using LLMs
  - Developed novel data insight generation tool using Langchain's MRKL agent with **Re-Act** framework & **Zero & Few-shot COT** prompting techniques; Integrated **RAG** on curated knowledge-base using MMR and SWR techniques
  - Implemented Auto-CoT for process self-improvement & prompt **injection detection** framework to enhance security.
  - Patented the process & showcased in *JCI's Tech-Challenge'24* finals as one of the top eight, out of 3,000+ submissions.
- **Asset & Points Mapping for Buildings:** Trained LLMs to predict standard names of points based on short-names & units.
  - Trained LLMs on Azure Databricks platform & developed solution to perform **RLHF** using techniques like **DPO & PPO**.
  - Developed features for **model explainability** for mapped classes & online training of deployed models to address *data-drift*.
  - Expanded solution for **multiple languages** enhancing overall sales & reducing onboarding time from 6 months to 1 month.
- **Selective Target Updation:** Developed a **new DL training method** to enable training DL-model exclusively for selective rare classes. Modified Focal Loss function to restrict loss back-propagation only for rare classes. Trained *EfficientDet* model and achieved *20% Gain in mAP* on COCO dataset's rare classes.
- **On-Devise AI for CV problems:** Developed industry acclaimed computer vision solutions, designed to run in constrained edge environment for realtime monitoring of Flaring, BOP & WorkerSafety-norms-violation in Oil & Gas industry.
  - Trained SOTA computer vision model like **EffcientDet**, *YOLOx, Faster R-CNN* for object Detection tasks; **Mask R-CNN**, DeepLab for Mask-Segmentation task; Performed hyper-parameter optimization during Transfer Learning.
  - Implemented **Knowledge Distillation** for model Optimization along with Quantization of models using OpenVino, ONNX and tensorRT for faster inferencing; Created custom Docker image for running inferencing on Jetson GPU edge.
  - Implemented **CI/CD pipeline** for edge deployement & Training pipeline deeply integrated with MLflow and Databricks for experiment tracking; Contributed **MLflow integration** to YOLOx repo having **9k+ stars.**
  - Productionized at 100+ sites having annual revenue of $0.5mil with BOP receiving *IoT Edge Computing Excellence Award 2021*, WorkerSafety receiving *Nascom AI GameChanger Award 2024* & Flare *granted patent* and *published in AdConIP.*
- Developed efficient **GPU Utilization Strategy** for optimizing GPU concurrent usage by multiple models in Edgeml® docker container; Implemented **CycleGAN** to convert synthetic images to real-looking images in unsupervised settings.

TwoPaksh Tech **Co-Founder** *Oct'19 - Sep'20*

- Created a **face-recognition** based **IoT surveillance system** to track and authenticate user entry in indoor gated premises. Optimized system performance by running **MTCNN + FaceNet** JS model on-device to outputs 128-d face-embedding, avoiding streaming video feed to cloud. System generated useful KPIs like footfall-heatmaps, user serving time etc.

Citigroup **Technical Analyst** *Aug'17 - Oct'19*

- Developed a low-cost, s3 object storage application using **Java SpringBoot/MVC** for microservices API & **Angular** for UI.
- Developed an ML powered **recommendation engine** for assistance in data entry from trade documents. Predicted the value for missing fields with **precision of 0.9**. Received Engineering Excellence, **Citi Applause** & **Citi Silver Award**.

Technical Proficiency & Publications

- **Technology:** Python, TensorFlow, Pytorch, OpenCV, NumPy, Pandas, Java, Docker, SQL, OpenVino, Angular, Azure/AWS
- **Theoretical:** Machine-Learning, Computer Vision, NLP, Deep-Learning, Generative-AI, ML System in Production
- **Papers:** *"Deep Learning based Flare Image Analytics at the Edge",* 2022 IEEE Int'l symp. on AdConIP, Vancouver, Canada.
- **Patents:** i.) Granted for developing Computer Vision based *"Flare Monitoring System and Method"* (US2023/0272910)
  ii.) Filed for *"Gen-AI Based Building System with Analysis & Contextual-Insight Generation"(Pending)*

Education

- **M.S. in Computer Science, The University of Chicago, IL**
  Coursework: Advanced Programming, Computer Systems, Algorithms, Discrete Mathematics
- **B.Tech. in Textiles, IIT-Delhi, India — GPA: 8.0/10** *Aug'13 - July'17*
  Relevant Coursework: Data Structures & Algorithms, Linear Algebra, Machine Learning, System Designing

Certifications

- Coursera: Deep Learning Specialization, ML Engineering for Production (MLOps) Specialization, CitiGroup: CFA-Level-2 Passed, Engineering Excellence level-2.