

# He Bai

200 Ring Rd, Waterloo  
ON, Canada N2L 3G1

☎ (+1) 548 333 6369

✉ [he.bai@uwaterloo.ca](mailto:he.bai@uwaterloo.ca)

📄 [www.richardbaihe.github.io](http://www.richardbaihe.github.io)

## Education

Start from **University of Waterloo**,  
09/2019 *Computer Science*, PhD,  
Supervisor: Ming Li.

09/2017– **Institute of Automation, Chinese Academy of Sciences**,  
07/2019 *Pattern Recognition and Intelligent System*, Master of Science,  
Supervisor: Yu Zhou, Chengqing Zong.  
Thesis Title: Spoken Language Understanding Methods for Task-oriented Dialogue System

## Research

- 10/2019– **Segment-aware Transformer**, *Ming Li*, University of Waterloo.
- 05/2020
  - Introducing a segment-aware transformer for language modeling and understanding.
  - Achieving the SOTA results in language modeling, and a new pre-trained model SegBERT.
- 10/2018– **Multi-turn Spoken Language Understanding**, *Chengqing Zong*, Chinese Academy of  
03/2019 Sciences.
  - Proposing a self-supervised dialogue coherence task for multi-turn spoken language understanding.
  - Model trained with our proposed method get significant improvement on KVRET dataset.
- 10/2017– **Cross-lingual Spoken Language Understanding**, *Mei-yuh Hwang*, Mobvoi.
- 03/2018
  - Transferring Chinese spoken language understanding dataset into English with reinforcement learning controlled machine translation method.

## Internship

- 07/2018– **Research Intern**, IBM RESEARCH.
- 10/2018
  - Chinese Openai-GPT model pre-training.
  - Matching based neural dialogue generation with GAN.
- 06/2018– **Summer Intern**, ALIBABA GROUP, ALIME .
- 07/2018
  - Query rewriting and contextual queries assignment.
- 09/2017– **Research Intern**, MOBVOI.
- 03/2018
  - Language transferring for spoken language understanding.
  - Paraphrase generation.
- 07/2016– **Summer Intern**, UNISOUND.
- 09/2016
  - Implementing the seq2seq based chit-chat bot with C++.

## Award

- 2019-2023 International Doctoral Student Awards, University of Waterloo
- 2018 Excellent Student Award, University of Chinese Academy of Sciences

2015 Meritorious Winner of Mathematical Contest in Modeling, Consortium for Mathematics and Its Applications

---

## Publications

- **He Bai**, Peng Shi, Jimmy Lin, Luchen Tan, Kun Xiong, Wen Gao, Ming Li. **Segatron: Segment-aware Transformer for Language Modeling and Understanding**. *arXiv preprint arXiv:2004.14996*. (under review)
- **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**. *arXiv preprint arXiv:2004.02251*. (under review)
- Peng Shi, **He Bai**, Jimmy Lin. **Cross-Lingual Training of Neural Models for Document Ranking**. *To appear in Findings of EMNLP (Conference on Empirical Methods in Natural Language Processing) 2020*. Online.
- **He Bai**, Yu Zhou, Jiajun Zhang and Chengqing Zong. **Memory Consolidation for Contextual Spoken Language Understanding with Dialogue Logistic Inference**. *In Proceedings of ACL (Annual Meeting of the Association for Computational Linguistics) 2019*. Florence, Italy, July.
- **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**. *In Proceedings of COLING (International Conference on Computational Linguistics) 2018*. Santa Fe, New Mexico, USA, August.

---

## Service

- Program committee member for the 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2021)
- Program committee member for the the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)
- Program committee member for the 28th International Conference on Computational Linguistics (COLING 2020)

---

## Interests

Research Natural language processing, Representation Learning  
Personal Traveling, Music