

**EDUCATION** University of Washington - *MS Computational Linguistics* 06/2018-05/2020 3.87/4  
Rensselaer Polytechnic Institute - *BS Computer Science* 08/2011-05/2015 3.43/4

**WORK EXPERIENCE** *Program Manager 2*-Microsoft Research & AI, Bing Applied Science & Metrics 11/2017-Present

- Rearchitected relevance metrics stack into a realtime streaming system through new human labeling tasks, novel query set sampling methods, and metric aggregation. Pipeline scaled to \$7m/year across 16 markets generating more than 60,000,000 judgements a year. Pipeline went from measuring user experience months old to seconds old. To scale the pipeline created a novel multi class stratified sampling method using the Horvits-Thompson estimator, work under review.
- Built family of MSMARCO Datasets and baselines including QnA, Passage Ranking, and Keyphrase Extraction which collectively have been used by over 3000 Researchers. Papers Cited over 250 times and competitive leaderboards with over 100 research labs. Baselines models built primarily in Pytorch and Tensorflow
- Research and analyzed Bing Relevance and user experience for executive team across various verticals, languages, and metrics using language embeddings, sampling methodologies and classification methods.

*Program Manager 2*-Microsoft Cloud + AI, Global Services C&E 05/2016-11/2017

- Designed, built and deployed Neural Machine Translation into software localization pipeline for entire division. Using the improved pipeline, renegotiated supplier contracts saving 20% annualizing to \$3m/year over the next 3 years.
- Optimized and expanded NLP system for customer feedback translation and categorization decreasing team effort by 80% using NLTK, Tensorflow, and scikit-learn.
- Drove end to end localization and internationalization for Azure and Visual Studio products including automation, bug triage, budgeting, vendor and release management.

*Program Manager*-Microsoft, Azure RemoteApp 05/2014-08/14 & 08/2015-05/2015

- Designed scalable multi cloud network architecture for cloud based Remote Desktop Service and drove large scale deployment to alpha customers and Microsoft.
- Created and scaled a predictive user demand modeling pipeline which optimized VM usage by 20%.

*NLP Researcher*-Basis Technology 01/2014-05/2014

- Grew Rossette Entity Resolution pipeline to decrease document analysis time, increase accuracy, and increase throughput while expanding language support to spanish, russian, french, and chinese.
- Optimized existing models improving accuracy in other language by 30%
- Formalized data creation, data injections, retrieval, formatting and analysis pipeline using primarily Bash, Java, and Python. Optimized models were KNN, SVMs, LSTMs and Decision Trees.

*Software Engineer*-Cisco Systems 05/2013-08/2013

- Developed JavaScript tools for bug tracking and HTML5 based screenshotting to allow developers to receive feedback during betas and private previews.

- Led a team of 12 other interns to develop a novel food ordering system for cafeterias which used containerized applications allowing efficient scaling.

*CoFounder/Software Engineer- Gapelia* 08/2013-05/2014

- Built an end to end digital multimodal publishing platform which allowed non technical users to create and publish beautiful digital magazines in a matter of minutes
- Created REST API in Java, front end in Express and deployed in AWS with tomcat and mongoDB.
- Won RPI 2014 Business Model Competition 1st place, 2014 Why not change the world grant, member of Harvard's ILAB

*Software Engineer-GE Capital* 05/2012-08/2012

- Developed a reusable web templating system which allowed GE employees to create websites that matched corporate guidelines around colors, fonts, icons, etc.
- Built out in Bootstrap and PHP

*Support Engineer-KIO Networks* 05/2011-08/2011

- Debugged and triaged customers networking issues in a secure datacenter.

*Support Engineer-Amerikano Experiences* 05/2009-06/2011

- Built up an event hosting business which would host parties for 16-22 year olds in Mexico City.
- Hosted 14+ events making 40,000 USDs profit and built up a sales group of 50 students/

*Developer Platform Engineer-Microsoft* 05/2010-08/2010

- Built proof of concept solutions for customers in the telecom and banking industries using windows products such as Windows Phone 7 and Xbox Kinect

## Research

### Publications

- MS MARCO: A Human Generated Machine Reading Comprehension Dataset- Neurips 2016,2018
- Open Domain Web Keyphrase Extraction Beyond Language Modeling-EMNLP-IJCNLP 2019 Oral Presentation
- Experiments in Inferring Social Networks of Diffusion-2015

### Patents

- Keyphrase Extraction Beyond Language Modeling - U.S. Appln. No. 16/460,853 - Filed July 2nd, 2019

### Datasets

- Optimal Freshness Crawl Under Politeness Constraints
- MSMARCO Question Answering and Natural Language Generation
- MSMARCO Passage Retrieval
- Open Domain Keyphrase Extraction
- Conversational Search with Privacy Minded Intent Representation

## Academic Activity

- NIST TREC 2019 Deep Learning-Track Coordinator
- NIST TREC 2020 Deep Learning-Track Coordinator
- ACM SIGIR/SIGKDD Africa Summer School on Machine Learning for Data Mining and Search 2019-Invited Lecturer for Deep Learning in Search
- AACM SIGIR/SIGKDD Africa Summer School on Machine Learning for Data Mining and Search 2020-Invited Lecturer for Deep Learning in Search
- LaTeX in AI Research at Neurips-Reviewer