Shang-Tse Chen

National Taiwan University
No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan

stchen@csie.ntu.edu.tw
http://www.csie.ntu.edu.tw/~stchen

Research Interests

Machine Learning, Data Mining, Security, and Algorithmic Game Theory

Current Position

Aug. 2024 — present **Associate Professor**
Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan

Education

2013 — 2019 **Ph.D. in Computer Science**
Georgia Institute of Technology, Atlanta, GA
Thesis: AI-infused Security: Robust Defense by Bridging Theory and Practice
Committee: Polo Chau (advisor), Nina Balcan (co-advisor), Wenke Lee, Le Song, Kevin Roundy, and Cory Cornelius

2006 — 2010 **B.S. in Computer Science Information Engineering**
National Taiwan University, Taipei, Taiwan

Selected Honors and Awards

2020 ACM Trans. Interactive Intelligent Systems (TiiS) 2018 Best Paper, Honorable Mention

2018 — 2019 IBM PhD Fellowship
For my Ph.D. research on "AI-infused Security: Robust Defense by Bridging Theory and Practice"

2018 KDD'18 Audience Appreciation Award, Runner Up
For "Shield: Fast, Practical Defense and Vaccination for Deep Learning using JPEG Compression"

2016 Symantec Fellowship Runner-Up

2016 KDD'16 Best Student Paper Award, Runner-Up
For "Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta"

2010 National Science Council Research Creativity Award
For my undergraduate research on "Link Prediction in Heterogeneous Networks"

2009 KDD Cup 2009 3rd Prize (slow track)
Out of 400+ submissions
KDD CUP is the most prestigious data mining contest

Industry Research Experience

Summer 2018 **Intel Labs**, Hillsboro, OR
Graduate Machine Learning Security Intern
Mentor: Cory Cornelius, Jason Martin
Explored regularization techniques as defense against adversarial attack.

Summer 2017 **Intel Labs**, Hillsboro, OR
Graduate Security Intern
Mentor: Cory Cornelius, Jason Martin
Developed *ShapeShifter*, the **1st physical adversarial attack** that fools Faster R-CNN object detectors

Summer 2016 **Symantec Research Labs**, Culver City, CA
Research Engineer Intern
Mentor: Kevin A. Roundy
Developed **patented** *Virtual Product*, a novel framework for enterprise cyber threat detection.

Summer 2015 **Pindrop Security**, Atlanta, GA
Research Intern
Mentor: Raj Bandyopadhyay
Improved phone fraud detection system significantly by 10 absolute percentage.

Academic Research Experience

2020 — 2024 **National Taiwan University**, Taipei, Taiwan
Assistant Professor, Department of Computer Science and Information Engineering

2013 — 2019 **Georgia Institute of Technology**, Atlanta, GA
Graduate Research Assistant, School of Computational Science and Engineering
Advisors: Polo Chau and Nina Balcan

2011 — 2013 **Academia Sinica**, Taipei, Taiwan
Graduate Research Assistant, Institute of Information Science
Advisors: Chi-Jen Lu and Hsuan-Tien Lin

2008 — 2010 **National Taiwan University**, Taipei, Taiwan
Undergraduate Research Assistant, Department of Computer Science and Information Engineering
Advisors: Shou-De Lin and Hsuan-Tien Lin

Publications

Refereed Conference Papers

Trap-MID: Trapdoor-based Defense against Model Inversion Attacks
Zhen-Ting Liu and Shang-Tse Chen
To appear in the Annual Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada. Dec. 2024.

Task Arithmetic can Mitigate Synthetic-to-Real Gap in Automatic Speech Recognition
Hsuan Su, Hua Farn, Fan-Yun Sun, Shang-Tse Chen, and Hung-yi Lee
To appear in the Conference on Empirical Methods in Natural Language Processing (EMNLP), Miami, FL. Nov. 2024.

Annealing Self-Distillation Rectification Improves Adversarial Training
Yu-Yu Wu, Hung-Jui Wang, and Shang-Tse Chen
International Conference on Learning Representations (ICLR), Vienna, Austria. May 2024..

Towards Large Certified Radius in Randomized Smoothing using Quasiconcave Optimization
Bo-Han Kung and Shang-Tse Chen
Annual AAAI Conference on Artificial Intelligence (AAAI), Vancouver, Canada. Feb. 2024.

AdvCAPTCHA: Creating Usable and Secure Audio CAPTCHA with Adversarial Machine Learning
Hao-Ping (Hank) Lee, Wei-Lun Kao, Hung-Jui Wang, Ruei-Che Chang, Yi-Hao Peng, Fu-Yin Cherng, and Shang-Tse Chen
Symposium on Usable Security and Privacy (USEC), San Diego, CA. Feb. 2024.

