

Aakanksha Sanctis

ars2352@columbia.edu | (732) 522-6403 | <https://www.linkedin.com/in/aakanksha-sanctis2020/> | <https://github.com/AakankshaSanctis>

EDUCATION

Columbia University

M.S in Computer Science | Machine Learning Track | GPA: 3.875/4.0

New York, NY
Sep 2022 - Dec 2023

Birla Institute of Technology and Science (BITS) Pilani

B.E. (Honors) Computer Science; M.Sc.(Hons.) Economics | CGPA: 9.1/10

Goa, India
Aug 2015 - Aug 2020

EXPERIENCE

Synopsys Inc. | Digital Implementation - Optimization Team (EDA Group)

ML Technical Intern

Sunnyvale, CA
May 2023 - Present

- Building a self-supervised transformer model with multivariate masked modeling using BeRT architecture for combinational circuits in chip design. Finetuning to predict gate sizes for combinational circuits with multiobjective optimization on power, performance and area

Columbia University | Prof Junfeng Yang

Research Assistant

New York
Nov 2022 - Present

- Injecting backdoors in language components of large vision-language models like CLIP using reasonable natural language rationale prompts for maximizing zero-shot attack accuracy and investigating robustness.

Amazon | International Machine Learning Team

Software Development Engineer II

Bangalore, India
Aug 2020 - Aug 2022

- Led the ML stack development for the Data Discovery project to rank and recommend 300k datasets for Amazon's biggest data lake using Learning to rank models with an Elasticsearch backbone. CTR increased from 68 to 75% with initiatives such as entity representation, keyphrase correlations, relevance modeling, query understanding and featureset explanations for search for increased interpretability.
- Created large scale automated metrics computation, monitoring and analytics visualization widget for team owned ML CI/CD platform for production ML models leveraging 10+ advanced AWS technologies. Developed features for multiple internal services for Spark based big data processing and model retraining workflows for ML automation.

Amazon | International Machine Learning Team

Applied Scientist Intern

Bangalore, India
Jan 2020 - Jul 2020

- Enhanced dual stage sourcing and selection ranking utilizing LightGBM models and devised an explainability pipeline based on top of SHAP for model predictions for deals in Amazon sales. Net business compliance increased by 30%.

Maastricht University | Institute of Data Science

Research Intern (Bachelor Thesis)

Maastricht, Netherlands
Sep 2019 - Dec 2019

- Constructed a low cost, accurate crowdsourcing methodology for lung cancer nodule annotation by non experts using segmented lung CT scans and optimized density based clustering algorithms. Additionally included a human in the loop based active learning workflow to train a 3D CNN nodule detection network using the collected annotations.

Deutsche Bank | Transaction Accounting Department

Summer Intern

Pune, India
May 2019 - Jul 2019

- Delivered a web application for all securities and derivative markets along with risk management in a fifteen member team. Frontend was dependent on Angular 5 along with a Java backend with Spring MVC framework and Hibernate

SKILLS

Languages: Python, Java, PySpark, SQL, C, C++, CUDA

Tools and Frameworks: Elasticsearch, Pytorch, Tensorflow, Ray, HuggingFace, Spark, Hadoop, Spring, Guice

AWS Technologies: EMR, EC2, Sagemaker, Lambda, Opensearch, S3, DynamoDB, Redshift, Glue, Step Functions, CloudWatch, Cloudformation, Quicksight, Amazon Athena

Relevant Coursework: [Machine Learning](#), [Neural Networks for Deep Learning](#), [Deep learning for Computer Vision](#), Natural Language Processing, [High Performance Machine Learning](#), [Advanced Spoken Language Processing](#), Information Retrieval, Data Structures and Algorithms, Database Management Systems, Operating Systems, Theory of Computation, Computer Architecture, Probability and Statistics, Applied Econometrics, [ML Safety](#)

Interests: Large Language Models, Robustness, Trustworthy AI, Deep Learning acceleration and model scaling, MLOps, Search and Recommendations

PROJECTS

Visual prompts using data augmentations for robust out-of-distribution image classification

Course Project (Columbia University) | Fall 2022 | [Neural Networks and Deep Learning](#)

- Tunable Visual prompting using data augmentation techniques such as DeepAugment, CutMix and CutOut for robust predictions for vision-language models like CLIP. Performance evaluations led to 2-3% improvement on different corruptions of CIFAR-10 dataset

Seeing is not believing: Privacy preserving facial manipulation using adversarial mask generation and diffusion models

Course Project (Columbia University) | Fall 2022 | [Deep learning for Computer Vision](#)

- Identified salient features in input images and generated adversarial masks using various techniques such as saliency gradient maps, GRAD-CAM and random patch masking. Created non-private representations of the input images using latent diffusion models, so that private information is not transmitted to downstream tasks such as FaceNet's recognition model.

Distributed Heterogeneous Training for Large Language Models using Ray and DeepSpeed

Course Project (Columbia University) | Spring 2023 | [High Performance Machine Learning](#)

- Conducted ablation studies for exploring efficient heterogeneous (CPU + GPU) distributed training for language models such as BeRT and RoBeRTa over different factors such as batch size, number of CPU/GPU parallel workers etc. Ray was used for parallelizing CPU processes and Deepspeed's ZeRO optimization was used for data parallelism along with mixed precision training for sentiment analysis.

OTHERS

- **Talks:** "Efficiently building ML workflows at scale for production pipelines" (*Analyticon Conference, Amazon*), "Personalized BI dashboards for hands-off-the-wheel ML workflows" (*HOTW Workshop, Amazon*)
- **Mentorship:** Mentored 3 interns in International Machine Learning team at Amazon with full time offers
- **Lead Organizer:** 1.) Search, Recommendation and Personalization Interest Group, 2.) Coding Excellence hour, *International Machine Learning, Amazon*
- **Teaching Assistant:** Advanced Analytics (Columbia University), Computer Programming (BITS Goa), Applied Econometrics (BITS Goa), Financial Management (BITS Goa)