

Dan Pechi

291 Manhattan Ave, Brooklyn, NY • 630.903.7853 • danpechi@gmail.com • danpechi.github.io/about/

- EDUCATION**
- Computational Linguistics, Cognitive Brain Science Major & Russian Studies Minor**
Tufts University, Medford, MA May 2019
Overall GPA: 3.6/4.0, Dean's List, Neubauer Scholar (Special Research Grant Recipient),
Institute for Global Leadership Research Fellow 2017
Independent Online Courses: Andrew Ng's Machine Learning, Geoff Hinton's Neural
Networks for Machine Learning
- Illinois Math and Science Academy, Aurora, IL** May 2015
Overall GPA: 3.8/4.0, over 200 service hours, National Merit Finalist, NSLI-Y Finalist
(Government-Funded Russia Exchange), Intermediate High Russian OPI Rating (2014)
- EXPERIENCE**
- UMass Lowell Text Machine Lab, Lowell, MA** September 2021-Current
Research Assistant to Anna Rumshisky
- Collaboratively researched optimizing vocabularization to efficiently pre-train language models, drawing from language acquisition research
 - Presented recent NLP papers as part of a weekly reading group
 - Compiled an extensive literature review of prompt-tuning methods alongside PhD students
- Opensignal, Boston, MA** February 2021-Current
R&D Data Scientist
- Engineered a ML model to classify IP's for a new business broadband market product, working with data engineers to QA and productionalize the classifier
 - Developed company's core ML algorithm for predicting device persistence, providing higher resolution data feeds for all downstream products
 - Led projects to integrate external data into our core products including acquisition from and negotiation with data vendors, QA, augmentation, and downstream monitoring
 - Used prompt-tuning to extract shipping delay information from Reddit forums in real time
 - Built a natural language processing model to assess impact of new fiber markets using live data feeds from Reddit forums
 - Rapidly prototyped a graph neural network algorithm for modeling geospatial market dynamics
 - Triaged existing data pipelines, and analyzed blue sky dataset integration
 - Trained other data scientists in AWS tools, as well as Pyspark, Sagemaker, and MLFlow
- Oliver Wyman Digital, New York, NY** July 2019-February 2021
Data Science Consultant (Financial Services, Transportation, Consumer & Retail)
- Developed and managed the creation of a demand forecasting tool at a major delivery services company, coordinating a team of data scientists and engineers
 - Built a promotion forecasting tool for a major US retailer using time series data to model customer behavior and promotion uplift, and identify ineffective offers
 - Built an assortment optimization proof-of-concept for a major craft store, leveraging SQL to pull data for store clustering, and build customer decision trees
 - Engineered features and fine-tuned a deep learning model in Pyspark and PyTorch to block scheduled flight times for a major airline to increase on-time performance and revenue
 - Developed a proposal to use graph neural networks for more accurate fraud detection in financial transaction data
 - Engineered logistic regression and random forest models of credit losses in collaboration with a major consumer credit reporting company to reflect updates in accounting standards for banks in Pyspark
 - Conducted stress testing for a banking client using multiple linear regression to

Dan Pechi

291 Manhattan Ave, Brooklyn, NY • 630.903.7853 • danpechi@gmail.com • danpechi.github.io/about/

- correlate business outcomes with macroeconomic factors and model risk
- Built a linear programming model that optimized the maintenance network of a major airline to increase system efficiency and reduce aircraft out-of-service events

Charles River Analytics, Cambridge, MA

AY 2017-2019

Part-Time Software Engineer

- Transferred a BERT language model to predict cyber-attacks from sociopolitical events in the news and on the darkweb
- Built a LSTM-based time-series model to predict the trajectory of missiles in a simulation
- Developed a dataset to identify bias in news sources to train a fake news classifier

Tufts University, Medford, MA

Fall 2018

Working With Corpora (COMP 150-1) Teaching Assistant

- Developed and led technical aspects of the course involving deep learning-based natural language processing for digital humanities applications
- Created a notebook and dataset for students to predict the gender of characters in Shakespeare plays using pre-trained word embeddings and deep neural networks

Tufts University, Medford, MA

AY 2017-2018

Introductory Russian (RUS 01&02) Teaching Assistant

- Led drill sections centered around vocabulary and grammar comprehension

Charles River Analytics, Cambridge, MA

Summer 2017&2018

Computational Russian Linguistics Internship

- Applied systemic functional grammar to statistical architectures to classify natural language on the internet
- Built machine learning classifiers in scikit-learn and visualized results to identify relationships between sociopolitical events and cyber-attacks
- Developed a transformer neural machine translation module to parse Russian language data on a single GPU
- In 2017 voted best intern on basis of supervisor and division presidents' evaluations and final intern presentation company-wide evaluation

Massachusetts Institute of Technology, Cambridge, MA

Spring 2016

Research Assistant to Dr. Carol Saivetz

- Translated Russian news articles from Novoe Vremya and Kommersant about Russian policy in Syria into English

Carnegie Mellon University, Pittsburgh, PA

Fall 2016

Research Assistant for the Digital Tolstoy Project

- Provided natural language processing skills to tag and analyze a Russian language corpus of Leo Tolstoy's collected works

University of Chicago Harris School of Public Policy, Chicago, IL

AY 2014-2015

Research Assistant to Dr. Steve Cicala

- Visualized power plant load output, relative humidity and Particulate Matter output data in Python, QGIS and ARCGIS to examine public health outcomes

COMPUTER SKILLS

Proficiency in Python, scikit-learn, Pytorch, SQL, C++, Java, HTML, NLTK, Google Apps, ARCGIS, QGIS, Elasticsearch, Mathematica, MongoDB, Excel, Powerpoint, Some Keras and Tensorflow

ACTIVITIES AND INTERESTS

Tufts Machine Intelligence Community President (<https://tufts-mic.github.io/MIC-site/>), Machine Intelligence Conference 2018 Workshop and Community Outreach Coordinator (machineintelligence.cc), Cooking, Piano, Biking