UnMask: Adversarial Detection and Defense Through Robust Feature Alignment
Scott Freitas, Shang-Tse Chen, Zijie J. Wang, and Duen Horng Chau
IEEE International Conference on Big Data (BigData). Atlanta, GA. Dec. 2020.
Github

ShapeShifter: Robust Physical Adversarial Attack on Faster R-CNN Object Detector
Shang-Tse Chen, Cory Cornelius, Jason Martin, and Duen Horng (Polo) Chau
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD). Dublin, Ireland. Sept. 2018.
Github | Video

Shield: Fast, Practical Defense and Vaccination for Deep Learning using JPEG Compression
Nilaksh Das, Madhuri Shanbhogue, Shang-Tse Chen, Fred Hohman, Siwei Li, Li Chen, Michael E. Kounavis, and Duen Horng (Polo) Chau
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). London, UK. Aug. 2018.
Github | Video | ***Audience Appreciation Award, Runner-Up***

Diversified Strategies for Mitigating Adversarial Attacks in Multiagent Systems
Maria-Florina Balcan, Avrim Blum, and Shang-Tse Chen (alphabetic order)
International Conference on Autonomous Agents and Multiagent Systems (AAMAS). Stockholm, Sweden. July 2018.

Predicting Cyber Threats with Virtual Security Products
Shang-Tse Chen, Yufei Han, Duen Horng (Polo) Chau, Christopher Gates, Michael Hart, and Kevin Roundy

Annual Computer Security Applications Conference (ACSAC). Orlando, FL. Dec. 2017.
Patented

Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta
Michael Madaio, Shang-Tse Chen, Oliver Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng (Polo) Chau, and Bistra Dilkina.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). San Francisco, CA. Aug. 2016.
Site | Github | Video | ***Best Student Paper Award, Runner-Up***

Communication Efficient Distributed Agnostic Boosting
Shang-Tse Chen, Maria-Florina Balcan, and Duen Horng (Polo) Chau
International Conference on Artificial Intelligence and Statistics (AISTATS). Cadiz, Spain. May 2016.

Boosting with Online Binary Learners for the Multiclass Bandit Problem
Shang-Tse Chen, Hsuan-Tien Lin, and Chi-Jen Lu
International Conference on Machine Learning (ICML). Beijing, China. June 2014.

An Online Boosting Algorithm with Theoretical Justifications
Shang-Tse Chen, Hsuan-Tien Lin, and Chi-Jen Lu
International Conference on Machine Learning (ICML). Edinburgh, Scotland. June 2012.
Code

Journal Articles and Book Chapters

Ensuring Bidirectional Privacy on Wireless Split Inference Systems
Chia-Che Sa, Li-Chen Cheng, Hsing-Huan Chung, Te-Chuan Chiu, Chih-Yu Wang, Ai-Chun Pang, and Shang-Tse Chen
IEEE Wireless Communications Magazine, 2024.

Chronodes: Interactive Multi-focus Exploration of Event Sequences
Peter J. Polack, Shang-Tse Chen, Minsuk Kahng, Kaya De Barbaro, Rahul Basole, Moushumi Sharmin, and Duen Horng (Polo) Chau
ACM Transactions on Interactive Intelligent Systems (TiiS) Special Issue on Interactive Visual Analysis of Human and Crowd Behaviors, 2018.
Video | ***Best Paper, Honorable Mention***

Exploratory Visual Analytics of Mobile Health Data: Sensemaking Challenges and Opportunities
Peter J. Polack, Moushumi Sharmin, Kaya de Barbaro, Minsuk Kahng, Shang-Tse Chen, and
Duen Horng (Polo) Chau
Mobile Health: Sensors, Analytic Methods, and Applications. Springer, 2017.

Constructing, Analyzing and Visualizing Social Networks: Exemplified by the Academia Social Network in Taiwan
Cheng-Te Li, Chun-Min Chang, Chien-Pang Liu, Shang-Tse Chen, and Shou-De Lin
Journal of Librarianship and Information Studies, 67:72-87, 2008.

Refereed Workshop, Poster, and Demo papers

Enhancing Targeted Attack Transferability via Diversified Weight Pruning
Hung-Jui Wang, Yu-Yu Wu, and Shang-Tse Chen
CVPR Workshop on Adversarial Machine Learning on Computer Vision: Robustness of Foundation Models (AdvML), Seattle, WA, June 2024.

Fair Robust Active Learning by Joint Inconsistency
Tsung-Han Wu, Hung-Ting Su, Shang-Tse Chen, and Winston H. Hsu
ICCV Workshop on Adversarial Robustness In the Real World (AROW). Paris, France. Oct. 2023.

Position Matters! Empirical Study of Order Effect in Knowledge-grounded Dialogue
Hsuan Su, Shachi H Kumar, Sahisnu Mazumder, Wenda Chen, Ramesh Manuvinakurike, Eda Okur, Saurav Sahay, Lama Nachman, Shang-Tse Chen, and Hung-yi Lee
ACL DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering. Toronto, Canada. July. 2023.

Extracting Knowledge For Adversarial Detection and Defense in Deep Learning
Scott Freitas, Shang-Tse Chen, and Duen Horng Chau
KDD Workshop on Learning and Mining for Cybersecurity (LEMINCS). Anchorage, AK, Aug. 2019.

Talk Proposal: Towards the Realistic Evaluation of Evasion Attacks using CARLA
Cory Cornelius, Shang-Tse Chen, Jason Martin, and Duen Horng Chau
Dependable and Secure Machine Learning (DSML). Portland, OR, June 2019.

