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" Advisor: Mark Dredze
Aug. 2019 – Dec. 2021	**Johns Hopkins University**	MSE, Computer Science. GPA: 4.0/4.0
Sept. 2013 – May 2017	**Northeastern University**	BS, Mathematics. Minors in Physics and Music. GPA: 3.9/4.0

Academic Research

Aug. 2019 – Aug. 2024	**Center for Language and Speech Processing (CLSP)** *Graduate Research Assistant* ▪ Developed health-oriented machine learning models that are robust across multiple environments (e.g., data platform, demographic composition, hospital system) ▪ Designed and deployed a web-based analytics dashboard for summarizing patient electronic communication data to aid in treatment of mood disorders
Aug. 2014 – Aug. 2019	**The Action Lab (P.I. Dagmar Sternad)** *Undergraduate Research Assistant* ▪ Engineered a new algorithm using Hidden Markov Models to precisely detect initiation of finger taps in noisy strain gauge time series data ▪ Co-supervised "Pitchers and Pianists" study at Boston Museum of Science from September 2015 through May 2016; educated 400+ visitors on human coordination and neural control

Industry Experience

Aug. 2024 – Present	**Netflix** *Machine Learning Scientist 4 – Content and Studio* ▪ Develop and improve predictive models that facilitate content-related purchasing, scheduling, and marketing decisions
Mar. 2021 – Aug. 2024	**Unforged** *Data Science and Machine Learning Consultant* ▪ Led specification and implementation of data science infrastructure for adolescent mental wellness platform (e.g., personalization, content moderation) ▪ Reviewed grant applications to ensure technical contributions are accurately described; identified overlap in core technology with competitors to ensure contributions are novel
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 ▪ Consulted on the development of an internal toolkit for detecting and characterizing distributional shift
June 2018 – June 2019	**Warner Media Applied Analytics** *Senior Quantitative Analyst* ▪ Developed language feature-extraction tools to model the relationship between thematic content in movie trailers and downstream effects on Wikipedia web traffic *Quantitative Analyst* ▪ Optimized the targeting of interest segments on Facebook in real time using contextual-bandits and factorization of audience overlap matrices

June 2017 – June 2018	**Legendary Entertainment**
	Quantitative Analyst

- Developed a multi-modal model to infer demographics of Reddit users and a collaborative filtering system to segment online communities
- Programmed an interactive tool to extract book titles mentioned on Reddit, scrape metadata from an online reading database, and visualize demographic-level trends
- Leveraged partial least squares regression to create a content- and marketplace-aware arbitrage model for the digital promotion of news articles

July 2016 – Dec. 2016	**True Fit Corporation**
	Scientist (Co-op)

- Designed a robust anomaly detection system to capture fraudulent retail transactions, reducing noise by 10% in recommendation engine training data
- Modeled e-commerce return rates to establish baselines for A/B testing

July 2015 – July 2016	**Legendary Entertainment**
	Quantitative Research Collaborator (Consultant)

- Led R&D of a conditional random field model for end-to-end named entity recognition on Twitter, allowing for dynamic query filtering based on temporal popularity fluctuations

Quantitative Research Analyst (Co-op)

- Created a command-line tool to acquire secondary market sales data and compile revenue reports, enabling 4 professional sports organizations to optimize ticket prices
- Trained Naïve Bayes model to quantify movie-going intent and infer sentiment within tweets

Publications

Harrigian, K., Tran, D., Tang, T., Gonzales, A., Nagy, P., Kharrazi, H., Dredze, M., Cai, C.X. "Improving the Identification of Diabetic Retinopathy and Related Conditions in the Electronic Health Record Using Natural Language Processing Methods." *Ophthalmology Science*. 2024.

Jeong, H. et al. "Recent Advances , Applications, and Open Challenges in Machine Learning for Health: Reflections from Research Roundtables at ML4H 2023 Symposium." *ArXiv*. 2024.

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.

Ayers, J., Zhu, Z., **Harrigian, K.**, Wightman, P., Dredze, M., Strathdee, S., Smith., D. "Managing HIV During the COVID-19 Pandemic: A Study of Help-Seeking Behaviors on a Social Media Forum." *AIDS and Behavior*. 2023.

Harrigian, K., Zirikly, A., Chee, B., Ahmad, A., Links, AR., Saha, S., Beach, MC., & Dredze, M. "Characterization of Stigmatizing Language in Medical Records." *In Proceedings of the 61st Meeting of the Association of Computational Linguistics (ACL)*. 2023.

Cai, C., Tran, D., Tang, T., Liou, W., **Harrigian, K.**, Scott, E., Nagy, P., Kharrazi, H., Crews, D., Zeger, S. "Health Disparities in Lapses in Diabetic Retinopathy Care." *Ophthalmology Science*. 2023.

Harrigian, K. & Dredze, M. "Then and Now: Quantifying the Longitudinal Validity of Self-disclosed Depression Diagnoses." *In Proceedings of the Computational Linguistics and Clinical Psychology Workshop (NAACL)*. 2022.

Harrigian, K. & 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. "On the State of Social Media Data for Mental Health Research." *In Proceedings of the Computational Linguistics and Clinical Psychology Workshop (NAACL)*. 2021.

Sherman, E., **Harrigian, K.**, Aguirre, C., & Dredze, M. "Towards Understanding the Role of Demographics in Deploying Social Media-Based Mental Health Surveillance Models." *In Proceedings of the Computational Linguistics and Clinical Psychology Workshop (NAACL)*. 2021.

Aguirre, C., **Harrigian, K.**, & Dredze, M. "Gender and Racial Fairness in Depression Research using Social Media." *In Proceedings of the 16th Conference of the European Chapter of the ACL (EACL)*. 2021.

