Alexandra DeLucia
Computer Science PhD Student

✉ aadelucia@jhu.edu ⌂ alexandradelucia.com 🐦 Alexir563 ⭘ AADeLucia

⌖ Baltimore, USA

Education

Johns Hopkins University, PhD in Computer Science   2019−Present
Advised by Prof. Mark Dredze in the Center for Language and Speech Processing (CLSP). GPA: 3.86/4.00

Johns Hopkins University, M.S.E in Computer Science   2019−2021

Rollins College, B.A. in Computer Science and Mathematics   2014−2018
Graduated *Summa cum laude* with honors. Recipient of *Outstanding Senior in Science and Mathematics Division* award.

Research Interests

My interests lie in modeling social media data for natural language processing machine learning tasks, primarily text generation, classification, and information extraction.

Keywords: Natural Language Processing, Machine Learning, Social Media, Large Language Models, Text Generation, Controllable Generation, Information Extraction

Research Experience

Select projects with publications and presentations.

Bernice: Language Modeling for Twitter, Johns Hopkins   2022
Trained a RoBERTa model from scratch on 2.5B tweets with a custom SentencePiece tokenizer. Compared the model to other pre-trained encoder models to evaluate effects of in-domain tokenizer and pre-training data.

> DeLucia, Wu, Mueller, Aguirre, Dredze, and Resnik: **"Bernice: A Multilingual Pre-trained Encoder for Twitter"**. EMNLP. 2022.

Decoding Methods for Narrative Generation, Johns Hopkins   2020
Fine-tuned GPT-2 for narrative generation and evaluated the effects of various decoding methods on story coherence, fluency, and interestingness. Also designed and conducted human evaluation surveys on Amazon Mechanical Turk.

> DeLucia, Mueller, Li, and Sedoc: **"Decoding Methods for Neural Narrative Generation"**. GEM Workshop at ACL-IJCNLP. 2021.

Forecasting Civil Unrest Events with Twitter, Johns Hopkins   2019−2023
Modeled a population using Twitter data to predict disruptive political events, such as riots and protests. Includes supporting work on dataset creation for classifying unrest-related tweets, qualitative case studies for specific protests, and tweet geolocation. Joint effort with Johns Hopkins University Applied Physics Laboratory's Crystal Cube project.

> Sech, DeLucia, Buczak, and Dredze: **"Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest"**. W-NUT at EMNLP. 2020.
>
> Chinta, Zhang, DeLucia, Dredze, and Buczak: **"Study of Manifestation of Civil Unrest on Twitter"**. W-NUT at EMNLP. 2021.
>
> Zhang, DeLucia, and Dredze: **"Changes in Tweet Geolocation over Time: A Study with Carmen 2.0"**. W-NUT at COLING. 2022.

Social Media and Public Health, Johns Hopkins   2022-2023
Various projects in a joint effort with the Food and Drug Administration (FDA) and JHU Bloomberg School of Public Health. Role comprises Reddit data collection, processing, and extraction of project-relevant documents through lexical and semantic search.

> Brenneke, Meacham, Bunting, DeLucia, and Proferes: **"r/AskAComputerScientist: Processing Reddit Data for the Social Sciences"**. CPDD. 2023.
>
> DeLucia, Poliak, Zu, Pitts, Navarro, Shojaie, Ayers, and Dredze: **"Automated Discovery of Perceived Health-related Concerns about E-cigarettes from Reddit"**. SRNT. 2023.

Characterization of Online Behavior, Applied Machine Learning Summer Research   Summer 2021
Fellowship at Los Alamos National Lab
Model various history-related communities on Reddit to identify features of conspiracy-focused subreddits. Specific tasks include classification of community type with graph neural networks and classification of user-shared link types from their surrounding information. Presented at INDE-2021: Beyond Misinformation: Towards a Research Agenda for Information Ecosystems, Network Dynamics and Emergent Epistemologies.

Detecting Similar Help Desk Tickets, Los Alamos National Lab   2019
Analyzed tickets from the Consult team in the High Performance Computing Division for trends and ticketing system improvements. Developed proof of concept tools to automate tasks such as category assignment and similar ticket recommendation.
   DeLucia and Moore: **"Analyzing HPC Support Tickets: Experience and Recommendations"**. Preprint. 2020.

***Revere*: Predicting HPC Job Failure**, Los Alamos National Lab   2017−2018
Created a proof of concept model to predict the outcome of a high performance compute (HPC) job based on engineered features from its corresponding system log messages. The features came from the fields of systems and natural language processing. This work is being converted into a production system at LANL, called *Revere*.
   DeLucia and Baseman: **"Work in Progress: Topic Modeling for HPC Job State Prediction"**. HPDC. 2018.

Campus and Community Involvement

Reviewing and Program Committees, Johns Hopkins   Present
Review for conferences and workshops.
   - Computational Linguistics Journal (CL). 2023.
   - Journal of Online Trust and Safety (JOTS). 2023.
   - Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) workshop. 2020-2023.
   - Machine Learning for Computing Systems (MLCS) workshop at Supercomputing. 2020.

GRACE Vice President, Johns Hopkins   2022−2023
Officer for Graduate Association of Women in Computer Science & Electrical and Computer Engineering (GRACE) club. Coordinate group lunches and social and professional events for women graduate students.

Research Mentor, Johns Hopkins   2020−Present
Co-advise undergraduate and master's students with Prof. Mark Dredze on a variety of research projects.
   - Jingyu "Jack" Zhang (2021−Present)
   - Abhinav Chinta (Spring 2021)
   - Justin Sech (2020−2021)

Instructor - Financial Computing Workshop, Johns Hopkins   Winter 2021
Co-taught a two week workshop for Financial Math Masters students. The workshop taught data science skills, such as collecting and manipulating data with python, to first-year masters students. Duties included teaching lessons, designing and grading homeworks, and helping students.

Teaching Assistant - Natural Language Processing, Johns Hopkins   Fall 2020
TA for Prof. Jason Eisner's NLP course. Duties included holding office hours, managing course assistants, grading exams, and helping design homework.

CLSP Student-Faculty Liaison Committee, Johns Hopkins   2020−2023
Convey concerns from students to a faculty representative and discuss possible solutions. Also organize town halls for Center for Language and Speech Processing community discussions.

CLSP Visit Weekend Committee, Johns Hopkins   2019−2020
Assist the planning and execution of the PhD prospective student visit weekend for Center for Language and Speech Processing.

Google IgniteCS Grant Program, Rollins College   2017
Teach basic coding concepts (MIT Scratch) to elementary school kids in the Orlando, FL school system.
Met once per week after school.