# Roy Finkelberg

Email: roy@gatech.edu
Github: RFinkelberg

## EDUCATION

**Georgia Institute of Technology** <div align="right">Atlanta, GA</div>
*Joint Bachelor and Master of Science in Computer Science*

- **M.S in Computer Science (Jan. 2020 - May. 2021)**: GPA: 4.0/4.0;  Specialization: Machine Learning
- **B.S in Computer Science (Aug. 2016 - Dec. 2019)**: GPA: 3.85/4.0 (Highest Honors)
  Concentrations: Artificial Intelligence, Embedded Computing

## WORK EXPERIENCE

**Facebook** <div align="right">Menlo Park, CA</div>
*Data Science - Infrastructure Strategy Intern* <div align="right">MAY 2020 - AUG. 2020</div>

- Conducted exploratory analyses to determine causes of poor evaluation metric quality for an internal search and knowledge discovery system
- Redesigned the system responsible for slicing search events into sessions to use linguistic features rather than static rules, reducing session fragmentation by 86% and improving downstream metric quality
- Refactored session slicing logic from a monolithic SQL query into a python-based Hive transformer, improving scalability and extensibility

**NVIDIA** <div align="right">Austin, TX</div>
*Data Science - AI Infrastructure Intern (RAPIDS AI)* <div align="right">MAY 2019 - AUG. 2019</div>

- Adapted deep representation learning methods (graph autoencoders) to develop scalable network analysis and link prediction methods for large cybersecurity networks
- Developed pure GPU implementations for ordinal feature encoding and data train/test split modules, giving up to 290x speedups over CPU implementations on large ($\sim 10^7$ row) datasets
- Constructed, profiled, and optimized fundamental data science primitives such as one-hot encoding and scalar-vector binary operations in Cython and Numba, improving performance by 1.5x on wide datasets

**Pindrop Security** <div align="right">Atlanta, GA</div>
*Software Engineering Intern - Research* <div align="right">MAY 2018 - AUG. 2018</div>

- Developed an automated testing harness for a cloud based machine learning platform
- Designed an abstract schema to streamline creation and integration of new models
- Created a standardized interface for reporting and viewing model performance metrics through Datadog, reducing manual Research Scientist intervention by $\sim 70$ hours per week
- Scaled Scikit-Learn's DBSCAN algorithm to $\sim 10^6$ dimensional feature vectors using Spotify's open source Annoy library

## SELECTED PROJECTS

**Dynamically Characterizing Hierarchical Information in Human Activity Recognition**
*Georgia Tech Computational Behavior Analysis Group* <div align="right">AUG. 2020 - MAY 2021</div>

- **MS Thesis:** Developed a framework to dynamically learn and characterize hierarchical feature information in visual human activity datasets, with the goal of providing additional weak supervision to resource constrained action recognition pipelines
- Designed and evaluated the performance of extensions to current state of the art recurrent architectures on hierarchically labeled time-series human activity data
- Demonstrated that modern activity recognition models implicitly learn salient action hierarchies without additional supervision, providing new avenues for developing interpretable deep action recognition systems

### Piazza Automated Related Question Recommender

*Graduate Researcher - Georgia Tech Contextual Computing Group* <span style="float:right">Aug. 2018 - Present</span>

- Developed a natural language understanding pipeline for a recommendation engine which leverages the collective memory of online forums to prevent duplicate posts. Results demonstrating 40% reduction in duplicate posts were published in ACM Learning @ Scale 2019 (34% acceptance rate)
- Conducted A/B tests of model performance and impact across 1000+ users
- Designed mixed-methods studies measuring the impact of the tool on student/instructor behavior in forums. Results showing statistically significant improvement in user efficiency were published in ACM Learning @ Scale 2020

### Towards Scalable Cybersecurity Network Analysis with Graph Autoencoders

*NVIDIA - RAPIDS AI* <span style="float:right">Aug. 2019</span>

- Investigated the use of autoencoder based methods for large scale cybersecurity network analysis
- Adapted existing graph autoencoder architectures in Tensorflow and PyTorch to static and dynamic cybersecurity networks
- Published as an internal NVIDIA whitepaper, demonstrating up to 4x performance increases and 9x speedups on link prediction tasks

### Model Based Intent Detection for Intelligent Prostheses

*Georgia Tech Exoskeleton Prosthetic and Intelligent Controls Lab* <span style="float:right">Aug. 2017 - Dec. 2017</span>

- Collected and analyzed biometric sensor data to determine features important to gait speed detection
- Preprocessed data and engineered features using Python's Scikit-Learn library
- Presented a preliminary offline gait speed detection model demonstrating the effectiveness of these features
- Assisted in the design of a Kivy GUI which interfaced with ROS to visualize and adjust a prosthetic's control parameters during operation

## Skills

- **Languages**: Python, Java, C

- **Tools/Technologies**: PyTorch, Numpy/Scipy, Pandas, Numba (with CUDA), Cython, Pytest, LaTeX, Software Integration, Experimental Design, Hardware Prototyping, Linux Environments

## Publications

[1] India Irish, **Finkelberg, Roy**, Daniel Nkemelu, Swar Gujrania, Aadarsh Padiyath, Sumedha Raman, Chirag Tailor, Rosa Arriaga, and Thad Starner. PARQR: Automatic Post Suggestion in the Piazza Online Forum to Support Degree Seeking Online Masters Students. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*, L@S '20, pages 125–134, New York, NY, USA, August 2020. Association for Computing Machinery.

[2] Noah Bilgrien, **Finkelberg, Roy**, Chirag Tailor, India Irish, Girish Murali, Abhishek Mangal, Niklas Gustafsson, Sumedha Raman, Thad Starner, and Rosa Arriaga. PARQR: Augmenting the Piazza Online Forum to Better Support Degree Seeking Online Masters Students. In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, pages 1–4, Chicago IL USA, June 2019. ACM.