Hai-Niu (Johnny) Xu

3700 Lancaster Avenue, Unit 614, Philadelphia, PA 19104 | (+1) 9172503774 | seacow@seas.upenn.edu

EDUCATION

King's College London *London, UK*
Ph.D. in Computer Science *Oct. 2023 -*

University of Pennsylvania *Philadelphia, PA*
M.S.E in Data Science *Aug. 2021- May. 2023*
- Course modules: Machine Reasoning; Machine Learning; Deep Learning; Principles of Deep Learning; Computational Linguistics; Machine Perception; Statistics for Data Science; Big Data Analytics
- Current GPA: 4.00 / 4.00

University of California, Davis *Davis, CA*
B.S. in Statistics with Highest Honour (Statistical Data Science Track) *Sept. 2016 - Jun. 2020*
- Thesis Topic: "Application of Curve Registration Methods on Analyzing Wearable Device Data."
- Supervisor: Professor Jane-Ling Wang
- Major GPA: 3.94 / 4.00 (overall GPA: 3.84 / 4.00)

PUBLICATION AND THESIS

1. Li Zhang*, **Hainiu Xu*,** Yue Yang, Shuyan Zhou, Manni Arora, Weiqiu You, Chris Callison-Burch. *Causal Reasoning of Entities and Events in Procedural Texts*, In Findings of EACL 2023. LINK

2. Tianyi Zhang, Issac Tham, Zhaoyi Hou, Jiaxuan Ren, Liyang Zhou, **Hainau Xu**, Harry Zhang, Lara J. Martin, Rotem Dror, Zoey Li, Heng Ji, Martha Palmer, Susan Brown, Reece Suchocki, Chris Callison-Burch. *Human-in-the-loop Schema Induction* (In Submission)

3. **Hainiu Xu,** "Application of Curve Registration Methods on Analyzing Wearable Device Data.", Undergraduate Thesis, University of California-Davis, LINK

RESEARCH EXPERIENCE

Research Assistant: Procedural Reasoning *Philadelphia, PA*
University of Pennsylvania, PennNLP, Supervisor: Prof. Chris Callison-Burch *May. 2022 -*
- Investigated methods for decomposing compound questions for procedural reasoning tasks under a few-shot setting.
- Conducted template-based question decomposition using GPT3 and T5.
- Created a metric for evaluating decomposition results for procedural reasoning tasks.
- Conducted image synthesis using procedural instructions with DALLE-mini and CLIP.
- Investigated methods for composing logical relationship between multi-hop questions and corresponding zero-hop premises.
- Conducted error analysis on SOTA models in entity state tracking.
- Wrote conference paper as co-first author.

Independent Research: Causal Inference on Entity States *Philadelphia, PA*
University of Pennsylvania, PennNLP, Supervisor: Prof. Chris Callison-Burch *May. 2022 -*
- Finetune LLMs as neural knowledge bank to query entity state information.
- Leverage Codex to generate interventional entity state information.
- Detect key frames that corresponds to entity state change.
- Leverage video key frames to ground LLM generated entity states.
- Construct causal graph using propensity score computed from machine-generated intervention.

Final Project: Autonomous Driving with Reinforcement Learning *Philadelphia, PA*
University of Pennsylvania, CIS522 Deep Learning *March. 2022 - May. 2022*
- Constructed end-to-end learning models using MIT's VISTA environment.
- Trained an autonomous driving agent using CNN and A2C successfully.
- Discovered that the input image's statistical properties greatly influence the training outcome.
- Designed experiments to compare the performance of the policy gradient method and A2C.

Honours Thesis: Curve Registration on Wearable Device Data *Davis, CA*
UC Davis, Department of Statistics, Supervisor: Professor Jane-Ling Wang *Jun. 2019 - Jun. 2020*
• Applied various time warping algorithms to functional data collected from wearable devices.
• Wrote a comprehensive literature review for the major warping methods.
• Compared time warping methods for wearable device data.
• Measured the effectiveness of warping methods with functional principal component analysis results on the

Research Training Group (RTG) Project: Analysis of Wearable Device Data *Davis, CA*
UC Davis Department of Statistics, National Science Foundation (NSF). *Dec. 2018-Jun. 2019*
Supervisor: Professor Jane-Ling Wang
• Studied theories of functional data analysis including smoothing, clustering, and functional principal componentanalysis.
• Conducted exploratory functional data analysis with wearable device data.
• Extracted user activity patterns using functional principal component analysis.

WORK EXPERIENCE

AI Engineer Intern *Beijing. China*
Schlumberger BGC, Supervisors: Dr. Ping Zhang, Dr. Qing Liu, Dr. Peng Jin *Sept. 2020 - May. 2021*
• Conducted text mining on drilling reports written in Chinese.
• Conducted fine-tuning on Chinese word vectors for the oil and gas industry.
• Studied the behaviour of fine-tuning algorithms under a small-sample fine-tuning setting.
• Conducted a comprehensive evaluation of the performance of state-of-the-art language models (BERT, ALBERT, ELECTRA, etc.) on representing words from drilling reports.
• Built mono-lingual and multi-lingual classification models for classifying Chinese drilling reports.
• Deployed the classification model as a web application using Flask.

TEACHING

Teaching Assistant *Philadelphia, PA*
University of Pennsylvania, CIS530 Computational Linguistics, Instructor: Prof. Mark Yatskar *Aug. 2022 -*
• Manage Piazza (course online QA platform)
• Make weekly quizzes
• Hold office hours
• Grade homework
• Mentor course projects

AATC Peer Tutor *Davis, CA*
UC Davis: Academic Assistant and Tutoring Centre *Sep. 2018 - Dec. 2018*
• Prepared for tutoring sessions by making interactive questions regarding the material and making up example questions.
• Provided individual tutoring for peers from STA 131A: Introduction to Probability Theory.

EXTRACURRICULAR ACTIVITIES

UC Davis Symphony Orchestra *Davis, CA*
Clarinet *Jan.2018 - Jun. 2020*
• Rehearsed and performed in numerous concerts at the Mondavi Centre.
• Performed as the principal clarinet in various concerts.

UC Davis University Concert Band *Davis, CA*
Principal Clarinet *Jun. 2016 - Dec. 2018*
• Rehearsed and performed in numerous concerts at the Mondavi Centre.
• Led the clarinet section rehearsal.

UC Davis Chamber Music– Clarinet Trio *Davis, CA*
Principal Clarinet *Jan. 2019 - Mar. 2019*
• Rehearsed and performed in numerous concerts at the Pizter Centre.
• Conducted and directed the rehearsal of the chamber music group.

SKILLS & INTERESTS

Language: Fluent in Mandarin and English (GRE: 329 (Verbal Reasoning: 161, Quantitative Reasoning: 168))
Programming Languages: Python, R, C, MATLAB, Bash
Machine Learning Libraries: PyTorch, Tensorflow, Huggingface Transformers, Numpy, Jax, Scikit-Learn, HyperOpt
Interests: Clarinet, Coding, Reading, Basketball, Table Tennis, Swimming