Xudong Shen

This version: December 11, 2022

CONTACT
- https://xudongolivershen.github.io
- https://www.linkedin.com/in/oliverxudongshen
- ✉ xudong.shen@u.nus.edu
- ☎ (+65) 88296866 or (+86) 18658143615

EDUCATION

Ph.D. in Integrative Sci. & Eng. (Comput. Sci. Focus), *National Univ. of Singapore* 2019-Present

B.A. in Naval. Arch. & Ocean Eng., *Zhejiang Univ.*, China 2015-2019

TECHNICAL SKILLS

Machine Learning (ML):

Experienced with the complete ML pipeline in Python (`numpy`, `scikit-learn`, `pandas`, `PyTorch`, `HuggingFace`,...). Experience with image (face), sensor (satellite remote sensing), sequential (skeleton time sequence), language, and FinTech (P2P lending) data; and vision, NLP, multimodal, and generative models.

Statistics:

Big data analysis and structural modelling using frequentist and Bayesian approaches using the R (`survival`, `rms`,...) and Python (`PyMC3`,...) stack. Experienced with survival modelling, financial default analysis, and social and network data analysis.

Development:

Experienced with dev tools `Git`, `Docker`, `Kubernetes`, `AWS`.

EXPERIENCE

Doctoral Researcher, **N-CRiPT Centre**, NUS, Singapore 2019–Present

(a) Designed and executed research in ML and finance: (*i*) mitigate bias in large ML systems, *e.g.*, face recognition, (*ii*) analyze million P2P lending data to estimate default risk and profitability, gender discrimination, and what drives gender discrimination.

(b) Actively contributed to community-driven research projects on biases and robustness in NLP and AI, such as BIG-Bench, BigScience, and NL-Augmenter.

(c) Communicated scientific insights to both academic and industry audience in 6 papers (2 first-authored, 3 published in leading journals and conferences) and 3+ talks.

(d) Taught the course *Computational Methods for Business Analytics* twice, totally 150+ hours to 200+ undergraduate students. Prepared course materials and weekly recap videos, and delivered regular Q&A sessions.

Research Intern, **Sea AI Lab**, Sea Limited, Singapore 2022–Present

(a) Drive research projects involving 5+ researchers from 2 institutes (NUS & Sea) on bias and safety issues of diffusion-based image generation ML models, *e.g.*, fair language representation and its impact on image generation.

Undergraduate Researcher, **Second Institute of Oceanography**, China Jul. 2017–Jul. 2019

(a) Worked on the full stack of image segmentation in satellite remote sensing, incl. sensor data preparation, image manual segmentation, neural net arch., and model testing.

Undergraduate Researcher, **Waseda University**, Japan Sep. 2018–Feb. 2019

(a) Analyzed the effectiveness of Japan's policy toward the elimination of Persistent Organic Pollutants (POPs), and proposed policy recommendations in China.

Undergraduate Researcher, **York University**, Canada May–Aug. 2018

(a) Worked on fluid dynamics simulation of a hydrodynamically focused printing process for printed electronics, using `COMSOL` and `FLUENT`, in collaboration with 5 other researchers.

PREPRINTS

1. Kaustubh D Dhole, Varun Gangal, Sebastian Gehrmann, Aadesh Gupta, ..., Xudong Shen, ...,: "NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation". In *arXiv preprint, arXiv:2112.02721* (2021).

2. Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, ..., Xudong Shen, ...: "Beyond the Imitation Game: Quantifying and Extrapolating the Capabilities of Language Models". In *arXiv preprint, arXiv:2206.04615* (2022).

3. Xudong Shen, Tianhui Tan, Tuan Q. Phan, Jussi Keppo: "Unwarranted Gender Disparity in Online P2P Lending: Evidence of Affirmative Action". In *arXiv preprint, arXiv:2210.07864* (2022).

PUBLICATIONS

1. Ziwei Xu, Xudong Shen, Yongkang Wong, Mohan Kankanhalli: "Unsupervised Motion Representation Learning with Capsule Autoencoders". In *NeurIPS* (2021).

2. Xudong Shen, Yongkang Wong, Mohan Kankanhalli: "Fair Representation: Guaranteeing Approximate Multiple Group Fairness for Unknown Tasks". In *IEEE Trans. PAMI* (2022).

3. Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, ..., Xudong Shen, ...: "Benchmarking Generalization via In-Context Instructions on 1,600+ Language Tasks". In *EMNLP* (2022).

INVITED
TALKS

Workshop on Enormous Language Models, *ICLR 2021*, Online	2021
N-CRiPT Seminar, *National University of Singapore*, Singapore	2022
International Conference on Smart Finance 2022, Online (Hong Kong)	2022

SERVICE

Referee: EMNLP 2022, ICLR 2023, WACV 2023.

ADDITIONAL
TRAINING

Master Kong Dream Scholarship Program, *Waseda University*, Japan	Sep. 2018-Feb. 2019
Globalink Research Internship, *York University*, Canada	May.-Aug. 2018
Erasmus+ Student Mobility, *Università di Trento*, Italy	Feb.-Jun. 2017