Vishnu Vinod

vishnuvinod2001 | vishnuvinod8 | vishnuvinod8.github.io | vishnuvinod8 | Vishnu Vinod

Education

Indian Institute of Technology Madras — *2019 - 2024*

Dual Degree (B.Tech + M.Tech) in Computer Science & Engineering — **8.8/10.0**

Publications

InvisibleInk: High-Utility & Low-Cost Text Generation with Differential Privacy — *NeurIPS 2025*

***Vinod, V.**, Pillutla, K., Thakurta, A.*

Preserving Expert-Level Privacy in Offline Reinforcement Learning — **Top 10% @ *TMLR 2025***

Sharma, N., **Vinod, V.***, Thakurta, A., Agarwal, A., Balle, B., Dann, C. & Raghuveer, A.* — *J2C Certification*

Generating Universal Adversarial Perturbations for Quantum Classifiers — *AAAI 2024*

Anil, G., **Vinod, V.*** & Narayan, A.*

Research Experience

Post-Baccalaureate Fellow, CeRAI, IIT Madras[†] — *Jul '24 - Present*

Mentored by **Prof. Krishna Pillutla**

InvisibleInk: Low-Cost Private Text Generation using LLMs

- LLM outputs can leak information provided at *inference time*; Differential Privacy (DP) mitigates this leakage.
- Prior work for DP text generation has high computational overhead ($\geq 100\times$) & low data yield rates ($\leq 1\%$).
- Introduced *InvisibleInk* for scalable text-generation from LLMs with DP guarantees; *Paper accepted at NeurIPS 2025*.
- Isolate & clip *only* sensitive information in model logits; Adapt truncated decoding to private text generation.
- Evaluation on medical, commercial, and legal datasets (MIMIC Notes/Yelp/TAB-ECHR); Additional ablation analyses.
- Reduced compute cost for DP text generation by **factor of 8x** (vs. prior SOTA), with no loss in generation quality.
- Near-optimal default hyperparameters for off-the-shelf usage; Released python software package `invink`.

Student Researcher, Google Research India — *Nov '23 - Apr '24*

Mentored by **Dr. Aravindan Raghuveer** & **Prof. Balaraman Ravindran**

Expert-Level Differentially Private Offline Reinforcement Learning

- Policies learnt by offline RL algorithms can leak the privacy of behavioural policies (experts) contributing the data.
- Prior work has strict assumptions: linear function approximators, tabular settings, trajectory-level DP guarantees.
- Proposed offline RL training paradigm with expert-level DP guarantees; compatible with non-tabular, deep RL.
- *Expert-consensus* filters *stable trajectory* prefixes for noise-free training; with *expert-level DP-SGD* for trajectory tails.
- Evaluated across environments; Filtering stable prefixes has consistent **performance gains across all settings** !
- Compatible with *all SOTA off-the-shelf* gradient-based offline RL and user-level privatization algorithms.

Research Intern, University of British Columbia — *May '22 - Aug '23*

Mentored by **Prof. Apurva Narayan**

Generating Universal Adversarial Perturbations for Quantum Classifiers

- Theorized and proved the existence of a novel class of "*additive*" UAPs for PQC-based quantum classifiers.
- Proposed *QuGAP-A* to generate additive UAPs for amplitude-encoded classical data using generative modelling.
- Proposed *QuGAP-U* to construct unitary UAPs for perturbing quantum data; Trained using a novel *fidelity-based loss*.
- Large scale experimental validation on multiple datasets (MNIST, FMNIST, TIM) for binary and 4-class classification.
- QuGAP-U achieved **full misclassification** at over **90% quantum state fidelity** (vs. 70% for prior SOTA).

Selected Projects

Visual Explanations for Drug-Target Affinity Prediction *Spring '24*

CS6024 - Algorithmic Approaches to Computational Biology under Prof. Manikandan Narayanan

- Explainable *Drug-Target Affinity Prediction* using GNNs to identify active regions in drug & protein molecules.
- Modified GradCAM to generate visual explanations; Occlusion-based quantitative evaluation of saliency maps.
- Proposed new metric to assess the conformity of drug and target explanations.

Empirical Study of Variance Reduced Methods in Machine Learning *Spring '23*

CS6515 - Stochastic Optimization under Prof. Prasanth LA

- Empirical study of variance-reduced optimizers (SAG/SAGA/SDCA/SVRG) in convex and non-convex settings.
- Studied effect of regularization and model depth on convergence rates for non-convex optimization.

Reinforcement Learning Methods *Spring '23*

CS6700 - Reinforcement Learning under Prof. Balaraman Ravindran

- Comparative study of SARSA & Q-Learning, and, SMDP & Intra-option Q-Learning across multiple environments.
- Implemented DQN & Actor-Critic methods on OpenAI Gym environments (Acrobot, CartPole, Mountain Car).

Geospatial Applications of Machine Learning *Summer '21*

Data Science Intern at GalaxEye Space Solutions Pvt. Ltd., mentored by Kishan Thakkar

- Building Footprint Extraction from multi-spectral satellite images (SpaceNet dataset) using semantic segmentation.
- Land-Use-Land-Cover (LULC) classification using a gradient boosting ensemble on the Sentinel-2 dataset.

Achievements

2025 Work recognized as a **top 10% paper @TMLR** and awarded the J2C Certification.
2024 Selected for Post-Baccalaureate Fellowship @ CeRAI and WSAI, IIT Madras.
2023 Selected for a 6-month student researcher internship @ Ad-Sciences team, Google Research India.
2022 Selected for 12-week MITACS Globalink Research Internship at the University of British Columbia.
2019 Secured **All India Rank 35** among 1.1 million candidates in the JEE (Main) examination (top 0.005%tile).
2019 Secured **All India Rank 90** among 200,000 candidates in the JEE (Adv.) examination (top 0.05%tile).
2017 Secured **All India Rank 21**; selected for KVPY 2017 Fellowship by the Dept. of Science and Technology, GoI.

Talks & Presentations

2025 **Spotlight Talk at CODS 2025** @ IISER Pune, Maharashtra, India.
2025 **Poster at NeurIPS 2025** @ San Diego Convention Centre, San Diego, CA, USA.
2025 **Poster at Conclave on AI Governance** @ IIT Madras, Pre-Summit event of the AI Impact Summit 2026
2025 **Talk at Academic Summit 2025** @ Microsoft Research India, Bangalore, India.
2024 **Poster at AAAI 2024** @ Vancouver Convention Centre, Vancouver, BC, Canada.

Co-Curriculars & Volunteering

2025 Volunteer @ 39th Annual Conference on Neural Information Processing Systems (NeurIPS), San Diego, USA.
2025 Teaching Assistant @ IIT Madras, Reinforcement Learning, lectured by Prof. Balaraman Ravindran.
2025 Teaching Assistant @ NPTEL, Introduction to Machine Learning, lectured by Prof. Balaraman Ravindran.
2024 Volunteer @ 38th Annual AAAI Conference on Artificial Intelligence, Vancouver, Canada.
2024 Teaching Assistant @ IIT Madras, Reinforcement Learning, lectured by Prof. Balaraman Ravindran.
2023 Teaching Assistant @ IIT Madras, Foundations of Machine Learning, lectured by Prof. Balaraman Ravindran.
2021 Academic Mentor @ Student Mentorship Cell, IIT Madras.
2021 Coordinator @ Shows & Exhibitions, Shaastra 2021, IIT Madras.
2020 Deputy Coordinator @ Placement & Internship Cell, IIT Madras.

*: equal contribution; †: Also affiliated with RBCDSAI and WSAI @ IIT Madras; SOTA: State–of–the–art