SHANTANU AGARWAL

@ a.shantanu08@gmail.com (413) 210 4842 ashantanu.com profileshantanu github.com/ashantanu

EDUCATION

MS, Computer Science *GPA 4.00*
CICS, University of Massachusetts Amherst *Feb 2021 - Dec 2022*
Courses: Intelligent Visual Computing, Neural Networks, Advanced NLP, Algorithms for Data Science, Probabilistic Graphical Models

BTech, Mathematics and Computing *CPI 9.22*
Indian Institute of Technology Guwahati *Jul 2013 - Jun 2017*
Minor in Product Design
Courses: Data Structures and Algorithms, Networks, Databases, Statistical Methods and Time Series Analysis, Advanced Statistical Algorithms, Probability Theory, Optimization, Discrete Maths, Scientific Computing, Stochastic Calculus, Product Design, Design Management

PUBLICATION

Thamizharasan, V., Liu, D., **Agarwal, S.**, Fisher, M., Gharbi, M., Wang, O., Jacobson, A. and Kalogerakis, E.. VecFusion: Vector Font Generation with Diffusion. **CVPR, 2024**
Raghu, D.*, **Agarwal, S.***, Joshi S., Mausam. End-to-End Learning of Flowchart Grounded Task-Oriented Dialogs. **EMNLP, 2021** ** Equal Contribution*

EXPERIENCE

Staff Engineer, AI, *Balbix, San Jose* *Jun 2024 – Present*
- Developed world's first cybersecurity AI Assistant, BIX, overseeing a team of 5 engineers. Spearheaded the implementation utilizing LangGraph, LangServe, Langfuse, AWS Bedrock, and PGVector. The release garnered coverage from major news outlets, including Washington Times and CSO Online
- Established comprehensive LLM evaluation framework for metrics-driven model selection, including custom benchmarks and automated training pipelines, leveraging Airflow, Ludwig, Unsloth, vLLM and MLFlow
- Designed an AI-driven cybersecurity reporting system to automate report generation and notifications

Sr. Engineer, AI, *Balbix, San Jose* *Feb 2023 – May 2024*
- Architected dashboard for clients to select and track progress on Vulnerability Remediation
- Late-Interaction retrieval methods to map software to appropriate Common Platform Enumeration
- Network Risk modeling for cybersecurity risk, propagation of risk and known attack patterns
- Shapley regression to estimate risk reduction potential of vulnerability resolution

AI/ML Engineer Intern, *Balbix, San Jose* *May 2022 – Aug 2022*
- Extensive EDA on publicly available cyber security datasets - CVE, CPE, CWE and ATT&CK
- Finetuned NLP model for classifying unstructured vulnerability text with 85% accuracy beating SoTA
- Deployed end-to-end ML pipeline for training and inference on cloud

Project Scientist, *Indian Institute of Technology Delhi* *Mar 2020 – Jul 2021*
- Crowdsourced a novel task oriented dialog dataset with 1,369 dialogs using Amazon MTurk
- Implemented dataset baselines using TF-IDF, MemN2N, GPT2 and retrieval based generative model; used two data split setting: seen and unseen flowchart; achieved BLEU score of 19.46 and 16.31 respectively

Associate - Equities Strategist, *Goldman Sachs, Bangalore* *Jun 2017 – Feb 2020*
- Built real-time tracking tool for cash impact of firm-wide trades and lockup under client protection rules
- Implemented a linear programming optimizer in Java for moving assets across different clearing locations and funding facilities to satisfy global client requirements while adhering to rules of each market

PROJECTS

Multi-Modal Multi-Task Learning for Ego4D Dataset, *Meta Reality Labs* *Skills: Pytorch, Python*

Implemented a multi-modal multi-task transformer for the Natural Language Query task on the Ego4D dataset; this project involved setting up data processing pipelines for videos, audio and text, extensive data analysis, implementing training and inference on the largest multi-modal egocentric video dataset

Deep Learning for Diamond Cutting, *UMass* *Skills: Pytorch, Python, Blender, 3D Computer Vision*

Synthesized a novel dataset and implemented 3D computer vision models based on PointNet++, MeshNet and MLP for predicting placements of fine cut diamond within a raw diamond with minimum impurities

Transductive Few-Shot Learning, *UMass* *Skills: Pytorch, Python, Machine Learning*

Used ResNet, EfficientNet and Vision Transformer for supervised, unsupervised and transductive unsupervised learning for few-shot image classification on miniImageNet dataset

SKILLS

Python, **PyTorch**, Tensorflow, JAX, C++, C, Java, MySQL, React, Next.js, Django
LangChain, LangGraph, MLFlow, AWS (Sagemaker, Bedrock), DSPy, Kubernetes, Docker, Airflow