Ashwinee Panda

website Email

Experience

24 - 25	Postdoc **UMD College Park** with Tom Goldstein
24	Applied Research **Capital One** Mixture of Expert Pretraining
20 - 24	PhD **Princeton University** advised by Prateek M. AI Safety
20	M.S. + B.S. **UC Berkeley** advised by Joey G. AI Systems

Research (selected 1st-author publications)

LoTA	**Ashwinee Panda**, Berivan I., Xiangyu Q., Sanmi K., Tsachy W., Prateek M. Lottery Ticket Adaptation: Mitigating Destructive Interference in LLMs *ICML 2024 ES-FoMO Oral, ICML 2024 FM-Wild Oral*
Auditing	**Ashwinee Panda***, Xinyu Tang*, Milad N., Chris C., Prateek M. Privacy Auditing of LLMs *ICML 2024 NextGenAISafety Oral*
DP-ZO	Xinyu Tang*, **Ashwinee Panda***, Milad N., Saeed M., Prateek M. Private Fine-tuning of LLMs with Zeroth-order Optimization *TPDP 2024 Oral*
DP-Scaling	**Ashwinee Panda***, Xinyu Tang*, Vikash S., Saeed M., Prateek M. A New Linear Scaling Rule for Private Adaptive Hyperparameter Optimization *ICML 2024 Poster*
Phishing	**Ashwinee Panda**, Chris C., Zhengming Z., Yaoqing Y., Prateek M. Teach LLMs to Phish: Stealing Private Information from LLMs *ICLR 2024 Poster*
DP-ICL	Tong Wu*, **Ashwinee Panda***, Tianhao Wang*, Prateek M. Privacy-Preserving In-Context Learning for LLMs *ICLR 2024 Poster*
DP-RandP	Xinyu Tang*, **Ashwinee Panda***, Prateek M. Differentially Private Image Classification by Learning Priors from Random Processes *NeurIPS 2023 Spotlight*
Neurotoxin	Zhengming Zhang*, **Ashwinee Panda***, Linyue S., Yaoqing Y., Prateek M., Joey G., Kannan R., Michael M. NeuroToxin: Durable Backdoors in Federated Learning *ICML 2022 Oral*
SparseFed	**Ashwinee Panda**, Saeed M., Arjun B., Supriyo C., Prateek M. SparseFed: Mitigating Model Poisoning Attacks in Federated Learning via Sparsification *AISTATS 2022 Poster*
FetchSGD	Daniel Rothchild*, **Ashwinee Panda***, Enayat U., Nikita I., Ion S., Vladimir B., Joey G., Raman A. FetchSGD: Communication-Efficient Federated Learning with Sketching *ICML 2020 Poster*

Invited Talks

Sep ’24	Privacy Auditing of LLMs *Google Privacy Seminar*
May ’24	Challenges in Adapting LLMs to Private Data Google Privacy Seminar (click for talk recording)
Nov ’23	New Privacy Attacks on Large Language Models *Sun Lab, Berkeley*
Nov ’23	Challenges in Data-Driven Alignment of Large Language Models *SPYLab, ETH Zurich*
Oct ’23	New Directions in Differentially Private Machine Learning *Meta CAS*
Sep ’23	Challenges in Data-Driven Alignment of Large Language Models *University of Maryland, College Park*
Sep ’23	Challenges in Augmenting Large Language Models with Private Data *SL^2 Lab, UIUC*
Sep ’23	Improving the Privacy Utility Tradeoff in Differentially Private Machine Learning with Prior Information *SECRIT, University of Michigan*
Apr ’23	Improving the Privacy Utility Tradeoff in Differentially Private Machine Learning with Public Data *Apple*
Mar ’23	Google Privacy Seminar (click for talk recording) *Google*
Jun ’22	Challenges and Directions in Privacy Preserving Machine Learning *Microsoft Research Cambridge*
May ’22	Towards Trustworthy Machine Learning *Meta AI*
Jan ’22	Federated Learning for Forecasting *Ohmconnect*
Nov ’21	Building Federated Learning Systems at Scale *Liftoff AI*
Nov ’21	Practical Defenses Against Model Poisoning Attacks Google (click for talk recording)

Research (Advised and workshop papers)

Safety — Xiangyu Qi, Ashwinee Panda, Kaifeng Lyu, Xiao Ma, Subhrajit Roy, Ahmad Beirami, Prateek M., Peter Henderson
Safety Alignment Should Be Made More Than Just a Few Tokens Deep

AdvVLM — Xiangyu Qi*, Kaixuan Huang*, **Ashwinee Panda**, Mengdi Wang, Prateek M.
Introducing Vision into Large Language Models Expands Attack Surfaces and Failure Implications
At *Thirty-Eighth AAAI Conference on Artificial Intelligence*

Phishing — **Ashwinee Panda**, Zhengming Z., Yaoqing Y., Prateek M.
Teach GPT to Phish: Neural Phishing Attacks on Large Language Models
At *40th International Conference on Machine Learning* AdvML Workshop

DP-Diffusion — Vikash S.*, **Ashwinee Panda***, Ashwini Pokle, Xinyu Tang, Saeed M., Mung Chiang, J Zico Kolter, Prateek M.
Differentially Private Generation of High Fidelity Samples From Diffusion Models
At *40th International Conference on Machine Learning* GenAI Workshop

DP-ICL — **Ashwinee Panda***, Tong Wu*, Tianhao Wang*, Prateek M.
Differentially Private In-Context Learning
At *NAACL 2023* TrustNLP Workshop

SoftPBT — **Ashwinee Panda**, Eric Liang, Richard Liaw, Joey G.
SoftPBT: Leveraging Experience Replay for Efficient Hyperparameter Schedule Search
Submitted to NeurIPS 2019

Service

Teaching

2023 — Teaching Assistant for COS/ECE 432 at Princeton
2019 — Course Staff for CS 189 (Machine Learning) at UC Berkeley
2018 — Undergraduate Student Instructor for CS 70 (Probability and Discrete Mathematics) and Course Staff for CS 189 at UC Berkeley
2017 — Course Staff for CS 70 at UC Berkeley

Peer Reviewing (* denotes Best Reviewer Award)

2024 — ICML 2024*, NeurIPS 2024
2023 — SATML 2023, ACL 2023, ICML 2023, NeurIPS 2023*, TMLR
2022 — ICML 2022, AISTATS 2022
2021 — ICML 2021, NeurIPS 2021
2020 — ICML 2020
2019 — ICLR 2019, NeurIPS 2019