

# HUA SHEN

Ph.D Student in Informatics - The PennState University - PA, USA

✉ [huashen218@psu.edu](mailto:huashen218@psu.edu); 🌐 <http://hua-shen.org>; 📷 [huashen218](#)

## EDUCATION

---

- Ph.D. in Informatics (GPA: 3.91)** *2019.08 - Present*  
Penn State University, PA, USA  
Advisor: Dr. Ting-Hao (Kenneth) Huang
- Ph.D. in Computer Science (Graduate Fellowship) (Transferred to PSU)** *2018.08 - 2019.08*  
Lehigh University, PA, USA  
Advisors: Dr. Ting Wang, Dr. Brian D. Davison.
- M.S. in Management Science and Engineering (Admission Exam Exemption)** *2013.09 - 2016.06*  
Renmin University of China, Beijing, China  
Advisor: Dr. Xun Liang
- B.S. in Information Security (Ranking: 2/35)** *2009.09 - 2013.06*  
University of Science and Technology Beijing, Beijing, China

## RESEARCH INTERESTS

---

My research interests lie in facilitating the **usefulness of responsible AI techniques** from **human-centered** perspective. I'm particularly interested in improving **interpretability and fairness** in NLP and ML techniques.

## PUBLICATIONS

---

### Peer-Reviewed Conference & Workshop Papers

- [P.8] **Hua Shen**, Ting-hao (Kenneth) Huang. Explaining the Road Not Taken. The 2021 ACM CHI Workshop on Operationalizing Human-Centered Perspectives in Explainable AI (CHI '21 HCXAI Workshop).
- [P.7] Jiaqi Wang, **Hua Shen**, Chacha Chen, Frank E. Ritter. Are Learners Satisfied with Their MOOC Experiences? Assessing and Improving Online Learners' Interactions. Asian CHI Symposium 2021 (AsianCHI '21).
- [P.6] **Hua Shen**, Ting-hao (Kenneth) Huang. How Useful Are the Machine-Generated Interpretations? A Human Evaluation on Guessing the Wrongly Predicted Labels. The 8th AAAI Conference on Human Computation and Crowdsourcing (HCOMP '20).
- [P.5] Xinyang Zhang, Ningfei Wang, **Hua Shen**, Shouling Ji, Ting Wang. Interpretable Deep Learning under Fire. Proceedings of the 29th USENIX Security Symposium (USENIX'20).
- [P.4] Ren Pang, **Hua Shen**, Xinyang Zhang, Shouling Ji, Yevgeniy Vorobeychik, Xiapu Luo, Alex X. Liu, Ting Wang. The Tale of Evil Twins: Adversarial Inputs versus Poisoned Models. The 27th ACM Conference on Computer and Communications Security (ACM CCS'20).
- [P.3] **Hua Shen**, Xun Liang. A Quantitative Analysis Decision System Based on Deep Learning and NSGA-II for FX Portfolio Prediction". The 31st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA-AIE'18).
- [P.2] **Hua Shen**, Xun Liang. A Time Series Forecasting Model Based on Deep Learning Integrated Algorithm with Stacked Autoencoders and SVR for FX Prediction. The 25th International Conference on Artificial Neural Networks (ICANN'16).
- [P.1] **Hua Shen**, Xun Liang, Mingming Wang. Emergency Decision Support Architectures for Bus Hijacking Based on Massive Image Anomaly Detection in Social Networks. The 2015 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC'15).

### Book

- [B.1] Xun Liang, Xiaoping Yang, **Hua Shen**. Social Commerce Theory and Practice[M], Tsinghua University Press, 2014, ISBN No. 9787302381129.

## PATENTS

---

- Xun Liang, **Hua Shen**, Run Cao. An innovative Emergency Discovery Method for Micro-blogging, Archived by State Intellectual Property Office of The P.R.C (SIPO), Patent Publication No. CN103577404A.
- Xiaofei Li, Xun Liang, Xiaoping Zhou, Xiaojing Shi, **Hua Shen**, Haiyan Zhang. A Cross-Platform Microblog Community Account Matching Method, Archived by State Intellectual Property Office of The P.R.C (SIPO), Patent Publication No. CN104765729A.

## RESEARCH EXPERIENCE AND PROJECTS

---

### Amazon Alexa Speech, Applied Scientist Intern, CA, USA

May 2021 - August 2021

Mentor: Dr.Yuguang Yang. Co-mentors: Dr.Guoli Sun, Dr.Ryan Langman.

Advisors: Dr.Andreas Stolcke, Dr. Jasha Droppo

Research Topics: *Fairness, Speaker Recognition Systems*.

- **Improving Gender Fairness in Speaker Recognition Systems**  
- Preparing a conference paper.