ADAGIO: Interactive Experimentation with Adversarial Attack and Defense for Audio
Nilaksh Das, Madhuri Shanbhogue, Shang-Tse Chen, Li Chen, Michael E. Kounavis, and Duen Horng (Polo) Chau
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) (demo). Dublin, Ireland. Sept. 2018.
Video

Physical Adversarial Attack on Object Detectors
Shang-Tse Chen, Cory Cornelius, Jason Martin, and Duen Horng (Polo) Chau
ACM KDD Project Showcase. London, UK. Aug. 2018.

Compression to the Rescue: Defending from Adversarial Attacks Across Modalities
Nilaksh Das, Madhuri Shanbhogue, Shang-Tse Chen, Fred Hohman, Siwei Li, Li Chen, Michael E. Kounavis, and Duen Horng (Polo) Chau
ACM KDD Project Showcase. London, UK. Aug. 2018.

TimeStitch: Interactive Multi-focus Cohort Discovery and Comparison
Peter J. Polack, Shang-Tse Chen, Minsuk Kahng, Moushumi Sharmin, and Duen Horng (Polo) Chau
IEEE Conference on Visual Analytics Science and Technology (VAST'15) (Poster). Chicago, IL,

Oct. 2015.
Video

Spotting Suspicious Reviews via (Quasi-)clique Extraction
Paras Jain, Shang-Tse Chen, Mozhgan Azimpourkivi, Duen Horng (Polo) Chau, and Bogdan Carbunar
IEEE Symposium on Security and Privacy (Oakland) (poster). SAN JOSE, CA. May 2015.

An Ensemble of Three Classifiers for KDD Cup 2009: Expanded Linear Model, Heterogeneous Boosting, and Selective Naive Bayes
Hung-Yi Lo, Kai-Wei Chang, Shang-Tse Chen, Tsung-Hsien Chiang, Chun-Sung Ferng, Cho-Jui Hsieh, Yi-Kuang Ko, Tsung-Ting Kuo, Hung-Che Lai, Ken-Yi Lin, Chia-Hsuan Wang, Hsiang-Fu Yu, Chih-Jen Lin, Hsuan-Tien Lin, and
Shou-de Lin
JMLR Workshop and Conference Proceedings, V.7, 57-64, 2009.
3rd Place of the KDD Cup'09 Slow Track

Technical Reports

Learning to Generate Prompts for Dialogue Generation through Reinforcement Learning
Hsuan Su, Pohan Chi, Shih-Cheng Huang, Chung Ho Lam, Saurav Sahay, Shang-Tse Chen, and Hung-yi Lee
arXiv:2302.05888, Feb. 2023.

Keeping the Bad Guys Out: Protecting and Vaccinating Deep Learning with JPEG Compression
Nilaksh Das, Madhuri Shanbhogue, Shang-Tse Chen, Fred Hohman, Li Chen, Michael E. Kounavis, and Duen Horng (Polo) Chau
arXiv:1705.02900, May 2017.

An Ensemble Ranking Solution to the Yahoo! Learning to Rank Challenge
Ming-Feng Tsai, Shang-Tse Chen, Yao-Nan Chen, Chun-Sung Ferng, Chia-Hsuan Wang, Tzay-Yeu Wen, and Hsuan-Tien Lin
National Taiwan University, Technical Report, Sept. 2010.

Teaching

Courses

Spring 2024 **Security and Privacy of Machine Learning**

Spring 2022 **Foundations of Artificial Intelligence**

Spring 2022 **Introduction to Medical Informatics**

Fall 2021 **Security and Privacy of Machine Learning**

Spring 2021 **Introduction to Medical Informatics**

Fall 2020 **Security and Privacy of Machine Learning**

Spring 2020 **Security and Privacy of Machine Learning**

Professional Activities

Area Chair

International Conference on Learning Representations (**ICLR**) 2025

AAAI Conference on Artificial Intelligence (**AAAI**) 2023 - 2024

Program Committee

International Conference on Machine Learning (**ICML**) 2018 - 2024

Annual Conference on Neural Information Processing Systems (**NIPS**) 2017 - 2024

AAAI Conference on Artificial Intelligence (**AAAI**) 2018 - 2022

Uncertainty in Artificial Intelligence (**UAI**) 2015 - 2019

SIAM International Conference on Data Mining (**SDM**) 2019

Asian Conference on Machine Learning (**ACML**) 2017 - 2019

Conference on Technologies and Applications of Artificial Intelligence (**TAAI**) 2014

Deep Learning and Security Workshop @ IEEE S&P (**DLS**) 2019

Reviewer

International Conference on Artificial Intelligence and Statistics (**AISTATS**) 2019

Deep Learning and Security Workshop @ IEEE S&P (**DLS**) 2018

SIAM International Conference on Data Mining (**SDM**) 2016 - 2017

ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**) 2017

Annual Network and Distributed System Security Symposium (**NDSS**) 2017

USENIX Security Symposium (**USENIX Security**) 2017

ACM Conference on Computer and Communications Security (**CCS**) 2017