

Mengnan Du

Texas A&M University
<https://mengnandu.com>
Phone: +1 979 267 5303
Email: dumengnan@tamu.edu

Education

- 2017.8 - Now **Ph.D Computer Science**, Texas A&M University, College Station, USA.
Advisor: Dr. Xia Ben Hu
- 2014.9 - 2017.3 **M.S. Electrical Engineering**, Beihang University, Beijing, China.
Advisor: Prof. Weihai Chen, Prof. Xingming Wu and Prof. Zhengguo Li
- 2010.9 - 2014.6 **B.E. Electrical Engineering**, North China Electric Power University, Beijing, China.
Advisor: Associate Prof. Guangsheng Liang

Professional Experience

- 2021.5 - 2021.8 **Microsoft Research**, Redmond, USA.
◦ Research Intern
◦ Mentors: Subho Mukherjee, Yu Cheng, Ahmed Hassan Awadallah
- 2020.5 - 2020.8 **Adobe Research**, San Jose, USA.
◦ Research Intern
◦ Bias detection and mitigation for few shot learning applications.
◦ Mentors: Varun Manjunatha, Tong Sun, Rajiv Jain, Jiuxiang Gu, Ruchi Deshpande, Franck Dernoncourt
- 2017.3 - 2017.8 **Intel Corporation**, Shanghai, China.
◦ Software Engineer Intern
◦ Worked on an AR (Augmented reality) project for Smart glasses.
◦ Developed image classification and speech recognition algorithms which can be deployed on Android system.

Research Interests

- 2017 -Till now Trustworthy Machine Learning (with a particular interest in the areas of Explainable, Fair, and Robust DNNs), and Natural Language Understanding

Publications (Google Scholar) Citations: 920+, h-index: 11, i10-index: 13, as of October 2021

33. **Mengnan Du**, Subhabrata Mukherjee, Guanchu Wang, Ruixiang Tang, Ahmed Hassan Awadallah, Xia Hu, "Fairness via Representation Neutralization", Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS), 2021.
32. **Mengnan Du**, Varun Manjunatha, Rajiv Jain, Ruchi Deshpande, Franck Dernoncourt, Jiuxiang Gu, Tong Sun and Xia Hu, "Towards Interpreting and Mitigating Shortcut Learning Behavior of NLU models", North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
31. Huiqi Deng, Na Zou, Weifu Chen, Guocan Feng, **Mengnan Du** and Xia Hu, "Mutual Information Preserving Back-propagation: Learn to Invert for Faithful Attribution", The 27rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2021.
30. Ruixiang Tang, **Mengnan Du**, Yuening Li, Zirui Liu and Xia Hu, "Mitigating Gender Bias in Captioning Systems", The Web Conference (WWW), 2021.
29. Huiqi Deng, Na Zou, **Mengnan Du**, Weifu Chen, Guocan Feng, Xia Hu, "A Unified Taylor Framework for Revisiting Attribution Methods", AAAI Conference on Artificial Intelligence (AAAI), 2021.
28. Yijun Bian, Qingquan Song, **Mengnan Du**, Jun Yao, Huanhuan Chen and Xia Hu, "Sub-Architecture Ensemble Pruning in Neural Architecture Search", accepted by IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2021.

