Soohyeon Choi

Department of Computer Science, University of Central Florida
12805 Pegasus Dr, HPA1-111, Orlando, FL, 32816, USA
Email: soohyeon.choi@ucf.edu
Linkedin: linkedin.com/in/soohyeon-choi

Summary

I am an experienced researcher in the field of Computer Science with a focus on Machine Learning, AI, and Security. I have an extensive background in code authorship attribution, programming language processing, and robustness measurement of Machine Learning applications. Holding a Ph.D. in Computer Science from the University of Central Florida, I served as a Research Officer at A*STAR Research Entities, where I continue to contribute to the field through my research work and publications.

Education

PH.D., Computer Science, University of Central Florida, Orlando, FL, USA (2021 – Current)
Advisor: Prof. David Mohaisen. Topic: Machine Learning, AI, LLM & Security
M.SC., Computer Science, South Dakota State Unversity, Brookings, SD, USA (2018 – 2021)
Advisor: Prof. Sung Shin. Topic: Sensor Network, Authentication & Security
B.SC., Computer Engineering, Keimyung University – Deagu, South Korea (2011 – 2017)

Research Interests

Software Engineering, Machine Learning, AI, Security, Code Authorship Attribution, Programming Language Processing, Large Language Models, and Robustness Measurement of Machine Learning Applications.

Professional Appointments

08/2021 – Current	Research Assistant	University of Central Florida	Machine Learning Security
12/2023 – 12/2024	Research Officer	A*STAR Research Entities (I2R)	Machine Learning Security
08/2019 – 12/2020	Research Assistant	South Dakota State University	Sensor Network Security
08/2018 – 05/2019	Teaching Assistant	South Dakota State University	CSc 300 Data Structures

Technical Publications and Manuscripts

1. **Soohyeon Choi**, Tan Yong Kiam, Mark Huasong Meng, Mohamed Ragab, Soumik Mondal, Khin Mi Mi Aung, and David Mohaisen, "**I Can Find You in Seconds! Leveraging Large Language Models for Code Authorship Attribution**", IEEE/ACM International Conference on Automated Software Engineering (ASE), 2025 (plan to submit).

2. **Soohyeon Choi**, Ali Al Kinoon, Aziz Alghamdi, Ahod Alguried and David Mohaisen, "**Attributing ChatGPT-Transformed Synthetic Code**", IEEE International Conference on Distributed Computing Systems (ICDCS), 2025 (under review).

3. Aziz Alghamdi, **Soohyeon Choi**, Ali Al Kinoon, Ahod Alguried and David Mohaisen, "**Unified API Call-based Detection of Android and IoT Malware**", IEEE International Conference on Distributed Computing Systems (ICDCS), 2025 (under review).

4. Ali Al Kinoon, **Soohyeon Choi**, Aziz Alghamdi, Ahod Alguried and David Mohaisen, "**Pandora's Box Reopened: New Insights into Android Permissions**", IEEE International Conference on Distributed Computing Systems (ICDCS), 2025 (under review).

5. Ahod Alguried, **Soohyeon Choi**, Ali Al Kinoon, Aziz Alghamdi and David Mohaisen, "**Fishing for Phishers: Learning-Based Phishing Detection in Ethereum Transactions**", IEEE International Conference on Distributed Computing Systems (ICDCS), 2025 (under review).

6. **Soohyeon Choi**, Rhongho Jang, Daehun Nyang and David Mohaisen, "**Untargeted Code Authorship Evasion with Seq2Seq-based Code Transformation**", IEEE Transactions on Dependable and Secure Computing (TDSC), 2025 (under review).

7. **Soohyeon Choi** and David Mohaisen, "**Attributing ChatGPT-generated Source Codes**", IEEE Transactions on Dependable and Secure Computing (TDSC), 2024.

8. **Soohyeon Choi**, Rhongho Jang, Daehun Nyang and David Mohaisen, "**Untargeted Code Authorship Evasion with Seq2Seq Transformation**", Computational Data and Social Networks: 12th International Conference (CSoNet), 2023.

9. **Soohyeon Choi**, Manar Mohaisen, Daehun Nyang and David Mohaisen, "**Revisiting the Deep Learning-based Eavesdropping Attacks via Facial Dynamics from VR Motion Sensors**", International Conference on Information and Communications Security (ICICS), 2023.

10. **Soohyeon Choi**, Malwan Omar, Daehun Nyang and David Mohaisen, "**Quantifying the performance of adversarial training on language models with distribution shifts**", Cybersecurity and Social Sciences (CySSS), 2022.

11. Abdulrahman Alabduljabbar, **Soohyeon Choi**, Runyu Ma, Rhongho Jang, Songqing Chen, and David Mohaisen, "**Understanding the Security of Free Content Websites by Analyzing their SSL Certificates: A Comparative Study**", Cybersecurity and Social Sciences (CySSS), 2022.

12. Malwan Omar, **Soohyeon Choi**, Daehun Nyang and David Mohaisen, "**Robust Natural Language Processing: Recent Advances, Challenges, and Future Directions**", IEEE Access, 2022.

