Niloofar Mireshghallah

cs.washington.edu/~niloofar niloofar@cs.washington.edu +1 (619) 888-9954

EDUCATION

Ph.D. in Computer Science
University of California San Diego, USA *Sep 2018-Apr 2023*
CGPA 3.90/4.00

M.S. in Computer Science
University of California San Diego, USA *Sep 2018-Jun 2020*
CGPA 3.90/4.00

B.Sc. in Computer Engineering
Sharif University of Technology, Iran *Sep 2014-Jun 2018*
CGPA 18.12/20.00

RESEARCH EXPERIENCE

Postdoctoral Scholar
University of Washington *May 2023-Present*
Advisors: Yejin Choi, Yulia Tsvetkov

Research Intern
Microsoft Semantic Machines *Jun 2022-Jul 2023*
Mentors: Richard Shin, Yu Su, Tatsunori Hashimoto, Jason Eisner

Graduate Research Assistant
Berg Lab, CSE Department, UC San Diego *Aug 2020-Apr 2023*
Advisor: Taylor Berg-Kirkpatrick

Research Intern
Microsoft Research, Algorithms *Jan 2022-Mar 2022*
Mentors: Sergey Yekhanin, Arturs Backurs

Research Intern
Microsoft Research, Language and Intelligent Assistance *Jun 2021-Sep 2021*
Mentors: Dimitrios Dimitriadis, Robert Sim

Research Intern
Microsoft Research, Knowledge Technologies and Intelligent Experiences *Jun 2020-Sep 2020*
Mentor: Robert Sim, Huseyin Inan

Graduate Research Assistant
ACT Lab, CSE Department, UC San Diego *Sep 2018-Sep 2020*
Advisor: Hadi Esmeailzadeh

Research Intern
Western Digital Co. Research and Development *Jun 2019-Sep 2019*
Mentor: Anand Kulkarni

Undergraduate Research Assistant
Computer Engineering Department, Sharif University of Technology *Sep 2016-Jun 2018*
Advisor: Hamid Sarbazi-Azad

AWARDS

Momental Foundation Mistletoe Research Fellowship (MRF) Finalist	*2023*
Rising Star in Adversarial Machine Learning (AdvML) Award Winner	*2022*
Rising Stars in EECS	*2022*
UCSD CSE Excellence in Leadership and Service Award Winner	*2022*
FAccT Doctoral Consortium	*2022*
Qualcomm Innovation Fellowship Finalist	*2021*
NCWIT (National Center for Women & IT) Collegiate Award Winner	*2020*
National University Entrance Exam in Math (Ranked 249^{th} of $223,000$)	*2014*
National University Entrance Exam in Foreign Languages (Ranked 57^{th} of $119,000$)	*2014*
National Organization for Exceptional Talents (NODET) (Admitted, $\sim$2% Acceptance Rate)	*2008*

PUBLICATIONS

Conference

1. L. Jiang, K. Rao, S. Han, A. Ettinger, F. Brahman, S. Kumar, **N. Mireshghallah**, X. Lu, M. Sap, Y. Choi, and N. Dziri, "WildTeaming at Scale: From In-the-Wild Jailbreaks to (Adversarially) Safer Language Models", *Advances in Neural Information Processing Systems (NeurIPS)*, Dec. 2024.

2. T. Chen, **N. Mireshghallah**, A. Asai, S. Min, J. Grimmelmann, Y. Choi, H. Hajishirzi, L. Zettlemoyer, and P. W. Koh, "CopyBench: Measuring Literal and Non-Literal Reproduction of Copyright-Protected Text in Language Model Generation", in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Nov. 2024.

3. M. Duan, A. Suri, **N. Mireshghallah**, S. Min, W. Shi, L. Zettlemoyer, Y. Tsvetkov, Y. Choi, D. Evans, and H. Hajishirzi, "Do Membership Inference Attacks Work on Large Language Models?", in *The First Conference on Language Modeling (COLM)*, Oct. 2024.

4. **N. Mireshghallah**, M. Antoniak, Y. More, Y. Choi, and G. Farnadi, "Trust No Bot: Discovering Personal Disclosures in Human-LLM Conversations in the Wild", in *The First Conference on Language Modeling (COLM)*, Oct. 2024.

