

Zhaoyi (Joey) Hou

website: joeyhou.github.io

email: joeyhou@seas.upenn.edu

Research Interest Artificial Intelligence, Machine Learning, Natural Language Processing
Commonsense Reasoning, Knowledge Representation, Intelligent System

Education **University of Pennsylvania** 2021 - Present
Data Science, Master of Science in Engineering

University of California, San Diego 2017 - 2021
Data Science, Bachelor of Science Cum Laude

Publications [4] Xiaochen Kev Gao, **Zhaoyi Hou**, Yifei Ning, Jingbo Shang, Vish Krishnan. *Towards Comprehensive Patent Approval Predictions: Beyond Traditional Document Classification*. [ACL 2022]

[3] 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]

[2] 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]

[1] Emily Nicole Corbett Manoogian, Adena Zadourian, ...(8 more) **Zhaoyi Hou**, Jason G. Fleischer, Shah Golshan, Pam R Taub, Satchidananda Panda. *Feasibility of Time-Restricted Eating and Impacts on Cardiometabolic Health in 24-Hour Shift Workers: The Healthy Heroes Randomized Clinical Trial*. [Cell Metabolism]

Research Experience **Computer and Information Science (CIS) at Penn Engineering** Sep 2022 - Present
NLP Researcher (advised by Prof. Chris Callison-Burch)
- Building a **human-in-the-loop script curation system** to validate and improve event-based scripts generated by pre-trained large language models (e.g. GPT-3).

Computer and Information Science (CIS) at Penn Engineering May 2022 - Present
NLP Researcher (advised by Li Zhang and Prof. Chris Callison-Burch)
- Building a **natural language generation** model to generate branching options in daily scripts;
- Building an **explanation generation** module to generate rationales for each option based on common sense.

Computer and Information Science (CIS) at Penn Engineering Jan 2022 - Aug 2022
NLP & Psychology Researcher (advised by Prof. Lyle Ungar)
- Defined and built a pipeline to extract **semantic mirroring** from mobile text conversations;
- Analyzed the correlation between **psychological conditions** and semantic mirroring in text messages.

Shang Data Lab at UCSD (SDLab) Jun 2020 - Apr 2021
NLP Researcher (advised by Prof. Jingbo Shang)
- Built a **text data ETL and classification** pipeline to handle **600,000** patent documents;
- Implemented a customized **BERT-based** text classification model and improved true negative rate (specificity) from **60% to 86%** (heavily unbalanced data with **84%** positive instance).

Salk Institute for Biological Studies Jul 2019 - Jun 2021
Data Science Researcher (advised by Prof. Satchidananda Panda)
- Built a **data ETL and analysis** pipeline for user behavior analysis (more than **500,000** records).

Projects	Amazon Alexa TaskBot Competition <i>Information Retrieval</i>	Nov 2021 - Apr 2022
	<ul style="list-style-type: none"> - Implemented the document retrieval module for the Alexa TaskBot competition; - Improved the retrieval success rate by 25% and advanced to the final list. 	
	Music Re-listen Prediction <i>Recommendation System</i>	Nov 2021 - Dec 2021
	<ul style="list-style-type: none"> - Built a click-through-rate(CTR) prediction model based on historical user behavior; - Built a feature engineering pipeline and deep factorization machine model to achieve 67% AUC in test data. 	
	Stack Overflow Question Quality Classification <i>Text Mining & Cloud Computing</i>	May 2021 - Jul 2021
Teaching	<ul style="list-style-type: none"> - 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. 	
	Food Text Parser <i>Natural Language Processing</i>	Jul 2020 - Apr 2021
	<ul style="list-style-type: none"> - Built an open-source text parsing pipeline to extract food content from a health monitor app; - Automatically correct typos and extract food-related phrases from the user's input text with 85% parsing accuracy. 	
	Machine Learning for Ophthalmological Diagnosis <i>Computer Vision & Healthcare</i>	Nov 2019 - Dec 2020
	<ul style="list-style-type: none"> - 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. 	
Community Involvement	Penn Engineering Online Computational Linguistics	Course Development Assistant Oct 2022 - Present
	<ul style="list-style-type: none"> - Developing homework and automated testing system for the online version of CIS5300 (Computational Linguistics). 	
	X Academy 2021 Intro to Machine Learning	Head Teaching Assistant Jul 2021 - Aug 2021
	<ul style="list-style-type: none"> - Designed the machine learning capstone project for the teenager-oriented summer camp in China. 	
	Halcolu Data Science Institute, UC San Diego Intro to Machine Learning	Student Tutor Mar 2021 - Jun 2021
Community Involvement	<ul style="list-style-type: none"> - Tutored CSE151A (Intro to Machine Learning) with Prof. Jingbo Shang. 	
	Halcolu Data Science Institute, UC San Diego Intro to Data Structure	Student Tutor Mar 2018 - Aug 2019
	<ul style="list-style-type: none"> - Tutored DSC20 (Intro to Data Structure) with Prof. Marina Langlois. 	
	Chinese Computer Community (Triple C) at UCSD <i>President</i>	May 2020 - Apr 2021
	<ul style="list-style-type: none"> - Project-based student community with more than 100 members and 15 student projects; - Designed the onboarding technical training for data science members. 	
Community Involvement	Big Data Digest - AI Scholar <i>Content Contributor</i>	Jul 2019 - Dec 2020
	<ul style="list-style-type: none"> - Translated the latest AI papers and research news from English to Chinese; - Promoted the understanding of AI technologies on Chinese social media. 	