Lawrence Chillrud

722 West 168th Street, New York, NY, 10032

📞 (845) 652-3798   ✉ lgc2139@cumc.columbia.edu   ⬛ github.com/lawrence-chillrud

Education

Columbia College, Columbia University in the City of New York   **August 2016 – May 2020**

Bachelor of Arts in Computer Science, Intelligent Systems   *GPA: 3.744*

Objective

I am eager to take my study of machine learning and algorithm design and analysis to new heights by doing hands-on, cross-disciplinary research at a rigorous computer science PhD program. I am keen to learn from and work with a group of passionate researchers and students. While my most recent research background has been in machine learning for environmental epidemiology and public health, I find ML and AI research with applications in medicine and healthcare extremely exciting, and would be thrilled to tackle a variety of problems in those domains.

Research Experience

Senior Programmer in Environmental Epidemiology   **October 2020 – Present**

Dept. of Environmental Health Sciences, Columbia Mailman School of Public Health   *New York, NY*

- Working under **Professor Marianthi-Anna Kioumourtzoglou** on a multi-disciplinary team to help adapt and extend Principal Component Pursuit (PCP), a robust dimensionality reduction technique from computer vision, for pattern recognition with environmental mixtures data to aid in epidemiological studies. In this work, we engineer and apply PCP to a variety of public health arenas and data, including air pollution, exposomic, and metabolomic data. In collaboration with **Professors John Wright** and **Jeff Goldsmith**, exploring convex and non-convex optimization approaches.
- Leveraging guassian processes to design faster cross-validated grid searches to assist in PCP parameter tuning.
- Working on a multi-disciplinary team to help develop a Bayesian Non-parametric Ensemble model for uncertainty characterization in $PM_{2.5}$ predictions across the contiguous United States.
- Using computer vision algorithms and CNN architectures in an attempt to quantitatively characterize changes in urban communities in response to policy developments and the COVID-19 pandemic.
- Conducting extensive code reviews for the papers of members in Professor Kioumourtzoglou's lab as needed. Reviewed over 9,000 lines of code to ensure reproducability and accuracy.
- Collecting, cleaning, and exploring various air pollution and public health datasets for different research applications and questions. Maintaining high quality documentation to accompany cleaned data.
- Aiding in the writing and editing of scientific papers and abstracts. Presenting work at various conferences.
- Exploring different methods for source apportionment, including PCP, Principal Components Analysis, Factor Analysis, Non-negative Matrix Factorization, and Autoencoders.
- Learning various methods and approaches for building environmental / epidemiological health models and analyses.

Environmental Epidemiology Research Assistant   **June 2020 – October 2020**

Dept. of Environmental Health Sciences, Columbia Mailman School of Public Health   *New York, NY*

- Worked under **Professor Marianthi-Anna Kioumourtzoglou** to help adapt and extend Principal Component Pursuit (PCP), a robust dimensionality reduction technique, for environmental mixtures data in epidemiological studies.
- Wrote and ran various experiments in R to assess PCP's performance across varying metrics and in varying settings, using both synthetic and real-world datasets, and compared results with existing methods and benchmarks.
- Interrogated PCP's computational efficiency and mathematical foundations in a multi-disciplinary research group.
- Aided in the development of a novel, user-friendly R package that will allow other researchers assessing multi-pollutant exposures in environmental epidemiological studies to implement PCP.

Natural Language Processing Research Assistant   **June 2020 – October 2020**

Dept. of Computer Science, Columbia University   *New York, NY*

- Conducted Natural Language Processing (NLP) research under **Professor Kathleen McKeown** to help develop, train, and evaluate a BERT-based fact-checking model to combat misinformation surrounding COVID-19 and climate-change.
- Built a COVID-19-specific dataset to train and test our fact-checking model. Designed software that crawled and scraped millions of online news articles for claims regarding COVID-19 before mapping them to relevant scientific papers.
- Aided in the writing and editing of an IRB protocol to receive approval for human annotators to help tag the dataset with information needed for the fact-checking model.
- Assisted in the development, implementation, and maintenance of a user-friendly, web-based annotation interface to facilitate annotations on the dataset. Worked with the popular NoSQL database MongoDB to do so.

- Explored various NLP approaches to the fact-checking problem and pipeline in Python, including but not limited to: BERT-based architectures, claim detection, unsupervised data augmentation, semi-supervised learning, transfer learning, named entity recognition, TF-IDF, few-shot learning, etc.
- Aided in the writing and editing of two publications related to the above work.

Publications, Abstracts, and Presentations

Manuscripts

1. Gibson EA, Zhang J, Yan J, **Chillrud LG**, Benavides JP, Colgan R, Nunez Y, Tao RH, Herbstman JB, Goldsmith J, Wright J, Kioumourtzoglou M-A. Principal Component Pursuit for Pattern Identification in Environmental Mixtures. In draft.

2. Rowland ST, **Chillrud LG**, Boehme AK, Wilson A, Rush J, Just AC, Kioumourtzoglou M-A. Can Weather Help Explain 'Why Now?': The Potential Role of Hourly Temperature as a Stroke Trigger. Under review at *Environmental Research.*

3. Wang G, **Chillrud LG**, McKeown KR. Evidence based Automatic Fact-Checking for Climate Change Misinformation. Submitted tp *SocialSens 2021*.

4. Wang G, **Chillrud LG**, Ananthram A, Subbiah M, McKeown KR. COVID-Check: Fact-Checking COVID-19 Misinformation with Scientific Evidence. Submitted to *EMNLP*.

