

# Xiang Pan

Homepage

Google Scholar

Github

Email: xiangpan@nyu.edu

Mobile: (+1) 646-549-7912

## EDUCATION

---

- **New York University** New York City, NY  
*PhD in Data Science* Sep. 2023 –
- **New York University** New York City, NY  
*Master of Science in Computer Science; GPA: 3.89* Jan. 2021 – Dec. 2022
- **Huazhong University of Science and Technology** Wuhan, China  
*Bachelor of Engineering in Computer Science; GPA: 3.86* Aug. 2016 – July. 2020

## RESEARCH

---

- **Math and Data, Advisor: Qi Lei** New York University, NY  
*Research* Jan. 2023 - Present
  - **A Theoretical Analysis of Multi-source Domain Adaptation with Meta Representation Learning:** Theoretically analyze the benefits of utilizing the approximately shared features.
  - **Disentangled Representation Learning for Test Time Domain Representation:** Propose a disentangled representation learning method to dynamically select the feature space for the target domain.
- **ML2 Lab, Advisor: He He** New York University, NY  
*Research* Sep. 2021 - Present
  - **Interpreting Robust and Spurious Features (EMNLP 2022):** Investigated types of spurious correlation in NLP with different behavior of debiasing methods to improve current understanding of spurious features in NLP.
  - **Learning Dynamic of Feature:** Investigated the learning dynamic of feature learning with different types of interactions, learning difficulties and noises.
- **University of Chicago, Advisor: Victor Veitch** Chicago, IL  
*Visiting Student* May. 2022 - Present
  - **Representation Learning for Posthoc Methods:** Studied invariant predictors from an aspect of causality, and developed a post-hoc projection method to construct an invariant representation.
- **IBM Question Answering Group, Advisor: Avi Sil** New York, NY  
*Research Internship* Mar. 2021 - May. 2021
  - **Task Transfer and Domain Adaptation for Zero-Shot Question Answering (DeepLo Workshop, NAACL 2022):** Proposed supervised post-training with auxiliary tasks in target domain to improve the zero-shot QA domain adaptation performance.
- **Zilliz, Summer Open Source Project** Remote  
*Research Internship, Opensource* Jun. 2021 - Oct. 2021
  - **Hyperparameter Optimization:** Proposed and implemented a hyperparameter tuning method for Milvus **Vector Database** based on Bayesian Optimization Hyperband Hyperparameter Optimization.
- **National University of Singapore, Advisor: WeeSun LEE** Singapore  
*Research Assistant, Full Time* Jan. 2020 - Jan. 2021
  - **The Exclusion Classifier for Feature Representation and Robust Learning:** Proposed a NON-ADV training method based on robust representation theory, and successfully achieved competitive defence results with Adversarial Training Methods.
  - **A Robust Framework for Targeted Sentiment Analysis:** Proposed BERT-Based Framework for Targeted Sentiment Analysis under the unseen target test setting and exploited Auxiliary Training and Adversarial Training methods for enhancing robustness toward unseen targets' domain.

## SELECTED PUBLICATIONS

---

- [1] Ziliang Samuel Zhong, Xiang Pan, and Qi Lei. A Theoretical Analysis of Multi-source Domain Adaptation with Meta Representation Learning. *Under Review*, 2023.
- [2] Tim G J Rudner, Xiang Pan, Yucen Lily Li, Ravid Shwartz-Ziv, and Andrew Gordon Wilson. Uncertainty-Aware Priors for Finetuning Pretrained Models. *Under Review*, 2023.
- [3] Nitish Joshi, Xiang Pan, and He He. Are all spurious features in natural language alike? an analysis through a causal lens. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9804–9817, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [4] Xiang Pan\*, Alex Sheng\*, David Shimshoni\*, Aditya Singhal\*, Sara Rosenthal, and Avirup Sil. Task transfer and domain adaptation for zero-shot question answering. In *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing*, pages 110–116, 2022.

## DEVELOPMENT AND COMPETITION

---

- **Amazon DGL, Advisor: Da Zheng** Remote  
*Opensource Projects* *May. 2021 - Sep. 2022*
  - **TGL**: Implemented CAPI for Temporal Graph Learning using C++, python, and OpenMP.
- **SonoScape** Wuhan, China  
*Leader of 5-member Development Intern Team* *Oct. 2017 - Jan. 2018*
  - **Editable Formula Parser**: Used C++ and QT implemented an Efficient Editable Formula Parser for embedded medical devices. Awarded **Second Prize**, Scientific Innovation, SonoScape.
- **Sim4Real**: Exploited Simulation Data for Realistic Semantic Segmentation Model.
- **NLG Model Inference Optimization, CLUEAI**: Applied lightseq and turbotransformer for efficient inference optimization and deployment of NLG models and achieved 10x faster inference speed.
- **The Analysis of Beibei Shopping Data, Seed Cup 2017, Second Prize (2/200+ teams)**: Designed and modeled the users' shopping behaviour state by State Machine to predict next item purchased.
- **Facial Expression Recognition and Transformation, Summer School 2018, National University of Singapore**: Implemented a Facial Expression Recognition (FER) system based on the **IBM Cloud Service**. and generated different features in the face by applying StarGAN. Scored A as the overall assessment.

## SERVICE AND AWARDS

---

- **CLUEAI and CLUEBenchmarks**: Constructing a benchmark suite for Chinese language understanding evaluation.
- **Reviewer**: EMNLP 2022, ICDM 2022, AAAI 2022, TALLIP
- **Two-time Winner, Merit Student Scholarship**. 2018, 2019
- **National Scholarship (National Top 0.2%)**. 2018
- **Outstanding Undergraduate Students (top 1% of 35000 in HUST)**. 2018
- **People's Scholarship**. 2018

## TECHNOLOGY STACK

---

- **HPC and Cloud Computing**: Slurm, Virtual Machine, Docker, GCP, AWS, NAS
- **Languages**: Python, C++, Matlab
- **Packages**: PyTorch, Tensorflow, Jax