

LECHEN ZHANG

📍 Urbana-Champaign, IL | ☎ +1 734 834 8529 | 📩 lechenz3@illinois.edu
🎓 Google Scholar | 🐾 GitHub | 🐦 Twitter | 🌐 Homepage

EDUCATION

University of Illinois Urbana-Champaign <i>PhD in Computer Science</i>	<i>Jan. 2026 – Dec. 2029 (Expected)</i>
University of Michigan, Ann Arbor <i>Master in Information Science GPA: 4.00/4.00 Distinguished Thesis</i>	<i>Aug. 2022 – May. 2024</i>
<ul style="list-style-type: none">• Advisor: Prof. Tal August Shanghai Jiao Tong University <i>Bachelor in Electrical and Computer Engineering</i> <ul style="list-style-type: none">• Related coursework: Computer Vision (A), Computer Architecture (A), Intro to Linguistics (A)	<i>Sep. 2019 – Aug. 2023</i>

WORK EXPERIENCE

Research Assistant <i>Advisor: Prof. David Jurgens and Prof. Lu Wang</i>	<i>May. 2024 – Aug. 2025</i> <i>University of Michigan</i>
--------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------

PUBLICATIONS

* indicates equal contribution

Peer-Reviewed Papers

- [10] **Skill-Aware Data Selection and Fine-Tuning for Data-Efficient Reasoning Distillation**
Lechen Zhang, Yunxiang Zhang, Wei Hu, Lu Wang
MATH-AI Workshohp @ NeurIPS 2025 | Project Leade [\[arXiv\]](#) [\[Code\]](#)
- [9] **Logit Arithmetic Elicits Long Reasoning Capabilities Without Training**
Yunxiang Zhang, Muhammad Khalifa, Lechen Zhang, Xin Liu, Ayoung Lee, Xinliang Frederick Zhang, Farima Fatahi Bayat, Lu Wang
ScalR Workshop @ COLM 2025 | Main Contributor [\[arXiv\]](#) [\[Code\]](#)
- [8] **VeriFact: Enhancing Long-Form Factuality Evaluation with Refined Fact Extraction and Reference Facts**
Xin Liu, Lechen Zhang, Sheza Munir, Yiyang Gu, Lu Wang
EMNLP 2025 | Main Contributor [\[ACL Anthology\]](#) [\[arXiv\]](#) [\[Code\]](#)
- [7] **FactBench: A Dynamic Benchmark for In-the-Wild Language Model Factuality Evaluation**
Farima Fatahi Bayat, Lechen Zhang, Sheza Munir, Lu Wang
ACL 2025 | Main Contributor [\[ACL Anthology\]](#) [\[arXiv\]](#) [\[Code\]](#) [\[Twitter\]](#)
- [6] **Towards Global AI Inclusivity: A Large-Scale Multilingual Terminology Dataset (GIST)**
Jiarui Liu*, Iman Ouzzani*, Wenkai Li*, Lechen Zhang, Tianyue Ou, Houda Bouamor, Zhijing Jin, Mona Diab
ACL 2025 Findings [\[ACL Anthology\]](#) [\[arXiv\]](#)
- [5] **Causally Modeling the Linguistic and Social Factors that Predict Email Response**
Yinuo Xu*, Hong Chen*, Sushrita Rakshit*, Aparna Ananthasubramaniam*, Omkar Yadav*, Mingqian Zheng*, Michael Jiang*, Lechen Zhang*, Bowen Yi*, Kenan Alkiek*, Abraham Israeli*, Bangzhao Shu*, Hua Shen*, Jiaxin Pei*, Haotian Zhang*, Miriam Schirmer*, David Jurgens (*Randomized Author Order*)
NAACL 2025 | Main Contributor [\[ACL Anthology\]](#)

[4] You don't need a personality test to know these models are unreliable: Assessing the Reliability of Large Language Models on Psychometric Instruments

Bangzhao Shu*, Lechen Zhang*, Minje Choi, Lavinia Dunagan, Lajanugen Logeswaran, Moontae Lee, Dallas Card, David Jurgens

NAACL 2024 Oral | Project Leader [\[ACL Anthology\]](#) [\[arXiv\]](#) [\[Slides\]](#) [\[Code\]](#) [\[Twitter\]](#)

Papers Under Review & Preprints

[3] Cross-Lingual Prompt Steerability: Towards Accurate and Robust LLM Behavior across Languages

Lechen Zhang*, Yusheng Zhou, Tolga Ergen, Lajanugen Logeswaran, Moontae Lee, David Jurgens

Under Review | Project Leader [\[arXiv\]](#) [\[Code\]](#)

[2] SPRIG: Improving Large Language Model Performance by System Prompt Optimization

Lechen Zhang, Tolga Ergen, Lajanugen Logeswaran, Moontae Lee, David Jurgens

Under Review | Project Leader [\[arXiv\]](#) [\[Slides\]](#) [\[Code\]](#) [\[Twitter\]](#)

