James Flemings

Email: jamesf17@usc.edu
Personal Website: https://james-flemings.github.io
Google Scholar: https://scholar.google.com/citations?user=V5-ATAYAAAAJ&hl=en

RESEARCH INTERESTS

My research broadly investigates privacy in language models. In particular, I'm interested in (1) principally understanding and measuring the privacy leakage of language models and (2) algorithmically controlling the privacy leakage of language models.

EDUCATION

Ph.D. Computer Science August 2022 – Current
University of Southern California
GPA: 3.83
Advisor: Murali Annavaram

B.S. Computer Science, Mathematics August 2017 – May 2022
Minor: Computer Systems Engineering
University of Alaska Anchorage
GPA: 3.94

AWARDS

- NSF Graduate Research Fellowship April 2023
- USC-Meta Center Top up Fellowship August 2022
- Google CS Research Mentorship Program (CSRMP) Scholar September 2021

RESEARCH EXPERIENCE

Student Researcher June 2025 – October 2025
Google
Mentor: Ren Yi; Federated Learning and Analytics Team
Topic: Privacy-conscious agentic language models

Research Scientist Intern May 2024 – August 2024
TikTok
Mentor: Zafar Takhirov; Privacy Innovation Lab
Topic: Characterizing context privacy and hallucination in language models

Center for the Study of Language and Information Program June 2022 – August 2022
Stanford University
Mentor: Christopher Potts
Topic: Building robust and interpretable AI with Interchange Intervention Training

Research Experiences for Undergraduates in Software Engineering June 2021 – August 2021
Carnegie Mellon University
Mentor: Heather Miller; Composable Systems Lab
Topic: Developing a novel testing suite to benchmark Federated Learning algorithms

PUBLICATIONS

1. **J. Flemings**, H. Gan, H. Li, M. Razaviyayn, M. Annavaram, "Differentially Private In-context Learning via Sampling Few-shot Mixed with Zero-shot Outputs," 2025. Under Review.

2. A. Mulrooney, D. Gupta, **J. Flemings**, H. Zhang, M. Annavaram, M. Razaviyayn, X. Zhang, "DP-GRAPE: Memory-Efficient Differentially Private Training with Gradient Random Projection," 2025, Under Review.

3. **J. Flemings**, W. Zhang, B. Jiang, Z. Takhirov, M. Annavaram, "Estimating Privacy Leakage of Augmented Contextual Knowledge in Language Models," In *Proceedings of the 2025 Conference of the Association for Computational Linguistics*, 2025

4. **J. Flemings**, M. Annavaram, "Differentially Private Knowledge Distillation via Synthetic Text Generation," In *Findings of the 2024 Conference of the Association for Computational Linguistics*, 2024.

5. **J. Flemings**, M. Razaviyayn, M. Annavaram, "Differentially Private Next-Token Prediction of Large Language Models," In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics*, 2024.

6. **J. Flemings**, W. Zhang, B. Jiang, Z. Takhirov, M. Annavaram, "Characterizing Context Memorization and Hallucination in Summarization," In *Towards Safe & Trustworthy Agents at Neurips*, 2024.

7. **J. Flemings**, M. Razaviyayn, M. Annavaram, "Adaptively Private Next-Token Prediction of Large Language Models," 2024, Under Review.

8. **J. Flemings**, M. Annavaram, "Differentially Private Knowledge Distillation via Synthetic Text Generation," In *PrivateNLP at ACL*, 2024.

9. **J. Flemings**, M. Razaviyayn, M. Annavaram, "Differentially Private Prediction of Large Language Models," In *The 5th Privacy-Preserving AI Workshop at AAAI*, 2024.

TALKS

1. "Differentially Private Prediction of Large Language Models." Tech Talk @ LinkedIn Research. July 2024.

2. "Differentially Private Prediction of Large Language Models." Tech Talk @ TikTok Privacy Innovation Lab. July 2024.

3. "Modular Monochromatic (3, t)-colorings". 52nd Southeastern International Conference on Combinatorics, Graph Theory & Computing. Florida Atlantic University. 2021. Link: `https://www.youtube.com/watch?v=qciRVyWc90M`

PROFESSIONAL SERVICE

Reviewer
ICLR 2025 TMLR 2025 ACL 2025

Program Committee Member and Reviewer
AAAI Workshop on Privacy Preserving Artificial Intelligence 2024, 2025
ACL Workshop on Large Language Model Memorization 2025
NAACL Workshop on Privacy in Natural Language Processing 2025

Artifact Evaluation Committee Member 2022
Principles and Practice of Parallel Programming Conference

TEACHING EXPERIENCE

Teaching Assistant August 2019 – December 2022
University of Southern California

- **Courses:** CSCI 350: Introduction to Operating Systems

University of Alaska Anchorage

- **Courses:** CSCI 311 Data Structures and Algorithms; CSCI 211: Computer Programming II

Summer Engineering Academies (SEA) Staff Member May 2019 – August 2019
University of Alaska Anchorage

- Facilitated the activities and learning of programming and robotics camps consisting of 20-30 kids from grades ranging from fourth to twelfth grade.

SKILLS

Programming Languages: C/C++, Python, Java, R, Bash
Tools and libraries: Git, GitHub, Tensorflow, PyTorch, Numpy, Pandas, Matplotlib

VOLUNTEER SERVICE

CURVE Mentor 2024
University of Southern California

- Mentoring three undergraduate students working on differentially private in-context learning and prompt optimization.

CSRMP Alumni Panel Discussion 2022
Google