# XINRAN ZHAO

Email: xzhaoar@connect.ust.hk │ GitHub: colinzhaoust │ Website: colinzhaoust.github.io

## EDUCATION

**The Hong Kong University of Science and Technology**　　　　　　　　　　　　　　　*2016 - 2020, Hong Kong*

- GPA = 3.958 (Top 1%, Academic Achievement Medal).
- Bachelor of Engineering, majoring in Computer Science.
- Academic Achievement Medal, First Class Honors, Continuous Undergraduate Scholarship, Dean's list for active semesters.
- Relevant Coursework:
  Explore & Visualize Data; OOP & Data Structures (Honor); Discrete Math Tools; Design & Analy. of Algori; Statistical Learning Models for Text and Graph Data (Graduate Level); Big Data Mining & Management; Deep Learning for Computer Vision; Matrix Algebra

**Cornell University**　　　　　　　　　　　　　　　*2018, NY, US*

- International Exchange Program in Computer Science.
- Relevant Coursework: Machine Learning; Natural Language Processing; Operating Systems

## RESEARCH EXPERIENCE

**WinoWhy: A Deep Diagnose of Essential Commonsense Knowledge for**
**Answering Winograd Schema Challenge (Link)**
In Proceedings of **ACL 2020**, Xinran Zhao*, Hongming Zhang*, and Prof. Yangqiu Song　　　　　　　*HKUST, HK*

- Proposes a new task, WinoWhy, which requires models to determine the plausibility of collected human reasoning from similar adversarial examples from humans and generative models. The task based on commonsense pronoun coreference choice questions further examines the model performance of understanding the commonsense reasoning instead of solely making correct prediction. Experiments on categorized human reasoning point out the limitation and possible future directions of current models.
- Personal Contribution: Proposed the idea in discussion; collected and rectified the dataset; implemented the models and baselines; wrote and reviewed the draft with collaborators.

**Leveraging Topic Relatedness for Argument Persuasion**
In Findings of **ACL 2021**., Xinran Zhao, Esin Durmus, Hongming Zhang, and Prof. Claire Cardie　　　　　*Cornell, NY*

- Proposes a bi-encoder model consisting of separate encoders for (1) the background and linguistic features of debaters ; (2) the language and the relations of the topics, to involve the information of debate topics to study more contextualized argument persuasiveness than the previous works that commonly focus on the debaters' characters and arguments alone.
- Personal Contribution: Proposed the idea in discussion; implemented the model and the baselines; performed qualitative and quantitative analysis; wrote and reviewed the draft with collaborators.

**Probing Toxic Content in Large Pre-Trained Language Models**
In Proceedings of **ACL 2021**, Nedjma Ousidhoum, Xinran Zhao, Tianqing Fang, Yangqiu Song, and Dit-Yan Yeung　　*HKUST, HK*

- Proposes a method to probe multilingual pretrained language models ( PTLMs) to quantify the potentially harmful content they convey with respect to a large-scale set of commonsense knowledge templates and social group names. We shed the light on how such negative content can be triggered within unrelated and benign contexts and explain how to take advantage of our methodology to assess and mitigate the toxicity transmitted by PTLMs.
- Personal Contribution: Proposed the idea in discussion; implemented the multilingual models and the baselines; wrote and reviewed the draft with collaborators.

**Learning Contextual Causality between Daily Events from Time-consecutive Images (Link)**
In **CVPR 2021**: Causality in Vision Workshop, Hongming Zhang, Yintong Huo, Xinran Zhao, Yangqiu Song, and Dan Roth　　*HKUST, HK*

- Designs a novel task studying contextual causal knowledge from videos and proposed a corresponding dataset containing human annotated events and causality inferred from video frame sequences. Then we propose a Vision-Contextual Causal (VCC) model that can utilize the images as context to better acquire causal knowledge from temporal consecutive videos.

**A Brief Survey and Comparative Study of Recent Development of**
**Pronoun Coreference Resolution (Link|Code)**
In **ArXiv**, Xinran Zhao, Hongming Zhang and Prof. Yangqiu Song　　　　　　　　　　　　　*HKUST, HK*

- Surveys in detail on the Pronoun Coreference Resolution (PCR) datasets (ordinary ones, hard ones requiring external knowledge, and others for special purposes) and models (from traditional ones to neural-based end-to-end models) commonly used for the task. Exhaustive experiments and comparisons show that some performance gain relies on the dataset similarity and knowledge graph based improved models relieve the problem of predicting infrequent cases.
- Personal Contribution: Implemented or replicated the models and the baselines; performed qualitative and quantitative analysis; wrote and reviewed the draft with collaborators.

