

# Wenhao Yu

CONTACTS	E-mail: wenhaoyu97@gmail.com	Websites: (1) <a href="#">Homepage</a> (2) <a href="#">Google Scholar</a>
WORKING EXPERIENCE	<b>Tencent AI Lab, Seattle, Washington, US</b> Senior Research Scientist in NLP	Aug. 2023 – Now
EDUCATION EXPERIENCE	<b>University of Notre Dame, IN, US</b> Ph.D. in Computer Science and Engineering <ul style="list-style-type: none"><li>• Advisor: Dr. Meng Jiang <a href="http://www.meng-jiang.com/">http://www.meng-jiang.com/</a></li><li>• Committee members: Dr. Heng Ji (UIUC), Dr. Scott Wen-tau Yih (FAIR),</li></ul> <b>Sichuan University, Chengdu, China</b> B.S. in Computer Science and Technology <ul style="list-style-type: none"><li>• Graduated with honors</li></ul>	Aug. 2019 – Jul. 2023 GPA: 3.95/4.0 Aug. 2015 – Jul. 2019 GPA: 3.84/4.0
RESEARCH INTERESTS	I am interested in <b>Language Model + Knowledge</b> (i.e., knowledge-enhanced methods) for solving various real-world applications, such as open-domain question answering, commonsense reasoning and natural language generation. Besides, I am also actively contributing to open-community project, such as StarCoder.	
INDUSTRY EXPERIENCE	<b>Research Intern at Bloomberg CTO office, New York City, New York, US</b> <i>Worked with Drs. Mark Dredze, David Rosenberg, Gideon Mann</i> <b>Research Intern at Allen Institute of AI (AI2), Seattle, Washington, US</b> <i>Worked with Drs. Ashish Sabharwal and Peter Clark</i> <b>Visiting Student Researcher at Microsoft Research, Redmond, Washington, US</b> <i>Worked with Drs. Chenguang Zhu and Michael Zeng</i> <b>Research Intern at Microsoft Research, Redmond, Washington, US</b> <i>Worked with Drs. Chenguang Zhu and Michael Zeng</i>	May 2023 - Aug. 2023 Sept. 2022 - May 2023 Sept. 2021 - Sept. 2022 June 2021 - Sept. 2021
HONORS AND AWARDS	<b>Bloomberg Ph.D. Fellowship</b> <b>Best Paper Award at SoCalNLP</b> NAACL 2022 Travel Award Notre Dame Outstanding Student Fellowship National Scholarship of China	2022 2022 2022 2019 2018
WORKSHOP ORGANIZATION	[W1] Chenguang Zhu, Shuohang Wang, Meng Jiang <b>Wenhao Yu</b> , Lu Wang, Huan Sun. “The First Workshop on Knowledge Augmented Methods for Natural Language Processing”, in AACL 2023. <a href="#">[Website]</a> <a href="#">[Blog]</a> <a href="#">[Twitter]</a> <ul style="list-style-type: none"><li>• With more than 70 people (virtual &amp; in-person) attending the workshop</li></ul>	
CONFERENCE TUTORIAL	[T2] Chenguang Zhu, Yichong Xu, Xiang Ren, Yuchen Lin, Meng Jiang, <b>Wenhao Yu</b> . “Knowledge-Augmented Methods for Natural Language Processing”, in ACL 2022. <ul style="list-style-type: none"><li>• With more than 200 people (virtual &amp; in-person) attending the tutorial</li><li>• With more than 200 people following our Github repository <a href="#">[Link]</a></li></ul> [T1] <b>Wenhao Yu</b> , Meng Jiang, Zhiting Hu, Qingyun Wang, Heng Ji, Nazneen Fatema Rajani. “Knowledge-Enriched Natural Language Generation”, in EMNLP 2021. <ul style="list-style-type: none"><li>• With more than 120 people (virtual &amp; in-person) attending the tutorial</li><li>• With more than 500 people following our Github repository <a href="#">[Link]</a></li></ul>	
PEER-REVIEWED PUBLICATIONS	[P14] <a href="#">[Language Model + Open-domain QA]</a> Zhihan Zhang, <b>Wenhao Yu</b> , Zheng Ning, Mingxuan Ju, Meng Jiang. “Exploring Contrast Consistency of Open-Domain Question Answering Systems on Minimally Edited Questions”, in TACL 2023.  [P13] <a href="#">[Large Language Model + Open-domain QA]</a> Noah Ziemis, <b>Wenhao Yu</b> , Zhihan Zhang, Meng Jiang. “Large Language Models are Built-in Autoregressive Search Engines”, in ACL 2023.  [P12] <a href="#">[Large Language Model + Open-domain QA]</a> <b>Wenhao Yu</b> , Dan Iter, Shuohang	

Wang, Yichong Xu, Mingxuan Ju, Soumya Sanyal, Chenguang Zhu, Michael Zeng, Meng Jiang. "Generate rather than Retrieve: Large Language Models are Strong Context Generators", in ICLR 2023.

