

# Hong Wang

3530 Phelps Hall, UC Santa Barbara, California, 93106  
hongwang600@ucsb.edu • <http://hongwang600.github.io> • Google Scholar

## ABOUT ME

I'm an Applied Scientist at Amazon AWS Connect. I obtained my Ph.D. in Computer Science at UC Santa Barbara. My research mainly focuses on Machine Learning and NLP, especially on the dialogue system, question answering and information retrieval. As for industrial experience, I have interned as Applied Scientist and Research Scientist at Amazon Alexa, Meta AI and Microsoft Research, where I accumulated much hands-on experience for large-scale data preparation, model design, training, evaluation and deployment. In addition, I have led a team of 5 PhD talents to participate the fierce Amazon Taskbot Challenge, competing with 10 university teams from 3 continents and successfully getting into the final round. Personally, I am always curious and passionate about applying ML to solve complex problems in an applied environment!

## EDUCATION

**University of California, Santa Barbara**, Santa Barbara, California, USA

- Ph.D. in Computer Science Sep 2018 – Jun 2023
  - Advisor: Prof. Xifeng Yan
  - Focus: Natural Language Processing, Deep Learning
  - Chancellor's Fellowship for three years

**Nanjing University**, Nanjing, Jiangsu, China

- B.S. in Computer Science Sep 2014 – Jun 2018
  - Research assistant in Learning And Mining from Data (LAMDA) Group
  - Excellent Bachelor Thesis Award
  - GPA Ranking 4/150

## WORKING EXPERIENCE

**Amazon Alexa**, Wakeword Group, Cambridge, MA

- Applied Scientist Intern Jun 2022 – Sep 2022
  - Supervisor: Dr. Gengshen Fu, Rajath Kumar
  - Project: Unified Wakeword and compact spoken language understanding under low-resource scenarios
  - Applied data synthesis approach to generate audio data for low-resource intents.
  - Explored data augmentation techniques to improve generation from low-resource to rich-resource intents.
  - Proposed techniques to further improve generalization such as using separate classification layers and conducting alignment between synthetic and real data.
  - Submitted a paper and delivered a demo for unified wakeword and intent detection on device.

**Amazon Alexa**, Wakeword Group, Cambridge, MA

- Applied Scientist Intern Jun 2021 – Sep 2021
  - Supervisor: Dr. Gengshen Fu, Dr. Bo Xiao
  - Project: Unified Wakeword and intent detection on device
  - Designed model architecture for unified Wakeword and intent detection.
  - Proposed new loss function to reduce false accepts for the streaming model.
  - Supported entity recognition for a finite set, which largely increases the coverage of NLU on device.
  - Achieved a promising coverage for intent and entity recognition when matching the precision as cloud model.
  - Submitted a patent and a paper.

**Meta AI**, Language and Translation Technologies (LATTE) Group, Menlo Park, CA

- Research Intern Jun 2020 – Sep 2020
  - Supervisor: Dr. Ves Stoyanov
  - Project: Memory augmented pretrained language model
  - Implemented product key memory (PKM) in Fairseq framework for language modeling, achieving better perplexity during pre-training while using similar computation cost.

**Microsoft Research Asia**, Machine Learning Group, Beijing, China

- Software Engineer Intern Oct 2017 – Jan 2018
  - Supervisor: Dr. Tao Qin
  - Built classification models for ads relevance and click prediction.
  - Investigated the rationality of data, extracted new features and designed new models.
  - Explored how to alleviate position bias problem in advertisement.

## SKILLS

- **Programming Language:** Python, C, C++
- **Machine Learning:** Tensorflow, Pytorch, Spark, Transformers, Fairseq, Dialogue System, Question Answering, Information Retrieval, Language Model Pretraining
- **Computer Science:** SQL, Compiler, Operating System, Distributed System, Networking, Edge Computing, Database, Data Structure, Algorithms

