Hyejun (June) Jeong

Amherst, MA +1 (413) 824-1648 hjeong@umass.edu
hyejunjeong.github.io linkedin.com/in/june-jeong

Research Interests

I study **security and privacy in AI systems**, including LLMs and autonomous AI agents. My current work focuses on identifying and mitigating threats to AI agents, such as vulnerabilities in their interaction pipelines, missing security properties, and **risks of persuasion or persona manipulation**. I have also conducted research on LLMs and Federated Learning (FL), with emphasis on fairness, bias similarity, and unlearning. More broadly, I am interested in **trustworthy and responsible AI** and in developing **privacy-preserving methods** for collaborative and agent-based learning systems.

Publications & Presentations

Peer-Reviewed

- **H. Jeong**, H. Son, S. Lee, J. Hyun, T.-M. Chung. "FedCC: Robust Federated Learning Against Model Poisoning Attacks." *SecureComm*, 2025. [Paper] [Code] [Slides]
- **H. Jeong**, T.-M. Chung. "Security and Privacy Issues and Solutions in Federated Learning for Digital Healthcare." *Future Data and Security Engineering (FDSE)*, 2022. [Paper]
- J.H. Yoo, **H. Jeong**, J. Lee, T.-M. Chung. "Open Problems in Medical Federated Learning." *International Journal of Web Information Systems (IJWIS)*, 2022. [Paper]
- J.H. Yoo, **H. Jeong** (co-first), J. Lee, T.-M. Chung. "Federated Learning: Issues in Medical Application." *FDSE*, 2021. [Paper]
- **H. Jeong**, J. An, J. Jeong. "Are You a Good Client? Client Classification in Federated Learning." *ICT Convergence (ICTC)*, 2020. [Paper] [Code]
- J.H. Yoo, H.M. Son, **H. Jeong**, et al. "Personalized Federated Learning with Clustering: Non-IID HRV Data." *ICTC*, 2020. [Paper]

Preprints / Under Review

- **H. Jeong**, M. Teymoorianfard, A. Kumar, A. Houmansadr, E. Bagdasarian. "Network-Level Prompt and Trait Leakage in Local Research Agents." *arXiv:2508.20282*, under review (USENIX 2026). [Paper] [Code] [Dataset]
- **H. Jeong**, S. Ma, A. Houmansadr. "Bias Similarity Measurement: A Black-Box Audit of Fairness Across 30 LLMs." *arXiv:2410.12010*, under review (ICLR 2026). [Paper] [Code]
- **H. Jeong**, S. Ma, A. Houmansadr. "SoK: Challenges and Opportunities in Federated Unlearning." Preprint, under review (IEEE Big Data 2025). [Paper][Slides] (NESD 2024, UConn)

Patent

- T.-M. Chung, J.H. Yoo, **H. Jeong**, H.J. Jeon. "Data Processing Method for Depressive Disorder Using AI Based on Multi-indicator." Patent No. 1024322750000.

Research Experience

Research Assistant, UMass Amherst 2023–Present

- Investigated security of AI agents; designed attacks to infer user prompts and persona traits from browsing traces, and released supporting datasets and tools.
- Developed cross-family bias comparison pipelines across 30+ LLMs; led multiple first-author manuscripts on fairness and bias similarity.
- Initiated and led a systematization-of-knowledge (SoK) project framing challenges and opportunities in federated unlearning.

Research Assistant, SKKU 2021–2023

- Studied defenses against backdoor and poisoning attacks in federated learning.
- Conducted research on privacy-preserving medical federated learning; co-authored several peer-reviewed publications.

Undergraduate Research Assistant, SBU 2019

- Aided in building a detection pipeline for GPS spoofing using a sensor and a camera.
- Implemented and validated the system through empirical testing and analysis.

Selected Projects

Exploring Model Inversion on Unlearned Samples 2024

Explored whether image samples removed through unlearning could be reconstructed by contrasting representations between original and unlearned models.

Federated Unlearning as Backdoor Mitigation 2023

Investigated unlearning defenses against backdoor attacks in FL. Led literature review, implemented experiments, and authored manuscript. [Code]

Malicious Client Detection in Federated Learning 2022

Proposed client classification method using model weight heatmaps to detect backdoors/data poisoning. Sole author of design, implementation, and write-up. [Code]

Covert C&C and Data Exfiltration 2020

Developed Python client/server for covert command-and-control and encrypted data exfiltration to an attacker-controlled AWS server. [Code]

Distributed Typosquatting Detector 2019

Built an application to detect typosquatting domains via headless Chrome scanning and automated reporting before the user is directed to the site. [Code]

Service & Affiliations

- **Ph.D. Mentor**, UMass Amherst Summer 2025
  Mentored undergraduates in an 11-week project on AI web agent security; guided research design, experimentation, and poster preparation [Poster].
- **Undergraduate Research Volunteer Program (URV) Mentor**, UMass Amherst 2023–2024
  Supervised undergraduates in semester-long URV projects. Supported research planning, experiments, and poster presentations at the URV Showcase.
- **Reviewer**, *IEEE Transactions on Information Forensics & Security (TIFS)* 2024-
- **Member**, UMass Amherst AI Security (AISEC) Lab 2025-
- **Member**, The Secure, Private Internet (SPIN) Research Group 2023-

Education

University of Massachusetts Amherst (UMass Amherst) Exp. 2027
Ph.D. in Computer Science Advisor: Amir Houmansadr, Eugene Bagdasaryan

SungKyunKwan University (SKKU), South Korea 2023
M.S. in Computer Science Advisor: Tai-Myoung Chung, GPA: 4.5/4.5

Stony Brook University (SBU) 2020
B.S. in Computer Science Security & Privacy Specialization, Dean's List (5x)

Teaching Experience

Teaching Assistant, CS 690: Trustworthy & Responsible AI Fall 2025
UMass Amherst. Organizing and grading group assignments, assisting with paper discussions, and mentoring teams on programming assignments and an AI security-focused final project.

Teaching Assistant, CS 360: Introduction to Computer & Network Security Spring 2025
UMass Amherst. Assisted with lectures; designed and graded weekly assignments (SHA-256 password cracking, web security, AI security); held office hours; and advised semester projects (proposal, experiments, and a research-style final report).

Tutor, KT Corp. Aivle School Feb–May 2022
South Korea. Tutored in AI model interpretation and CS fundamentals; supported projects in ML/DL, NLP, and web app development with Django.

Teaching Assistant, Global Capstone Design Course. Spring 2022
SKKU. Guided teams through ideation → prototyping → evaluation; projects applied AI techniques to build deployable products.

Undergraduate Teaching Assistant, Web Design and Programming. Spring 2018
SBU. Guided web design wireframing and documentation across SDLC phases; graded assignments and held recitation sections.

Honors & Awards

Dean's List, Stony Brook University (5 semesters)
Graduate Research Assistantship, UMass Amherst (2023–Present)

Technical Skills

Languages: Python, Java, C, LaTeX, JavaScript, PHP, SQL, R
Frameworks/Tools: PyTorch, TensorFlow, Django, Git, Docker
Areas: Security & Privacy, Federated Learning, LLMs, Unlearning, Deep Learning

Updated September 12, 2025