

Keith Harrigian, PhD

keithharrigian@gmail.com | kharrigian.github.io | github.com/kharrigian

Education

- Aug. 2019 – Aug. 2024 **Johns Hopkins University** | PhD, Computer Science
Thesis: “Towards Robust Natural Language Processing to Promote Health Equity”
- Sept. 2013 – May 2017 **Northeastern University** | BS, Mathematics
Capstone: “Geolocation Inference on Reddit”

Academic Research

- Aug. 2019 – Aug. 2024 **Center for Language and Speech Processing** *Graduate Research Assistant* | P.I. Mark Dredze
- Developed health-oriented machine learning models that are robust across multiple environments (e.g., data platform, demographic composition, hospital system)
- Aug. 2014 – Aug. 2019 **Action Lab** *Undergraduate Research Assistant* | P.I. Dagmar Sternad
- Engineered a new algorithm using Hidden Markov Models to precisely detect initiation of finger taps in noisy strain gauge time series data

Industry Experience

- Aug. 2024 – Present **Netflix** | *Machine Learning Scientist 4 (Content and Studio Data Science)*
- Develop and improve predictive models that facilitate content-related purchasing, scheduling, and marketing decisions
- June 2023 – Aug. 2023 **Netflix** | *Graduate Machine Learning Intern (Content Demand Modeling)*
- Investigated whether audio-visual representations of long-form multimedia content (i.e., movies, television series) can be used to better forecast audience size
- June 2017 – June 2019 **Warner Media** | *Senior Quantitative Analyst*
- Built an interactive tool to extract book titles mentioned on Reddit, scrape metadata from an online reading database, and visualize demographic-level trends
- July 2016 – Dec. 2016 **True Fit** | *Scientist Co-op*
- Designed a robust anomaly detection system to capture fraudulent retail transactions, reducing noise by 10% in recommendation engine training data

Selected Publications

- Harrigian, K.**, Tang, T., Gonzales, A., Cai, C.X., Dredze, M. “An Eye on Clinical BERT: Investigating Language Model Generalization for Diabetic Eye Disease Phenotyping.” *Machine Learning for Health (Findings)*. 2023.
- Harrigian, K.**, et al. “Characterization of Stigmatizing Language in Medical Records.” *In Proceedings of the 61st Meeting of the Association of Computational Linguistics (ACL)*. 2023.
- Harrigian, KZ** & Dredze, M. “The Problem of Semantic Shift in Longitudinal Monitoring of Social Media.” *In Proceedings of the 14th ACM Web Science Conference*. 2022.
- Harrigian, K.**, Aguirre, C., & Dredze, M. “Do Models of Mental Health Based on Social Media Generalize?” *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP): Findings*. 2020.

Selected Honors and Awards

- Oct. 2016 *Rhodes Scholar Nominee* | *Marshall Fellowship Finalist*
Nominated by faculty for outstanding academic merit and ambassadorial ability
- Apr. 2016 *Outstanding Student Research Winner (Computer and Information Sciences)*
Best undergraduate poster in Computer and Information Science at Northeastern RISE 2016

Technical Skills

- Programming Languages Python, Bash, SQL, R, Stan, MATLAB, C
- Computing Libraries pandas, NumPy, SciPy, Matplotlib, PyTorch, scikit-learn, Gensim, tomotopy, NLTK