

EDUCATIONAL BACKGROUND

- Sep/2025 - Current. **University of Chicago** Chicago, IL, USA
M.A. in Computational Social Science GPA: 3.73/4.00
- Jan/2021 - Jun/2025. **University of California, San Diego** San Diego, CA, USA
B.S. in Cognitive Science (Machine Learning Track) with Highest Distinction GPA: 3.61/4.00
- *B.S. in Applied Mathematics*
Minor in History; Minor in Linguistics

Related Courses: Neural Networks and Deep Learning (A+); Computational Linguistics (A+); Machine Learning with Few Labels (A); Advanced Machine Learning Methods (A+); Modeling & Data Analysis (A).
Honors: **Provost Honors** (6 terms), **Triton Research & Experiential Learning Scholars** (2 terms)

RESEARCH INTERESTS

I study what language models do to human meaning and judgment as they increasingly sit between people: what they compress, add, or distort as they take part in creative expression and cultural transmission, and how they reshape attention and evaluation when people and institutions rely on them. I aim to build human-centered AI and evaluations that capture what traditional benchmarks miss.

PUBLICATIONS

- **Zehan Li**, Yutong Zhu, Siyang Wu, Honglin Bao, James A. Evans. [Narrative Flattening: How Post-Training Compresses Thematic, Affective, and Stylistic Variation in LLM Fiction](#). **EMNLP 2026**, Main Conference, **Oral**.
- **Zehan Li**, Yuxuan Wang, Ali El Lahib, Ying-Jieh Xia, Xinyu Pi. [Simulated Ignorance Fails: A Systematic Study of LLM Behaviors on Forecasting Problems Before Model Knowledge Cutoff](#). **IJCAI 2026**, Main Track.
- Ali El Lahib, Ying-Jieh Xia, **Zehan Li**, Yuxuan Wang, Xinyu Pi. [Temporal Leakage in Search-Engine Date-Filtered Web Retrieval: A Retrospective Forecasting Case Study](#). **ACL 2026**, Main Conference.
- **Zehan Li**, Ruhua Pan, Xinyu Pi. [Beyond LLMs: A Linguistic Approach to Causal Graph Generation from Narrative Texts](#). **WNU @ NAACL 2025** (Workshop).

RESEARCH EXPERIENCES

- What happens to meaning when LLMs translate across languages and cultures? Jun/2026 - Current
*Project Lead. With **Junsol Kim** & Prof. **James A. Evans**, University of Chicago*
 - **Built a per-dimension semantic atlas of cross-lingual transmission** across lexical, grammatical, and pragmatic dimensions (e.g., kinship, evidentiality, honorifics) and six model families.
 - Showed that translation compresses meaning in one direction and forces new commitments in the other, and that surface-form loss is not semantic loss; now tracing how meaning drifts as texts are relayed across multiple languages.
- What happens when AI sits on both sides of a hiring decision? Jun/2026 - Current
*Project Lead. With Prof. **Jiaxin Pei** (UT Austin) & Prof. **Xinyu Liang** (INSEAD)*
 - **Built a three-level audit of LLM résumé screening**, from a single résumé to a multi-candidate prompt to an autonomous agent, on the same real résumés and multiple frontier LLMs, placing every design choice on a common rerun-to-chance scale.
 - Found that screening decisions are highly unstable and shaped by many factors: reversing the display order of 100 résumés in one prompt replaces 74% of the top-3 shortlist (vs. 26% on rerun), and 30% of résumés pass under some LLMs but fail under others.

- How do LLMs remember, and reshape, the scientific literature? Oct/2025 - Current
Researcher. Supervised by Prof. Bernard Koch, University of Chicago
 - **Found that LLM citation memory is organized by textual co-occurrence** (semantic similarity, co-citation) rather than citation-graph topology or the social processes that generate citations.
 - Designed a citation in-painting task across more than 10 LLMs on ~100K SciSciNet v2 papers, using Infini-gram corpus frequencies to separate model behavior from data composition.
 - Now comparing LLM-generated bibliographies against real human citations for the same papers.
- Why do LLM-generated stories feel narratively flat? Jan/2026 - May/2026
Project Lead. Supervised by Prof. James A. Evans, Knowledge Lab, University of Chicago
 - Resulting publication: [EMNLP 2026 \(Oral\)](#). **Led a matched story-continuation study of how post-training reshapes LLM fiction**, comparing Base, SFT, DPO, and RLVR checkpoints against human continuations across multiple story corpora.
 - Developed computational measures of thematic motion, affective intensity, and stylistic diversity, showing that post-training progressively compresses narrative variation into more uniform and neutral outputs.
- Can prompts reliably rewind LLM knowledge for forecasting evaluation? Jul/2025 - Jan/2026
Project Lead. Supervised by Prof. Zhiting Hu, UCSD
 - Resulting publications: [IJCAI 2026](#) and [ACL 2026](#). **Provided the first systematic test of whether Simulated Ignorance (SI) can approximate True Ignorance (TI) in retrospective LLM forecasting evaluation**, deploying Qwen3-32B and Llama models on cluster servers.
 - Across 477 competition-level questions and 9 models, found that SI fails systematically: cutoff instructions leave a 52% performance gap, chain-of-thought cannot suppress post-cutoff knowledge, and reasoning-optimized models are *worse*, showing that prompts cannot reliably “rewind” model knowledge.
- Quantifying Conceptual Density and Vacuity in Text Aug/2024 - Jun/2025
Undergrad Honors Thesis. Supervised by Prof. Zhuowen Tu & Alex Warstadt, UCSD
 - **Developed a theoretical model for quantifying text specificity and vacuity** using linguistic properties as primary dimensions, proposing a metric system for conceptual density in written communication.
 - Benchmarked outputs from multiple LLMs to compare their conceptual density and informational richness against human-written text.
- Can linguistic properties improve causal graph generation from narratives? Feb/2024 - Sep/2024
Project Lead. Supervised by Prof. Zhiting Hu, UCSD
 - Resulting publication: [WNU @ NAACL 2025](#). **Developed an open-source linguistics-based model for causal graph generation**, achieving 82% interpretability, a 9% improvement over direct GPT prompting baselines.
 - Built a feature extraction pipeline using spaCy and nltk that leverages linguistic properties (e.g., time boundness) for causal link identification, yielding a 15% accuracy gain over LLM-only approaches based on human evaluation.

EXTRACURRICULAR EXPERIENCES

- Metaculus AI Forecasting Tournament (Fall 2025 Cycle) Oct/2025 - Dec/2025
 - Developed an autonomous AI forecasting agent that competed in the Metaculus AI Bot Tournament, achieving **15th place globally** and a \$1,500 prize.
 - Engineered a well-calibrated probabilistic prediction system, demonstrated by close alignment between predicted confidence and actual resolution rates across the full probability range.

PROFESSIONAL EXPERIENCES

- Instructional Assistant (2 courses): Deep Learning & Supervised Machine Learning Sep/2024 - Jun/2025
UCSD Cognitive Science Department. Under Prof. Zhuowen Tu
 - Led weekly 80-minute discussion sessions for 100+ students across two quarters, covering topics from regression and classification to GPT architectures and prompt engineering.
 - Provided individualized support through office hours and detailed assignment feedback.

OTHER INFORMATION

- Computer Skills: Python, Java, SQL, R, MATLAB, JavaScript, L^AT_EX, HTML/CSS, Docker
- Languages: Chinese (native), English (proficient), French(Intermediate)