Janvi Thakkar

(+44) 7308908696 | janvi.thakkar22@imperial.ac.uk | jvt3112.github.io | jvt3112 | janvi-thakkar-9b6004170

Education

Imperial College London — *London, United Kingdom*

Masters in Computing, Artificial Intelligence and Machine Learning (Pass with Distinction) — *Sep 2023*

Indian Institute of Technology (IIT) Gandhinagar — *Gandhinagar, India*

B.Tech. with Honors in Computer Science and Engineering (Cumulative Performance Index (CPI): 8.99/10) — *Jul 2022*

Publications

Janvi Thakkar, Giulio Zizzo, Sergio Maffeis **Differentially Private and Adversarially Robust Machine Learning: An Empirical Evaluation**, 5th AAAI Workshop on Privacy-Preserving Artificial Intelligence (**PPAI-AAAI'24**) [pdf]

Janvi Thakkar, Giulio Zizzo, Sergio Maffeis **Elevating Defenses: Bridging Adversarial Training and Watermarking for Model Resilience**, 2nd workshop on Deployable AI in Conjunction with AAAI 2024 (**DAI-AAAI'24**) [pdf]

Janvi Thakkar, Devvrat Joshi **FedSpectral+: Spectral Clustering using Federated Learning**, 3rd workshop on Graphs and more Complex structures for Learning and Reasoning in Conjunction with AAAI 2023 (**GCLR-AAAI'23**) [pdf]

Janvi Thakkar, Devvrat Joshi **k-Means SubClustering: A Differentially Private Algorithm with Improved Clustering Quality**, Workshop on Privacy Algorithms in Systems (PAS), International Conference on Information and Knowledge Management (**CIKM 2022**) [pdf]

Experience

Graduate Software Engineer | Ocado Technology, London, UK — *Sep 2023 - present*

- Leveraged Google BigQuery and Cloud Storage APIs to architect a components for data platform, seamlessly integrated with Datastore.
- Enabled self-service asset creation and service account management in GCP via a user-friendly portal. Automated resource management with internally generated Terraform code, supporting over 10 cross-functional teams. Incorporated Azure OAuth for enhanced security.
- Utilized Datadog and dbt's observability tool for comprehensive data monitoring, ensuring robust data health.
- Orchestrated GitLab CI pipelines across multiple projects, streamlining development workflows and enhancing project reliability.

Data Scientist Intern | Decimal Point Analytics, India — *May 2021 - Jul 2021*

- Enhanced performance of text summarization models on proprietary company data. Fine-tuned an abstractive model, BART, and an extractive model, BertSum. Achieved Rogue-1 score of 54.53 on BertSum and 49.83 on BART model.
- Conceptualized and implemented an API using the Flask framework, utilized by 150+ employees within the news summarization department

Summer Research Associate | Lancaster University, UK — *May 2020 - Jul 2020*

- Learned models on-line for a large number of agents (swarm) and used it for on-line planning under strong real-time constraints.
- Implemented the Monte Carlo Tree Search (MCTS), a simulation-based approach for the continuous action space for making sequential decision.
- Evaluated the technique in an "infiltration game," where an agent's goal is to reach the target guarded by an unknown swarm without pretraining.

Projects

Unified Defense: Adversarial Training, Watermarking & Privacy [MSc Thesis] | Imperial — *Mar 2023 - Sep 2023*

- Proposed a unified defense strategy to simultaneously address the security and privacy challenges faced by machine learning models.
- Designed a novel framework to integrate adversarial training with adversarial watermarks to fortify against evasion attacks, providing SOTA results over the existing techniques.
- Addressed the privacy concerns raised for the DP-Adv technique and provided evidence demonstrating its efficacy in maintaining data privacy

Analysis of Image Generation using Scenegraph [Publication] | IIT Gandhinagar — *Jan 2021 - Apr 2021*

- Analyzed and optimized the coherence between the text-to-scenegraph and image-to-scenegraph conversion pipelines to ensure consistency in the final image outputs.
- Identified 3 limitations within the current methodology and suggested a potential future direction for improvement.

Geometrical Homogeneous Clustering for Image Data [Publication] | IIT Gandhinagar — *Jan 2021 - Apr 2021*

- Proposed a novel approach to reduce an image dataset using a Geometrical Homogenous Clustering (GHCIDR).
- Acquired an accuracy of 99.35 % and 81.10%, and a training data reduction of 87.27% and 32.34%, on MNIST, and CIFAR10 respectively.

Snappy - Command line tool for Snapshot Management [Repository] | IIT Gandhinagar — *Sep 2020 - Dec 2020*

- Created a command-line tool for the Linux file system that supports snapshot management
- Used traditional split-mirror-based approach with some improvisation over the data storage required while creating the copy.

Skills

Programming Languages: Python, HTML, CSS; **Comfortable:** C, JAVA, C++, SQL, Bash

Tools and Frameworks: Git, Terraform, LaTeX, PyTorch, TensorFlow, Continuous Integration (CI), dbt, Google BigQuery, GCP Services, AppScript

Proficiencies: Privacy and Security in ML, Federated Learning, Time-management, Writing, Team Building

Relevant Coursework: Natural Langugare Processing, Mathematics for Machine Learning, Computer Vision, Reinforcement Learning, Robot Learning, Deep Learning, Machine Learning for Imaging, Data Science, Data Structures and Algorithm, Machine Learning