

MIN-CHUN (MICHAEL) CHEN

(425)588-8312 ◇ hiiammichaelo520@gmail.com ◇ LinkedIn ◇ Personal Site ◇ GitHub

RESEARCH INTERESTS

Security of agentic AI systems: prompt- and skill-level backdoors, and the limits of static auditing for LLM agents.
Related interests in machine learning privacy (membership inference) and machine unlearning.

EDUCATION

University of Maryland, Baltimore County, Baltimore, MD 2026/02 - Present

Ph.D. in Computer Science

Advisor: Prof. Keke Chen, Trustworthy and Intelligent Computing Lab (TAIC)

University of Maryland, Baltimore County, Baltimore, MD 2024/08 - 2026/12 (expected)

M.S. in Computer Science

National Taiwan University, Taipei, Taiwan 2019/09 - 2023/06

B.S. in Information Management

PUBLICATIONS

Peer-Reviewed

[1] J. He, **MC. Chen**, X. Chen, X. Fang, Y. Gu, K. Chen. *Membership Inference Attacks on In-Context Examples in LLM-based Recommender Systems*. 20th ACM Conference on Recommender Systems (RecSys 2026), main track.

Under Review

[2] N. Chetlapalli*, Y. Liao*, **MC. Chen***, K. Chen. *Diverse by Reasoning: Harnessing the Wisdom of LLM Crowds for Future Prediction*. IEEE International Conference on Big Data (BigData 2026).

[3] Q. Yang, Y. Gu, J. He, **MC. Chen**, K. Chen. *Privacy Déjà Vu: Re-emergence of Sensitive Samples in Continual Fine-tuning*. In revision, AAAI 2027.

*Equal contribution.

PRESENTATIONS

MC. Chen (speaker). *Membership Inference Attacks on In-Context Examples in LLM-based Recommender Systems*. Oral presentation, ACM RecSys 2026, Minneapolis, MN, Sept 29 - Oct 1, 2026 (forthcoming).

MC. Chen. Poster presentation of TAIC lab research, NAIRR 2026 Annual Meeting.

RESEARCH EXPERIENCE

Trustworthy and Intelligent Computing Lab (TAIC), UMBC — *Graduate Research Assistant* 2025/06 - Present

- **Dissertation work** — **security of agent skills**. Studying third-party natural-language instruction packages that execute with operator-level authority in agentic systems, and the failure of artifact-scoped auditing to secure them
- Extended the **AgentDojo** benchmark with a skill-execution capability and built an evaluation harness around a self-hosted **Llama-3.1-8B-Instruct** backbone
- Designed a conditional-trigger backdoor attack in which the malicious artifact and its activating condition arrive through separate channels, and benchmarked it against an **InjecAgent** baseline
- Developing an out-of-band intent-integrity checker that evaluates agent trajectories against observable substrate facts rather than model-reported justifications

- Co-designed the experiments and defense evaluation for a membership-inference study on in-context examples in LLM recommender systems (RecSys 2026)

TEACHING EXPERIENCE

CMSC 421 — Principles of Operating Systems, UMBC — *Grader*

Spring 2025, Fall 2025

- Replaced the course's manual grading workflow with an automated grading script that extracts the relevant sections of each submission, cutting turnaround time and standardizing scoring across graders
- Returned the script's per-submission logs to students, letting them verify their own scores and reducing regrade requests
- Graded projects and exams for 40+ students per semester, writing rubrics and per-student feedback

INDUSTRY EXPERIENCE

CHT Security — *Software Engineering Intern*

2022/07 - 2023/01

- Built an infrastructure-automation tool managing firewall policies across **Google Cloud Platform** with **Terraform**, replacing manual console changes with reviewable code
- Integrated **GitHub Actions** CI/CD to detect policy changes in JSON and apply them to GCP automatically
- Automated periodic drift detection across the GCP environment in **Python**, alerting the team via the Slack API

TECHNICAL SKILLS

Languages Python, C++, JavaScript, Bash, SQL

ML PyTorch, Hugging Face Transformers

Cloud & Infra Google Cloud Platform, Terraform, Docker, GitHub Actions, Git, Linux

SELECTED PROJECTS

Matchmaking System with Graph Neural Networks

2025/03 - 2025/05

- Modeled player synergy for team-based games with **GNNs** over a player-interaction graph built from third-party API data

Real-Time Surgery Auxiliary System

2022/02 - 2022/12

- Trained a **real-time object detection** model to track surgical instruments and organs during laparoscopic surgery; shipped as an iOS app in **SwiftUI**