[1] Real or Robotic? Assessing Whether LLMs Accurately Simulate Qualities of Human Responses in Dialogue

Jonathan Ivey*, Shivani Kumar*, Jiayu Liu*, Hua Shen*, Sushrita Rakshit*, Rohan Raju*, Haotian Zhang*, Aparna Ananthasubramaniam*, Junghwan Kim*, Bowen Yi*, Dustin Wright*, Abraham Israeli*, Anders Giovanni Møller*, Lechen Zhang*, David Jurgens (*Randomized Author Order*)

Under Review | Project Leader [\[arXiv\]](#) [\[Code\]](#) [\[Twitter\]](#)

RESEARCH EXPERIENCE

Concluded Projects

Improving LLMs' general performance by System Prompt Optimization

Feb. 2024 – May. 2025

University of Michigan

Advisor: David Jurgens

- Design an edit-based genetic system prompt optimizer *SPRIG* that generally improves LLM performance across 47 diverse benchmarks.
- Discover strong generalization capability of system prompt optimization across tasks, models, languages, and even out-of-domain challenges, and its complementary effect with existing task-specific optimizers.
- Develop new RL strategies to efficiently explore and expand the design space of system prompts.
- Lead the whole project independently, completing all aspects from research ideation to paper writing.

Factuality Evaluation pipeline and benchmark in real-world scenarios

May. 2024 – Nov. 2024

University of Michigan

Advisor: Lu Wang

- Develop a retrieval-based factuality evaluation pipeline for long-form text that is more fine-grained, efficient and aligns better with human than existing work.
- Build a benchmark of prompts that are factually challenging to LLMs by filtering LMSYS-1M dataset, clustering representative prompts, and selecting based on their scores on the designed evaluation pipeline.
- Lead experiments on open-source models, reproduce 3 existing studies as baselines, and implement parallel optimization for a 10x speedup.
- Contribute extensively to paper writing and post-submission tasks, including drafting key sections, analyzing results, creating visuals, managing code repository, and preparing rebuttals.

Assessment of LLM Simulation Ability of Human Responses in Dialogue

Jul. 2024 – Oct. 2024

University of Michigan

Advisor: David Jurgens

- Implement 15 evaluation metrics for LLM simulation quality across lexical, syntactic, semantic, and style features.
- Lead the collection of 50 instruction prompts and generate 1M dialogue simulation results from 9 LLMs.
- Set up annotation platform for the whole lab to collect human annotations as a baseline.
- Lead the project as the main contributor to coding, paper writing and post-submission tasks.

Modeling Intent, Expectation, and Responsiveness in Email Conversations

Mar. 2024 – Jun. 2024

University of Michigan

Advisor: David Jurgens

- Preprocess raw Email data and build an email relationship network to sample and construct a dataset for analysis.
- Evaluate LLM's ability to infer Email Intent by fine-tuning RoBERTa and running zero-shot inference on Llama-3.
- Serve as the main contributor to annotating, coding, paper writing and post-submission tasks.

Robustness of LLMs' personality under Psychometric Instruments

Sep. 2023 – Dec. 2023

University of Michigan

Advisor: David Jurgens

- Build evaluation dataset and metrics that measures the robustness of various LLMs' personalities under spurious prompt variation and rephrased statements, and evaluate on 17 different LLMs.
- Experiment the personality and robustness shifts under different conditions, such as injecting personalities through prompts, and fine-tuning LLMs (Llama2, Flan-T5, etc.) on various corpora (Bible, 4chan, r/Donald, etc.).
- Lead the project and contribute to most coding, experiments, writing, and post-submission tasks such as the rebuttal, code repository, Twitter thread, and related presentations.

PRESENTATIONS

EMNLP 2025 Poster (Suzhou) – *VeriFact: Enhancing Long-Form Factuality Evaluation with Refined Fact Extraction and Reference Facts.* [\[ACL Anthology\]](#)

NAACL 2024 Oral (Mexico City) – *You don't need a personality test to know these models are unreliable: Assessing the Reliability of Large Language Models on Psychometric Instruments.* [\[Slides\]](#) [\[ACL Anthology\]](#)

PROFESSIONAL SERVICES

Conference Reviewer

- ACL 2025
- MATH-AI Workshohp @ NeurIPS 2025
- ScalR Workshop @ COLM 2025
- NAACL 2025
- EMNLP 2024 (*Outstanding Reviewer*), EMNLP 2025

SKILLS

Programming: Python, C/C+++, C#, Java, Go, SQL, MATLAB, R, Kotlin, LaTeX

Frameworks: PyTorch, Tensorflow, Transformers, Accelerate, DeepSpeed, PEFT, NLTK, Scikit-Learn

Languages: Chinese (Native), English (Fluent), Japanese (Basic)