Wasi Uddin Ahmad

Applied Scientist, AWS AI Labs
(+1)434-202-9102 ⋄ wasicse90@gmail.com
https://wasiahmad.github.io

Current Research Focus

Training large language models for code completion and code in-filling, repository-level code generation, and building embedding models to facilitate their applications to code.

Education

Ph.D. in **Computer Science** 09.2017 - 09.2021
University of California, Los Angeles (UCLA)
CGPA: 3.78 on a scale of 4.00
Advisor: Dr. Kai-Wei Chang

Master of **Computer Science** 08.2015 - 08.2017
University of Virginia (UVA)
CGPA: 4.00 on a scale of 4.00
Advisor: Dr. Kai-Wei Chang

B.Sc. in **Computer Science and Engineering** 01.2008 - 02.2013
Bangladesh University of Engineering and Technology (BUET)
CGPA: 3.81 on a scale of 4.00
Position: Ranked 8th in a class of 142 students

Publications [Google Scholar]

(* indicates equal contribution)

1. Wu, D., **Ahmad, W. U.**, Zhang, D., Ramanathan, M. K., & Ma, X. (2024). RepoFormer: Selective Retrieval for Repository-Level Code Completion. In Proceedings of ICML.
2. Zhang, D.*, **Ahmad, W. U.***, Tan, M., Ding, H., Nallapati, R., Roth, D., Ma, X., & Xiang, B. (2024). Code Representation Learning At Scale. In Proceedings of ICLR.
3. Ding, Y.*, Wang, Z.*, **Ahmad, W. U.***, Ramanathan, M. K., Nallapati, R., Bhatia, P., Roth, D., & Xiang, B. (2024). CoCoMIC: Code Completion By Jointly Modeling In-file and Cross-file Context. In Proceedings of LREC-COLING.
4. Wu, D., **Ahmad, W. U.**, & Chang, K. W. (2024). On Leveraging Encoder-only Pre-trained Language Models for Effective Keyphrase Generation. In Proceedings of LREC-COLING.
5. Wu, D., **Ahmad, W. U.**, & Chang, K. W. (2023). Rethinking Model Selection and Decoding for Keyphrase Generation with Pre-trained Sequence-to-Sequence Models. In Proceedings of EMNLP.
6. Ding, Y.*, Wang, Z.*, **Ahmad, W. U.***, Ding, H., Tan, M., Jain, N., Ramanathan, M. K., Nallapati, R., Bhatia, P., Roth, D., & Xiang, B. (2023). CrossCodeEval: A Diverse and Multilingual Benchmark for Cross-File Code Completion. In Proceedings of the NeuRIPS Track on Datasets and Benchmarks.
7. Wei, X., Gonugondla, S., **Ahmad, W. U.**, Wang, S., Ray, B., & Others. (2023). Greener yet Powerful: Taming Large Code Generation Models with Quantization. In Proceedings of ESEC/FSE.

8. Jain, N.*, Zhang, D.*, **Ahmad, W. U.***, Wang, Z., Feng, N., Li, X., Tan, M., Nallapati, R., Ray, B., Bhatia, P., Ma, X., & Xiang, B. (2023). ContraCLM: Contrastive Learning For Causal Language Model. In Proceedings of ACL.

9. Hasan, T., Bhattacharjee, A., **Ahmad, W. U.**, Li, Y. F., Kang, Y. B., & Shahriyar, R. (2023). CrossSum: Beyond English-Centric Cross-Lingual Abstractive Text Summarization for 1500+ Language Pairs. In Proceedings of ACL.

10. Chi, J., **Ahmad, W. U.**, Tian, Y., & Chang, K. W. (2023). PLUE: Language Understanding Evaluation Benchmark for Privacy Policies in English. In Proceedings of ACL.

11. **Ahmad, W. U.**, Tushar, M. G. R., Chakraborty, S., & Chang, K. W. (2023). AVATAR: A Parallel Corpus for Java-Python Program Translation. In Findings of the ACL.

12. Athiwaratkun, B., Gouda, S. K., Wang, Z., Li, X., Tian, Y., Tan, M., **Ahmad, W. U.**, & Others. (2023). Multi-lingual Evaluation of Code Generation Models. In Proceedings of ICLR.

13. **Ahmad, W. U.**, Chakraborty, S., Ray, B., & Chang, K. W. (2023). Summarize and Generate to Back-translate: Unsupervised Translation of Programming Languages. In Proceedings of EACL.

14. Parvez, M. R., Chi, J., **Ahmad, W. U.**, Tian, Y., & Chang, K. W. (2023). Retrieval Enhanced Data Augmentation for Question Answering on Privacy Policies. In Proceedings of EACL.

15. Bhattacharjee, A., Hasan, T., **Ahmad, W. U.**, & Shahriyar, R. (2023). BanglaNLG: Benchmarks and Resources for Evaluating Low-Resource Natural Language Generation in Bangla. In Findings of the ACL: EACL.