5. M. Zhang, T. He, T. Wang, **N. Mireshghallah**, B. Chen, H. Wang, and Y. Tsvetkov, "LatticeGen: A Cooperative Framework which Hides Generated Text in a Lattice for Privacy-Aware Generation on Cloud", in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL Findings)*, Aug. 2024.

6. T. Sorensen, J. Moore, J. Fisher, M. Gordon, **N. Mireshghallah**, C. M. Rytting, A. Ye, L. Jiang, X. Lu, N. Dziri, T. Althoff, and Y. Choi, "A Roadmap to Pluralistic Alignment", in *The Fourty-first International Conference on Machine Learning (ICML)*, Jul. 2024.

7. **N. Mireshghallah**, H. Kim, X. Zhou, Y. Tsvetkov, M. Sap, R. Shokri, and Y. Choi, "Can LLMs Keep a Secret? Testing Privacy Implications of Language Models via Contextual Integrity Theory", in *Proceedings of the Twelfth International Conference on Learning Representations (ICLR **Spotlight**)*, May 2024.

8. X. Tang, R. Shin, H. A. Inan, A. Manoel, **N. Mireshghallah**, Z. Lin, S. Gopi, J. Kulkarni, and R. Sim, "Privacy-Preserving In-Context Learning with Differentially Private Few-Shot

Generation", in *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, May 2024.

9. **N. Mireshghallah**, J. Mattern, S. Gao, R. Shokri, and T. Berg-Kirkpatrick, "Smaller Language Models are Better Black-box Machine-Generated Text Detectors", in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Mar. 2024.

10. J. Forristal, **N. Mireshghallah**, G. Durrett, and T. Berg-Kirkpatrick, "A Block Metropolis-Hastings Sampler for Controllable Energy-based Text Generation", in *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, Dec. 2023.

11. **N. Mireshghallah**, N. Vogler, J. He, O. Florez, A. El-Kishky, and T. Berg-Kirkpatrick, "Non-Parametric Temporal Adaptation for Social Media Topic Classification", in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Dec. 2023.

12. J. Mattern, **N. Mireshghallah**, Z. Jin, B. Scholkop, M. Sachan, and T. Berg-Kirkpatrick, "Membership Inference Attacks against Language Models via Neighbourhood Comparison", in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL Findings)*, Jul. 2023.

13. **N. Mireshghallah**, R. Shin, Y. Su, T. Hashimoto, and J. Eisner, "Privacy-Preserving Domain Adaptation of Semantic Parsers", in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, Jul. 2023.

14. **N. Mireshghallah**, A. Backurs, H. A. Inan, L. Wutschitz, and J. Kulkarni, "Differentially Private Model Compression", *Advances in Neural Information Processing Systems (NeurIPS)*, Dec. 2022.

15. **N. Mireshghallah**, K. Goyal, A. Uniyal, T. Berg-Kirkpatrick, and R. Shokri, "Quantifying privacy risks of masked language models using membership inference attacks", in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Dec. 2022.

16. **N. Mireshghallah**, A. Uniyal, T. Wang, D. Evans, and T. Berg-Kirkpatrick, "Memorization in NLP Fine-tuning Methods", in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP),* ***Oral Presentation***, Dec. 2022.

17. **N. Mireshghallah**, V. Shrivastava, M. Shokouhi, T. Berg-Kirkpatrick, R. Sim, and D. Dimitriadis, "UserIdentifier: Implicit User Representations for Simple and Effective Personalized Sentiment Analysis", in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Jul. 2022.

18. H. Brown, K. Lee, **N. Mireshghallah**, R. Shokri, and F. Tram'er, "What Does it Mean for a Language Model to Preserve Privacy?", in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, Jun. 2022.

19. **N. Mireshghallah**, K. Goyal, and T. Berg-Kirkpatrick, "Mix and Match: Learning-free Controllable Text Generation using Energy Language Models", in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (ACL)*, May 2022.

20. **N. Mireshghallah** and T. Berg-Kirkpatrick, "Style Pooling: Automatic Text Style Obfuscation for Improved Classification Fairness", in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP),* ***Oral Presentation***, Nov. 2021.

21. T. Koker, **N. Mireshghallah**, T. Titcombe, and G. Kaissis, "U-Noise: Learnable Noise Masks for Interpretable Image Segmentation", in *2021 IEEE International Conference on Image Processing (ICIP)*, Sep. 2021.

