Umar Ali Syed

Google
111 8th Ave
New York, NY 10011

Phone: (860) 748-7724
usyed@google.com
http://www.umarsyed.com

Education

- **Princeton University**, Princeton, NJ. August 2004 - August 2009
  Ph.D., Computer Science.
- **Cornell University**, Ithaca, NY. January 2001 - December 2001
  M.Eng., Computer Science.
- **University of Florida**, Gainesville, FL. August 1996 - December 2000
  B.Sc., Computer Engineering.

Positions held

- **Google** September 2011 - present
  Research scientist.
- **University of Pennsylvania** September 2009 - August 2011
  Postdoctoral researcher, Department of Computer and Information Science.
- **Microsoft Research** June 2008 - September 2008
  Research internship, Search Labs.
- **AT&T Labs Research** June 2007 - August 2007
  Research internship, Speech Group.
- **Gene Network Sciences** February 2003 - May 2004
  Bioinformatician.
- **Cornell University** January 2001 - February 2003
  Research assistantship, Department of Computer Science.

Post-graduate honors

Oral presentation, Neural Information Processing Systems (NeurIPS), 2007, 2010 — *Approximately 10% of accepted papers are invited to give an oral presentation at the conference.*

Google Student Award Winner, Machine Learning Symposium, New York Academy of Sciences, 2007, 2009 — *Awarded to five presenters at the symposium.*

Wallace Memorial Fellowship in Engineering, Princeton University, 2008 — *Awarded annually to two Princeton engineering graduate students.*

Best Student Paper Runner-up, Uncertainty in Artificial Intelligence (UAI), 2007

Under-graduate honors

National Merit Scholarship, Florida Bright Futures Scholarship, Robert C. Byrd Scholarship, United Technologies Scholarship

Publications

Amin, K., Babakniya, S., Bie, A., Kong, W., Syed, U., Vassilvitskii, S. (2025)* Escaping collapse: The strength of weak data for large language model training. *Advances in Neural Information Processing Systems 38 (NeurIPS 2025).*

Busa-Fekete, R., Syed, U. (2025)* Near-optimal algorithms for private estimation and sequential testing of collision probability. *Proceedings of the Twenty-Eighth International Conference on Artificial Intelligence and Statistics (AISTATS 2025).*

Berke, A., Ghazi, B., Bacis, E., Kamath, P., Kumar, R., Lassonde, R., Manurangsi, P., Syed, U. (2025) How unique is whose web browser? The role of demographics in browser fingerprinting among US users. *Proceedings on Privacy Enhancing Technologies (PoPETs 2025).*

Amin, K., Bie, A., Kong, W., Kurakin, A., Ponomareva, N., Syed, U., Terzis, A., Vassilvitskii, S. (2024)* Private prediction for large-scale synthetic text generation. *Findings of the Association for Computational Linguistics (EMNLP 2024).*

Kong, W., Medina, A. M., Ribero, M., Syed, U. (2024)* DP-Auditorium: A large-scale library for auditing differential privacy. *Proceedings of the IEEE Symposium on Security and Privacy (S&P 2024).*

Bacis, E., Bilogrevic, I., Busa-Fekete, R., Herath, A., Sartori, A., Syed, U. (2024)* Assessing web fingerprinting risk. *Companion Proceedings of the ACM Web Conference (WWW 2024).*

Abernethy, J., Schapire, R. E., Syed, U. (2024)* Lexicographic optimization: Algorithms and stability. *Proceedings of the Twenty-Seventh International Conference on Artificial Intelligence and Statistics (AISTATS 2024).*

Busa-Fekete, R., Medina, A. M., Syed, U., Vassilvitskii, S. (2023)* Label differential privacy and private training data release. *Proceedings of the Fortieth International Conference on Machine Learning (ICML 2023).*

Bravo-Hermsdorff, G., Busa-Fekete, R., Ghavamzadeh, M., Medina, A. M., Syed, U. (2022)* Private and communication-efficient algorithms for entropy estimation. *Advances in Neural Information Processing Systems 35 (NeurIPS 2022).*

Esfandiari, H., Mirrokni, V., Syed, U., Vassilvitskii, S. (2021)* Label differential privacy via clustering. *Proceedings of the Twenty-Fifth International Conference on Artificial Intelligence and Statistics (AISTATS 2022).*

Epasto, A., Medina, A. M., Avery, S., Bai, Y., Busa-Fekete, R., Carey, C., Gao, Y., Guthrie, D., Ghosh, S., Ioannidis, J., Jiao, J., Lacki, J., Lee, J., Mauser, A., Milch, B., Mirrokni, V., Ravichandran, D., Shi, W., Spero, M., Sun, Y., Syed, U., Vassilvtiskii, S., Wang, S. (2021) Clustering for private interest-based advertising. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2021).*

Medina, A. M., Syed, U., Vassilvitskii, S., Vitercik, E. (2021)* Private optimization without constraint violations. *Proceedings of the Twenty-Fourth International Conference on Artificial Intelligence and Statistics (AISTATS 2021).*

Syed, U., Vassilvitskii, S. (2017)* SQML: Large-scale in-database machine learning with pure SQL. *Proceedings of the 2017 Symposium on Cloud Computing (SoCC 2017).*

