Yuxuan (Effie) Li

liyuxuan@google.com | https://effie-li.github.io

- Research focus: machine and human cognition, mechanistic interpretability

Education

2019 – 2024   **Stanford University, PhD** in Cognitive Psychology.

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

Research Positions

2025 -   **Research Scientist** @ Google DeepMind

2024 summer   **Research Intern** @ Meta

2023 summer   **Research Intern** @ Allen Institute for AI

2017 – 2019   **Research Specialist** @ UPenn

2016 summer   **Research Intern** @ Columbia Business School

Research Topics and Publications

Visual reasoning and agentic planning

2025   **Video models are zero-shot learners and reasoners.** *paper*, *website*

T. Wiedemer, **Y. Li**, P. Vicol, S. Gu, N. Matarese, K. Swersky, B. Kim, P. Jaini, & R. Geirhos.

2025   **EgoToM: Benchmarking Theory of Mind Reasoning from Egocentric Videos.** *paper*

Y. Li, V. Veerabadran, M.L. Iuzzolino, B.D. Roads, A. Celikyilmaz, & K. Ridgeway.

2023   Understanding representations pretrained with auxiliary losses for embodied agent planning. *paper*

Y. Li, & L. Weihs. *NeurIPS 2023 Generalization in Planning Workshop.*

Learning, generalization, and interpretability

2025   **Just-in-time and distributed task representations in language models.** *paper*

Y. Li, D. Campbell, S.C.Y. Chan, & A.K. Lampinen. *NeurIPS 2025 Mechnistic Interpretability Workshop (spotlight).*

2025   Latent learning: episodic memory complements parametric learning by enabling flexible reuse of experiences. *paper*

A.K. Lampinen, M. Engelcke, **Y. Li**, A. Chaudhry, & J.L. McClelland.

2025   Learning to decompose: Human-like subgoal preferences emerge in transformers learning graph traversal.

Y. Li, & J.L. McClelland. *Under review.*

2023   Representations and computations in transformers that support generalization on structured tasks. *paper*, *code*

Y. Li, & J.L. McClelland. *Transactions on Machine Learning Research.*

Human psychology and neuroscience

2025   Representation biases: will we achieve complete understanding by analyzing representations? *paper*

A.K. Lampinen, S.C.Y. Chan, **Y. Li**, & K. Hermann.

2024	EEG decoders track memory dynamics. *paper, code*
	Y. Li, J.K. Pazdera, & M.J. Kahana. *Nature Communications.*
2023	The Penn Electrophysiology of Encoding and Retrieval Study. *paper*
	M.J. Kahana, L.J. Lohnas, K. Healey, . . ., **Y. Li**, . . ., & C.T. Weidemann. *JEP: LMC.*
2023	A weighted constraint satisfaction approach to human goal-directed decision making. *paper, code*
	Y. Li, & J.L. McClelland. *PLOS Computational Biology.*
2022	EEG biomarkers of free recall. *paper*
	B.S. Katerman, **Y. Li**, J.K. Pazdera, C. Keane, & M.J. Kahana. *NeuroImage.*
2018	Assessing the role of accuracy-based feedback in value-driven attentional capture. *paper*
	M.A. Grubb, & **Y. Li.** *Attention, Perception, & Psychophysics.*

Talks and Presentations

Dec 2024	**Li, Y.** Emergent task decomposition and subgoal choices in transformers. *Mind, Brain, Computation and Technology Seminar Series, Stanford University.*
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	NeurIPS, CVPR, TMLR, CogSci, CCN
TA	Neural network models of cognition, brain decoding, Experimental methods, developmental psychology, introduction to computing, mathematical foundations of computing