22. **N. Mireshghallah**, H. A. Inan, M. Hasegawa, V. Rühle, T. Berg-Kirkpatrick, and R. Sim, "Privacy Regularization: Joint Privacy-Utility Optimization in Language Models", in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, Jun. 2021.

23. **N. Mireshghallah**, M. Taram, A. Jalali, A. T. Elthakeb, D. Tullsen, and H. Esmaeilzadeh, "Not All Features Are Equal: Discovering Essential Features for Preserving Prediction Privacy", in *Proceedings of The Web Conference 2021 (WWW)*, Apr. 2021.

24. A. T. Elthakeb, P. Pilligundla, **N. Mireshghallah**, A. Cloninger, and H. Esmaeilzadeh, "Divide and Conquer: Leveraging Intermediate Feature Representations for Quantized Training of Neural Networks", in *The Thirty-seventh International Conference on Machine Learning (ICML)*, Jul. 2020.

25. **N. Mireshghallah**, M. Taram, A. Jalali, D. Tullsen, and H. Esmaeilzadeh, "Shredder: Learning Noise Distributions to Protect Inference Privacy", in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Mar. 2020.

Journal

1. A. T. Elthakeb, P. Pilligundla, **N. Mireshghallah**, A. Yazdanbakhsh, and H. Esmaeilzadeh, "ReLeQ: A Reinforcement Learning Approach for Automatic Deep Quantization of Neural Networks", in *IEEE Micro*, Sep. 2020.

2. **N. Mireshghallah**, M. Bakhshalipour, M. Sadrosadati, and H. Sarbazi-Azad, "Energy-Efficient Permanent Fault Tolerance in Hard Real-Time Systems", in *IEEE Transactions on Computers*, Apr. 2019.

Workshop

1. N. G. Brigham, C. Gao, T. Kohno, F. Roesner, and **N. Mireshghallah**, "Breaking News: Case Studies of Generative AI's Use in Journalism", in *Socially Responsible Language Modelling Research (SoLaR) Workshop at NeurIPS*, Dec. 2024.

2. **N. Mireshghallah**, A. M. Kassem, O. Mahmoud, H. Kim, Y. Tsvetkov, Y. Choi, S. Saad, and S. Rana, "Alpaca against Vicuna: Using LLMs to Uncover Memorization of LLMs", in *Red Teaming GenAI Workshop at NeurIPS*, Dec. 2024.

3. I. C. Ngong, J. P. Near, and **N. Mireshghallah**, "Differentially private learning needs better model initialization and self-distillation", in *Socially Responsible Language Modelling Research (SoLaR) Workshop at NeurIPS*, Dec. 2024.

4. X. Zhou, H. Kim, F. Brahman, L. Jiang, H. Zhu, X. Lu, F. Xu, B. Y. Lin, Y. Choi, **N. Mireshghallah**, R. L. Bras, and M. Sap, "HAICOSYSTEM: An Ecosystem for Sandboxing Safety Risks in Human-AI Interactions", in *Safe & Trustworthy Agents Workshop at NeurIPS*, Dec. 2024.

5. K. Lee, A. F. Cooper, C. A. Choquette-Choo, K. Liu, M. Jagielski, **N. Mireshghallah**, L. Ahmed, J. Grimmelmann, D. Bau, C. De Sa, *et al.*, "Machine Unlearning Doesn't Do What You Think", in *Generative AI and Law (GenLaw) Workshop, ICML 2024*, Jul. 2024.
6. P. Basu, T. Singha Roy, R. Naidu, Z. Muftuoglu, S. Singh, and **N. Mireshghallah**, "Benchmarking Differential Privacy and Federated Learning for BERT Models", in *Machine Learning for Data Workshop at ICML 2021*, Jun. 2021.
7. R. Naidu, A. Priyanshu, A. Kumar, S. Kotti, H. Wang, and **N. Mireshghallah**, "When Differential Privacy Meets Interpretability: A Case Study", in *Responsible Computer Vision Workshop at CVPR 2021*, Jun. 2021.
8. A. Uniyal, R. Naidu, S. Kotti, S. Singh, P. J. Kenfack, **N. Mireshghallah**, and A. Trask, "DP-SGD Vs. PATE: Which Has Less Disparate Impact on Model Accuracy?", in *Machine Learning for Data Workshop at ICML 2021*, Jun. 2021.
9. T. Farrand, **N. Mireshghallah**, S. Singh, and A. Trask, "Neither Private Nor Fair: Impact of Data Imbalance on Utility and Fairness in Differential Privacy", in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS 2020), Privacy-Preserving Machine Learning in Practice workshop (PPMLP)*, Nov. 2020.

