He Bai (Richard)

200 Ring Rd, Waterloo
ON, Canada N2L 3G1
✆ *(+1) 548 333 6369*
✉ *he.bai@uwaterloo.ca*
🖝 *https://uwaterloo.ca/scholar/h32bai*

Education

Start from 09/2019	**University of Waterloo**, *Computer Science*, PhD, Supervisor: Ming Li.
09/2017– 07/2019	**Institute of Automation, Chinese Academy of Sciences**, *Pattern Recognition and Intelligent System*, Master of Science, Supervisor: Yu Zhou, Chengqing Zong.

Research Experience

May 2022– Aug. 2022	**ML Research Intern**, APPLE, CUPERTINO. ○ My research project would be the large language model pre-training for text generation. We will answer this research question: can multiple experts outperform a super expert in the context of GPT3 pre-training?
Sep. 2021– Mar. 2022	**NLP Research Intern**, BAIDU USA, SUNNYVALE. ○ Speech-text pre-training for speech editing and prompt-based one-shot speech synthesis.
May 2021– Jul. 2021	**Research Intern**, MICROSOFT RESEARCH, MONTREAL. ○ Language model trained with curriculum learning and hypernym prediction task.
Jul. 2018– Oct. 2018	**NLP Research Intern**, IBM RESEARCH, BEIJING. ○ Chinese Openai-GPT model pre-training.

Publications

- **He Bai**, Renjie Zheng, Xintong Li, Junkun Chen, Liang Huang. A^3T**: Alignment-Aware Acoustic and Text Pretraining for Speech Synthesis and Editing. *ICML* 2022 (full paper).**

- **He Bai**, Tong Wang, Alessandro Sordoni, Peng Shi. **Better Language Model with Hypernym Class Prediction. *ACL* 2022 (full paper).**

- Peng Shi, Rui Zhang, **He Bai**, Jimmy Lin. **Cross-Lingual Training with Dense Retrieval for Document Retrieval. *EMNLP-MSR* 2021 (workshop paper).**

- **He Bai**, Peng Shi, Jimmy Lin, Luchen Tan, Kun Xiong, Wen Gao, Jie Liu, Ming Li. **Semantics of the Unwritten: The Effect of End of Paragraph and Sequence Tokens on Text Generation. *ACL-SRW* 2021 (workshop paper).**

- **He Bai**, Peng Shi, Jimmy Lin, Yuqing Xie, Luchen Tan, Kun Xiong, Wen Gao, Ming Li. **Segatron: Segment-aware Transformer for Language Modeling and Understanding. *AAAI* 2021 (full paper).**

- Peng Shi, **He Bai**, Jimmy Lin. **Cross-Lingual Training of Neural Models for Document Ranking.** *Findings of **EMNLP** 2020 (short paper).*

- **He Bai**, Yu Zhou, Jiajun Zhang and Chengqing Zong. **Memory Consolidation for Contextual Spoken Language Understanding with Dialogue Logistic Inference. *ACL* 2019 (short paper).**

- **He Bai**, Yu Zhou, Jiajun Zhang, Liang Zhao, Mei-Yuh Hwang and Chengqing Zong. **Source Critical Reinforcement Learning for Transferring Spoken Language Understanding to a New Language. *COLING* 2018 (full paper).**

Program Committees and Peer Review

2021-now	**Association for Computational Linguistics Rolling Review (ARR).**
2022	**International Conference on Machine Learning (ICML).**

2022   **International Conference on Computational Linguistic (COLING)**.

2022   **Workshop on Multilingual Information Access (MIA)**.

2021   **Annual Meeting of the Association for Computational Linguistics (ACL)**.

2020   **Annual Meeting of the Association for Computational Linguistics (ACL)**.

2020   **AAAI Conference on Artificial Interlligence (AAAI)**.

2020   **International Conference on Computational Linguistics (COLING)**.

2019   **Conference on Empirical Methods in Natural Language Processing (EMNLP)**.

Insterests

Research   Language Sequence Modeling, Unsupervised Learning, Multilingual NLP

Personal   Traveling, Cooking