Om Dipakbhai Thakkar

omthkkr@google.com www.omthakkar.com

WORK EXPERIENCE

Senior Research Scientist, Google, Mountain View, CA 05/21 - Present
Research Scientist, Google, Mountain View, CA 12/20 - 04/21
Research Software Engineer, Google, Mountain View, CA 09/19 - 12/20
Working in the team of Françoise Beaufays on privacy-preserving data analysis.

Visiting Graduate Student, University of California, Berkeley, CA 01/19 - 05/19
Participated in the program Data Privacy: Foundations and Applications.

Software Engineering Intern, Google, Mountain View, CA 05/18 - 08/18
Worked with Úlfar Erlingsson on using adaptivity to improve DP Stochastic Gradient Descent. Empirically showed that the novel technique achieves the utility of the state-of-the-art with up to 30x speed-up in time.

Visiting Student Researcher, University of California, Berkeley, CA 08/17 - 12/17
Worked with Dr. Dawn Song on designing a practical DP optimization algorithm that works for all standard convex losses, can leverage any off-the-shelf optimizer, and has a competitive hyperparameter-free variant.

Software Engineering Intern, Google, Seattle, WA 05/17 - 08/17
Worked with Brendan McMahan on devising adaptive strategies for eliminating hyperparameter tuning for DP federated learning, achieving utility similar to systems with tuned hyperparameters.

Machine Learning Engineer (Intern), CoreOS Team, Apple, Cupertino, CA 05/16 - 08/16
Designed a scalable general-purpose DP recommendation system based on collaborative filtering. Developed a model from scratch showing positive results.

EDUCATION

Ph.D., Computer Science 08/14 - 09/19
Boston University(BU), Boston, MA GPA 4.00/4.00
The Pennsylvania State University (Penn State), University Park, PA[1] GPA 3.97/4.00
Advisor: Dr. Adam Smith

B.Tech., Information and Communication Technology 07/10 - 05/14
Dhirubhai Ambani University (DAU), Gujarat, India GPA 8.57/10.00

RESEARCH INTERESTS

Privacy-preserving Deep Learning.

PATENTS

Leveraging Intermediate Checkpoints To Improve The Performance of Trained Differentially Private Models. *Filed US Patent 63/376,528.*
Om Thakkar, Arun Ganesh, Virat Shejwalkar, Abhradeep Thakurta, and Rajiv Mathews.

Detecting Unintended Memorization in Language-Model-Fused ASR Systems. *Published US Patent 2023/0335126.*
W. Ronny Huang, Steve Chien, Om Thakkar, and Rajiv Mathews.

[1] Transferred to BU in 01/18.

Generating and/or Utilizing Unintentional Memorization Measure(s) for Automatic Speech Recognition Model(s). *Published US Patent 2023/0317082.*
Om Thakkar, Hakim Sidahmed, W. Ronny Huang, Rajiv Mathews, Françoise Beaufays, and Florian Tramèr.

Server Efficient Enhancement of Privacy in Federated Learning. *Published US Patent 2023/0223028.*
Om Thakkar, Peter Kairouz, Brendan McMahan, Borja Balle, and Abhradeep Thakurta.

Phrase Extraction for ASR Models. *Published US Patent 2023/0178094.*
Ehsan Amid, Om Thakkar, Rajiv Mathews, and Françoise Beaufays.

Leveraging Public Data in Training Neural Networks with Private Mirror Descent. *Published US Patent 2023/0103911.*
Ehsan Amid, Arun Ganesh, Rajiv Mathews, Swaroop Ramaswamy, Shuang Song, Thomas Steinke, Vinith Suriyakumar, Om Thakkar, and Abhradeep Thakurta.

Ascertaining And/or Mitigating Extent of Effective Reconstruction, of Predictions, From Model Updates Transmitted in Federated Learning. *Published US Patent 2022/0383204.*
Om Thakkar, Trung Dang, Swaroop Ramaswamy, Rajiv Mathews, and Françoise Beaufays.

Mixed Client-Server Federated Learning. *Published US Patent 2022/0293093.*
Françoise Beaufays, Swaroop Ramaswamy, Rajiv Mathews, Om Thakkar, and Andrew Hard.

PUBLICATIONS

Unless specifically indicated, all publications have authors listed in the alphabetical order of last names.

Efficiently Train ASR Models that Memorize Less and Perform Better with Per-core Clipping. *In Interspeech, 2024.*
Lun Wang, Om Thakkar, Zhong Meng, Nicole Rafidi, Rohit Prabhavalkar, and Arun Narayanan. *(In order of contribution)*

Quantifying Unintended Memorization in BEST-RQ ASR Encoders. *In Interspeech, 2024. (Accepted for an oral presentation)*
Virat Shejwalkar, Om Thakkar, and Arun Narayanan. *(In order of contribution)*

Unintended Memorization in Large ASR Models, and How to Mitigate It. *In ICASSP, 2024.*
Lun Wang, Om Thakkar, and Rajiv Mathews. *(In order of contribution)*

Noise Masking Attacks and Defenses for Pretrained Speech Models. *In ICASSP, 2024.*
Matthew Jagielski, Om Thakkar, and Lun Wang.