Preprint

1. X. Lu, M. Sclar, S. Hallinan, **N. Mireshghallah**, J. Liu, S. Han, A. Ettinger, L. Jiang, K. Chandu, N. Dziri, *et al.*, "AI as Humanity's Salieri: Quantifying Linguistic Creativity of Language Models via Systematic Attribution of Machine Text against Web Text", *ArXiv preprint arXiv:2410.04265*, Oct. 2024.
2. M. H. Garcia, A. Manoel, D. M. Diaz, **N. Mireshghallah**, R. Sim, and D. Dimitriadis, "Flute: A scalable, extensible framework for high-performance federated learning simulations", *ArXiv preprint arXiv:2203.13789*, 2022.
3. **N. Mireshghallah**, M. Taram, P. Vepakomma, A. Singh, R. Raskar, and H. Esmaeilzadeh, "Privacy in Deep Learning: A Survey", *ArXiv preprint arXiv:2004.12254*, Apr. 2020.

Patent

1. J. M. Eisner, E. C. Shin, **N. Mireshghallah**, T. B. Hashimoto, and Y. Su, "Privacy-preserving generation of synthesized training data", US Patent App. 18/321,460, Google Patents, Jun. 2024.
2. **N. Mireshghallah** and H. Esmaeilzadeh, "METHODS OF PROVIDING DATA PRIVACY FOR NEURAL NETWORK BASED INFERENCE", US Patent 11,288,379, Mar. 2020.
3. **N. Mireshghallah**, H. Esmaeilzadeh, and M. Taram, "METHOD AND SYSTEM OF LEARNING NOISE ON INFORMATION FROM INFERENCES BY DEEP NEURAL NETWORK", US Patent 009062-8413, Oct. 2019.

INVITED TALKS

Stanford University
NLP Seminar *Jan. 2025*
Privacy, Copyright and Data Integrity: The Cascading Implications of Generative AI

University of California, Los Angeles
Guest lecture for CS 269 - Computational Ethics, LLMs and the Future of NLP *Jan. 2025*

A False Sense of Privacy: Semantic Leakage and Non-literal Copying in LLMs

NeurIPS Conference
Red Teaming GenAI workshop *Dec. 2024*
A False Sense of Privacy: Semantic Leakage and Non-literal Copying in LLMs

NeurIPS Conference
Panelist *Dec. 2024*
PrivacyML: Meaningful Privacy-Preserving Machine Learning tutorial

Johns Hopkins University
CS Department Seminar *Dec. 2024*
Privacy, Copyright and Data Integrity: The Cascading Implications of Generative AIs

Future of Privacy Forum
Panelist *Nov. 2024*
Technologist Roundtable for Policymakers: Key Issues in Privacy and AI

University of Utah
Guest lecture for the School of Computing CS 6340/5340 NLP course *Nov. 2024*
Can LLMs Keep a Secret?

UMass Amherst
NLP Seminar *Oct. 2024*
Membership Inference Attacks and Contextual Integrity for Language

Northeastern University
Khoury College of Computer Sciences Security Seminar *Oct. 2024*
Membership Inference Attacks and Contextual Integrity for Language

Stanford Research Institute (SRI) International
Computational Cybersecurity in Compromised Environments (C3E) workshop *Sep. 2024*
Can LLMs keep a secret? Testing privacy implications of Language Models via Contextual Integrity

LinkedIn Research
Privacy Tech Talk *Sep. 2024*
Can LLMs keep a secret? Testing privacy implications of Language Models via Contextual Integrity

National Academies (NASEM)
Forum on Cyber Resilience *Aug. 2024*
Oversharing with LLMs is underrated: the curious case of personal disclosures in human-LLM conversations

ML Collective
DLCT reading group *Aug. 2024*
Privacy in LLMs: Understanding what data is imprinted in LMs and how it might surface!

