Trenton Chang

Website: https://trentonchang.org
Email: ctrenton@umich.edu
LinkedIn: trenton-chang

Research Interests

My research interests are in machine learning and causal inference inspired by healthcare use cases. To date, my work has focused on analyzing how sources of bias in real-world data can impact the performance and equity of AI models.

Education

University of Michigan — Ann Arbor, MI
Ph.D. Candidate in Computer Science & Engineering, GPA: 4.00 — 2021–present
Advisor: Jenna Wiens

Stanford University — Stanford, CA
M.S. in Computer Science—Artificial Intelligence track, GPA: 4.05 — 2020–2021

Stanford University — Stanford, CA
B.A. in American Studies, *with distinction* & Phi Beta Kappa, GPA: 3.98 — 2016–2020

Experience

Microsoft Research — Redmond, WA
Mentors: Adith Swaminathan & Tobias Schnabel, Augmented Learning & Reasoning Group — May 2024-Aug 2024

Publications & Preprints

[P1] **T. Chang**, J. Wiens, T. Schnabel, and A. Swaminathan, "Measuring Steerability in Large Language Models", in *Workshop on Safe Generative AI, Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024, to appear.

[P2] **T. Chang**, L. Warrenburg, S.-H. Park, R. B. Parikh, M. Makar, and J. Wiens, "Who's Gaming the System? A Causally-Motivated Approach for Detecting Strategic Adaptation", in *Proceedings of the Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024, to appear.

[P3] W. Chen, **T. Chang**, and J. Wiens, "LobsterNet: Estimating Conditional Average Treatment Effects Under Treatment Non-compliance", 2024, under review.

[P4] **T. Chang**, M. Nuppnau, Y. He, K. Kocher, T. S. Valley, M. W. Sjoding, and J. Wiens, "Racial differences in laboratory testing as a mechanism for bias amplification for AI models in healthcare: the emergency department as a case study ", in *PLOS Global Public Health*, Oct. 2024.

[P5] **T. Chang** and J. Wiens, "From Biased Selective Labels to Pseudo-Labels: An Expectation-Maximization Framework for Learning from Biased Decisions", in *Proceedings of the 41st International Conference on Machine Learning*, Jul. 2024.

[P6] E. A. Chi, A. Paranjape, A. See, C. Chiam, **T. Chang**, K. Kenealy, S. K. Lim, A. Hardy, C. Rastogi, H. Li, A. Iyabor, Y. He, H. Sowrirajan, P. Qi, K. R. Sadagopan, N. Minh Phu, D. Soylu, J. Tang, A. Narayan, G. Campagna, and C. Manning, "Neural generation meets real people: Building a social, informative open-domain dialogue agent", in *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Association for Computational Linguistics, Sep. 2022.

[P7] **T. Chang**, M. W. Sjoding, and J. Wiens, "Disparate Censorship: A Plausible, Underexplored Mechanism for Model Performance Gaps in Clinical Machine Learning", in *7th Machine Learning for Healthcare Conference*, Proceedings of Machine Learning Research, Aug. 2022.

[P8] **T. Chang** and D. Y. Fu, "Lost in Transmission: On the Impact of Networking Corruptions on Video Machine Learning Models", Jun. 2022. arXiv: 2206.05252 [cs.CV].

[P9] E. A. Chi, C. Chiam, **T. Chang**, S. K. Lim, C. Rastogi, A. Iyabor, Y. He, H. Sowrirajan, A. Narayan, J. Tang, H. Li, A. Paranjape, and C. D. Manning, "Neural, neural everywhere: Controlled generation meets scaffolded, structured dialogue", in *Alexa Prize Socialbot Grand Challenge 4 Proceedings*, Jul. 2021.

[P10] **T. Chang**, D. Y. Fu, Y. Li, and C. Ré, "Beyond the Pixels: Exploring the Effect of Video File Corruptions on Model Performance", in *2020 European Conference in Computer Vision, Workshop on Adversarial Robustness in the Real World*, Aug. 2020.

Press Appearances & Media Outreach

[M1] D. Smith, "Accounting for bias in medical data helps prevent AI from amplifying racial disparity", *Michigan Engineering News*, Oct. 2024.