Why Is Public Pretraining Necessary for Private Model Training? *In ICML, 2023.*
Arun Ganesh, Mahdi Haghifam, Milad Nasr, Sewoong Oh, Thomas Steinke, Om Thakkar, Abhradeep Thakurta, and Lun Wang.

Measuring Forgetting of Memorized Training Examples. *In ICLR, 2023.*
Matthew Jagielski, Om Thakkar, Florian Tramèr, Daphne Ippolito, Katherine Lee, Nicholas Carlini, Eric Wallace, Shuang Song, Abhradeep Thakurta, Nicolas Papernot, and Chiyuan Zhang. *(In order of contribution)*

Extracting Targeted Training Data from ASR Models, and How to Mitigate It. *In Interspeech, 2022. (Accepted for an oral presentation)*
Ehsan Amid*, Om Thakkar*, Arun Narayanan, Rajiv Mathews, and Françoise Beaufays. *(*Equal contribution)*

Detecting Unintended Memorization in Language-Model-Fused ASR. *In Interspeech, 2022. (Accepted for an oral presentation)*
W. Ronny Huang, Steve Chien, Om Thakkar, and Rajiv Mathews. *(In order of contribution)*

Public Data-Assisted Mirror Descent for Private Model Training. *In ICML, 2022.*
Ehsan Amid, Arun Ganesh, Rajiv Mathews, Swaroop Ramaswamy, Shuang Song, Thomas Steinke, Vinith Suriyakumar, Om Thakkar, and Abhradeep Thakurta.

A Method to Reveal Speaker Identity in Distributed ASR Training, and How to Counter It. *In ICASSP, 2022.*
Trung Dang, Om Thakkar, Swaroop Ramaswamy, Rajiv Mathews, Peter Chin, and Françoise Beaufays. *(In order of contribution)*

The Role of Adaptive Optimizers for Honest Private Hyperparameter Selection. *In AAAI, 2022. (Accepted for an oral presentation)*
Shubhankar Mohapatra, Sajin Sasy, Gautam Kamath*, Xi He*, and Om Thakkar*. *(*Alphabetical order)*

Differentially Private Learning with Adaptive Clipping. *In NeurIPS, 2021.*
Galen Andrew, Om Thakkar, Swaroop Ramaswamy, and Brendan McMahan. *(In order of contribution)*

Revealing and Protecting Labels in Distributed Training. *In NeurIPS, 2021.*
Trung Dang, Om Thakkar, Swaroop Ramaswamy, Rajiv Mathews, Peter Chin, and Françoise Beaufays. *(In order of contribution)*

Practical and Private (Deep) Learning without Sampling or Shuffling. *In ICML, 2021.*
Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu.

Evading the Curse of Dimensionality in Unconstrained Private GLMs. *In AISTATS, 2021.*
Shuang Song, Thomas Steinke, Om Thakkar, and Abhradeep Thakurta.

Privacy Amplification via Random Check-Ins. *In NeurIPS, 2020.*
Borja Balle, Peter Kairouz, Brendan McMahan, Om Thakkar, and Abhradeep Thakurta.

Guaranteed Validity for Empirical Approaches to Adaptive Data Analysis. *In AISTATS, 2020.*
Ryan Rogers, Aaron Roth, Adam Smith, Nathan Srebro, Om Thakkar, and Blake Woodworth.

Towards Practical Differentially Private Convex Optimization. *In S&P, 2019.*
Roger Iyengar, Joseph P. Near, Dawn Song, Om Thakkar, Abhradeep Thakurta and Lun Wang.

Model-Agnostic Private Learning. *In NeurIPS, 2018. (Accepted for an oral presentation)*
Raef Bassily, Om Thakkar, and Abhradeep Thakurta.

Differentially Private Matrix Completion Revisited. *In ICML, 2018. (Accepted for a long talk)*
Prateek Jain, Om Thakkar, and Abhradeep Thakurta.

Max-Information, Differential Privacy, and Post-Selection Hypothesis Testing. *In FOCS, 2016.*
Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar.

WORKSHOP PAPERS

Differentially Private Parameter-Efficient Fine-tuning for Large ASR Models. *In DLSP (S&P, 2024), and TPDP 2024 .*
Hongbin Liu, Lun Wang, Om Thakkar, Abhradeep Thakurta, Arun Narayanan. *(In order of contribution)*

Recycling Scraps: Improving Private Learning by Leveraging Intermediate Checkpoints. *In PPAI (AAAI, 2023). (Accepted for an oral presentation)*
Virat Shejwalkar, Arun Ganesh*, Rajiv Mathews*, Om Thakkar*, Abhradeep Thakurta*. *(*Alphabetical order)*

Public Data-Assisted Mirror Descent for Private Model Training. *In TPDP (ICML, 2022).*
Ehsan Amid, Arun Ganesh, Rajiv Mathews, Swaroop Ramaswamy, Shuang Song, Thomas Steinke, Vinith Suriyakumar, Om Thakkar, Abhradeep Thakurta.