Carnegie Mellon University
Invited Talk *Jun. 2024*
Alpaca against Vicuna: Using LLMs to Uncover Memorization of LLMs

Generative AI and Law workshop, Washington DC
Invited Talk *Apr. 2024*
What is differential privacy? And what is it not?

Meta AI Research
Invited Talk *Apr. 2024*
Membership Inference Attacks and Contextual Integrity for Language

Georgia Institute of Technology
Guest lecture for the School of Interactive Computing *Apr. 2024*
Safety in LLMs: Privacy and Memorization

University of Washington
Guest lecture for CSE 484 and 582 courses on Computer Security and Ethics in AI *Apr. 2024*
Safety in LLMs: Privacy and Memorization

Carnegie Mellon University
Guest lecture for LTI 11-830 course on Computational Ethics in NLP *Mar. 2024*
Safety in LLMs: Privacy and Memorization

Simons Collaboration
TOC4Fairness Seminar *Mar. 2024*
Membership Inference Attacks and Contextual Integrity for Language

University of California, Santa Barbara
NLP Seminar Invited Talk *Mar. 2024*
Can LLMs Keep a Secret? Testing Privacy Implications of LLMs

University of California, Los Angeles
NLP Seminar Invited Talk *Mar. 2024*
Can LLMs Keep a Secret? Testing Privacy Implications of LLMs

University of Texas at Austin
Guest lecture for LIN 393 course on Social Applications and Impact of NLP *Feb. 2024*
Can LLMs Keep a Secret? Testing Privacy Implications of LLMs

Google Brain
Google Tech Talk *Feb. 2024*
Can LLMs Keep a Secret? Testing Privacy Implications of LLMs

University of Washington
Allen School Colloquium *Jan. 2024*
Can LLMs Keep a Secret? Testing Privacy Implications of LLMs

University of Washington
eScience Institute Seminars *Nov. 2023*
Privacy Auditing and Protection in Large Language Model

CISPA Helmholtz Center for Security
Invited Talk *Sep. 2023*
What does privacy-preserving NLP entail?

Max Planck Institute for Software Systems
Next 10 in AI Series *Sep. 2023*
Auditing and Mitigating Safety Risks in LLMs

Mila / McGill University
Invited Talk *May 2023*
Privacy Auditing and Protection in Large Language Models

EACL 2023
Tutorial co-instruction *May 2023*
Private NLP: Federated Learning and Privacy Regularization

Cohere for AI
C4AI Community Talk *May 2023*

Auditing and Mitigating Safety Risks in Large Language Models

LLM Interfaces Workshop and Hackathon
Invited Talk *Apr. 2023*
Learning-free Controllable Text Generation

NDSS Conference
Keynote talk for EthiCS workshop *Feb. 2023*
How much can we trust large language models?

Google
Federated Learning Seminar *Feb. 2023*
Privacy Auditing and Protection in Large Language Models

University of Texas Austin
Invited Talk *Oct. 2022*
How much can we trust large language models?

Johns Hopkins University
Guest lecture for CS 601.670 course on Artificial Agents *Sep. 2022*
Mix and Match: Learning-free Controllable Text Generation

KDD Conference
Adversarial ML workshop *Aug. 2022*
How much can we trust large language models?

Microsoft Research Cambridge
Invited Talk *Mar. 2022*
What Does it Mean for a Language Model to Preserve Privacy?

PriSec ML Interest Group
Invited Talk *Feb. 2022*
What Does it Mean for a Language Model to Preserve Privacy?

University of Maine
Guest lecture for COS435/535 course on Information Privacy Engineering *Dec. 2021*
Improving Attribute Privacy and Fairness for Natural Language Processing

National University of Singapore
Invited Talk *Nov. 2021*
Style Pooling: Automatic Text Style Obfuscation for Fairness

Big Science for Large Language Models
Invited Panelist *Oct. 2021*
Privacy-Preserving Natural Language Processing

Research Society MIT Manipal
Cognizance Event Invited Talk *Jul. 2021*
Privacy and Interpretability of DNN Inference

Alan Turing Institute
Privacy and Security in ML Seminars *Jun. 2021*
Low-overhead Techniques for Privacy and Fairness of DNNs