Balkanski, E., Syed, U., Vassilvitskii, S. (2017)* Statistical cost sharing. *Advances in Neural Information Processing Systems 30 (NeurIPS 2017).*

Nazerzadeh, H., Paes Leme, R., Rostamizadeh, A., Syed, U. (2016)* Where to sell: Simulating auctions from learning algorithms. *Proceedings of the Seventeenth ACM Conference on Economics and Computation (EC2016).*

Heidari, H., Mahdian, M., Syed, U., Vassilvitskii, S., Yazdanbod, S. (2016)* Pricing a low-regret seller. *Proceedings of the Thirty-Third International Conference on Machine Learning (ICML 2016).*

Choromanski, K., Rostamizadeh, A., Syed, U. (2015)* An optimal online algorithm for retrieving heavily perturbed statistical databases in the low-dimensional querying model. *Proceedings of the Twenty-Fourth ACM International Conference on Information and Knowledge Management (CIKM 2015).*

DeSalvo, G., Mohri, M., Syed, U. (2015)* Learning with deep cascades. *Proceedings of the Twenty-Sixth International Conference on Algorithmic Learning Theory (ALT 2015).*

Cortes, C., Kuznetsov, V., Mohri, M., Syed, U. (2015)* Structural maxent models. *Proceedings of the Thirty-Second International Conference on Machine Learning (ICML 2015).*

Kuznetsov, V., Mohri, M., Syed, U. (2014)* Multi-class deep boosting. *Advances in Neural Information Processing Systems 27 (NeurIPS 2014).*

Amin, K., Rostamizadeh, A., Syed, U. (2014)* Repeated contextual auctions with strategic buyers. *Advances in Neural Information Processing Systems 27 (NeurIPS 2014).*

Cortes, C., Mohri, M., Syed, U. (2014)* Deep boosting. *Proceedings of the Thirty-First International Conference on Machine Learning (ICML 2014).*

Asuncion, A., de Haan, J., Mohri, M., Patel, K., Rostamizadeh, A., Syed, U., Wong, L. (2014)* Corporate learning at scale: Lessons from a large online course at Google. *Learning at Scale 2014.*

Amin, K., Rostamizadeh, A., Syed, U. (2013)* Learning prices for repeated auctions with strategic buyers. *Advances in Neural Information Processing Systems 26 (NeurIPS 2013).*

Amin, K., Kearns, M., Syed, U. (2011)* Graphical models for bandit problems. *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Seventh Conference (UAI 2011).*

Amin, K., Kearns, M., Syed, U. (2011)* Bandits, query learning, and the haystack dimension. *Proceedings of the Twenty-Fourth Annual Conference on Learning Theory (COLT 2011).*

Fabrikant, A., Syed, U., Rexford, J. (2011) There's something about MRAI: Timing diversity can exponentially worsen BGP convergence. *Proceedings of the Thirtieth IEEE International Conference on Computer Communications (INFOCOM 2011).*

Syed, U., Taskar, B. (2010) Semi-supervised learning with adversarially missing label information. *Advances in Neural Information Processing Systems 23 (NeurIPS 2010).*

Syed, U., Schapire, R. E. (2010) A reduction from apprenticeship learning to classification. *Advances in Neural Information Processing Systems 23 (NeurIPS 2010).*

Brautbar, M., Kearns, M., Syed, U. (2010)* Private and third-party randomization in risk-sensitive equilibrium concepts. *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010).*

Syed, U., Slivkins, A., Mishra, N. (2009) Adapting to the shifting intent of search queries. *Advances in Neural Information Processing Systems 22 (NeurIPS 2009).*

Syed, U., Bowling, M., Schapire, R. E. (2008) Apprenticeship learning using linear programming. *Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML 2008).*

Syed, U., Williams, J. D. (2008) Using automatically transcribed dialogs to learn user models in a spoken dialog system. *Proceedings of the Forty-Sixth Annual Meeting of the Association for Computational Linguistics (ACL 2008).*

Syed, U., Yona, G. (2008) Enzyme function prediction with interpretable models. *Methods in Molecular Biology: Computational Systems Biology*, edited by J. McDermott, K. Montgomery, R. Bumgarner and R. Samudrala. Humana Press. [Book chapter]

Syed, U., Schapire, R. E. (2007) A game-theoretic approach to apprenticeship learning. *Advances in Neural Information Processing Systems 20 (NeurIPS 2007).*

Syed, U., Schapire, R. E. (2007) Imitation learning with a value-based prior. *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Third Conference (UAI 2007).*

Syed, U., Yona, G. (2003) Using a mixture of probabilistic decisions trees for direct prediction of protein function. *Proceedings of the Seventh Annual International Conference on Research in Computational Biology (RECOMB 2003).*

*Denotes alphabetical author order.

Teaching assistantships

Introduction to Computer Organization (Cornell, 2001), Introduction to Web Design (Cornell, 2001), Theoretical Computer Science (Princeton, 2005), Computer Science for Non-Majors (Princeton, 2006, 2007), Machine Learning (Princeton, 2008), Machine Learning (NYU, 2012).