

# Houjun Liu

(650) 282-0675 [houjun@stanford.edu](mailto:houjun@stanford.edu) [jemoka.com](http://jemoka.com)

## Education

**Stanford University** • Undergraduate, Computer Science *September 2023 -*  
Concentration in Artificial Intelligence. Advised by Prof. Mykel Kochenderfer, GPA 4.169.

### Activities

Team Lead, Stanford Student Space Initiative, Mars; we're sending a rover to Antarctica!

### Selected Coursework

*Graduate Level* (CS239) Advanced Sequential Decision Making, (CS224n) Natural Language Processing with Deep Learning, (CS242) Programming Languages, (CS254) Computational Complexity.

*Undergraduate Level* (CS143) Compilers, (CS111) Operating Systems Principles, (MATH53) Differential Equations, (CS154) Theory of Computation, (CS107) Computer Systems.

## Experience

**Stanford Artificial Intelligence Laboratory** • Stanford, CA *2023 - present*  
Undergraduate Research Associate, NLP Group.

I work on developing adaptive, scalable reasoning transformers with Prof. Chris Manning.

I develop efficient models for **Stanza**, a classic NLP system used by millions.

**Stanford Intelligent Systems Laboratory** • Stanford, CA *2024 - present*  
Undergraduate Research Associate.

I work on using RL+RLAIF to red-team and improve LLMs with Prof. Mykel Kochenderfer.

My projects have **won \$500k+ from the Schmidt Futures innovation grants**.

**Carnegie Mellon University, TalkBank** • Pittsburgh, PA *2022 - present*  
Consulting Research Engineer.

I'm an NIH-funded consulting research engineer at CMU TalkBank.

I help maintain TalkBank, the world's largest auditory language pathology dataset.

I'm the lead maintainer of **Batchalign**, a speech analysis tool with 1000+ MAU.

**Shabang Systems, LLC** • [shabang.io](http://shabang.io) *2019 - present*  
Managing Partner.

I started a software development workshop and creative agency.

**Conduction** • [conduction.com](http://conduction.com) *2019 - 2023*  
Co-Founder, Lead Developer.

I created a **task management app** with 4,000+ MAU worldwide.

I led effective marketing campaigns for 5,000+ monthly impressions

**Dragonfruit AI** • Menlo Park, CA *2021 - 2022*  
Consulting Machine Learning Engineer.

I trained YOLO models with a custom data flywheel for loss prevention.

I pushed 4 major features into production, created 33% training workflow.

## Publications

- 2025 **H Liu, J Bauer, CD Manning. Drop Dropout on Single Epoch Language Model Pretraining.**  
We suggest that dropout is useless in language model pretraining.  
*ACL Findings 2025. Vienna, Austria.*
- H Liu, J Bauer, K D'Oosterlinck, C Potts, CD Manning. MSCAW-coref: Multilingual, Singleton and Conjunction-Aware Word-Level Coreference Resolution.**  
We propose an entity tracking architecture which reduces tracking complexity from  $O(n^4)$  to  $O(n^2)$ .  
*CRAC at EMNLP 2024. Miami, Florida.*
- MR Schlichting, V Rasmussen, H Alazzez, H Liu, K Jafari, AF Hardy, DM Asmar, MJ Kochenderfer. LeRAAT: LLM-Enabled Real-Time Aviation Advisory Tool.**  
In collaboration with Airbus, we build an intelligent tool to alert pilots of ECAM-generated failures.  
*arXiv preprint.*
- 2024 **AF Hardy\*, H Liu\*, B Lange, MJ Kochenderfer. ASTPrompter: Weakly Supervised Automated Language Model Red-Teaming to Identify Likely Toxic Prompts.**  
We propose a reinforcement-learning scheme which identifies likely-to-occur toxic sequences.  
*Under Review.*
- H Liu, AM Weakley, J Zhang, X Liu. A Transformer Approach for Cognitive Impairment Classification and Prediction.**  
We propose a new transformers-based architecture for time-series data imputation and use it on NACC Alzheimer's data.  
*Workshop in Health Intelligence at AAAI 2024, Vancouver, BC.*  
*Published in Alzheimer's Dementia and Associated Disorders.*
- H Liu, B MacWhinney. Morphosyntactic analysis for CHILDES.**  
We highlight the difficulties of morphosyntactic analysis for children's speech pathology.  
*Language Development Research.*
- 2023 **H Liu, B MacWhinney, D Fromm, A Lanzi. Automation of Language Sample Analysis.**  
We discuss Batchalign, a tool that automates annotation required for language sample analysis.  
*Journal of Speech, Language, and Hearing Research.*
- AM Lanzi, AK Saylor, D Fromm, H Liu, B MacWhinney, ML Cohen. DementiaBank: Theoretical rationale, protocol, and illustrative analyses.**  
We discuss DementiaBank, a new effort to collect pathological speech for patients with dementia.  
*American Journal of Speech-Language Pathology.*

## Invited Talks

- "Understanding and improving language model capabilities through implicit design and explicit reasoning", Singapore National University (2025).
- "Deep-Learning Analysis of Longitudinal Alzheimer's Data", National Alzheimer's Dementia Research Center (2023).
- "Automated Context-Aware Lexicography Using Large Online Encyclopedias", Intelligent Computing (2022).

Service

ICLR 2025 Workshop on Human-AI Coevolution • Organizer	2025
AAAI 2025, Special Track on AI Alignment • Reviewer	2025
Alzheimer's Research & Therapy • Reviewer	2025
Journal of Aerospace Information Systems • Reviewer	2024

Teaching

Decision Making Under Uncertainty (CS238) • TA, Stanford	Fall Quarter, 2024
AIFS AI Bridge • Co-Instructor, Head TA, UC Davis AIFS	Summer Quarter, 2024

Awards	Neo Scholar (2025), NIH Research Consultant (NIH grant #HD082736).
--------	--

Technical Skills	PyTorch, TensorFlow, Docker, Kubernetes, Julia, Rust, Mathematica, MATLAB, Clojure, Common Lisp, Haskell, Eigen UART/i2C development, React, Svelte.
------------------	--

Human Languages	English (Native), Mandarin (Native), Spanish (Intermediate)
-----------------	---