Stanford

Amir Zur

Ph.D. Student in Linguistics, admitted Autumn 2025

Publications

PUBLICATIONS

- **Causal Abstraction: A Theoretical Foundation for Mechanistic Interpretability** *JOURNAL OF MACHINE LEARNING RESEARCH*
  Geiger, A., Ibeling, D., Zur, A., Chaudhary, M., Chauhan, S., Huang, J., Arora, A., Wu, Z., Goodman, N., Potts, C., Icard, T.
  2025; 26