27. Weijie Fu, Meng Wang, **Mengnan Du**, Ninghao Liu, Shijie Hao and Xia Hu, "Differentiated Explanation of Deep Neural Networks with Skewed Distributions", accepted by IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.
26. Fangsheng Wu, **Mengnan Du**, Chao Fan, Ruixiang Tang, Yang Yang, Ali Mostafavi, Xia Hu, "Understanding Social Biases behind Location Names in Contextual Word Embedding Models", accepted by IEEE Transactions on Computational Social Systems (TCSS), 2021.
25. Ninghao Liu, **Mengnan Du**, Ruocheng Guo, Huan Liu and Xia Hu, "Adversarial Attacks and Defenses: An Interpretation Perspective", accepted by SIGKDD Explorations, 2021.
24. Fan Yang, Ninghao Liu, **Mengnan Du**, and Xia Hu, "Generative Counterfactuals for Neural Networks via Attribute-Informed Perturbation", accepted by SIGKDD Explorations, 2021.
23. Sina Mohseni, Fan Yang, Shiva Pentala, **Mengnan Du**, Yi Liu, Nic Lupfer, Xia Hu, Shuiwang Ji, and Eric Ragan, "Machine Learning Explanations to Prevent Overtrust in Fake News Detection", International AAAI Conference on Web and Social Media (ICWSM), 2021.
22. **Mengnan Du**, Ninghao Liu, and Xia Hu, "Techniques for Interpretable Machine Learning", Communications of the ACM (CACM), 2020. [Highlighted on the cover page of the January 2020 issue.](#)
21. **Mengnan Du**, Fan Yang, Na Zou, and Xia Hu, "Fairness in Deep Learning: A Computational Perspective", IEEE Intelligent Systems, 2020.
20. **Mengnan Du***, Shiva Pentala*, Yuening Li, Xia Hu, "Towards Generalizable Deepfake Detection with Locality-aware AutoEncoder", The 29th ACM International Conference on Information and Knowledge Management (CIKM), 2020. (*Equal contribution)
19. **Mengnan Du**, Ninghao Liu, Fan Yang, and Xia Hu, "Learning Credible DNNs via Incorporating Prior Knowledge and Model Local Explanation", Knowledge and Information Systems (KAIS), 2020.
18. Ruixiang Tang, **Mengnan Du**, Ninghao Liu, Fan Yang, and Xia Hu, "An Embarrassingly Simple Approach for Trojan Attack in Deep Neural Networks", The 26rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020.
17. Fan Yang, Ninghao Liu, **Mengnan Du**, Kaixiong Zhou, Shuiwang Ji and Xia Hu, "Deep Neural Networks with Knowledge Instillation", SIAM International Conference on Data Mining (SDM), 2020.
16. Nur Hafieza Ismail, Ninghao Liu, **Mengnan Du**, Zhe He and Xia Hu, "A Deep Learning Approach for Identifying Cancer Survivors Living with Post-Traumatic Stress Disorder on Twitter", BMC Medical Informatics and Decision Making (BMC), 2020.
15. **Mengnan Du**, Ninghao Liu, Fan Yang, and Xia Hu, "Learning Credible Deep Neural Networks with Rationale Regularization", IEEE International Conference on Data Mining (ICDM), 2019.
[Best Paper Candidate](#)
14. **Mengnan Du**, Ninghao Liu, Fan Yang, Shuiwang Ji, and Xia Hu, "On Attribution of Recurrent Neural Network Predictions via Additive Decomposition", The Web Conference (WWW), 2019.
[Best Paper Candidate](#)
13. Fan Yang*, Shiva Pentala*, Sina Mohseni*, **Mengnan Du***, Hao Yuan*, Rhema Linder, Eric Ragan, Shuiwang Ji and Xia Hu, "XFake: Explainable Fake News Detector with Visualizations", The Web Conference (WWW), demo track, 2019. (*Equal contribution)
12. Ninghao Liu, **Mengnan Du** and Xia Hu, "Representation Interpretation with Spatial Encoding and Multimodal Analytics", ACM International Conference on Web Search and Data Mining (WSDM), 2019.
11. Nur Hafieza Ismail, **Mengnan Du**, Diego Martinez and Zhe He, "Multivariate Multi-step Deep Learning Time Series Approach in Forecasting Parkinson's Disease Future Severity Progression", The 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB), 2019.
10. Nur Hafieza Ismail, Ninghao Liu, **Mengnan Du**, Zhe He, and Xia Hu, "Identification of Cancer Survivors Living with Post-Traumatic Stress Disorder (PTSD) on Social Media", The World Congress on Medical and Health Informatics (MEDINFO), poster, 2019.
9. Yuening Li, Xiao Huang, Jundong Li, **Mengnan Du**, and Na Zou, "SpecAE: Spectral Autoencoder for Anomaly Detection in Attributed Networks", ACM International Conference on Information and Knowledge Management (CIKM), 2019. [INFORMS 2019 Best Refereed Paper Finalists](#)
8. Yuening Li, Ninghao Liu, Jundong Li, **Mengnan Du**, and Xia Hu, "Deep Structured Cross-Modal Anomaly Detection", International Joint Conference on Neural Networks (IJCNN), 2019.

7. **Mengnan Du**, Ninghao Liu, Qingquan Song, and Xia Hu, "Towards Explanation of DNN-based Prediction with Guided Feature Inversion", The 24rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2018.
6. **Mengnan Du**, Xingming Wu, Weihai Chen, Zhengguo Li, "Supervised Training and Contextually Guided Salient Object Detection", Digital signal processing (DSP), 2017.
5. **Mengnan Du**, Xingming Wu, Weihai Chen, Jianhua Wang, "Exploiting multiple contexts for saliency detection", Journal of Electronic Imaging (JEI), 2016.
4. Xingming Wu, **Mengnan Du**, Weihai Chen, Jianhua Wang, "Salient object detection via region contrast and graph regularization", SCIENCE CHINA Information Sciences, 2016.
3. Xingming Wu, **Mengnan Du**, Weihai Chen, Zhengguo Li, "Exploiting Deep Convolutional Neural Network and Patch-level CRFs for Indoor Semantic Segmentation", in IEEE Conference on Industrial Electronics and Applications (ICIEA), 2016.
2. **Mengnan Du**, Xingming Wu, Weihai Chen, Jianhua Wang, "Fusing region contrast and graph regularization for saliency detection", Chinese Control and Decision Conference (CCDC), 2015.
1. Guangsheng Liang, **Mengnan Du**, Zihao Zhou, et al. "Attitude control based on complementary filter for two-wheel self-balance car", Measurement Control Technology, 2015. (in chinese)