[M2] C. Ross, B. Trang, and M. Aguilar, "What does generative AI mean for health care? We asked the experts", *STAT+*, May 2023.

[M3] T. Render, "Decisive Differences in Healthcare AI", *Discover Rackham*, Oct. 2022.

[M4] Michigan AI Lab [@michigan_AI], "AI, Healthcare, and Humanities with Trenton Chang", Aug. 2022.

Invited Talks & Presentations

[T1] **T. Chang**, "Bias in, bias out: Analyzing sources of bias in machine learning for healthcare", in *Borgwardt Group, Max-Planck-Institut für Biochemie, internal talk*, Jul. 2024.

[T2] **T. Chang**, "Measuring and mitigating the impact of biases in laboratory testing on machine learning models", in *NIH Office of Data Science Strategy AI Supplement Program PI Meeting*, Feb. 2024.

[T3] **T. Chang**, "Mitigating the effects of label-bias: An expectation-maximization approach", in *Michigan AI Symposium*, Oct. 2023.

[T4] **T. Chang**, "Recognizing and addressing biases in machine learning for healthcare", in *Ann Arbor Machine Learning Meetup (Ann Arbor SPARK)*, Oct. 2023.

[T5] **T. Chang**, "Disparate censorship: A plausible, underexplored mechanism for model performance gaps in clinical machine learning", in *Michigan AI Symposium*, Dec. 2022.

[T6] **T. Chang** and D. Ganelin, "Machine learning bias in criminal justice", in *Computer Science Teachers of America Conference*, Jul. 2021.

Teaching & Mentoring

- Graduate Student Instructor, EECS 598-009 (Causality and machine learning), University of Michigan (2023)
  - Delivered 80-min. lecture on fairness in machine learning from a causal perspective.
  - Assisted in writing homework questions, solutions, and grading for course on causal inference with 23 M.S. and Ph.D. stduents.
- Workshop Organizer, Discover Engineering, University of Michigan (2023)
  - Recruited 9 volunteer instructors and designed workshop introducing high school students to computer science and an interactive exploration of the limitations and capabilities of ChatGPT, reaching 4 cohorts of approx. 10 students each.
- Workshop Organizer, Xplore Engineering: "How do Computers Think?", University of Michigan (2023)
  - Recruited 12 volunteer instructors and designed workshop introducing 4th - 7th grade students to computer science and an activity analyzing the robustness of image classification models, reaching 6 cohorts of approx. 10 students each.
- Volunteer Instructor, AI4ALL, University of Michigan (2022)
  - Co-taught project on n-gram based text generation and sentiment analysis to 9 high school stduents.
- Research Mentor, ACM Stanford (2021)
  - Advised two undergraduate students' accepted submission to the Google Big-Bench benchmark of tasks for evaluating large language models.
- Instructor, Inspirit AI (2020, 2021)
  - Wrote and taught project on the usage of AI in criminal justice decisions for high school students.

- Residential Counselor, Artificial Intelligence Course, Stanford Pre-Collegiate Studies (2019)
  - Mentored proejcts in AI ranging from computer vision to price prediction for 2 cohorts of approx. 15 students each.

Service

- University of Michigan AI Blog Co-Coordinator (2024, *current*)
- Workflow Chair, Machine Learning for Health (ML4H) Symposium (2024, *current*)
- University Relations Chair, Computer Science & Engineering Graduate Student Organization, University of Michigan (2023-2024)
- Panelist, Summer Research Opportunity Program, University of Michigan (2023)
- AI Lab Graduate Admissions Committee Volunteer, Department of Computer Science & Engineering, University of Michigan (2022, 2024)
- Reviewing: AISTATS, ML4H, MLHC, NeurIPS, ICLR, KDD (workshop). **NeurIPS Research2Clinics 2021 Best Reviewer Award.**

Awards

- CSE Honors Competition Finalist (AI Lab Representative), University of Michigan (2024)
- Team 2nd Prize (Stanford Chirpy Cardinal), Alexa Socialbot Grand Challenge (2021)