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.

Harrigian, K., Guo, D., Park, S., & Sternad, D. "Pitchers and Pianists: A Large-scale Study on Discrete and Rhythmic Timing." *In Preparation*.

Harrigian, K. "Geocoding Without Geotags: A Text-based Approach for reddit." *In Proceedings of the 4th Workshop on Noisy User-generated Text (EMNLP)*. 2018.

Gundogdu, A., Sanghvi, A., & **Harrigian, K**. "Recognizing Film Entities in Podcasts." *In Proceedings of the 1st Workshop on Machine Learning and Data Mining for Podcasts (KDD)*. 2018.

Posters and Talks

Harrigian, K. "Fighting Bias With Bias: Challenges and Opportunities for Artificial Intelligence in Healthcare." *Alzheimer's Association AI Working Group*. Invited Talk. 2023.

Harrigian, K. "Characterization of Stigmatizing Language in Medical Records." *The 1st International Workshop on Ethics and Bias of Artificial Intelligence in Clinical Applications*. Keynote Talk. 2023.

Sternad, D., Guo, D., & **Harrigian, K**. "Pitchers and Pianists: Timing in Discrete and Rhythmic Motor Skills." *New England Sequencing and Timing Meeting*. Talk. 2017.

Harrigian, K., Sanders, N., Foster, J., & Sanghvi, A. "When Anonymity is Not Anonymous: Gender Inference on Reddit." Won Outstanding Student Research (Computer and Information Sciences). *Northeastern Research, Innovation, and Scholarship Expo*. Poster. 2016.

Harrigian, K., Kuznetsov, N., Sternad, D. "Effects of tDCS on Precision of Finger Force Control and Rhythmic Tapping Movements." *Northeastern Research, Innovation, and Scholarship Expo*. Poster. 2015.

Honors and Awards

Oct. 2016	*Marshall Fellowship Finalist* Nominated by faculty for outstanding academic merit and ambassadorial ability
Oct. 2016	*Rhodes Scholar Nominee* Nominated by faculty for scholarly merit, social commitment, and leadership
Apr. 2016	*Outstanding Student Research (Computer and Information Sciences)* Best undergraduate poster in Computer and Information Science at Northeastern RISE 2016
Dec. 2015	*Barry Goldwater Scholarship Nominee* Research Proposal: Extreme Learning Machine for Localization of EEG in Parkinson's Patients

Grants

Apr. 2015	*Undergraduate Research and Creative Endeavors Award* $1000 to research effect of metric structure strength on motor learning of temporal rhythms
Apr. 2014	*Lawrence Award for Undergraduate Scholastic Excellence in Physics* $250 scholarship awarded to student(s) with the highest GPA in class year
Sept. 2013	*Northeastern College of Science Dean's Scholarship* $80,000 scholarship awarded to top incoming undergraduates

Academic Service

Aug. 2019 – June 2020	**Northeastern Honors Program** *Alumni Advisor* ▪ Provide career and course guidance to two Northeastern University computer science undergraduate students
Sept. 2015 – May 2017	**Northeastern College of Science** *Peer Advising Coach and Ambassador* ▪ Met weekly with a first-year physics undergraduate student to instill successful academic habits; curated a study schedule to address time-management issues

Sept. 2013 – Jan. 2016	**Northeastern Student Government Association**
	Chair of Elections
	▪ Raised voter turnout by 25% to a record high for campus of 18,000 undergraduates
	▪ Reformed referendum process by increasing accountability and transparency of legislature

Community Service

Apr. 2014 – Apr. 2018	**Boston Athletic Association**
	Team Captain (Recycling)
	▪ Led recycling operations for the Boston Marathon Finish Area
	▪ Supervised team of 40+ volunteers in collection of recyclable goods and trash
Jan. 2009 – Aug. 2014	**Golden Retriever Club of Greater Los Angeles Rescue**
	Volunteer and Foster
	▪ Served as caretaker for over 40 dogs; assisted in their transportation to medical appointments
	▪ Expedited revenue collection at several fundraisers via PayPal

Reviewing Service

Journals	Journal of Medical Internet Research (JMIR)
Conferences	Computational Linguistics and Clinical Psychology Workshop (CLPsych); International Conference on Linguistics (COLING); Meeting of the Association of Computational Linguistics (ACL); ACL Rolling Review (ARR)

Teaching

Spring 2021	**Deep Learning**. Johns Hopkins University. Teaching Assistant. Graduate and Undergraduate.

Advising

Mar. 2023 – Aug. 2024	**Yahan (Zoe) Li**. Johns Hopkins University. Masters Student.
Mar. 2022 – Dec. 2023	**Hyun Joo Rosalyn Shin**. Johns Hopkins University. Masters Student.
Apr. 2020 – Jan. 2021	**Narayani Wagle**. Johns Hopkins University. Undergraduate Student.
Jan. 2019 – June 2019	**Aniruddah Tapas**. Warner Media Applied Analytics. Co-op Student.
Oct. 2018 – Dec. 2018	**Ryan Oakley**. Warner Media Applied Analytics. Co-op Student.
Jan. 2018 – June 2018	**Ahmet Gundogdu**. Warner Media Applied Analytics. Co-op Student.

Technical Skills

Programming Languages	Python (Advanced), Bash (Intermediate), SQL (Intermediate), R (Functional), Stan (Functional), MATLAB (Functional), C (Functional)
Computing Libraries	pandas, NumPy, SciPy, Matplotlib, PyTorch, scikit-learn, Gensim, tomotopy, NLTK
Miscellaneous	Git (Intermediate), AWS (Functional)
Certifications	National Institutes of Health Office of Extramural Research (Human Subjects)