

Wenhao Yu

CONTACT INFORMATION	355 Fitzpatrick Hall Department of Computer Science and Engineering University of Notre Dame, IN 46556, US	Phone: (+1) (574) 292-6516 Web: wyu97.github.io E-mail: wyu1@nd.edu
RESEARCH INTERESTS	<ul style="list-style-type: none">• Applied Machine Learning• Language Model Pre-training• Knowledge-Enhanced Methods• Natural Language Generation	
EDUCATION EXPERIENCE	University of Notre Dame, IN, US Ph.D. in Computer Science and Engineering <ul style="list-style-type: none">• Advisor: Professor Meng Jiang Sichuan University, Chengdu, China B.S. in Computer Science and Technology <ul style="list-style-type: none">• Graduated with honors	Aug. 2019 – <i>present</i> GPA: 3.95/4.0 Aug. 2015 – Jul. 2019 GPA: 3.84/4.0
INDUSTRY EXPERIENCE	Microsoft Research, Redmond, Washington, US <i>Research Intern; Working with Dr. Chenguang Zhu</i> <ul style="list-style-type: none">• Worked on enhancing language model pre-training with dictionary. IBM Greater China Group Lab, Beijing, China <i>Applied Research Intern; Worked with Dr. Minquan Gao</i> <ul style="list-style-type: none">• Worked on tabular data extraction.	June 2021 - Oct. 2021 Jan. 2019 - Mar. 2019
HONORS AND AWARDS	First Place at the 14th Notre Dame CSE Annual Poster Contest Notre Dame CSE Select Fellowship Tang Lixin Excellent Student Scholarship National Scholarship of China	2020 2019 2018 2017
CONFERENCE TUTORIAL	[T1] Wenhao Yu , 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">• With more than 120 people (virtual & in-person) attending the tutorial• With more than 300 people following our Github repository [Link]	
PEER-REVIEWED FIRST-AUTHOR PUBLICATIONS	[C5] [Language Generation] Wenhao Yu , Chenguang Zhu, Tong Zhao, Zhichun Guo, Meng Jiang. “Sentence-permuted Paragraph Generation”, in EMNLP 2021. [C4] [Language Generation + Entity Knowledge] Wenhao Yu , Xiangyu Dong, Chenguang Zhu, Meng Jiang. “Injecting Entity Types into Entity-Guided Text Generation”, in EMNLP 2021 (selected as oral presentation). [C3] [Language Model + Domain Knowledge] Wenhao Yu , Lingfei Wu, Yu Deng, Qingkai Zeng, Ruchi Mahindru, Sinem Guven, Meng Jiang. “Technical Question Answering across Tasks and Domains”, in NAACL 2021. [C2] [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. [C1] [Language Model + Graph Knowledge] Wenhao Yu , Mengxia Yu, Tong Zhao, Meng Jiang. “Identifying Referential Intention with Heterogeneous Contexts”, in TheWebConf 2020 (formerly known as WWW).	
PUBLICATIONS UNDER REVIEW	[P5] [Language Generation + Knowledge Graph] Wenhao Yu , Chenguang Zhu, Lianhui Qin, Zhihan Zhang, Tong Zhao, Meng Jiang “Diverse Generative Reasoning on Commonsense Knowledge Graphs”, under review at ACL 2022. [P4] [Language Model + 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”, under review at ACL 2022. [P3] [Language Generation + Knowledge] Wenhao Yu , Chenguang Zhu, Zaitang, Li, Zhiting Hu, Qingyun Wang, Heng Ji, Meng Jiang. “A Survey of Knowledge-	

	Enhanced Text Generation”, under review at ACM Computing Surveys (ACM CSUR).
	[P2] [Language Model + 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”, under review at ACL 2022.
	[P1] Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, Meng Jiang. “Counterfactual Graph Learning for Link Prediction”, under review at ICLR 2022.
PEER-REVIEWED CO-AUTHOR PUBLICATIONS	[C3] [Language Model + Taxonomy Graph] Qingkai Zeng, Jinfeng Lin, Wenhao Yu , Jane Cleland-Huang, Meng Jiang. “Enhancing Taxonomy Completion with Concept Generation via Fusing Relational Representations”, in KDD 2021.
	[C2] Zhichun Guo, Chuxu Zhang, Wenhao Yu , John Herr, Olaf Wiest, Meng Jiang, Nitesh V. Chawla. “Few-Shot Graph Learning for Molecular Property Prediction”, in TheWebConf 2021 (formerly known as WWW).
	[C1] [Language Model + Entity] Qingkai Zeng, Wenhao Yu , Mengxia Yu, Tianwen Jiang, Tim Weninger, Meng Jiang “Tri-Train: Automatic Pre-Fine Tuning between Pre-Training and Fine-Tuning for SciNER”, in EMNLP Findings 2020.
INVITED TALK	HIT, Improving language diversity in paragraph generation, Nov. 2021 Google Research, Knowledge-enriched natural language generation, Nov. 2021 EMNLP Tutorial, Knowledge-enriched natural language generation, Nov. 2021 Microsoft Research, Improving language model with dictionary, Sept. 2021 Notre Dame, Improving language model pre-training with dictionary, Sept. 2021
SELECTED PROJECT EXPERIENCE	Improving Language Model Pre-training with Dictionary <ul style="list-style-type: none"> Proposed to enhance language model pre-training by leveraging definitions of the rare words in dictionaries (e.g., Wiktionary). We designed two novel self-supervised pre-training tasks on word and sentence-level alignment between input text sequence and rare word definitions. It can significantly improve the understanding of rare words and boost model performance on various NLP downstream tasks (Submitted to ACL 2022). Sentence-permuted Paragraph Generation <ul style="list-style-type: none"> Proposed to permute the sentence orders to improve the content diversity of multi-sentence paragraph. It maximizes the expected log-likelihood of output paragraph with respect to all possible sentence orders. It can generate more diverse outputs with a higher quality than various SoTA (EMNLP 2021 main). Knowledge-enhanced Text Generation <ul style="list-style-type: none"> Proposed a multi-step decoder that injects the entity types into the process of entity mention generation. It effectively embeds the entity’s meaning into hidden states, making the generated words precise. (EMNLP 2021 oral.) Wrote a survey paper on knowledge-enhanced text generation. The main content includes two parts: (i) general methods and architectures for integrating knowledge into text generation; (ii) specific techniques and applications according to different forms of knowledge data. (Submitted to ACM CSUR.)
PROFESSIONAL SERVICES	Conference PC/Reviewer <ul style="list-style-type: none"> EMNLP 2020 WSDM 2021 AAAI 2021 IJCAI 2021 ACL 2021 EMNLP 2021 AAAI 2022 IJCAI 2022 WSDM 2022 KDD 2022 ARR (ACL) Journal Invited Reviewer: <ul style="list-style-type: none"> TKDD IJIS BigData TNNLS WWWJ KBS