13. **Soohyeon Choi**, "**Improved secure and low computation authentication protocol for wireless body area network with ecc and 2d hash chain**", M.Sc. thesis, South Dakota State University, 2021.

14. Xiaozhu Jin, **Soohyeon Choi**, Sangwon Shin, and Sung Shin, "**Human activity recognition based on wearable flex sensor and pulse sensor**", Asia Pacific International Conference on Information Science and Technology (APIC-IST), 2020.

15. **Soohyeon Choi**, Xiaozhu Jin, Sangwon Shin, and Sung Shin, "**Secure and low computation authentication protocol for wireless body area network with ecc and 2d hash chain**", The ACM International Conference on Research in Adaptive and Convergent Systems (RACS), 2020.

16. Chungyup Lee, **Soohyeon Choi**, Jung Y. Kim, and Kwanghee Won, "**Instance segmentation in urban scenes using inverse perspective mapping of 3D point clouds and 2D images**", The ACM International Conference on Research in Adaptive and Convergent Systems (RACS), 2019.

17. Sangwon Shin, **Soohyeon Choi**, Kwanghee Won, and Sung Shin, "**Preprocessed symmetric rsa authentication for wireless body area networks in space**", The ACM International Conference on Research in Adaptive and Convergent Systems (RACS), 2019.

Representative research project

1. **Code Authorship with LLMs**: This project aims to explore the capabilities of Large Language Models (LLMs) as tools for code authorship attribution. Specifically, we seek to leverage the advancements of LLMs in natural language processing to tackle challenges in identifying the authors of source code, even across different programming languages and styles. We also investigate the ability of LLMs to attribute code authorship using zero-shot and few-shot prompting without relying on extensive labeled datasets, demonstrating promising results in both zero-shot code verification and few-shot in-context learning. The project also examines the robustness of LLMs against adversarial attacks and their generalization across languages.

2. **ChatGPT Code Transformation**: This project aims to explore ChatGPT's programming capabilities, acknowledging its appeal as a programming tool while emphasizing potential ethical and security risks. We investigate ChatGPT's code transformation ability and assess the effectiveness of a dedicated authorship attribution technique for its generated code. The experiments reveal that ChatGPT can transform code styles, posing challenges to existing attribution methods. Feature-based attribution proves effective even on ChatGPT-transformed code, while a binary classification model remains accurate, achieving up to high accuracy. These findings contribute insights into ChatGPT's code transformation and highlight the efficacy of certain attribution techniques for its generated code. The outcomes of this project are going to be submitted to AAAI 2023.

3. **ChatGPT Code Authorship**: This project aims to identify an AI assistant (ChatGPT) to prevent unethical and security issues. The vast amount of data AI assistants are trained on may cause AI-generated source codes to have diverse styles, even when generated in response to the same question. As a result, current approaches to code authorship attribution are ineffective as they heavily rely on stylistic analysis. Toward this, a feature-based approach is proposed to classify AI-generated codes successfully. The proposed approach has achieved a classification accuracy of over 85%, which is significantly higher than the accuracy of the naive approach of around 8%. The outcomes of this project are being submitted to IEEE TDSC 2023.

4. **Code Authorship Evasion**: This project aims to generate transformed codes from the original codes, which can deceive code authorship attribution models with a machine learning-based Sequence to Sequence (Seq2Seq) model. The existing approach, which uses Monte-Carlo Tree Search (MCTS), is disadvantaged since it requires expensive resource usage and processing time due to the nature of MCTS. The machine learning-based Seq2Seq model could perform the same code transformation while requiring fewer resources and time. As a result, the seq2Seq model (known as StructCoder) achieved 85% of the transform success rate while preserving the semantics and syntax of the original code. The outcomes of this project is being submitted to IEEE TDSC 2023.

5. **Robustness Measurement**: This project aims to measure the robustness of a user classification model which uses speech-associated facial dynamics captured by motion sensors in VR headsets under variations of users' ethnicity/race and gender. When an attacker and victim share the same ethnicity/race and gender, it becomes easier to deceive the classification model as they are likely to have similar face shape features, which are heavily influenced by ethnicity/race and gender, compared to when they have different ethnicity/race and gender. Under the scenario of the same features, it is easier to deceive the user classification model than in the case of different features. The outcomes of this project are submitted to ICICS 2023.

Skills

1. Well experienced - Python, C/C++, Java, Latex
2. Some experienced - Linux, Unity, SQL
3. Basic experienced - JSP, JavaScript, HTML, C#

Services and Activities

- External reviewer for IEEE Transactions on Dependable and Secure Computing (2025)
- External reviewer for IEEE Transactions on Dependable and Secure Computing (2024)
- Conference web chair for IEEE Conference on Communications and Network Security (2023)
- External reviewer for IEEE Transactions on Dependable and Secure Computing (2023)
- External reviewer for IEEE Transactions on Mobile Computing (2022)
- External reviewer for Computer Networks (2022)

AWARDS

- Keimyung University Alumni Scholarship, Keimyung University (July. 2020).
- UPE Academic Achievement Award, Upsilon Pi Epsilon International Honor Society (Sep. 2019).