16. Haque, M. M. A., **Ahmad, W. U.**, Lourentzou, I., & Brown, C. (2023). FixEval: Execution-based Evaluation of Program Fixes for Programming Problems. In International Workshop on APR.

17. Wu, D., **Ahmad, W. U.**, Dev, S., & Chang, K. W. (2022). Representation Learning for Resource-Constrained Keyphrase Generation. In Findings of the ACL: EMNLP.

18. Bhattacharjee, A.*, Hasan, T.*, **Ahmad, W. U.**, Samin, K., Islam, M. S., Iqbal, A., Rahman, M. S., & Shahriyar, R. (2022). BanglaBERT: Language Model Pretraining and Benchmarks for Low-Resource Language Understanding Evaluation in Bangla. In Findings of the ACL: NAACL.

19. Huang, K. H., **Ahmad, W. U.**, Peng, N., & Chang, K. W. (2021). Improving Zero-Shot Cross-Lingual Transfer Learning via Robust Training. In Proceedings of EMNLP.

20. Parvez, R. M., **Ahmad, W. U.**, Chakraborty, S., Ray, B., & Chang, K. W. (2021). Retrieval Augmented Code Generation and Summarization. In Findings of the ACL: EMNLP.

21. **Ahmad, W. U.**, Li, H., Mehdad, Y., & Chang, K. W. (2021). Syntax-augmented Multilingual BERT for Cross-lingual Transfer. In Proceedings of ACL.

22. **Ahmad, W. U.***, Chi, J.*, Le, T., Norton, T., Tian, Y., & Chang, K. W. (2021). Intent Classification and Slot Filling for Privacy Policies. In Proceedings of ACL.

23. **Ahmad, W. U.**, Bai, X., Lee, S., & Chang, K. W. (2021). Select, Extract, and Generate: Neural Keyphrase Generation with Layer-wise Coverage Attention. In Proceedings of ACL.

24. Hasan, M.*, Muttaqueen, T.*, Ishtiaq, A. A., Mehrab, K.S., Haque, M. M. A., Hasan, T., **Ahmad, W. U.**, Iqbal, A., & Shahriyar, R. (2021). CoDesc: A Large CodeDescription Parallel Dataset. In Findings of ACL.

25. **Ahmad, W. U.***, Chakraborty, S.*, Ray, B., & Chang, K. W. (2021). Unified Pre-training for Program Understanding and Generation. In Proceedings of NAACL-HLT.

26. **Ahmad, W. U.**, Peng, N., & Chang, K. W. (2021). GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction. In Proceedings of AAAI.
27. Chakraborty, S.*, Tafseer, M. T.*, & **Ahmad, W. U.** (2021). Simple or Complex? Learning to Predict Readability of Bengali Texts. In Proceedings of AAAI.
28. **Ahmad, W. U.***, Chi, J.*, Tian, Y., & Chang, K. W. (2020). PolicyQA: A Reading Comprehension Dataset for Privacy Policies. In Findings of the ACL: EMNLP.
29. **Ahmad, W. U.**, Chakraborty, S., Ray, B., & Chang, K. W. (2020). A Transformer-based Approach for Source Code Summarization. In Proceedings of ACL.
30. **Ahmad, W. U.**, Zhang, Z., Ma, X., Chang, K. W., & Peng, N. (2019). Cross-lingual Dependency Parsing with Unlabeled Auxiliary Languages. In Proceedings of CoNLL.
31. **Ahmad, W. U.**, Chang, K. W., & Wang, H. (2019). Context Attentive Document Ranking and Query Suggestion. In Proceedings of SIGIR.
32. **Ahmad, W. U.***, Zhang, Z.*, Ma, X., Hovy, E., Chang, K. W., & Peng, N. (2019). On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing. In Proceedings of NAACL-HLT.
33. Duong, D., **Ahmad, W. U.**, Eskin, E., Chang, K. W., & Li, J. J. (2019). Word and sentence embedding tools to measure semantic similarity of Gene Ontology terms by their definitions. Journal of Computational Biology.
34. **Ahmad, W. U.**, Chang, K. W., & Wang, H. (2018). Intent-aware query obfuscation for privacy protection in personalized web search. In Proceedings of SIGIR.
35. Yu, P., **Ahmad, W. U.**, & Wang, H. (2018). Hide-n-Seek: An Intent-aware Privacy Protection Plugin for Personalized Web Search. In Proceedings of SIGIR.
36. **Ahmad, W. U.**, Chang, K. W., & Wang, H. (2018). Multi-task learning for document ranking and query suggestion. In Proceedings of ICLR.
37. **Ahmad, W. U.**, & Chang, K. W. (2018). A Corpus to Learn Refer-to-as Relations for Nominals. In Proceedings of LREC.
38. **Ahmad, W. U.**, Rahman, M. M., & Wang, H. (2016). Topic model based privacy protection in personalized Web search. In Proceedings of SIGIR.

