

# KANGDA WEI

L.F. Peterson Building, Room 330, 435 Nagle St, College Station, TX 77843

(919) 904-5280 ◊ kangda@tamu.edu ◊ hansonwei666@gmail.com

## RESEARCH INTEREST

---

Natural Language Processing, Deep Learning, Discourse Analysis, Large Language Model

## EDUCATION

---

### Texas A&M University

*Doctor of Philosophy in Computer Science*

College Station, TX

Aug 2023 - Present

- **Graduate Research Assistant:** Advised by Professor Ruihong Huang.

### The University of North Carolina at Chapel-Hill

*Master of Science in Computer Science*

Chapel-Hill, NC

Aug 2022 - May 2023

- **Graduate Teaching Assistant:** COMP431 Internet Service and Protocol, COMP211 System Fundamentals

### The University of North Carolina at Chapel-Hill

*B.S. in Computer Science, B.S. in Statistics and Operational Research*

Chapel-Hill, NC

Aug 2019 - May 2022

- **GPA:** 3.812/4.0

- **Relevant coursework:** Foundation of Programming, Data Structures, Algorithms and Analysis, Computational Photograph, Optimization, Structured Prediction, Machine Learning & Deep Learning, Programming Language, Time Series, Simulation Analysis, Stochastic Modeling, Linear Algebra, Computer Organization, Reinforcement Learning

### Online courses

- Stanford CS224n (NLP), Machine Learning, Deep Learning Specialization Series taught by Andrew Ng on Coursera

## RESEARCH EXPERIENCE

---

### Prompting Multi-hop Question Answering Task With Pre-trained Language Model

*Johns Hopkins University*

Baltimore, MD

*Visiting Research Assistant, Mentor: Prof. Benjamin van Durme*

May 2022 - May 2023

- Explored the effect of decomposition on machine reading with an exhaustive set of variants across a range of models over the high-level Question Decomposition Meaning Representation (QDMR) BREAK dataset.
- Discovered that question decomposition is not helpful for machine reading but rather harmful.
- Conducted a qualitative error analysis, showing that machine reading using question decomposition struggle due to compound error and question decomposition in bad formats.
- Published paper at *Proceedings of the Empirical Methods in Natural Language Processing 2023*.

### Test-time Adaptation of Language-guided Classifiers

*The University of North Carolina at Chapel Hill*

Chapel Hill, NC

*Research Assistant, Mentor: Prof. Shashank Srivastava*

Apr 2022 - May 2023

- Present a framework for test-time adaptation of language explanation-guided classifiers towards a specific task during inference.
- Achieved 8% higher classification accuracy by utilizing label aggregation with language model (LM) for test-time adaptation, and three times better accuracy comparing to baselines with zero-shot learning.
- Further improved the accuracy by 20% by incorporating self-learning by fine-tuning pre-trained LM on noisy labeled data.
- Conducted qualitative analysis for framework's interpretability.
- Simplify and improve the label aggregation technique by replacing hand-written labeling functions with LM.
- Published paper at *Findings of the Empirical Methods in Natural Language Processing 2023*.

### Compositional Generalization for Kinship Prediction through Data Augmentation

*The University of North Carolina at Chapel Hill*

Chapel Hill, NC

*Research Assistant, Mentor: Prof. Shashank Srivastava*

Feb 2021 - Mar 2022

- Evaluated empirically the utility of data augmentation and intermediate structured representations towards compositional generalization for the task of kinship prediction from a story.
- Tested the impact of incorporating data augmentation and intermediate structured data on model's performance. Data augmentation boosted generalization performance by around 20% on average relative to a baseline model from prior work.
- Found that predicting and using intermediate kinship graphs led to a deterioration in the generalization of kinship prediction.
- Published paper at *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*.

## A Multilingual COVID-19 Question Answering System

University of California, Santa Barbara

Visiting Research Assistant, Mentor: Prof. William Wang

Santa Barbara, CA

May 2021 - Aug 2021

- Established a multilingual COVID-19 Question Answering system using mBERT and XLM-Roberta models.
- Focused primarily on building and training the Reading Comprehension part of the QA system, including collecting and processing the data, and performing machine translation on large-scale natural language dataset.
- Performed large-scale neural machine translation on entire COVID-19 dataset to acquire data in Chinese and French to alleviate the data scarcity problem in foreign languages.
- Reached an F1 score of 60.5 for the reading comprehension model of the final QA system.

## PUBLICATIONS

- **Kangda Wei**, Dawn Lawrie, Benjamin Van Durme, Yunmo Chen, Orion Weller. When Do Decompositions Help for Machine Reading? *Proceedings of the Empirical Methods in Natural Language Processing 2023*
- **Kangda Wei**, Sayan Ghosh, Rakesh Menon, and Shashank Srivastava. TALC: Test-time Adaptation of Language-guided Classifier. *Findings of the Empirical Methods in Natural Language Processing 2023*
- **Kangda Wei**, Sayan Ghosh, and Shashank Srivastava. 2022. Compositional Generalization for Kinship Prediction through Data Augmentation. In *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*, pages 1319, Seattle, United States. Association for Computational Linguistics.
- \*Songhe Wang, **\*Kangda Wei**, Lei Lin, Weizi Li. Spatial-temporal Analysis of COVID-19's Impact on Human Mobility: the Case of the United States, in the *20th and 21st Joint COTA International Conference of Transportation Professionals*. \*Co-author: equal contribution

## NOTABLE PROJECTS

### Email Plug-in Startup

Jan 2022 - May 2022

- Built a classification model from scratch for email classification using PyTorch and Huggingface Transformer.
- Increased the model accuracy from 0.32 to 0.84 with limited amount of annotated data.

### Cyber-infrastructure: Web-based Application for Sharing Neural Imaging Data

Sep 2020 - Jan 2021

- Built a website where users can register and login to access the neural imaging data of patients.
- Practiced the basis of front-end and back-end engineering using Python Flask and Firebase.

## SKILLS

- **Programming Languages/Frameworks:** Python, Java, JavaScript, C, HTML, CSS, MATLAB, R, PyTorch, Linux
- **Software & Tools:** PyCharm, VS Code, Spyder, Jupyter Notebook, Tableau, Latex, Anaconda, RStudio
- **Language:** Chinese (native), English (proficient)