Yuxuan (Effie) Li

liyuxuan@stanford.edu | https://effie-li.github.io

- 5+ years of interdisciplinary research experience in deep learning, cognitive science, neuroscience
- Research interests: human-like AI, machine/human cognition, mechanistic interpretability

Education

2019 – 2024 **Stanford University, PhD** in Cognitive Psychology (expected Dec 2024). Transcript.
- Weiland Fellow, School of Humanities and Sciences
- Alumna, Center for Mind, Brain, Computation and Technology

2013 – 2017 **Trinity College, BS** in Computer Science and Psychology. *summa cum laude.*

Research Experience

2024 Research intern @ **Meta**. Mentor: Karl Ridgeway.
- Use multi-modal (vision+language) representation learning to understand real-world human behavior

2023 PhD research intern @ **Allen Institute for AI**. Mentor: Luca Weihs.
- Designed and implemented self-supervised goal-directed pretraining objectives
- Systematically evaluated representation learning for agent planning in realistic environments (preprint)

2019 – now PhD researcher @ **Stanford University PDP Lab**. Advisor: James McClelland.
- Reverse-engineered how transformers learn graph traversal, evaluated models' emergent human-like subgoal choices, iterative subgoal selection, and interpretability
- Established mechanistic insights on how transformers achieve task decomposition and multi-task learning on simple algorithmic tasks; Proposed a new sequence encoding method that boosts length generalization in transformers (TMLR paper, code)
- Built hippocampus-inspired recurrent memory modules for deep RL agents (report, code)
- Conducted behavioral studies and built models of human planning processes (paper, code)

2017 – 2019 Research specialist @ **UPenn Computational Memory Lab**. Advisor: Michael Kahana.
- Developed a novel data sampling method and trained neural decoders from large-scale EEG time series data, yielding new insights into human memory (NatComm paper, code)

Technical Skills

Coursework Graduate coursework in deep learning, reinforcement learning, deep multi-task and meta-learning, machine learning, computational neuroscience

Programming **Python, R,** some experience with MATLAB, HTML/CSS/JavaScript (jquery, jspsych)

Packages **Deep learning** (pytorch, pytorch-lightning, allenact, einops), **experiment/server management** (wandb, beaker), **machine learning** (scikit-learn), **data analysis** (scipy, numpy, pandas; tidyr, dplyr, lme4), **data visualization** (matplotlib; ggplot2), **cognitive (neuro)science** (mne, ptsa; rtdists)

Other LaTeX, statistics (linear modeling, generalized linear modeling, mixed-effects models), representation analysis, online behavioral platforms (Amazon MTurk, Prolific)

Publications and Preprints

2024 **Li, Y.**, & McClelland, J.L. Emergent human-like path preferences and implicit subgoal selection in transformers learning graph traversal. *Cognitive Computational Neuroscience.*

2024 **Li, Y.**, Pazdera, J.K., & Kahana, M.J. EEG decoders track memory dynamics. *Nature Communications.*

2023 **Li, Y.**, & Weihs, L. Understanding representations pretrained with auxiliary losses for embodied agent planning. *NeurIPS 2023 Generalization in Planning Workshop.*

2023 **Li, Y.**, & McClelland, J.L. Representations and computations in transformers that support generalization on structured tasks. *Transactions on Machine Learning Research.*

2023 Kahana, M.J., Lohnas, L.J., Healey, K., . . ., **Li, Y.**, . . ., & Weidemann, C.T. The Penn Electrophysiology of Encoding and Retrieval Study. *JEP: LMC*.

2022 **Li, Y.**, & McClelland, J.L. A weighted constraint satisfaction approach to human goal-directed decision making. *PLOS Computational Biology*.

2022 Katerman, B.S., **Li, Y.**, Pazdera, J.K., Keane, C., & Kahana, M.J. EEG biomarkers of free recall. *NeuroImage*.

2018 Grubb, M.A., & **Li, Y.** Assessing the role of accuracy-based feedback in value-driven attentional capture. *Attention, Perception, & Psychophysics*.

Talks and Presentations

Mar 2024 **Li, Y.** Emergent structured computation from learning and its implications for cognitive science and AI. *Microsoft Research Lab – Redmond.*

Nov 2023 **Li, Y.** Systematic generalization and emergent structures in transformers trained on structured tasks. *FriSem seminar, Department of Psychology, Stanford University.*

Apr 2022 **Li, Y.** A weighted constraint satisfaction approach to human goal-directed decision making. *Cognitive Tools Lab, University of California, San Diego.*

Feb 2021 **Li, Y.** Model-based reinforcement learning and the reinforcement learning framework for human behavior. *TA Lecture in PSYCH 209, Stanford University.*

2020, 2021 **Li, Y.** Building online psychology experiments with jsPsych: a tutorial. *TA Lecture in PSYCH 251, Stanford University.*

Honors and Awards

2022 – 2024 Ric Weiland Graduate Fellowship in the Humanities & Sciences. Stanford University.

2013 – 2017 Phi Beta Kappa, Dean's Scholar (top 5%), Faculty Honors, Holland Scholar. Trinity College.

Teaching and Services

Reviewer Cognitive Science Society, 2022 –

Committee Cognitive Neuroscience Seminar Organizing Committee, Stanford Psychology, 2021 – 2022

TA Department of Psychology, Stanford University. For graduate courses: neural network models of cognition, brain decoding, experimental methods, developmental psychology

TA Department of Computer Science, Trinity College. For undergraduate courses: introduction to computing, mathematical foundations of computing