Julius Broomfield

jbroomfield@gatech.edu – /juliusbroomfield – juliusbroomfield.github.io – /juliusbroomfield

Education

Georgia Institute of Technology — Atlanta, GA
B.S. in Computer Science, Minor in Philosophy — Expected May 2026
Concentrations: Artificial Intelligence, Theory
Relevant Coursework: Machine Learning, Randomized Algorithms, Probabilistic Combinatorics, Advanced Linear Algebra, Artificial Intelligence

Research Experience

SALT Lab @ Stanford University — October 2025 – Present
Research Collaborator

- Improving vision language models' (**VLMs**) ability to recognize and respond to physical-social **norms** in videos

CLAWS Lab @ Georgia Tech — August 2024 – Present
Machine Learning Researcher, Dr. Srijan Kumar

- Leading research in evaluating standpoint bias in **LLMs**; developed an **evaluation benchmark**; developing a Mixture of Standpoints methods to improve models' ability to surface **latent user needs** in **underspecified** tasks
- Led research investigating **multimodal persona** embodiment in **VLMs**; developed a benchmark and conducted model **evaluations** indicating weaknesses in grounding visual representations

FAR.AI — January 2025 – May 2025
Research Collaborator, Dr. Kellin Pelrine

- Studied **fine-tuning** vulnerabilities in **LLMs**; developed low-resource language **data poisoning** attacks on **LLMs**

Supervised Program for Alignment Research (SPAR) — February 2024 – May 2025
Machine Learning Researcher, Dr. Kellin Pelrine

- Developed **novel adversarial attacks** to expose non-generalizing safety behaviors in **LLMs** under structural perturbations
- Designed and implemented a **guardrail** that normalizes input structures to improve robustness across **modalities** and **languages**

Selected Publications

Jailbreak-Tuning: Models Efficiently Learn Jailbreak Susceptibility **(EMNLP 2025)**
Murphy B., Bowen D., Mohammadzadeh S., **Broomfield J.**, Gleave A., Pelrine K.

A Thousand Words or An Image: Studying the Influence of Persona Modality in VLMs **(ICML R2-FM 2025)**
Broomfield J.*, Sharma K.*, Kumar S.

The Structural Safety Generalization Problem **(ACL Findings 2025)**
Broomfield J.*, Gibbs T.*, Kosak-Hine E.*, Ingebretsen G.*, Nasir T., Zhang J., Iranmanesh R., Pieri S., Rabbany R., Pelrine K.

Decompose, Recompose, and Conquer: VLMs are Vulnerable to Multi-Image Attacks **(NeurIPS RBFM 2025)**
Broomfield J., Ingebretsen G., Iranmanesh R., Pieri S., Kosak-Hine E., Gibbs T., Rabbany R., Pelrine K.

Industry Experience

LinkedIn — Sunnyvale, CA
Data Science Intern: Job Seeker Experience Team — May 2025 - August 2025

- Developed a **topic discovery service** enabling automated trend discovery in user feedback, reducing manual analysis time from **10 hrs/week → 30 min**
- Designed a zero-shot **taxonomy induction** and **self-labeling loop**; deployed to **9 agent products** and processed **85K** texts across **3** teams across various use cases
- Conducted topic alignment evaluations, achieving **97.7%** label agreement with human taxonomy and **95%** topic homogeneity

Microsoft — Redmond, WA
Software Engineering Intern: Copilot Infrastructure Team — May 2024 - August 2024

- Led **migration** of an enterprise **MLOps platform** powering machine learning inference endpoints across M365 workloads
- Refactored monolithic application infrastructure into **18 microservices**, enabling independent rollout cadence, versioning, and failure isolation across Azure production environments
- Published documentation and internal engineering blog detailing integration steps for teams building microservices on the platform

Microsoft — Redmond, WA
Software Engineering Intern: M365 Core Team — June 2023 - August 2023

- Developed an **OSINT privacy** assessment tool to automatically report users' self-disclosure risks across **3** social media platforms, identifying an average of **14.7** potential privacy exposures per profile
- Integrated a **NER model** to surface **PII** indicators in posted content, achieving ∼**90%** detection accuracy
- Led technical scoping for a 5 person team, defining **15+ core features** and technical requirements

Technical Skills

Machine Learning: Python, Pandas, Numpy, PyTorch, Scikit-learn, OpenCV, BeautifulSoup, Pillow, spaCy, W&B, Seaborn
Software Engineering: Flask, FastAPI, Git, Docker, Kubernetes, gRPC, Slurm, REST APIs, C#, C/C++, Bash, CI/CD, SQL