Split Learning Workshop
Invited Talk *Mar. 2021*
Shredder: Learning Noise Distributions to Protect Inference Privacy

University of Massachusetts Amherst
Machine Learning and Friends Lunch *Oct. 2020*
Privacy and Fairness in DNN Inference

OpenMined Privacy Conference
Invited Talk *Sep. 2020*
Privacy-Preserving Natural Language Processing

Microsoft Research AI
Breakthroughs Workshop *Sep. 2020*
Private Text Generation through Regularization

National Center for Women & IT (NCWIT)
Collegiate Award Ceremony Winner Talk *Jul. 2020*
Shredder: Learning Noise Distributions to Protect Inference Privacy

TA EXPERIENCE

CSE Department of UC San Diego

CSE 151A (Undergraduate Machine Learning)	*Fall 2021*
CSE 251A (Graduate Machine Learning)	*Winter 2021*
CSE 276C (Graduate Mathematics for Robotics)	*Fall 2020*
CSE 141 (Undergraduate Computer Architecture)	*Spring 2020*
CSE 240D (Graduate Accelerator Design for Deep Learning)	*Winter & Fall 2019*

CE Department of Sharif University
Digital Electronics, Computer Architecture, Signals and Systems, Probability and Statistics, Numerical Methods *2016-2018*

DIVERSITY & INCLUSION

NAACL 2025 D&I co-chair	*2025*
Women in ML (WiML) at NeurIPS Mentor	*2024*
Widening NLP (WiNLP) co-chair	*2022-2024*
NAACL 2022 D&I co-chair	*2022*
Mentor at ICLR	*2021*
Mentor for the Women in Machine Learning (WiML) workshop at NeurIPS	*2020*
Mentor for the Graduate Women in Computing (GradWIC) at UCSD	*2020-2023*
Course instructor for the OpenMined Privacy Course	*2020*
Mentor for the UC San Diego Women Organization for Research Mentoring (WORM) in STEM	*2019-2023*
Mentor for the USENIX Security Undergraduate Mentorship Program	*2020*
Volunteer at the Women in Machine Learning Workshop Held at NeurIPS	*2019*
Invited Speaker at the Women in Machine Learning and Data Science (WiMLDS) NeurIPS Meetup	*2019*
Mentor for the UCSD CSE Early Research Scholars Program (CSE-ERSP)	*2018*

STUDENT MENTORING & ADVISING

Skyler Hallinan, BS/MS student at University of Washington → PhD student at USC	*Jun 2024–Present*
Tong Chen, PhD student at University of Washington	*Feb 2024–Present*
Sara Kodeiri, MS student at University of Waterloo	*Jan 2024–Present*
Sabrina Mokhtari, MS student at University of Waterloo	*Jan 2024–Present*
Rui Xin, PhD student at University of Washington	*Dec 2023–Present*

Yash More, MS student at McGill University *Dec 2023–Aug 2024*
Aly Kassem, MS student at University of Windsor → Research Engineer at Dynamo AI *Jul 2023–Present*
Michael Duan, BS student at University of Washington → Software Engineer at Uber *Jun 2023–Present*
Grace Brigham, BS student at University of Washington *Mar 2023–Present*
Chongjiu Gao, BS/MS student at University of Washington *Mar 2023–Present*
Supriti Vijay, BS at Manipal Institute → MS student at Carnegie Mellon University *Mar 2023–Jul 2023*
Justus Mattern, BS student at RWTH Aachen → Y Combinator *Dec 2022–Jul 2023*
Ivoline Ngong, PhD student at University of Vermont *Sep 2022–Present*
Jarad Forristal, BS at University of Texas Austin → PhD Student at UC San Diego *Aug 2022–Jul 2023*
Archit Uniyal, BE at UIET Panjab University → PhD student at University of Virginia *Mar 2021–Dec 2022*
Aman Priyanshu, BS at Manipal Institute → MS student at Carnegie Mellon University *Mar 2021–Jul 2023*
Rakshit Naidu, BS at Manipal Institute → PhD student at Georgia Tech *Mar 2021–Jul 2023*
Priyam Basa, BS at Manipal Institute → MS student at University of Washington *Mar 2021–Sep 2021*
Zumrut Muftuogl, PhD student at Yıldız Teknik Üniversitesi *Mar 2021–Sep 2021*
Teddy Koker, Research Engineer at Lightning AI → PhD Student at MIT EECS *Apr 2020–Feb 2021*
Tom Farrand, Deep Learning Engineer at IBM *Feb 2020–Sep 2020*
Jaydeep Borkar, BE at Pune University → PhD student at Northeastern University *Jan 2020–Present*

