



Megha Srivastava

Ph.D. Student in Computer Science, admitted Autumn 2020

Publications

PUBLICATIONS

- **Training large language models on narrow tasks can lead to broad misalignment.** *Nature*
Betley, J., Warncke, N., Szyber-Betley, A., Tan, D., Bao, X., Soto, M., Srivastava, M., Labenz, N., Evans, O.
2026; 649 (8097): 584-589
- **Shared Autonomy for Proximal Teaching**
Srivastava, M., Iranmanesh, R., Cui, Y., Gopinath, D., Sumner, E., Silva, A., Dees, L., Rosman, G., Sadigh, D., ASSOC COMPUTING MACHINERY
IEEE.2025: 232-241
- **Do Users Write More Insecure Code with AI Assistants?**
Perry, N., Srivastava, M., Kumar, D., Boneh, D., ACM
ASSOC COMPUTING MACHINERY.2023: 2785-2799
- **Question Generation for Adaptive Education**
Srivastava, M., Goodman, N., Assoc Computat Linguist
ASSOC COMPUTATIONAL LINGUISTICS-ACL.2021: 692-701
- **Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning**
Srivastava, M., Heidari, H., Krause, A., Assoc Comp Machinery
ASSOC COMPUTING MACHINERY.2019: 2459-68