



Yijia Shao

(+86) 17705770688 | shaoyj@pku.edu.cn
Personal Website: shaoyijia.github.io

EDUCATION

- Peking University**, Beijing Sep 2019 - July 2023 (expected)
BS in Data Science, Yuanpei College **GPA: 3.94/4.00 (1/305)**

RESEARCH EXPERIENCE

- Wangxuan Institute of Computer Technology, Peking University**, Beijing Oct 2020 - Present
Student Research Assistant instructed by Prof. Bing Liu (Distinguished Professor of UIC)
 - Out-of-Distribution Detection**: Design SOTA algorithm for OOD detection through pseudo OOD data generation. The proposed algorithm is *10 times faster* than contrastive learning methods.
 - Continual Post-Training of Language Models**: Improve pre-trained language models for domain-specific data through a guided integration of the specialized and the general knowledge. Design algorithm to pre-train in a continual learning setting.
- Plus Lab, University of California, Los Angeles**, Los Angeles, CA (remote) July 2022 - Present
Visiting Summer Intern instructed by Prof. Nanyun (Violet) Peng
 - Commonsense Evaluation of Dialogue Systems**: Work on designing commonsense evaluation metric which leverages symbolic representations and the Commonsense Knowledge Base. Create dataset for dialogue commonsense evaluation via crowdsourcing.
- Data, Knowledge, and Intelligence, Microsoft Research Asia**, Beijing Jan 2022 - July 2022
Research Intern mentored by Dr. Mengyu Zhou
 - Metadata Modeling for Tabular Data**: Design and implement algorithm to analyze the metadata of tables to enhance the understanding of tabular data in downstream tasks. Deploy the model to Azure Synapse to improve pivot chart recommendation.
 - FormLM for Online Form Understanding**: Design FormLM, a transformer-based model with modules for structural encoding, and collect the first open online form dataset through Cosmos to conduct continual pre-training.
 - Awarded a certificate of “**Stars of Tomorrow**” for the accomplishment in the internship.

PROFESSIONAL EXPERIENCE

- Peking University**, Beijing Fall 2022
Teaching Assistant of Introductory Topics in Data Science with Prof. Tiejun Li
- Tensorflow Lite, Google**, Beijing July 2021 - Sep 2021
Engineering Intern hosted by Tiezhen Wang and Chao Mei
 - Model Maker UI Prototype**: Design and build a cross-platform UI prototype above Tensorflow Lite APIs mainly using Flutter and Dart, receiving a *peer bonus* from the product manager.
 - Contribute code to **Tensorflow core** in C++ as an additional accomplishment.

PUBLICATIONS (*: Equal Contribution)

- Yijia Shao**, Mengyu Zhou, Yifan Zhong, Tao Wu, Hongwei Han, Gideon Huang, and Dongmei Zhang. FormLM: Recommending Creation Ideas for Online Forms by Modeling Semantic and Structural Information. Under submission.
- Zixuan Ke, **Yijia Shao**, Haowei Lin, Hu Xu, Lei Shu, and Bing Liu. Domain-Aaptive Pre-Training of Language Models via Guided Adaptation. Under submission.
- Zixuan Ke, Haowei Lin, **Yijia Shao**, Hu Xu, Lei Shu, and Bing Liu. Continual Post-Training of Language Models for Few-Shot Learning. Under submission.
- Mengyu Wang*, **Yijia Shao***, Haowei Lin, Wenpeng Hu, and Bing Liu. CMG: A Class-Mixed Generation Approach to Out-of-Distribution Detection. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Cham, 2022

LEADERSHIP

- Data Science Association of Yuanpei College, Peking University**, Beijing Jun 2021 - May 2022
President
 - Yuanpei College Intelligent Campus Project**: Colead the development of Yuanpei Underground Management System using Django framework which automates the resource management (*e.g.*, room booking, notification, *etc.*) and *is used by all students* in our college.

HONORS AND AWARDS

- May 4th Scholarship (the highest honor for students in Peking University), 2021
- National Scholarship (top 1%), 2020
- First prize in 12th Chinese Mathematics Competition Final (the only woman in the list), 2020
- First Class Scholarship for Freshmen of Peking University, 2019

SKILLS, INTEREST & OTHERS

- Programming Languages**: Python, C/C++, Dart, Bash, SQL, HTML, CSS, JavaScript
- Frameworks & Tools**: PyTorch, Scikit-learn, Django, Git, CUDA, Azure Cosmos
- Languages**: Chinese (native), English (proficient, GRE 339+4.5, TOEFL 113)
- Interest**: Travelling, Swimming, Dancing (traditional Chinese dance, Latin dance, Waltz)