Archit Uniyal

Email: a.uniyal@virginia.edu
Phone: (+1) 4348330881
GitHub
LinkedIn
Website

Skills

Python, C++, C, SQL, Java, Javascript, research, Differential Privacy, Privacy-preserving ML, Natural Language Processing, Computer Vision, Pruning Algorithms, Pytorch, Keras, Tensorflow

Education

University of Virginia, Charlottesville, VA
Masters in Computer Science — August 2022 – December 2024
CGPA: 4.0

University Institute of Engineering and Technology, Panjab University, Chandigarh
B.E in Computer Science — August 2018 – July 2022

Publications

An Empirical Analysis of Memorization in Fine-tuned Autoregressive Language Models
*Mireshghallah, F., **Uniyal, A.**, Wang, T., Evans, D. K., Berg-Kirkpatrick, T.*
Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022

Quantifying Privacy Risks of Masked Language Models Using Membership Inference Attacks
*Mireshghallah, F., Goyal, K., **Uniyal, A.**, Berg-Kirkpatrick, T., Shokri, R.*
Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022

DP-SGD vs PATE: Which Has Less Disparate Impact on Model Accuracy?
***Uniyal, A.**, Naidu, R., Kotti, S., Singh, S., Kenfack, P. J., Mireshghallah, F., Trask, A.*
ML4Data Workshop at ICML 2021

GC-NET for classification of glaucoma in the retinal fundus image
*Juneja, M., Thakur, N., Thakur, S., **Uniyal, A.**, Wani, A., Jindal, P.*
Machine Vision and Applications 31, 38 (2020)

DC-Gnet for detection of glaucoma in retinal fundus imaging
*Juneja, M., Thakur, S., Wani, A., **Uniyal, A.**, Thakur, N., Jindal, P.*
Machine Vision and Applications 31, 34 (2020)

Work Experience

Research Scientist Intern at Oracle, Burlington, MA, USA — *May 2023 - August 2023*

Conducted research on privacy in large language models (LLMs)
Coined a new concept of **entity-relationship privacy** in LLMs and registered it for a patent at Oracle

Research Intern at Samsung Research Institute, Noida, India — *January 2022 – July 2022*
Conducted research on improving the existing human activity recognition algorithms in wearable trackers
Conducted research on integrating GAN-based models for better quality of videos in samsung smartphones

Researcher at Openmined — *November 2020 – July 2022*
Conducted research on the disparate impact of DP-SGD and PATE on minority groups in the dataset.

Research Intern at Accelerating Visions , Dehradun, Uttarakhand — *May 2020 – July 2020*
Built an image search engine powered by deep learning.
Built a custom architecture using image captioning network to perform image search.
Built a Differentially Private image captioning architecture.

Research Intern at Design Innovation Centre, Panjab University — *June 2019 – July 2019*
Proposed a custom architecture for image classification.
Proposed an algorithm to calculate DDLS and ISNT of a fundus image using image segmentation.

Projects

An Empirical Analysis of Memorization in Fine-tuned Autoregressive Language Models
This project focuses on the memorization in large language models due to fine-tuning.
We analyze different fine-tuning methods and observe that fine-tuning the head of the models makes it most vulnerable to attacks.

Quantifying Privacy Risks of Masked Language Models Using Membership Inference Attacks
This project puts forward an efficient membership inference attack on masked language models (MLM) based on the likelihood ratio hypothesis testing that involves an additional reference MLM to more accurately quantify the privacy risks of memorization in MLMs.

PrivateClassImbalance
Conducted experiments on DP-SGD and PATE for epsilon values 0.5, 5 and 15 on MNIST and SVHN datasets.
Conducted an ablation study for determining the optimum number of teachers in PATE to get better performance in terms of fairness and privacy.

Indian Sign Language
Built an interface to detect alphabets of Indian Sign Language and convert it into text.
InceptionResnetV3 was used as our base architecture to classify the alphabets.
LSTMs were used to generate a sequence of text, which is then fed to the google text to speech API to generate an audio.