Preprints

1. Shahgir, H. S., Sayeed, K. S., Bhattacharjee, A., **Ahmad, W. U.**, Dong, Y., & Shahriyar, R. (2024). IllusionVQA: A Challenging Optical Illusion Dataset for Vision Language Models. arXiv preprint arXiv:2403.15952.
2. Wu, D., **Ahmad, W. U.**, & Chang, K. W. (2022). Pre-trained Language Models for Keyphrase Generation: A Thorough Empirical Study. arXiv preprint arXiv:2212.10233.
3. Hasan, M.*, Mehrab, K.S.*, **Ahmad, W. U.**, & Shahriyar, R. (2021). Text2App: A Framework for Creating Android Apps from Text Descriptions. arXiv preprint arXiv:2104.08301.
4. **Ahmad, W. U.**, Bai, X., Peng, N., & Chang, K. W. (2018). Learning Robust, Transferable Sentence Representations for Text Classification. arXiv preprint arXiv:1810.00681.

Industrial Experience

AWS AI Labs, Santa Clara, CA 10.2021 - Present
Applied Scientist II, Amazon CodeWhisperer

Facebook AI, Menlo Park, CA 06.2020 - 09.2020
Research Intern, Language and Translation Technology

Yahoo Research, Sunnyvale, CA 06.2019 - 09.2019
Research Intern, Ad Quality Science

Microsoft AI & Research, Redmond, WA 06.2018 - 09.2018
Research Intern, Business Applications Group

Walmart Labs, Reston, VA 06.2016 - 08.2016
Research Intern, Wireless Fraud Prevention

REVE Systems, Dhaka, Bangladesh 02.2013 - 10.2013
Software Development Engineer, Android

Teaching Experience

University of California, Los Angeles
Graduate Teaching Assistant

- Introduction to Computer Science II. Instructor: David Smallberg. 01.2019 - 06.2019
- Introduction to Machine Learning. Instructor: Kai-Wei Chang. 01.2018 - 03.2018

University of Virginia
Graduate Teaching Assistant

- Advanced Machine Learning. Instructor: Kai-Wei Chang. 01.2017 - 05.2017
- Natural Language Processing. Instructor: Kai-Wei Chang. 08.2016 - 12.2016
- Advanced Software Development. Instructor: Tom Horton. 08.2015 - 05.2016
- Capstone Practicum I & II. Instructor: Aaron Bloomfield. 08.2015 - 05.2016

Ahsanullah University of Science & Technology
Lecturer, Department of Computer Science 11.2013 - 08.2015

Research/Invited Talks

- Representation Learning for Low-resource Language Processing, AWS AI, May 2021
- Representation Learning for Low-resource Language Processing, Microsoft, April 2021
- Context Attentive Document Ranking and Query Suggestion, SIGIR, July 2019
- Multi-Task Learning, Machine Learning Seminar, UCLA, November 2017

Professional Services

Senior Area Chair

- 2024: NAACL

Senior Program Committee/Area Chair/Action Editor

- 2024: ACL, AAAI, ARR

- 2023: AAAI, ARR

Program Committee/Reviewer

- 2024: CIKM, COLM, ICML, IJCAI, SIGIR, LREC-COLING, ICLR
- 2023: EMNLP, NeuRIPS, ICML, SIGIR, ACL, IJCAI, ICLR, ARR, EACL
- 2022: EMNLP, NeuRIPS, ICML, SIGIR, IJCAI, KDD, ARR, LREC, AAAI, WSDM, ICLR
- 2021: EMNLP, NeuRIPS, SIGIR, ACL-IJCNLP, IJCAI, NAACL, EACL, AAAI
- 2020: EMNLP, ICML, IJCAI, AAAI, LREC
- Others: NAACL 2019, EMNLP 2018 (secondary reviewer)

Journal Reviewer

- Journal of Artificial Intelligence Research
- Engineering Applications of Artificial Intelligence
- ACM Transactions on Information Systems
- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Big Data
- IEEE Transactions on Knowledge and Data Engineering

Honors, Awards, and Scholarships

- SIGIR Student Travel Grant — 2016, 2019
- ICLR Travel Award — 2018
- Graduate Division Fellowships, UCLA — 2017 - 2018
- William L Ballard Jr Endowed Graduate Fellowship, UVA — 2017
- Graduate Division Fellowships, UVA — 2015 - 2016
- Dean's List Award, BUET — 2008-09 to 2010-11
- University Merit Scholarship, BUET — 2007-08 to 2010-11

References

Dr. Kai-Wei Chang
Associate Professor
Department of Computer Science
University of California, Los Angeles
Email: kwchang@cs.ucla.edu