Aman Priyanshu

Graduate Student at Carnegie Mellon University

amanpriyanshu.github.io @ apriyans@andrew.cmu.edu github.com/AmanPriyanshu
Google Scholar twitter.com/AmanPriyanshu6

Education

Present | **Carnegie Mellon University** — **Pittsburgh, PA, USA**
Aug 2023 | MSIT — Privacy Engineering

Jul 2023 | **Manipal Institute Of Technology, MAHE** — **Karnataka, India**
Jul 2019 | B.Tech Information Technology *with Minors in Big Data Analytics*

Research Experience

Present | **Privacy Engineering Research** [🌐] — **Carnegie Mellon University, USA**
Aug 2023 | *Independent Study | Advisor: Professor Norman Sadeh*
Project: For prompt-engineering geared towards usable privacy & security.

Present | **OpenMined | Research Team** [🌐] — **Remote / United Kingdom**
Aug 2023 | *Project Lead and Collaborator | Collaborators: Dr. Niloofar Mireshghallah*
Project: The impact of epsilon differential privacy on LLM hallucinations.

Aug 2022 | **Concordia University** [🌐] — **Montreal, Canada**
Jun 2022 | *MITACS Globalink Research Intern | Advisors: Professor Wahab Hamou-Lhadj*
Project: Exploring machine learning for anomaly detection toolkit.

Mar 2023 | **ICT Department Research | Manipal Institute of Technology** [🌐] — **Manipal, India**
Jan 2021 | *Undergraduate Research Assistant | Mentor: Professor Balachandra M & Professor Nisha P Shetty*
Project: Explored dynamics of graph-based differential privacy, federated learning, and social network analysis.

Professional Experience

Present | **OpenAI Red Teaming Network | Independent Contractor** — **Remote / San Francisco, CA, USA**
Jan 2024 | Participated in OpenAI led red teaming efforts to assess the risks and safety profile of OpenAI models and systems.

Aug 2023 | **Eder Labs R&D Private Limited | Privacy Engineer Intern** — **Remote / Delaware, USA**
Aug 2022 | Worked on differentially private synthetic data generation for high-content tabular data and relation database systems. Conducted research towards vertical federated learning on financial data.

May 2022 | **DynamoFL | Federated Learning Intern** — **Remote / California, USA**
Mar 2022 | Worked on federated recommendation systems for privately secure federated aggregation.

Publications

S=In Submission, C=Conference, W=Workshop, P=Poster, (* = Equal Contribution)

[S.2] **Are Chatbots Ready for Privacy-Sensitive Applications? An Investigation into Input Regurgitation and Prompt-Induced Sanitization** [Preprint]
Aman Priyanshu, Supriti Vijay, Ayush Kumar, Rakshit Naidu, Niloofer (Fatemeh) Mireshghallah
[In Submission]

[S.1] **Whispers of Wisdom or Wandering Minds? Exploring Hallucinations in Large-Language Models through DP Training and Epsilon Correlations**
Aman Priyanshu, Supriti Vijay, Vrinda Kohli, Rakshit Naidu, Niloofer (Fatemeh) Mireshghallah
[In Submission]

[W.7] **Efficient Hyperparameter Optimization for Differentially Private Deep Learning** [PDF]
Aman Priyanshu, Rakshit Naidu, Fatemehsadat Mireshghallah, Mohammad Malekzadeh
Privacy Preserving Machine Learning Workshop at ACM CCS'21 **[PPML@ACM CCS'21]**

[W.6] **When Differential Privacy Meets Interpretability: A Case Study** [PDF]
Rakshit Naidu, Aman Priyanshu, Aadith Kumar, Sasikanth Kotti, Haofan Wang, Fatemehsadat Mireshghallah
Responsible Computer Vision Workshop at CVPR'21 **[RCV@CVPR'21]**

[W.5] **FedPandemic: A Cross-Device Federated Learning Approach Towards Elementary Prognosis of Diseases During a Pandemic** [PDF]
Aman Priyanshu and Rakshit Naidu
MLPCP & DPML Workshops at International Conference on Learning Representations, 2021 [**MLPCP & DPML@ICLR'21**]

[W.4] **Continual Distributed Learning for Crisis Management** [PDF]
Aman Priyanshu, Mudit Sinha, Shreyans Mehta
3rd Workshop on Continual and Multimodal Learning for Internet of Things at IJCAI'21 [**W-CML4IoT@IJCAI'21**]

