JOONGWON KIM

+1(267) 591-1256 ⋄ Seattle, WA
jwonkim@cs.washington.edu ⋄ danieljkim0118.github.io ⋄ linkedin.com/in/daniel-joongwon-kim

EDUCATION

Ph.D. in Computer Science, University of Washington 2022-2027 (Expected)
Advisor: Hannaneh Hajishirzi Research Area: NLP, ML
Bachelor of Computer Science, University of Pennsylvania 2018 - 2022
Minor in Mathematics GPA: 3.95

PUBLICATIONS

[1] TaskWeb: Selecting Better Source Tasks for Multi-task NLP
- **Joongwon Kim**, Akari Asai, Gabriel Ilharco, Hannaneh Hajishirzi
- arXiv preprint (under submission) [paper]

[2] Induce, Edit, Retrieve: Language Grounded Multimodal Schema for Instructional Video Retrieval
- Yue Yang, **Joongwon Kim**, Artemis Panapogolou, Mark Yatskar, Chris Callison-Burch
- CVPR 2022 @ ODRUM (spotlight talk) [paper]

[3] BiSECT: Learning to Split and Rephrase Sentences with Bitexts
- **Joongwon Kim***, Mounica Maddela*, Reno Kriz, Wei Xu, Chris Callison-Burch
- Proceedings of EMNLP 2021 (long paper) [paper] [video] [poster]

AWARDS

NSF Graduate Research Fellowship 2022-27
Paul G. Allen School CSE Fellowship (given to select first-year Ph.D. students) 2022-23

RESEARCH PROJECTS

Solving Complex Reasoning Tasks with Open-Source LLMs *Ongoing*
- Designing a new, unique method to solve complex reasoning tasks by first generating a high-level plan which decomposes each problem into multiple easier subproblems, and then executing relevant tools to solve each subproblem.
- Building a pipeline by only using various open-source large language models such as Llama-2, CodeLlama, WizardCoder, etc. Using techniques such as paged attention and parallelism to increase generation speed.

Selecting Better Source Tasks for Multi-task NLP
- Conducted large-scale experiments of pairwise transfer learning between 22 different NLP tasks with various models and adaptation methods. Performed more than 10,000 transfer experiments on GPU cluster.
- Used results to 1) devise a new method for computing task similarity, and 2) improve multi-task learning.
- Resulting work available as a preprint [1].

Language Grounded Multimodal Schema for Video Retrieval
- Induced schema representations of goal-oriented (how-to) tasks from multimodal sources. Proposed methods to modify existing schemas to address unseen tasks. Improved performance on instructional video retrievals. Work published in CVPR 2022 ORDUM, spotlight talk [2].

Learning to Split and Rephrase Sentences with Bitexts
- Curated a multilingual Split and Rephrase corpus using machine translation over parallel corpora. Developed a sentence splitter with controllable generation and performed evaluations. Collaborated with Georgia Tech. Work published in EMNLP 2021 [3].

TEACHING

Computational Linguistics (CIS 530) Jan 2021 - May 2021
-Helped to teach a graduate-level NLP class of 150 students by managing office hours and Piazza forum.

Data Structures and Algorithms (CIS 121) Jan 2020 - Dec 2020
-Led weekly office hours and recitations of 20-30 students on fundamental algorithms/data structure concepts.