Mayee Chen

Curriculum Vitae

Education

09/19 – **Stanford University**, *PhD Candidate in Computer Science*
- Advisor: Christopher Ré

09/15 – 06/19 **Princeton University**, *B.S.E in Operations Research and Financial Engineering (ORFE), certificate in Applications of Computing*, GPA: 3.962/4
- Graduated Summa Cum Laude
- Senior Thesis: *A Quantum Version of the Multiplicative Weights Algorithm* (recipient of the Ahmet S. Çakmak Thesis Prize)
- Thesis advisor: Elad Hazan

Research Interests

I'm interested in using theoretical tools to understand and improve on modern machine learning techniques. My current research is in understanding representation geometry and modeling data. In particular, I have been studying how to induce better representations in contrastive learning, and how to better model noisy training data and unstructured signals through Weak Supervision and probabilistic graphical models.

Publications and Preprints

- **Ask Me Anything: A simple strategy for prompting language models.**
  Simran Arora, Avanika Narayan, Mayee Chen, Laurel Orr, Neel Guha, Kush Bhatia, Ines Chami, Christopher Ré.
  In submission, 2022.
- **Reducing Reliance on Spurious Features in Medical Image Classification with Spatial Specificity.**
  Khaled Saab, Sarah M. Hooper, Mayee F. Chen, Michael Zhang, Daniel Rubin, Christopher Ré.
  Machine Learning for Healthcare (MLHC), 2022.
- **Shoring Up the Foundations: Fusing Model Embeddings and Weak Supervision**
  Mayee F. Chen*, Daniel Y. Fu*, Dyah Adila, Michael Zhang, Frederic Sala, Christopher Ré.
  Uncertainty in Artificial Intelligence (UAI), 2022. **Best Student Paper Runner-Up Award, Oral Presentation.**
- **Perfectly Balanced: Improving Transfer and Robustness of Supervised Contrastive Learning**
  Mayee F. Chen*, Daniel Y. Fu*, Avanika Narayan, Michael Zhang, Zhao Song, Kayvon Fatahalian, Christopher Ré.
  International Conference on Machine Learning (ICML), 2022.
- **TABi: Type-Aware Bi-Encoders for Open-Domain Entity Retrieval**
  Megan Leszczynski, Daniel Y. Fu, Mayee F. Chen, Christopher Ré.
  Findings of ACL, 2022.
- **The Details Matter: Preventing Class Collapse in Supervised Contrastive Learning**
  Mayee F. Chen*, Daniel Y. Fu*, Michael Zhang, Kayvon Fatahalian, Christopher Ré.
  AAAI Workshop on Artificial Intelligence with Biased or Scarce Data, 2022. **Best Paper Award.**
- **An Adversarial Model of Network Disruption: Maximizing Disagreement and Polarization in Social Networks.**
  Mayee F. Chen and Miklos Z. Racz.
  IEEE Transactions on Network Science and Engineering (TNSE), 2021.
- **Mandoline: Model Evaluation under Distribution Shift**
  Mayee F. Chen*, Karan Goel*, Nimit Sohoni*, Fait Poms, Kayvon Fatahalian, and Christopher Ré.
  International Conference on Machine Learning (ICML), 2021.
- **Comparing the Value of Labeled and Unlabeled Data in Method-of-Moments Latent Variable Estimation**
  Mayee F. Chen*, Benjamin Cohen-Wang*, Steve Mussmann, Frederic Sala, and Christopher Ré.
  AISTATS, 2021.
- **Fast and Three-rious: Speeding Up Weak Supervision with Triplet Methods**
  Daniel Y. Fu*, Mayee F. Chen*, Frederic Sala, Sarah M. Hooper, Kayvon Fatahalian, and Christopher Ré.
  International Conference on Machine Learning (ICML), 2020.

Awards and Honors

2021 NSF GRFP Honorable Mention

2019 Ahmet S. Çakmak Prize, *Princeton University*, for innovative senior thesis research.

2018 Phi Beta Kappa, *Princeton University*, one of 28 early inductees.

2017 Tau Beta Pi Engineering Honor Society, *Princeton University*

2017 Shapiro Prize for Academic Excellence, *Princeton University*, awarded to top 2-3% of the class.

Work Experience

2016 – 19 **Grader for Computer Science Department**, *Princeton University*

- Algorithms and Data Structures (lead grader), Functional Programming, Reasoning about Computation, Introduction to Machine Learning, and Economics and Computing

06/18–08/18 **Quantitative Trading Intern**, *IMC Trading*, Chicago, IL, Fixed Income, Currencies, and Commodities Desk

05/17–08/17 **Software Engineering Intern**, *Google*, Mountain View, CA, Advertiser Platform Team
Worked on AdWords Next Overviews, frontpage data analytics for ads campaigns

05/16–08/16 **Engineering Practicum Intern**, *Google*, Mountain View, CA, Cloud/Cluster/Kernel team
Worked on an infrastructure tool for pushing configuration and data updates to services within Google

Talks

2022 Snorkel AI Machine Learning Whiteboard Talk: Liger: Fusing weak supervision with foundation model embeddings

2021 MedAI Talk Series: Correcting distribution shift in the ML model evaluation process

2021 DAWN Research Workshop: Mandoline: Model Evaluation under Distribution Shift

2020 Google x Stanford Summit: Labeled vs Unlabeled Data in Latent Variable Graphical Models

Coursework

Relevant graduate courses:

- Information Theoretic Lower Bounds in Data Science, Principles of Data-Intensive Systems, Convex Optimization II, Randomized Algorithms

Relevant undergraduate courses:

- *ORFE Courses:* Probability Theory (graduate-level course), Optimization, High Frequency Trading, Decision Modeling for Business Analytics, Monte Carlo Simulation, Strategy and Information, Financial Mathematics, Analysis of Big Data, Probability and Stochastics, Microeconomic Theory, Statistics
- *Computer Science Courses:* Optimization for Machine Learning (graduate-level seminar), Computer Networks, Operating Systems, Economics and Computing, Introduction to Machine Learning, Information Security, Human-Computer Interfaces, Neural Networks, Functional Programming, Reasoning About Computation, Programming Systems, Algorithms and Data Structures

Leadership and Activities

Reviewer for NeurIPS and ICML (2021, 2022), AISTATS (2022). Subreviewer for UAI and KDD (2020).

At Stanford University:

- WiML PhD Application Mentorship Program
- CS PhD Admissions Committee (2020, 2021)
- CS Student Applicant Support Program Volunteer (2020, 2021)
- Graduate WiCS Mentor (2021)
- XTRM Kpop Cover Group: dance captain (2019–2021), Alliance Dance Team (2019–2020)

Skills

Advanced: Python, C, Java Intermediate: Go, OCaml, R, Dart, PyTorch Basic: Matlab, Julia