## PUBLICATIONS

### CONFERENCES

- [1] Jing Qian\*, Hong Wang\*, Zekun Li, Shiyang Li, Xifeng Yan, “Limitations of Language Models in Arithmetic and Symbolic Induction”, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, Toronto, Canada, Jul 2023.
- [2] Zekun Li, Wenhui Chen, Shiyang Li, Hong Wang, Jing Qian, Xifeng Yan, “Dialogic: Controllable Dialogue Simulation with In-Context Learning”, in *Findings of the Conference on Empirical Methods in Natural Language Processing (Findings of EMNLP 2022)*, Abu Dhabi, Dec 2022
- [3] Zekun Li\*, Hong Wang\*, Alon Albalak, Yingrui Yang, Jing Qian, Shiyang Li, Xifeng Yan, “Making Something out of Nothing: Building Robust Task-oriented Dialogue Systems from Scratch”, in *Proceedings of Alexa Prize TaskBot*, Jun 2022.
- [4] Jing Qian, Hong Wang, Mai ElSherief, Xifeng Yan, “Lifelong Learning of Hate Speech Classification on Social Media”, in *Proceedings of 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2021)*, online, Jun 2021.
- [5] Hong Wang\*, Wenhan Xiong\*, William Yang Wang, “Progressively Pretrained Dense Corpus Index for Open-Domain Question Answering”, in *Proceedings of the 16th conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, online, Apr 2021.
- [6] Wenhui Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, William Wang, “HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data”, In *Findings of the Conference on Empirical Methods in Natural Language Processing (Findings of EMNLP 2020)*, online, Nov 2020.
- [7] Wenhui Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyu Zhou and William Yang Wang, “TabFact: A Large-scale Dataset for Table-based Fact Verification”, In *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*, Addis Ababa, Ethiopia, Apr 2020.
- [8] Hong Wang, Xin Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guo, Shiyu Chang and William Wang, “Self-Supervised Learning for Contextualized Extractive Summarization”, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, Florence, Italy, Jul 2019.
- [9] Wenhan Xiong, Jiawei Wu, Hong Wang, Vivek Kulkarni, Mo Yu, Xiaoxiao Guo, Shiyu Chang and William Wang, “TweetQA: A Social Media Focused Question Answering Dataset”, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, Florence, Italy, Jul 2019.
- [10] Hong Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guo, Shiyu Chang and William Wang, “Sentence Embedding Alignment for Lifelong Relation Extraction”, in *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Minneapolis, MN, USA, Jun 2019.
- [11] Hong Wang, Hong Qian and Yang Yu, “Noisy derivative-free optimization with value suppression”, in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*, New Orleans, LA, USA, Feb 2018.

### PREPRINTS

- [1] Shiyang Li, Jianshu Chen, Yelong Shen, Zhiyu Chen, Xinlu Zhang, Zekun Li, Hong Wang, Jing Qian, Baolin Peng, Yi Mao, Wenhui Chen, Xifeng Yan, “Explanations from Large Language Models Make Small Reasoners Better”, technical report, Oct 2022
- [2] Wenhan Xiong, Mo Yu, Xiaoxiao Guo, Hong Wang, Shiyu Chang, Murray Campbell and William Wang, “Simple yet Effective Bridge Reasoning for Open-Domain Multi-Hop Question Answering”, MRQA: Machine Reading for Question Answering workshop at EMNLP-IJCNLP 2019, Hong Kong, China, Nov 2019.

- [3] Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra and William Wang, “Fine-tune Bert for DocRED with Two-step Process”, technical report, Sep 2019.
- [4] Hong Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guo, Shiyu Chang and William Wang, “Meta Reasoning over Knowledge Graphs”, technical report, Jun 2019.

#### PATENTS

- [1] Hong Wang, Gengshen Fu, Bo Xiao and Tao Zhang, “Unified wakeword and language understanding processing”, patent submitted, Sep 2021.

#### AWARDS & HONORS

<b>Chancellor’s Fellowship</b> , University of California, Santa Barbara	Sep 2022
<b>Chancellor’s Fellowship</b> , University of California, Santa Barbara	Sep 2020
<b>Chancellor’s Fellowship</b> , University of California, Santa Barbara	Sep 2018
<b>Outstanding Undergraduate Award</b> , Nanjing University	Jun 2018
<b>Excellent Bachelor Thesis Award</b> , Nanjing University	Jun 2018
<b>National Scholarship</b> , Ministry of Education, China	2016 – 2017
<b>National Endeavor Scholarship</b> , Ministry of Education, China	2014 – 2016

#### SERVICE

##### COMMITTEE MEMBER & REVIEWER

- Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP-19), 2019
- The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), 2020
- The 58th annual meeting of the Association for Computational Linguistics (ACL-20), 2020
- The Conference on Empirical Methods in Natural Language Processing (EMNLP-20), 2020
- The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACL-IJCNLP-20), 2020
- The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021
- The 16th conference of the European Chapter of the Association for Computational Linguistics (EACL-21), 2021.
- The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT-21), 2021
- The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 21), 2021
- The 10th CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC2021), 2021
- The thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021), 2021
- ACL Rolling Review November, 2021
- ACL Rolling Review December, 2021
- The Tenth International Conference on Learning Representations (ICLR 2022), 2022
- The 13th Edition of its Language Resources and Evaluation Conference (LREC 2022), 2022
- The 39th International Conference on Machine Learning (ICML 2022), 2022
- The thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022), 2022
- The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022), 2022
- The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL-23), 2023
- The 40th International Conference on Machine Learning (ICML 2023), 2023
- The 61st Annual Meeting of the Association for Computational Linguistics (ACL-23), 2023
- 29TH ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2023)

#### TEACHING

- Teaching Assistant, CS165B, Machine Learning, Fall 2019, UCSB
- Teaching Assistant, CS165B, Machine Learning, Spring 2020, UCSB
- Teaching Assistant, CS165B, Machine Learning, Spring 2021, UCSB

#### TALKS & PRESENTATIONS

<b>Sentence Embedding Alignment for Lifelong Relation Extraction.</b>	May 2019
▪ IBM invited online talk	

[Last updated on 2023-06-13]