Zhaoyi (Joey) Hou

website: joeyhou.github.io email: joey.hou@pitt.edu

Research Interest Natural Language Processing, LLM Evaluation, Commonsense Reasoning

Education

University of Pittsburgh 2023 - Present
Computer Science, Doctor of Philosophy, Graduate Student Researcher
Research Advisor: Dr. Xiang Lorraine Li

University of Pennsylvania 2021 - 2023
Data Science, Master of Science in Engineering
Research Advisor: Dr. Chris Callison-Burch

University of California, San Diego 2017 - 2021
Data Science, Bachelor of Science Cum Laude
Research Advisor: Dr. Jingbo Shang

Publications

[10] **Zhaoyi Joey Hou**, Bowei Alvin Zhang, Yining Lu, Bhiman Kumar Baghel, Anneliese Brei, Ximing Lu, Meng Jiang, Faeze Brahman, Snigdha Chaturvedi, Haw-Shiuan Chang, Daniel Khashabi, Xiang Lorraine Li. *CreativityPrism: A Holistic Benchmark for Machine Creativity* [In Submission]

[9] **Zhaoyi Joey Hou**, Alejandro Ciuba, Xiang Lorraine Li. *Improve LLM-based Automatic Essay Scoring with Linguistic Features* [Innovation and Responsibility in AI-Supported Education (Spotlight Paper) - AAAI2025]

[8] **Zhaoyi Joey Hou**, Adriana Kovashka, Xiang Lorraine Li. *Leveraging Large Models for Evaluating Novel Content: A Case Study on Advertisement Creativity* [Accepted to EMNLP2025]

[7] **Zhaoyi Hou**, Li Zhang, Chris Callison-Burch. *Choice-75: A Dataset on Decision Branching in Script Learning* [LREC-COLING2024]

[6] Tianyi Zhang*, Li Zhang*, **Zhaoyi Hou**, Ziyu Wang, Yuling Gu, Peter Clark, Chris Callison-Burch, Niket Tandon. *PROC2PDDL: Open-Domain Planning Representations from Texts* [2nd Natural Language Reasoning and Structured Explanations Workshop - ACL2024]

[5] Alyssa Hwang*, Bryan Li*, **Zhaoyi Hou***, Dan Roth. *Large Language Models as Sous Chefs: Revising Recipes with GPT-3*

[4] Tianyi Zhang*, Isaac Tham*, **Zhaoyi Hou***, Jiaxuan Ren, Liyang Zhou, Hainiu Xu, Li Zhang, Lara J. Martin, Rotem Dror, Sha Li, Heng Ji, Martha Palmer, Susan Brown, Reece Suchocki, and Chris Callison-Burch. *Human-in-the-Loop Schema Induction* [ACL2023]

[3] Xiaochen Kev Gao, **Zhaoyi Hou**, Yifei Ning, Jingbo Shang, Vish Krishnan. *Towards Comprehensive Patent Approval Predictions: Beyond Traditional Document Classification* [ACL2022]

[2] Caitlin A. Stamatis, Jonah Meyerhoff, Tingting Liu, **Zhaoyi Hou**, Garrick Sherman, Brenda L. Curtis, Lyle H. Ungar, David C. Mohr. *The Association of Language Style Matching in Text Messages with Symptoms of Affective Psychopathologies* [Procedia Computer Science]

[1] Artemis Panagopoulou, Manni Arora, ...(6 more), **Zhaoyi Hou**, Alyssa Hwang, Lara Martin, Sherry Shi, Chris Callison-Burch, Mark Yatskar. *QuakerBot: A Household Dialog System Powered by Large Language Models* [Alexa Prize TaskBot Challenge Proceedings]

(*Equal contribution)

Work Experience

Amazon Web Service May 2025 - Aug 2025
Applied Scientist Intern (AWS Q Console)
- Synthesized tool-use conversation with real-world failure cases;
- Designed and implemented a comprehensive evaluation framework capable of detecting signals

beyond user satisfaction, e.g., tool failure, agent hallucination when using tool, etc.

United Imaging Intelligence May 2023 - Aug 2023
LLM Research Intern
- LoRA fine-tuned Llama-2 with HuggingFace, using medical domain textbooks.
- Implemented a retrieval-augmented generation (RAG) question-answering (QA) pipeline based on Llama-2 and the medical domain knowledge base.
- Outperformed existing open-sourced models (60% accuracy) in the United States Medical Licensing Examination (USMLE) benchmark.

Projects

Amazon Alexa TaskBot Competition Nov 2021 - Apr 2022
Information Retrieval
- Implemented the document retrieval module for the Alexa TaskBot competition;
- Improved the retrieval success rate by 25% and advanced to the final list.

Stack Overflow Question Quality Classification May 2021 - Jul 2021
Text Mining & Cloud Computing
- Built a text data ETL and analysis pipeline for 60,000 Stack Overflow question texts;
- Built an XGBoost classification model with AWS SageMaker and deployed it as an AWS SageMaker Endpoint;
- Achieved 87% accuracy in the "High-Quality Question" classification task.

Machine Learning for Ophthalmological Diagnosis Nov 2019 - Dec 2020
Computer Vision & Healthcare
- Built an image classification pipeline to pre-diagnose common eye diseases with a convolutional neural network;
- Achieved 75% accuracy for classifying involutional ptosis, thyroid eye disease, and normal eyes.

Awards

Best Problem-Solution Sep 2024
Annual Doctoral Guild Poster Slam
School of Computing and Information at University of Pittsburgh

HDSI Undergraduate Scholarship Dec 2019
Halıcıoğlu Data Science Institute
University of California, San Diego

Service

Reviewer
Association of Computational Linguistics Rolling Review (ARR) 2024, 2025
Empowering Machine Learning and Large Language Models with Domain and Commonsense Knowledge (AAAI-MAKE 2024) 2024

Teaching

University of Pittsburgh
Teaching Assistant Jan 2024 - Apr 2024
- CS1503 (Mathematical Foundation for Machine Learning)
- CS1671 (Human Language Technologies)

Penn Engineering Online
Course Development Assistant Oct 2022 - May 2023
- CIS5300 (Computational Linguistics)

Halıcıoğlu Data Science Institute, UC San Diego
Student Tutor Mar 2018 - Aug 2019, Mar 2021 - Jun 2021
- CSE151A (Intro to Machine Learning)
- DSC20 (Intro to Data Structure)