### Penn State University, Research Assistant, PA, USA

Aug 2019 - Present

Advisor: Dr.Ting-Hao (Kenneth) Huang

Research Topics: *Interpretability, Human-AI Collaboration, Crowdsourcing*

- **Investigating Gaps between NLP Explanations and End-User Demands**  
We surveyed 218 NLP explanation papers collected from top-tier NLP venues from 2015-2020. We summarized 12 common forms of NLP explanations. We find users are commonly interested in counterfactual explanations where exiting NLP algorithms do not cover well. This work is published in CHI'21 HCXAI.
- **Usefulness Evaluation for AI Interpretation by General Users**  
We investigate whether or not showing machine-generated visual interpretations helps users understand the incorrectly predicted labels by image classifiers. We generate a set of saliency maps (e.g., SmoothGrad, ExtremalPerturb, GradCAM) for ResNet model on ImageNet dataset. The human evaluation is finished by Amazon MTurk crowdsourcing method. The work is published at HCOMP'20.
- **Crowd-Powered Conversational AI System Explanation**  
We are developing a dialogue system with frontend as Facebook Messenger and backend powered by Amazon Mechanical Turk. We plan to conduct interpretation related analysis on the dialogue system. I use full-stack techniques including *Python, Pytorch* for deep learning models, *SQLAlchemy* for database, *HTML, CSS, JavaScript, ReactJS* for website development.
- **Towards Effective and Efficient Evaluation of Interpretation Fidelity**  
It is challenging to evaluate whether an interpretation of model behavior is faithful. Current evaluation approaches suffer from high time consumption or low generalization issues. We propose a novel fidelity evaluation framework which is accurate but more efficient and general.

Mentor: Dr.Ting Wang

Research Topics: *Interpretability, Robustness, Adversarial ML*

- **i-Algebra: Towards Interactive Interpretability of Neural Networks**  
Existing interpretable methods mostly provide interpretability in a static manner, without accounting for the end-users' response, resulting in their poor usability in practice. We propose interactive framework, i-Algebra, for interpreting deep neural networks.
- **Assessing Interpretation Robustness by Attacking both Interpretation and Models**  
We propose a new class of attacks that generate adversarial inputs misleading both interpretations and classifiers. We attacked different types of saliency maps, including GradSaliency, CAM, etc. The evaluation is tested on real applications like skin cancer and ImageNet datasets. The work is accepted by USENIX'20.
- **A Tale of Evil Twins: Adversarial Inputs versus Poisoned Models**  
ML models are inherently susceptible to two manipulations: adversarial inputs and poisoned models. We conduct systematic study of two attack vectors within a unified framework by developing a new attack model that jointly optimizes adversarial inputs and poisoned models. We further discuss potential countermeasures against such attacks. This work is accepted by ACM CCS'20.

Mentor: Dr.Frank Ritter

Research Topics: *Human-Computer Interaction, MOOC*

- **Assessing and Improving MOOC Online Learners' Interactions**

Existing MOOC platforms suffer from interactions and collaborated learning problems. In this project, we investigated MOOC learners' experience with current popular MOOC platforms starting with a survey. We also conducted a task analysis on three popular MOOC platforms to analyze how they include key interaction functions. We propose a set of comprehensive guidelines, called IN-MOOC, to facilitate interpersonal interactions on MOOC platforms. This work is accepted by AsianCHI'21.

**Lehigh University, Research Assistant, PA, USA**

Aug 2018 - Aug 2019

Advisors: Dr. Brian D. Davison, Dr.Ting Wang

Research Topics: *Interpretability, Robustness, NLP.*

- **Analyzing QA Model Robustness by Interpreting Adversarial Attacks**

We aim to analyze robustness of Question Answering (QA) model by interpreting its adversarial attack. I designed an interpretation method for QA model to interpret layer-wise and token-wise attention-based importance scores. I analyzed SQuAD datasets (with adversaries as ADDSent, ADDAny, etc.) on QA models (BiDAF, Match-LSTM) using Pytorch. This work was presented at the CRA-W Grad Cohort Workshop'19.

## MEDIA COVERAGE

---

- **PSU News.** Jessica Hallman. October 26, 2020  
**Users don't understand computer explanations for image labeling errors.**  
[\[PSU News Link\]](#) [\[TechXplore Link\]](#) [\[ClickTech News Link\]](#)

## SERVICE

---

**PC Member:** ACL' 21 (First-time Reviewer Mentor: Dr. Marco Tulio Ribeiro)

**Conference Review:** KDD' 20,19, ICDM' 19, ASONAM' 19, CIKM' 18, BigData' 18

**Journal Review:** Transactions on Intelligent Systems and Technology (TIST)' 19, Transactions on Information Systems (TIS)' 19

## TEACHING ASSISTANT EXPERIENCE

---

**The Pennsylvania State University**

- **Application Development Studio I** (undergraduate, 50 students) Fall 2020
- **Algorithmic Methods** (undergraduate, 25 students) Fall 2020 and Spring 2021

**Renmin University of China**

- **Calculus I and Calculus II** (undergraduate, 50 students) Spring 2015 and Fall 2015
- **Linear Algebra** (undergraduate, 30 students) Fall 2014
- **Probability Theory and Mathematical Statistics** (undergraduate, 40 students) Fall 2013 and Spring 2014

## AWARDS AND HONORS

---

- Graduate Student Travel Grant**, IST, PennState. (Virtual Conference Registration) 2020
- Travel Grant Award**, CRA-W Grad Cohort Workshop. 2019
- Graduate Fellowship**, Lehigh University. 2018-2019
- Outstanding Master Degree Candidate**, RenMin University of China 2015
- Outstanding Merit Student**, Beijing Municipal. 2013
- Excellent Graduate**, Beijing Municipal. 2013
- Outstanding-Dissertation Award**, USTB. 2013
- National Endeavor Fellowship**, Nationwide(twice). 2011 and 2012
- Alumni Funding Scholarship**, USTB. 2012
- Vice Chairman**, The 22nd Student Union of USTB. 2011-2012