

Haotian Chen

Phone Number: +86 18019207701, Email: htchen18@fudan.edu.cn

Education

Fudan University (FDU), School of Computer Science

Shanghai, China

A Ph.D. candidate in Data Science (Information Extraction).

2018.9 – 2024.6 (Expected Time)

- Honors: **Outstanding Student of FDU, Scholarship of Academic Excellence (2018-2019, 2019-2020, 2020-2021).**
- Reviewers: TKDE; EMNLP 2023; MM2023.

Dalian University of Technology (DLUT), School of Ocean Science and Technology

Dalian, Liaoning, China

Bachelor in Ocean EE

2014.9 - 2018.7

- **Rank: 1/74 (1%)**
- Honors: **Outstanding Graduates of Liaoning Province (top 1% of Liaoning Province, top 3% of China), Scholarship of Academic Excellence (top 5% of DLUT; 2015, 2016, 2017).**

Research Interests

My research goal is to make AI systems to be trustworthy, robust, align with human values, and perform well in real-world applications (e.g., social goods, self-driving, healthcare, human-AI interaction, etc.). My research experience focuses on analyzing the decision rules (understanding ability) of models in legal judgment prediction (LJP) and relation extraction (RE) and improving their performance, robustness, and generalization ability in text generation, time series, LJP, and RE.

Selected Papers

- [1] **Haotian Chen**, Bingsheng Chen, Xiangdong Zhou. Did the Models Understand Documents? Benchmarking Models for Language Understanding in Document-Level Relation Extraction. **ACL 2023**. [\[paper\]](#)[\[code\]](#)
- [2] **Haotian Chen**, Lingwei Zhang, Yiran Liu, Yang Yu. Rethinking the Development of Large Language Models from the Causal Perspective: A Legal Text Prediction Case Study. **AAAI 2024**. [\[arxiv\]](#)[\[code\]](#)
- [3] **Haotian Chen**, Houjing Guo, Bingsheng Chen, Xiangdong Zhou. OODREB: Benchmarking the State-of-the-Art methods for Out-Of-Distribution Generalization on Relation Extraction. **WWW 2024**.
- [4] **Haotian Chen**, Han Zhang, Houjing Guo, Shuchang Yi, Bingsheng Chen, Xiangdong Zhou. SALAS: Supervised Aspect Learning Improves Abstractive Multi-Document Summarization through Aspect Information Loss. **ECML-PKDD 2023**. [\[paper\]](#)[\[code\]](#)
- [4] **Haotian Chen**, Han Zhang, Houjing Guo, Shuchang Yi, Bingsheng Chen, Xiangdong Zhou. Recovering Missing Key Information: An Aspect-guided Generator for Abstractive Multi-document Summarization. **DASFAA 2023**.
- [5] Yunzhong Lou, Xueyang Li, **Haotian Chen**, Xiangdong Zhou. BRep-BERT: Pre-training Boundary Representation BERT with Sub-graph Node Contrastive Learning. **CIKM 2023**.
- [6] **Haotian Chen**, Bingsheng Chen, Xiangdong Zhou. A Novel Entity-Aware Generate-then-Extract Paradigm for Cross-Document Relation Extraction. **Submitted to KDD2024**.
- [8] **Haotian Chen**, Shifei Xiao, Xueyang Li, Yunzhong Lou, Xiangdong Zhou. Data Know the Answers: Adaptively Selecting Contrastive Learning Strategies for Time Series in a Multi-Modal Way. **Submitted to IJCAI2024**.
- [9] **Haotian Chen**, Houjing Guo, Xiangdong Zhou. Taking the Essence and Discarding the Dross: Approaching Evidence Information by Identifying Semantic Magnitude for Out-of-Distribution Relation Extraction. **Submitted to KDD2024**.

Selected Research Experience

Microsoft Research Asia, Machine Learning Group

Beijing, China

Advised by Xiao Yang [[GitHub](#)]

2023.11 - Present

1) Contribute to developing an open-source code framework aiming to apply LLM (e.g., GPT-4) agent to perform automatic research and development (R&D) in real-world scenarios. 2) Identify research questions through the implementation and then formalize questions and expose the bottleneck of current LLMs. 4) Demonstrate the promising future of LLM-driven automatic R&D and propose the effective technical methods (knowledge-augmented evolving strategies).

Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University

Beijing, China

Advised by Assistant Professor [Yang Yu](#) [[Google Scholar](#)]

2021.06 - 2023.04

1) Investigate the decision rules of transformer-based AI judges by feature attribution and find that they make legal judgment predictions according to irrelevant information in the given case descriptions. 2) Causally analyze the AI judges' nature and argue that the lack of human knowledge, the imbalance of training data, and the incomprehensiveness of testing methods impede them from learning causal relationships. 3) Propose two methods to infuse knowledge into data and model's architecture, respectively. 4) Further propose 8 kinds of legal-specific attacks to complete the testing methods. 5) Experimentally demonstrate that knowledge improves models' robustness and performance. (Accepted by **AAAI 2024**)

School of Computer Science, Fudan University

Shanghai, China

Independent Research

2020.09 - Present

- **Benchmarking Models for Language Understanding in Document-level RE (ACL 2023)**

1) Investigate the decision rules of SOTA document-level RE (DocRE) methods by integrated gradient and find that they extract relations based on non-causal information (e.g., entity names) instead of evidence information. 2) Argue by SCM that the selection bias causes the problem during the pre-training process of language models. 3) Propose 6 kinds of RE-specific attacks to expose the weak robustness and understanding ability of SOTA DocRE methods. 4) Further propose a dataset with human-annotated word-level evidence and introduce mean average precision to evaluate models' understanding ability and select robust, trustworthy, and well-performed models.

- **Benchmarking the SOTA Methods in RE for Out-of-Distribution Generalization**

1) Demonstrate that SOTA methods in RE perform poorly under our proposed distributional shifts. 2) Propose a benchmark named (OODREB) for selecting the model that can be deployed in real-world applications.

Awards

- Merit Student of DLUT (**top 5% of DLUT**; 2014-2015, 2015-2016, 2016-2017);
- Dalian Advanced Mathematics Competition Second Prize (**top 3% of Dalian**);
- DLUT Electronic Design Contest Second Prize (top 20% of DLUT);
- DLUT Academic Science and Technology Works Competition Second Prize (**top 10% of DLUT**; 2016, 2017);
- Excellent Volunteer of DLUT (**top 5% of DLUT**).

Other Information

- Passed **Level I of the Chartered Financial Analyst (CFA) program** (pass rate is 27%, verified by CFA Institute).
- Proficient in python, pytorch, latex, markdown, and C++.
- Working experience (**6 years**) in **Public Relations Department of Student Union** (3 years in DLUT and 3 years in FDU). Each period of 3 years comprises 1 year for member, 1 year for vice-president, and 1 year for president.
- I like running, which leads to a more positive state of my mind and makes each day count. I'm also good at Taekwondo.