[P11] [Document Retrieval, Commonsense QA] **Wenhao Yu**, Chenguang Zhu, Zhihan Zhang, Shuohang Wang, Zhuosheng Zhang, Meng Jiang. "Retrieval Augmentation for Commonsense Reasoning: A Unified Approach", in EMNLP 2022.

[P10] [Document Retrieval, Knowledge Graph] **Wenhao Yu**, Mingxuan Ju (equal contribution), Tong Zhao, Chuxu Zhang, Yanfang Ye. "Knowledge Graph Enhanced Passage Reader for Open-domain Question Answering", in EMNLP 2022.

[P9] [Entity Memory, Open-domain QA] Zhihan Zhang, **Wenhao Yu**, Chenguang Zhu, Meng Jiang. "A Unified Encoder-Decoder Framework with Entity Memory", in EMNLP 2022. (selected as oral presentation, 166 out of 1392 accepted papers)

[P8] [Document Retrieval + Knowledge Graph] Ziniu Hu, Yichong Xu, **Wenhao Yu**, Shuohang Wang, Ziyi Yang, Chenguang Zhu, Kai-Wei Chiang, Yizhou Sun. "Empowering Language Models with Knowledge Graph Reasoning for Open-Domain Question Answering", in EMNLP 2022.

[P7] [Multi-task Learning] Zhuosheng Zhang, Shuohang Wang, Yichong Xu, Yuwei Fang, **Wenhao Yu**, Yang Liu, Hai Zhao, Chenguang Zhu, Michael Zeng. "Task Compass: Scaling Multi-task Pre-training with Task Prefix", in EMNLP 2022.

[P6] [Language Model Pre-training + English Dictionary] **Wenhao Yu**, Chenguang Zhu, Yuwei Fang, Donghan Yu, Shuohang Wang, Yichong Xu, Meng Jiang. "Dict-BERT: Enhancing Language Model Pre-training with Dictionary", in ACL 2022.

[P5] [Document Retrieval + Knowledge graph] Donghan Yu, Chenguang Zhu, Yuwei Fang, **Wenhao Yu**, Shuohang Wang, Yichong Xu, Xiang Ren, Yiming Yang, Michael Zeng. "KG-FiD: Infusing Knowledge Graph in Fusion-in-Decoder for Open-Domain Question Answering", in ACL 2022.

[P4] [Survey, Knowledge Graph, Text Retrieval] **Wenhao Yu**, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, Meng Jiang. "A Survey of Knowledge-Enhanced Text Generation", in ACM Computing Survey 2022 (CSUR, [IF:10.28](#)).

[P3] [Language Generation, Entity Knowledge] Xiangyu Dong, **Wenhao Yu**, Chenguang Zhu, Meng Jiang. "Injecting Entity Types into Entity-Guided Text Generation", in EMNLP 2021 (selected as oral presentation).

[P2] [Document Retrieval, Transfer Learning] **Wenhao Yu**, Lingfei Wu, Yu Deng, Qingkai Zeng, Ruchi Mahindru, Sinem Guven, Meng Jiang. "Technical Question Answering across Tasks and Domains", in NAACL 2021.

[P1] [Joint Retrieval and Generation] **Wenhao Yu**, Lingfei Wu, Qingkai Zeng, Shu Tao, Yu Deng, Meng Jiang. "Crossing Variational Autoencoders for Answer Retrieval", in ACL 2020 (selected as oral presentation).

**PH.D THESIS** **Wenhao Yu**. "Retrieval-augmented Generation across Heterogeneous Knowledge", accepted to NAACL Ph.D thesis proposal track 2022.

**MENTORED STUDENTS** **Tong Chen**, Tecent AI intern, Ph.D student at the University of Washington  
**Zhihan Zhang**, on [P9], now Ph.D student at the University of Notre Dame  
**Zaitang Li**, on [P4], now Ph.D student at the Chinese University of Hong Kong  
**Xiangyu Dong**, on [P3], now software engineer at the Bloomberg LP

**PROFESSIONAL SERVICES** Conference Reviewer and Journal Invited Reviewer:

- ICLR                      • ICML                      • Neurips                      • AACL                      • IJCAI
- ACL                      • EMNLP                      • NAACL                      • CSUR                      • TACL