# Tong Zhou

✉ zhou.tong1@northeastern.edu | ☎ (734)-358-1431
**Google Scholar** | **LinkedIn** | **Homepage**

## Education

**Northeastern University, Boston, MA, USA** Sep. 2021 – present
Ph.D. in Electrical & Computer Engineering

**University of Michigan, Ann Arbor, MI, USA** Sep. 2019 – May 2021
M.S. in Electrical & Computer Engineering **GPA: 3.81/4.0**

**Xidian University, Xi'an, Shaanxi, China** Sep. 2015 – Jul. 2019
B.S. in Electrical Engineering **GPA: 3.80/4.0**

## Research Interests

**Trustworthy AI · Efficient ML · Generative AI · Privacy**

## Research Experience

**Research Assistant @ Xiaolin Xu's Lab** Sep. 2021 – present
***Advisor:* *Prof. Xiaolin Xu*** *Northeastern University*
Developing efficient and secure frameworks for machine learning models, while also exploring innovative solutions to address vulnerabilities in these models. My goal is to advance the field of efficient and trustworthy AI.

Selected Projects

1. **Plant Unforgeable Watermarks for Large Language Model for Reliable Detection (NeurIPS'24)**
   - This project addresses the urgent need for regulatory measures in response to the increasing misuse of advanced generative models, specifically focusing on LLM. With a focus on identifying the origin of generated content, our proposed framework ensures public and reliable detection of watermarks, immune to forging attempts by malicious parties.
2. **Adaptive DNN Models for Efficient Private Inference in Edge Computing (ICCAD'24)**
   - Designed weight-level and feature-level soft masks to achieve weight pruning and non-linear activation function pruning.
   - Transformed soft masks into multiple binary masks, enabling adaptive adjustment of communication and computation workloads based on specific energy budgets.
   - Achieved optimal accuracy through a progressive training approach with increasingly dense binary masks, enabling models to adapt to diverse resource constraints by applying different masks.
3. **Restrict Unauthorized Model Transfers at the Architecture Level (ICLR'24)**
   - Introduced an architecture-level defense against unauthorized transfers, ensuring optimal performance on source tasks while degrading performance on unauthorized tasks, regardless of attacker data access.
   - Developed a zero-cost proxy-based binary predictor to accelerate Neural Architecture Search (NAS), incorporating task characteristics for efficient architecture assessment and enabling cross-task search with rank-based fitness scoring.
4. **Accelerate Private Inference via Automatic ReLU Pruning (ICCV'23)**
   - Tackled challenges associated with private inference techniques employing cryptographic primitives, where elevated computation and communication costs, especially with non-linear operators like ReLU, posed significant obstacles.
   - Engineered a parameterized discrete indicator function to achieve precise ReLU pruning, effectively mitigating the impact of non-linear operators. Subsequently, replaced ReLU with its polynomial approximation to uphold high model accuracy.

**Research Assistant @ Jiande Chen's Lab** Nov. 2020 – Apr. 2021
***Advisor: Prof. Jiande Chen*** *University of Michigan*
Developed deep learning models for feature extraction from electrocardiogram data to detect food intake phases, aiming to assist in treating obesity and diabetes.

**Research Assistant @ Laboratory of Integrated Brain Imaging** May 2020 – Oct. 2020
***Advisor: Prof. Zhongming Liu*** *University of Michigan*
Enhanced segmentation performance for Transmission Electron Microscopy (TEM) images by integrating a self-attention mechanism into the U-Net architecture.

## Selected Publications (*indicates equal contribution)

- Bileve: Securing Text Provenance in Large Language Models Against Spoofing with Bi-level Signature
  **Tong Zhou**, Xuandong Zhao, Xiaolin Xu, and Shaolei Ren
  The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS), 2024.
- ArchLock: Locking DNN Transferability at the Architecture Level with a Zero-Cost Binary Predictor
  **Tong Zhou**, Shaolei Ren, and Xiaolin Xu
  The Twelfth International Conference on Learning Representations (ICLR), 2024.
- AutoReP: Automatic ReLU Replacement for Fast Private Network Inference
  **Tong Zhou***, Hongwu Peng*, Shaoyi Huang*, Yukui Luo, Xiaolin Xu, Caiwen Ding, *et al.*
  International Conference on Computer Vision (ICCV), 2023.
- NNSplitter: An Active Defense Solution to DNN Model via Automated Weight Obfuscation
  **Tong Zhou**, Yukui Luo, Shaolei Ren, Xiaolin Xu
  International Conference on Machine Learning (ICML), 2023.
- ObfuNAS: A Neural Architecture Search-based DNN Obfuscation Approach
  **Tong Zhou**, Shaolei Ren, Xiaolin Xu
  IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2022.
  Best Paper Nomination

## Work Experience

**Applied Scientist Intern @ Amazon** May 2024 – Aug. 2024
***Mentors: Xiang Cui and Guannan Liang***

This project aims to develop a unified model to improve account takeover detection by leveraging multiple data sources.

- Generated and engineered sequence, categorical, and numerical features from click data, introducing learnable feature importance to prioritize key features to better learn fraud patterns.
- Designed and implemented a Unified Multi-Modality Transformer with a Multi-Source Cross-Attention Mechanism, enabling the model to handle diverse features seamlessly and capture dependencies across multiple data tables without requiring structural changes.
- Boosted model performance under a multi-task setting by integrating an additional tag source.

## Technical Skills

**Programming**: Python, MATLAB, C, Julia
**Frameworks & Others**: PyTorch, TensorFlow, PySpark, Pandas, Scikit-learn, OpenCV

## Selected Awards

| | |
|---|---|
| **NeurIPS Scholar Award** | 2024 |
| **ICML Travel Grant** | 2023 |
| **COE Outstanding Graduate Student Award**, Northeastern University | 2023 |
| **IEEE/ACM William J. McCalla ICCAD Best Paper Nomination** | 2022 |
| **COE Dean's Fellowship Award**, Northeastern University | 2021 |
| **Outstanding Graduate Award (top 5%)**, Xidian University | 2019 |
| **First Prize Scholarship (top 3%)**, Xidian University | 2016 - 2018 |

## Professional Service

**Volunteer:** ICML 2023, New England Hardware Security Workshop 2023
**Conference Reviewer:** ICLR 2025, AISTATS 2025, NeurIPS 2024, HOST 2023, ICCD 2022
**Journal Reviewer:** Transactions on Information Forensics & Security, IEEE Systems Journal