Workshop Papers

3. Sina Mohseni, Fan Yang, Shiva Pentiyala, **Mengnan Du**, Yi Liu, Nic Lupfer, Xia Hu, Shuiwang Ji, Eric Ragan, "Trust Evolution Over Time in Explainable AI for Fake News Detection", CHI Workshop on Fair & Responsible AI, 2020.
2. Haofan Wang, Zifan Wang, **Mengnan Du**, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, Xia Hu, "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks", IEEE CVPR Workshop on Fair, Data-Efficient and Trusted Computer Vision, 2020.
1. Nur Hafieza Ismail, Ninghao Liu, **Mengnan Du**, Zhe He, and Xia Hu, "Using Deep Neural Network to Identify Cancer Survivors Living with Post-Traumatic Stress Disorder on Social Media ", The 4th Interpretable Workshop on Semantics-Powered Data Mining and Analytics (SEPDA), 2019.

Book Chapter

1. Nur Hafieza Ismail, **Mengnan Du**, Xia Hu, "Social Media and Psychological Disorder", Social Web and Health Research, Springer, 2019. 171-192.

Preprints

4. Huiqi Deng, Na Zou, **Mengnan Du**, Weifu Chen, Guocan Feng, Xia Hu, "A General Taylor Framework for Unifying and Revisiting Attribution Methods", arXiv:2105.13841, 2021.
3. Yuening Li, Zhengzhang Chen, Daochen Zha, **Mengnan Du**, Denghui Zhang, Haifeng Chen, Xia Hu, "Learning Disentangled Representations for Time Series", arXiv:2105.08179, 2021.
2. Ruixiang Tang, **Mengnan Du**, and Xia Hu, "Deep Serial Number: Computational Watermarking for DNN Intellectual Property Protection", arXiv:2011.08960, 2020.
1. Fan Yang, **Mengnan Du**, and Xia Hu, "Evaluating Explanation Without Ground Truth in Interpretable Machine Learning", arXiv:1907.06831, 2019.

Teaching Experience

2021 Fall **Guest Lecturer at Texas A&M University**, *In person*.

- IDIS 344: Distributor Information & Control Systems, Section: 503, 504, 505, 506
- Department of Engineering Technology & Industrial Distribution
- Lecture contents: Introduction to Recommender Systems

2020 Spring **Guest Lecturer at Texas A&M University**, *Using zoom*.

- IDIS 344: Bio-Seminar, ECE Department
- Lecture contents: Towards Effective Interpretation of Deep Neural Networks: Algorithms and Applications

Conference and Invited Talks

- 2021 "Towards Interpreting and Mitigating Shortcut Learning Behavior of NLU models", NAACL-21
- 2020 "Towards Generalizable Deepfake Detection with Locality-aware AutoEncoder", CIKM-20
- 2019 "Learning Credible Deep Neural Networks with Rationale Regularization", ICDM-19
- 2019 "On Attribution of Recurrent Neural Network Predictions via Additive Decomposition", WWW-19
- 2018 "Towards Explanation of DNN-based Prediction with Guided Feature Inversion", KDD-18

Services

Conference Program Committee Member:

- International Conference on Learning Representations (ICLR), 2022
- AAAI Conference on Artificial Intelligence (AAAI), 2022, 2021, 2020
- Neural Information Processing Systems (NeurIPS), 2021, 2020
- International Conference on Machine Learning (ICML), 2021
- Annual Meeting of the Association for Computational Linguistics (ACL), 2021, 2020
- Empirical Methods in Natural Language Processing (EMNLP), 2021, 2020
- North American Chapter of the Association for Computational Linguistics (NAACL), 2021
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020
- AMIA Informatics Summit (AMIA), 2020

Reviewer of (in charge of reviewing one submission):

- The IEEE Visualization Conference (VIS), 2021
- ICML Workshop on Uncertainty & Robustness in Deep Learning (UDL), 2020
- IEEE International Conference on Robotics and Automation (ICRA), 2019

Journal Reviewer of:

- ACM Computing Surveys (CSUR)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Transactions on Big Data (TBD)
- IEEE Signal Processing Letters (SPL)
- Journal of Electronic Imaging (JEI)
- IEEE Intelligent Systems
- Neurocomputing
- Applied AI Letters
- IEEE Access

Major Honors & Awards

- 2020 Graduate Research Excellence Award, Texas A&M University
- 2020 NortonLifeLock Graduate Fellowship 2020 finalist
- 2020 CIKM Student Travel Award
- 2019 ICDM 2019 Best Paper Candidate
- 2019 Texas A&M Institute of Data Science Graduate Travel Grant
- 2019 ICDM Student Travel Award
- 2019 INFORMS 2019 Best Refereed Paper Finalists
- 2019 WWW 2019 Best Paper Shortlist
- 2019 TAMU CSE Student Travel Award covering the attendance of the WWW 2019
- 2018 KDD Student Travel Award