ORGANIZED EVENTS

Privacy Session Chair at SAGAI workshop at IEEE S&P *2025*
Co-organizer of the Generative AI and Law (GenLaw) workshop at ICML *2024*
Co-organizer of the Privacy Regulation and Protection in Machine Learning workshop at ICLR *2024*
Co-organizer of the Private NLP workshop at ACL *2024*
Co-organizer of the Privacy-Preserving AI (PPAI) workshop at AAAI *2024*
Co-organizer of the Generative AI and Law (GenLaw) workshop at ICML *2023*
Co-organizer of the Widening NLP (WiNLP) workshop at EMNLP *2023*
Co-organizer of the Generative AI + Law (GenLaw) workshop at ICML *2023*
Co-organizer of the Private NLP Tutorial at EACL *2023*
Co-organizer of the Ethics in NLP Birds of a Feather session at EMNLP *2022*
Privacy & Fairness Roundtable lead at AFCP workshop at NeurIPS *2022*
Co-organizer of the Broadening Collaborations in ML workshop at NeurIPS *2022*
Co-organizer of the Widening NLP (WiNLP) workshop at EMNLP *2022*
Co-organizer of the Private NLP workshop at NAACL *2022*
Co-organizer of the Federated Learning for NLP workshop at ACL *2022*
Co-organizer of the Widening NLP (WiNLP) workshop at EMNLP *2021*
Co-leader for the "Machine Learning for Privacy: An Information Theoretic Perspective" Break-out session at the Women in Machine Learning (WiML) Un-workshop Held at ICML *2021*
Co-organizer of the Privacy-Preserving Machine Learning (PPML) Workshop at MICCAI *2021*
Co-organizer of the Distributed Private Machine Learning (DPML) Workshop at ICLR *2021*
Co-organizer of the SoCal joint Machine Learning and Natural Language Processing Symposium *2021*
Co-leader for the "Feminist Perspectives for Machine Learning & Computer Vision" Break-out session at the Women in Machine Learning (WiML) Un-workshop Held at ICML *2020*

PROFESSIONAL SERVICES

Reviewer for ICLR Workshop Proposals *2025*
PC member for ACM CCS *2025*
Area Chair for NAACL *2025*

Reviewer for AISTATS	*2025*
Reviewer for TACL	*2020-Present*
Area Chair for EMNLP	*2024*
PC member for ACM CCS	*2024*
Reviewer for EACL	*2024*
Reviewer for CHI	*2024*
Reviewer for AAAI	*2023, 2024*
Reviewer for NeurIPS	*2020-2024*
Reviewer for ICML	*2020-2024*
Reviewer for ICLR (Outstanding Reviewer Award in 2021)	*2021-2025*
Reviewer for FAccT	*2023, 2024*
Ethics PC member for EMNLP	*2022*
Reviewer for NeurIPS Workshop Proposals	*2023*
Reviewer for INLG	*2023*
Reviewer for TMLR Journal	*2022-Present*
PC member for the AFCP workshop at NeurIPS	*2022*
PC member for the TSRML Workshop at NeurIPS	*2022*
Reviewer for NAACL Student Research Workshop (SRW)	*2022*
Reviewer for IEEE S&P magazine	*2021-Present*
Reviewer for CCS Poster Sessions	*2021*
Shadow PC member for IEEE Security and Privacy Conference Winter	*2021*
Artifact Evaluation Program Committee Member for USENIX Security	*2021*
Security & Privacy Committee Member and Session Chair for Grace Hopper Celebration (GHC)	*2020*
Reviewer for ACM TACO Journal	*2020-Present*
Reviewer for IEEE TC Journal	*2020-Present*
Program Committee Member for Latinx in AI Research Workshop at ICML	*2020*
Program Committee Member for the Workshop on Human Interpretability in ML at ICML	*2020*
Program Committee Member for the ML for Computer Architecture and Systems Workshop at ISCA	*2020*
Artifact Evaluation Program Committee Member for ASPLOS	*2020*