Abstracts

5. **Chillrud LG**, Gibson EA, Nunez Y, Colgan R, Tao RH, Zhang J, Yan J, Wright J, Goldsmith J, Kioumourtzoglou M-A. Principal Component Pursuit for Pattern Recognition from Incomplete Environmental Data. Accepted to *ENAR 2022*, Houston, March 27-30, 2022.

6. **Chillrud LG**, Gibson EA, Nunez Y, Colgan R, Tao RH, Zhang J, Yan J, Wright J, Goldsmith J, Kioumourtzoglou M-A. Principal Component Pursuit for Exposure Pattern Recognition: An Application to Persistent Organic Pollutants and Leukocyte Telomere Length. *ISEE 2021*, NYC, August 23-26.

7. Rowland ST, **Chillrud LG**, Boehme AK, Wilson A, Rush J, Just AC, Kioumourtzoglou M-A. Can Weather Help Explain 'Why Now?': The Potential Role of Hourly Temperature as a Stroke Trigger. *ISEE 2021*, NYC, August 23-26.

8. Tao RH, Nunez Y, **Chillrud LG**, Rowland ST, Boehme AK, Kioumourtzoglou M-A. Source-specific Fine Particulate Matter and Hospitalization due to Myocardial Infarction. *ISEE 2021*, NYC, August 23-26.

9. Rowland ST, Makkar A, Benavides JP, **Chillrud LG**, Coull B, Fiore A, Henze D, Martin R, Milly GP, Donkelaar Av, Parks RM, Paisley J, Kioumourtzoglou M-A. Uncertainty characterization in $PM_{2.5}$ Predictions Across the Contiguous US. *ISEE 2021*, NYC, August 23-26.

10. Benavides JP, Nunez Y, **Chillrud LG**, Gibson EA, Kioumourtzoglou M-A. Pre- and Postnatal Urban Exposure Patterns and Childhood Neurobehavior. *Exposome Data Challenge*, ISGlobal, April 28-30, 2021.

Technical Skills

Languages with Advanced Proficiency: Python, R, Java, LaTeX
Languages with Basic Proficiency: C, C++, HTML, CSS, MATLAB
ML python libraries: TensorFlow, Keras, PyTorch, Scikit-learn, Hugging Face Transformers, NumPy, SciPy, Pandas, Matplotlib, Seaborn
Developer Tools: iTerm, Vim, tmux, RStudio, Jupyter Notebook, Eclipse, Google Cloud Platform, Homebrew, Conda
Operating Systems: UNIX, macOS
Version Control Systems: GitHub
Database Systems: MongoDB, NoSQL
Certifications: HIPAA, CITI, Human Subjects Protection
Experienced Writing and Editing: academic papers, IRB protocols

Relevant Coursework

- Machine Learning
- Natural Language Processing
- Computational Genomics
- Artificial Intelligence
- Analysis of Algorithms
- Computer Science Theory
- Advanced Programming
- Fundamentals of Computer Systems

- Data Structures & Algorithms
- Number Theory
- Cryptography
- Linear Algebra
- Calculus I, II, & III
- Discrete Mathematics
- Geochemistry
- Death Valley Geology

- Organic Chemistry I
- General Chemistry I & II
- General Chemistry Lab
- Introduction to Linguistics
- Probability and Statistics*
- Real Analysis I*
- Analysis and Optimization*
- Ordinary Differential Equations*

* To be taken in the Fall 2021 / Spring 2022 semesters as a student at the **Columbia University School of Professional Studies**, as a part of their *Postbaccalaureate Studies* program.

Other Projects

Scraping Georgia Jails for Georgia Get Out The Vote   **November 2020**
- Wrote a Python web-crawler to scrape jails for Georgia's *Get Out The Vote* Campaign to help send voter registration and ballot information to every incarcerated person in Georgia's jails before the 2020 Georgia State Senate runoff elections. Made the code publicly available on GitHub and helped edit other volunteers' web-crawlers.

RoBERTa for Claim Detection   **October 2020**
- Wrote a Python class called `ClaimDetective` that allows the user to rank a list of sentences (i.e. potential claims) in order of most check-worthy to least check-worthy, i.e., the priority with which each sentence should be fact-checked.
- `ClaimDetective` was built by fine-tuning RoBERTa to identify and rank claims that are worth fact-checking. Implemented with `PyTorch` and `Scikit-learn` and made publicly available on GitHub.

Automatic Diagnosis of COVID-19 Chest X-rays with Neural Nets   **May 2020**
- Designed, trained, and evaluated a convolutional neural network that classified chest x-ray images into one of four classes: Viral Pneumonia, Bacterial Pneumonia, COVID-19, and Healthy. Implemented in Python with `TensorFlow` and `Keras`.
- Wrote a 31-page report in LaTeX summarizing the results of the final model. The report included a detailed error analysis, along with an interpretability section that aimed to help clinicians and front-line workers learn what aspects of the chest x-rays the model was using to successfully diagnose patients.

SARS-CoV-2 Sequence Analysis   **May 2020**
- Performed a preliminary RNA secondary structure analysis of SARS-CoV-2's spike (S) gene coding region of its genome across inter- and intraspecies datasets to identify conserved structures that could be targeted when making a vaccine.
- Identified 15 potentially conserved structures within the interspecies dataset, as well as 43 at the intraspecies level.
- Wrote a custom Python script to perform the RNA secondary structural analysis identifying conserved sites. Made code available on GitHub, and wrote a Bioinformatics-style Applications Note in LaTeX to summarize the project.

References

References are available upon request. Transcript is available upon request.