**Seek to Embed ASER: A Large-scale Eventuality Knowledge Graph ([Link](#))**
In **WWW 2020**, An acknowledged contributor advised by Xin Liu and Prof. Yangqiu Song                    *HKUST, HK*

- Designs a novel embedding model combining text embedding and graph embedding algorithms to learn the node representation for ASER, which has text as nodes and eventuality relations as edges. Provides useful signals for link prediction, unknown event resolution, and representation captioning on the graph, with a model utilizing BERT embeddings and LSTM.
- Applied ASER to solve the commonsense reasoning problems (Section 6.2) and beat the state-of-the-art at that time.

**PCR4ALL: A Comprehensive Evaluation Benchmark for**
**Pronoun Coreference Resolution**
Submitted to **EMNLP 2021**, Xinran Zhao, Hongming Zhang and Prof. Yangqiu Song                    *HKUST, HK*

- Proposes a novel benchmark, PCR4ALL, which bridges the gap between document-level (e.g., CoNLL-2012) and sentence-level (e.g., Winograd Schema Challenge) pronoun coreference questions to evaluate the systems from different angles (i.e., knowledge sources, domains, frequency, bias) to check their reliability in real applications. Experiments demonstrate a notable gap between existing PCR models and a reliable PCR system since they are often optimized on a single target type and lack generalizablity.
- Personal Contribution: Proposed the idea in discussion; implemented the code for dataset alignment and data labeling; implemented the models and baselines; wrote and reviewed the draft with collaborators.

**Towards Reference-free Text Simplification Evaluation**
To be Submitted to **AAAI 2022**, Xinran Zhao and Prof. Dit-Yan Yeung                    *HKUST, HK*

- Previous text simplification (TS) metrics rely heavily on the exact token match and human-annotated simplified sentences. We present BETS, the first neural network-based reference-free TS metric that leverages BERT and large-scale paraphrasing datasets to evaluate paired original and simplified sentences directly. Comparative study shows that BETS correlates better than existing metrics with human judgments. The relief of reference and controllable coefficients further improve the applicability and transferability of TS models.
- Personal Contribution: Proposed the idea; implemented the models and baselines; wrote and reviewed the draft with collaborators.

**Mining Course Structure for Course Recommendation**
Undergraduate Research Opportunity, advised by Prof. Raymond Wong                    *HKUST, HK*

- Builds a course recommendation and academic planner model based on course dependency graph (generated from school major requirements) and past students' performance records. Serves as part of the unreleased official HKUST undergraduate course planner, which provides requirement fulfillment checker and study path recommendation service for university students.

**A Comparative Study on the Sentimental Characteristics of Chinese and Western Tourists (in Chinese)**
Under review, advised by Yue Xi and Prof. Yushi Jiang                    *Chengdu, China*

- Builds an NLP supported model to detect the comparative behavior differences on sensitiveness and attentiveness of Airbnb tourists from various cultures. Paper was presented in the Annual Conference of JMS China Marketing Science 2019.

## WORK EXPERIENCE

**An Online Learning Platform with NLP Supported Teaching Assistant**
Research Assistant, supervised by Prof. Dit-Yan Yeung                    *Sep 2020 - Feb 2021, HK*
- Seeks to build an online learning platform with NLP supported widgets, including a question answering module that retrieves video chunks from lecture videos to answer students' queries and a quiz generator that automatically generates basic multiple choice quizzes from auto-rectified and masked lecture subtitles. This work is as a follow-up for my undergraduate Final Year Project.

**Narrow Band Communication**
Engineer Intern, Sichuan Jinhutong Tech. Stock Company                    *June 2017 - Sep 2017, Chengdu*
- Assisted in the development of a smart office system for local China Mobile Headquarters based on Narrow Band Communication, writing C++ for Integrated Circuit components.

## SKILLS

**Programming:** Python, C++, Java, JavaScript, Matlab, Pytorch, Keras/Tensorflow, and etc.

**Languages:** English (Proficient, with TOEFL=112), Mandarin (Native).

**Debate:** HKUST Madarin Debate Team: Championship for Bayarea Debate Invitational; Top 8 for the Sixth and Seventh International Mandarin Debate Invitational in Singapore.

**Interests:** Archery, Gym, Engraving, Basketball, Traditional Poem Writing