[W.3] **NERDA-Con: Extending NER models for Continual Learning — Integrating Distinct Tasks and Updating Distribution Shifts** [PDF]
Supriti Vijay*, Aman Priyanshu*
Updatable Machine Learning Workshop, ICML, 2022 [**UpML@ICML'22**]

[W.2] **F-BRIM: A Semi-Supervised Approach for Bias Mitigation with Activation-Weighted Neuron Regularization** [PDF]
Supriti Vijay*, Aman Priyanshu*, Ashalatha Nayak
Artificial Intelligence for Social Good Workshop, AAAI'23 [**AI4SG@AAAI'23**]

[W.1] **"Something Something Hota Hai!" An Explainable Approach towards Sentiment Analysis on Indian Code-Mixed Data.** [PDF]
Aman Priyanshu*, Sudarshan Sivakumar*, Supriti Vijay*, Aleti Vardhan*, Nipuna Chhabra*
Workshop on Noisy User-generated Text (W-NUT), EMNLP, 2021 [**W-NUT@EMNLP'21**]

[J.2] **FedBully: A Cross-Device Federated Approach for Privacy Enabled Cyber Bullying Detection using Sentence Encoders** [PDF]
Nisha P Shetty, Balachandra Muniyal, Aman Priyanshu, Vedant Rishi Das
Journal: Journal of Cyber Security and Mobility, 2021 [**Journal of Cyber Security and Mobility**]

[J.1] **Finding an elite feature for (D)DoS fast detection-Mixed methods research** [PDF]
Josy Elsa Varghese, Balachandra Muniyal, Aman Priyanshu
Journal: Computers & Electrical Engineering, Volume: 98, Pages: 107705, 2021 [**Computers & Electrical Engineering, Vol. 98**]

[C.1] **Stance Classification with Improved Elementary Classifiers Using Lemmatization (Grand Challenge)** [PDF]
Aman Priyanshu, Vedant Rishi Das, Shashank Rajiv Moghe, Harsh Rathod, Sai Sravan Medicherla, Mini Shail Chhabra, Sarthak Shastri
2020 IEEE Sixth International Conference on Multimedia Big Data [**IEEE BigMM'20**]

[P.3] **ARLIF-IDS–Attention augmented Real-Time Isolation Forest Intrusion Detection System** [PDF]
Aman Priyanshu, Sarthak Shastri, Sai Sravan Medicherla
Poster session at the 43rd IEEE Symposium on Security and Privacy, 2021 [**IEEE S&P'21**]

[P.2] **#maskUp: Selective Attribute Encryption for Sensitive Vocalization for English language on Social Media Platforms.** [PDF]
Supriti Vijay*, Aman Priyanshu*
Eastern European Machine Learning Summer School 2023 [**EEML'23**]

[P.1] **AdaptKeyBERT: An Attention-Based approach towards Few-Shot & Zero-Shot Domain Adaptation of KeyBERT** [PDF]
Aman Priyanshu*, Supriti Vijay*
Eastern European Machine Learning Summer School 2023 [**EEML'23**]

Honours and Awards

Space Theme Category Winner, HackCMU, Sept 2023 [🌐] Winner in the Space Theme Category at HackCMU

AAAI Undergraduate Consortium Scholar, 2023 [🌐] One out of 12 people awarded worldwide. Included mentorship program with funded registration, travel, meal and accommodation to AAAI'23

Eastern European Machine Learning Summer School (EEML), Kosiche, Slovakia, 2023 [🌐] Invited to attend series of lectures and tutorials on ML & RL. Also, awarded travel funding of 1250 EUR

Research Week with Google India 2023 Invited to attend a series of lectures on the latest advances in ML and DL. Also, awarded travel funding of 12,000 INR

Second Runners-Up - ShowYourSkill (Coursera), Jun 2022 [🌐] Came second runners-up in #ShowYourSkill in the Research & Reports Track, creating a NLP augmented Machine Learning Application for women safety

MAHE Research Grant 2022 Recognized among the top 10 researchers at MIT, Manipal for the year 2022-2023 & awarded 10,000 INR in funds for a research proposal

