# Kai **Zhang**

727 West 40th St, Baltimore, MD 21211

✉ kzhan118@jhu.edu  |  📞 +1 (443) 226 3132  |  🐙 Leadlegend  |  in kaizhang-jhu

## **Edu**cation

### **MS, Computer Science**
Johns Hopkins University

*2024 - Now*
*Baltimore, MD*

- Research Interest: Large Language Models  |  BioNLP  |  Knowledge Graph
- Course: Convex Optimization  |  Natural Language Processing

### **BS in Computer Science and Technology (Turing Class)**
Peking University

*2018 - 2023*
*Beijing, China*

- Thesis: Exploring Few-Shot Learning of Large Language Models on Document-level Relation Extraction
- Course: Machine Learning  |  Computer Architecture  |  Database  |  Operating System
- Honor:
  ‣ The Third Prize of Peking University Scholarship
  ‣ summa cum laude

## **Pro**fessional Experience

### **Research Intern**
Tencent Ai Lab

*2021 - 2022*
*Beijing, China*

- Participated in construction and maintenance of multi-lingual universal-domain knowledge graph Topbase
- Implemented and optimized an entity linking model based on knowledge graph Topbase via distributed parallel development on multi-GPU, using data parallel and gradient parallel to improve contrastive learning effectiveness and inference accuracy.
- Designed a domain-specific semi-supervised named-entity recognition model based on contrastive learning paradigm Self-Tuning, introduced training signal annealing and self-distillation for better performance and achieved around 0.70 F1-score in sports and education domains with very low demand for annotated data.
- Implemented a controlled text generation model based on Chinese GPT-2 and UER-py, which takes keyword as input and outputs descriptive advertising text; Optimized model's keyword generation coverage and output diversity through in-domain pretraining and Mention Flags. The project received "Tencent Monthly Innovation Award" and has been applied to Tencent advertisement business of online reading platform.

`Knowledge Graph`   `Information Extraction`   `Pre-training`

## **Res**earch Experience

### **Exploring Few-Shot Learning of Large LMs on Document-level Relation Extraction**
Supervisor: Associate Prof. Yansong Feng, Wangxuan Institute of Computer Technology, Peking University

*2023*
*Beijing, China*

- Reviewed the few-shot learning (FSL) performance of large language models (LLMs) on mainstream NLP tasks, and investigated key factors contributing to models' generalization ability, especially their pre-training phases such as instruction tuning and prompt learning.
- Studied the limitations of document-level relation extraction (DocRE) on supervised learning settings, and explored the challenges and benefits of conducting DocRE task on FSL setting.
- Explored the influences of LLMs' DocRE generalization ability by FSL ablation experiments on scientific LLMs suite Pythia, especially the number of samples and model parameter amount.
- Validated the facilitating effect of positive correlation between pre-training corpus and inference data on DocRE task, and conducted experiments to check the correlation saliency for different models.

### **Interpretable Neural Network for Drug Response Prediction**
Supervisor: Assistant Prof. Sheng Wang, Paul G. Allen School of CSE, University of Washington

*2022*
*Seattle, WA*

- Investigated the interpretability of neural networks, a critical problem in BioNLP, especially the way of encoding feature and information among neurons in models and how to comprehend it.
- Reproduced DrugCell, a canonical interpretable model for drug response prediction on cancer cell-line and optimized the model's inference efficiency and prediction accuracy.
- Designed a new interpretable model architecture for drug response prediction: Readable Neural Networks, which extracted contextual text embeddings of Gene Ontology terms from PubMed literatures through distant supervision.

### **Development of Commonsense-based Question Generation Models**
Supervisor: Associate Prof. Yunfang Wu, Institute of Computational Linguistics, PKU

*2020*
*Beijing, China*

- Independently designed and implemented a seq-to-seq question generation model, leveraging prior knowledge from knowledge graph to enhance model performance and the quality of generated output.
- Reviewed development of pre-trained NLG methods (BERTsum, BART, ProphetNet, etc.), especially focusing on text summarization, and designed feasible ways to introduce pretraining paradigm into question generation task.

## **Ski**lls

| | |
|---|---|
| **Languages** | Python  |  C/C++  |  Java  |  HTML/CSS  |  Bash  |  SQL |
| **Developer Tools** | Docker  |  Git  |  Google Cloud Platform  |  $L^AT_EX$ |