Congzheng Song

csong4@apple.com
1 Apple Park Way, Cupertino, CA
https://csong27.github.io

Education

2020	*Ph.D.* in Computer Science
	Cornell University Ithaca, NY
	Advisor: Vitaly Shmatikov
	Thesis: *Measuring the Unmeasured: New Threats to Machine Learning Systems*
2019	*M.S.* in Computer Science
	Cornell University Ithaca, NY
2016	*B.S. with Summa Cum Laude* in Computer Science
	Emory University Atlanta, GA
	Advisor: Ymir Vigfusson
	Honor Thesis: *Using Deep Recurrent Neural Networks to Estimate Influenza Prevalence from Mobile Phone Records*

Experience

2021 –	*Machine Learning Research Engineer*
	Apple, AI/ML Cupertino, CA
	Private federated learning
2017 – 2020	*Graduate Research Assistant*
	Cornell Tech New York, NY
	Research topics: security and privacy in machine learning
Summer 2020	*Applied Scientist Intern*
	Amazon.com Seattle, WA
	Mentor: Suleiman Ali Khan
Fall 2019	*Research Intern*
	Google Research, Brain Team Mountain View, CA
	Mentor: Ananth Raghunathan
Summer 2019	*Research Intern*
	Petuum Inc Pittsburgh, PA
	Mentors: Najmeh Sadoughi, Pengtao Xie

Publications and Manuscripts † denotes equal contribution

2022	Eugene Bagdasaryan, **Congzheng Song**, Rogier van Dalen, Matt Seigel, Áine Cahill
	Training a Tokenizer for Free with Private Federated Learning. In *arXiv preprint*
2021	Roei Schuster, **Congzheng Song**, Eran Tromer, Vitaly Shmatikov
	You Autocomplete Me: Poisoning Vulnerabilities in Neural Code Completion. In *the 30th USENIX Security Symposium, Vancouver, Canada*

2021 Ymir Vigfusson, Thorgeir A. Karlsson, Derek Onken, **Congzheng Song**, Atli F. Einarsson, Nishant Kishore, Rebecca M. Mitchell, Ellen Brooks-Pollock, Gudrun Sigmundsdottir, Leon Danon
Cellphone Traces Reveal Infection-associated Behavioral Change. In *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*

2020 **Congzheng Song**, Alexander Rush, Vitaly Shmatikov
Adversarial Semantic Collisions. In *the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Punta Cana, Dominican Republic*

2020 **Congzheng Song**, Ananth Raghunathan
Information Leakage in Embedding Models. In *the 27th ACM Conference on Computer and Communications Security (CCS), Orlando, Florida*

2020 **Congzheng Song**, Shanghang Zhang, Najmeh Sadoughi, Pengtao Xie, Eric P. Xing
Generalized Zero-Shot Text Classification for ICD Coding. In *the 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI), Yokohama, Japan*

2020 **Congzheng Song**, Reza Shokri
Membership Encoding for Deep Learning. In *the 15th ACM ASIA Conference on Computer and Communications Security (AsiaCCS), Taipei, Taiwan*

2020 **Congzheng Song**, Vitaly Shmatikov
Overlearning Reveals Sensitive Attributes. In *International Conference on Learning Representation (ICLR), Addis Ababa, Ethiopia*

2019 **Congzheng Song**, Vitaly Shmatikov
Auditing Data Provenance in Text Generation Models. In *the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Anchorage, Alaska*

2019 Luca Melis[†], **Congzheng Song**[†], Emiliano De Cristofaro, Vitaly Shmatikov
Exploiting Unintended Feature Leakage in Collaborative Learning. In *the 40th IEEE Symposium on Security and Privacy (Oakland), San Francisco, California*

2018 Tyler Hunt, **Congzheng Song**, Reza Shokri, Vitaly Shmatikov, Emmett Witchel
Chiron: Privacy-preserving Machine Learning as a Service. In *arXiv preprint*

2018 **Congzheng Song**, Vitaly Shmatikov
Fooling OCR Systems with Adversarial Text Images. In *arXiv preprint*

2018 **Congzheng Song**[†], Yiming Sun[†]
Kernel Distillation for Fast Gaussian Processes Prediction. In *NeurIPS Workshop on All of Bayesian Nonparametrics (BNP@NeurIPS), Montreal, Canada*

2018 Vitaly Shmatikov, **Congzheng Song**
What Are Machine Learning Models Hiding?. In *the 11th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs), Barcelona, Spain*

2017 **Congzheng Song**, Thomas Risternpart, Vitaly Shmatikov
Machine Learning Models that Remember Too Much. In *the 24th ACM Conference on Computer and Communications Security (CCS), Dallas, Texas*

2017 Reza Shokri, Marco Stronati, **Congzheng Song**, Vitaly Shmatikov
Membership Inference Attacks against Machine Learning Models. In *the 38th IEEE Symposium on Security and Privacy (Oakland), San Jose, California*

2017 Safoora Yousefi, Fatemeh Amrollahi, Mohamed Amgad, Coco Dong, Joshua E. Lewis, **Congzheng Song**, David A. Gutman, Sameer H. Halani, Jose Enrique Velazquez Vega, Daniel J. Brat, Lee A.D. Cooper
Predicting Clinical Outcomes from Large Scale Cancer Genomic Profiles with Deep Survival Models. In *Scientific Reports 7 (Nature)*

2016 Safoora Yousefi, **Congzheng Song**, Nelson Nauata, Lee Cooper
Learning Genomic Representations to Predict Clinical Outcomes in Cancer. In *International Conference on Learning Representation Workshop (ICLRW), San Juan, Puerto Rico*

Awards

2021	USENIX Security Distinguished Paper Award
2020	Digital Life Initiative Doctoral Fellowship
2018	The Caspar Bowden PET Award
2016	Trevor Evans Award
2016	Deborah Jackson Award

Activities

2022	*Reviewer*, ICLR, ICML
2021	*Reviewer*, NeurIPS
2021	*PC Member*, PrivateNLP Workshop at NAACL
2021	*PC Member*, Security and Safety in Machine Learning Systems Workshop at ICLR
2021	*PC Member*, Distributed and Private Machine Learning Workshop at ICLR
2020	*PC Member*, Privacy Preserving Machine Learning Workshop at NeurIPS
2020	*PC Member*, PrivateNLP Workshop at EMNLP
2020	*PC Member*, Towards Trustworthy ML Workshop at ICLR
2020	*PC Member*, PrivateNLP Workshop at WSDM

Teaching

Fall 2020	*Teaching Assistant*, Cornell Tech	New York, NY
	CS 5435: Security and Privacy Concepts in the Wild	
Fall 2016	*Teaching Assistant*, Cornell University	Ithaca, NY
	CS 3410: Computer System Organization and Programming	
Fall 2013	*Teaching Assistant*, Emory University	Atlanta, GA
	Chem 141: General Chemistry I	