Mitacs Globalink Research Internship, Canada A 12-week paid internship program with Canadian universities and faculty, accompanied by a 15,000 CAD scholarship

Runners-Up - BobHacks 2021 (MetaBob API), Sept 2021 [🌐] Came runners-up in BobHacks, building a pattern recognition API on top of the MetaBob API

First Prize - Code Innovation Series, associated with GitHub, Aug 2021 [🌐] Won the Code Innovation Series Hackathon organized by Manipal Institute of Technology, employing Document-Embedding for contextual similarity analysis

First Prize - HackRx by Bajaj Finserv, July 2021 [🌐] HackRx is the Annual Hackathon hosted by Bajaj Finserv. Developed a solution using Deep Learning and Classical Image processing for face verification and profile-rank estimation, outperforming classic methods. Also created an API for the same.

First Prize - ACM UCM Datathon, UC Merced, May 2021 [🌐] Won the ACM UCM Datathon, created DeCrise, an online platform aggregating public support/utility services for fast-response during crises or disasters.

Runners-Up - Paper Presentation, IEEE SBM Manipal, April 2021 [🌐] Presented a preliminary investigation on integrating Federated Learning with Continual Learning for Crisis Management.

First Prize - Community & Civic Engagement Track, CalHacks Hackathon, April 2021 [🌐] Won at CalHacks Hackathon by UC Berkeley under the Community & Civic Engagement track. Developed Voix, a social media platform with privacy-enabled ML to promote community ideas while conserving user identity.

Runners-Up - IEEE BigMM Data Challenge, IEEE Grand-Challenge, August 2020 [🌐] Achieved runners-up position in IEEE Grand-Challenge for harassment detection on tweets using Elementary Classifiers. Invited to present a paper at IEEE Sixth International Conference on Multimedia Big Data (BigMM).

Scholarship Recipient - Intel Labs, January 2020 [🌐] Received the Intel Edge AI Scholarship Program. Focused on Machine Learning Implementation on the Edge.

Select Research Projects

An Empirical Study of Input Regurgitation: Probing In-context Example Leakage in Chatbot Responses for Privacy-Sensitive Applications May'23

Collaborators: Niloofar Mireshghallah, Rakshit Naidu, and Supriti Vijay

- Investigated LLM-powered chatbots across diverse sectors, including healthcare, personal assistants, and hiring processes for PII leakage over in-context/few-shot samples.
- Revealed a concerning 19% PII leakage in credit risk, 73% omission of non-binary records in hiring, and 32.4% hallucination in medical data generation.

AdaptKeyBERT [] Oct'22 - Nov'22

Collaborators: Supriti Vijay

- Python library for training keyword extractors with LLM bases by incorporating the concept of regularized attention
- Evaluated over 2 benchmarks on Few-Shot & Zero-Shot Domain Adaptation. Increase of 12.58% and 7.88% F1-Score for FSL and ZSL in the FAO-780 dataset and an increase of 9.95% and 3.81% in the CERN-290 dataset.

Leadership Roles

Research Society MIT *Expertise Sub Head — AI* [★] Jun'20-Aug'22

- Led research projects, organized paper presentations, interdisciplinary events and funding proposals
- Mentored 30+ students in the AI and CS domains - devised lesson plans as a guide to Privacy-Preserving Machine Learning

Cryptonite *Technical Head* [★] Jan'20-Aug'22

- Technical lead for Cryptonite — cybersecurity team of Manipal Institute of Technology.
- Conducted research on PPML and AI-security under the student project. Grew from a rank 2000+ on CTFTimes to top 10 nationally.

Academic Service

Sub-Reviewer EMNLP '23, EMNLP '22, Blackbox NLP'21

Volunteer EACL'23, AAAI'23

Skills

Programming Languages	Python, Java, SQL, Shell Scripting(Git & Bash)
Frameworks	PyTorch, Tensorflow, NLTK, Huggingface, FastAPI, Flask
Tools	Docker, PostMan
Languages	English, Hindi, French
Relavant Coursework	Prompt Engineering, AI Governance, Law of Computer Technology, Differential Privacy, Deep Learning, CS229: Machine Learning by Stanford (Stanford), Data Structures and Algorithms, Design and Analysis of Algorithms, Object Oriented Programming, Probability and Statistics, Computer Networks, Operating Systems, Database Management