Practical and Private (Deep) Learning without Sampling or Shuffling. *In TPDP (ICML, 2021).*
Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu.

The Role of Adaptive Optimizers for Honest Private Hyperparameter Selection. *In TPDP (ICML, 2021).*
Shubhankar Mohapatra, Sajin Sasy, Gautam Kamath*, Xi He*, Om Thakkar*. *(*Alphabetical order)*

Training Production Language Models without Memorizing User Data. *In PPML (NeurIPS, 2020). (Accepted for an oral presentation)*
Swaroop Ramaswamy*, Om Thakkar*, Rajiv Mathews, Galen Andrew, Brendan McMahan, and Françoise Beaufays. *(In order of contribution. *Equal contribution)*

Privacy Amplification via Random Check-Ins. *In TPDP (CCS, 2020).*
Borja Balle, Peter Kairouz, Brendan McMahan, Om Thakkar, and Abhradeep Thakurta.

Understanding Unintended Memorization in Federated Learning. *In PrivateNLP (NAACL, 2021), TPDP (CCS, 2020), and PPML (NeurIPS, 2020).*
Om Thakkar, Swaroop Ramaswamy, Rajiv Mathews, and Françoise Beaufays. *(In order of contribution)*

Characterizing Private Clipped Gradient Descent on Convex Generalized Linear Problems. *In TPDP (CCS, 2020) (accepted for an oral presentation) and PPML, 2020 (NeurIPS, 2020).*
Shuang Song, Om Thakkar, and Abhradeep Thakurta.

PROFESSIONAL SERVICES

Program committee member for TPDP (2020, 2022).

Reviewer for journals: Information Sciences 2024, T-IFS (2019, 2021-2022), JPC (2019, 2022), JSSAM 2021, TSC 2020, JMLR 2018.

Reviewer for conferences: ICML (2018, 2021-2024), ASRU 2023, NeurIPS (2019-2023), RANDOM 2023, AISTATS 2022, S&P (2017, 2019, 2022), PETS (2017-2021), IJCAI 2019, CCS (2018-2019), STOC (2016, 2018), ACSAC 2017, FOCS 2017, WABI 2015.

Reviewer for NIST's The Unlinkable Data Challenge: Advancing Methods in Differential Privacy.

TEACHING EXPERIENCE

Teaching Assistant, CMPSC 465 Data Structures and Algorithms, Penn State	01/17 - 05/17
Teaching Assistant, CMPSC 360 Discrete Mathematics for Computer Science, Penn State	01/15 - 05/15
Teaching Assistant, IT 114 Object Oriented Programming, DAU	01/14 - 05/14
Teaching Assistant, IT 105 Introduction to Programming, DAU	08/13 - 12/13

AWARDS AND ACHIEVEMENTS

Received travel awards for S&P 2019, NeurIPS 2018, ICML 2018, a registration award for FOCS 2014, and a GSO Conference Travel Grant for Summer 2018.

Ranked 127th in the ACM - Inter Collegiate Programming Contest, Asia Amritapuri Region, 2012.

Ranked in top 500 in the IEEEXtreme Programming Competition (editions 4.0, and 5.0).

Won state- and district-level yoga competitions, and participated at national and international levels.

POSITIONS OF RESPONSIBILITY

Webmaster of the Theory group webpage, Penn State	08/14 - 05/17
Captain of the Cricket team, DAU	04/13 - 04/14
Vice-Chairperson, IEEE Student Branch, DAU	01/13 - 11/13
Treasurer, IEEE Student Branch, DAU	07/12 - 12/12

NOTABLE PROJECTS

Efficient Pattern Matching incorporating Modifications in a Genome 01/15 - 05/15
Guide: Dr. Paul Medvedev (Penn State) *Team Size - 2*
Constructed an algorithm for efficiently re-computing pattern matches in case of any modification in a genome, as well as constructed an appropriate data structure for storing occurrences of each pattern match.

Distributed Data Lookup in a Peer-to-Peer (P2P) File System 08/14 - 12/14
Guide: Dr. Guohong Cao (Penn State) *Team Size - 2*
Implemented a P2P file sharing system, and a distributed data lookup for it using TCP/IP.

GRADUATE COURSEWORK

Approximation Algorithms, Graphs of Bounded Widths, Probabilistic Algorithms, Computational Complexity, Cryptography, Error Correcting Codes, Mathematical Logic, Mathematical Neuroscience, Sublinear Algorithms, Algorithms in Bioinformatics, Foundations of Data Privacy, Algorithm Design and Analysis, Distributed Systems.