LAWRENCE CHILLRUD

Senior Programmer

Columbia University Mailman School of Public Health
722 West 168th Street, New York, NY 10032
(845) 652-3798
lgc2139@cumc.columbia.edu
github.com/lawrence-chillrud
lawrence-chillrud.github.io

EDUCATION

9/2021 – 5/2022 **Columbia University** School of Professional Studies
Post-baccalaureate Studies Program: Non-degree *New York, NY*
GPA: 4.00 / 4.00

9/2016 – 5/2020 **Columbia University** Columbia College
Bachelor of Arts in Computer Science, Intelligent Systems *New York, NY*
GPA: 3.74 / 4.00

RESEARCH EXPERIENCE

10/2020 – Present **Senior Programmer** Columbia Mailman School of Public Health
Environmental Health Sciences *PI: Dr. Marianthi-Anna Kioumourtzoglou*

- Continued work on Principal Component Pursuit w/ Drs. John Wright & Jeff Goldsmith.
- Investigating convex & non-convex approaches to matrix decomposition, dim reduction.
- Leveraging Gaussian processes to design faster cross-validated grid searches.
- Developing Bayesian non-parametric ensemble model for uncertainty characterization.
- Utilizing computer vision algorithms to track changes in urban communities.
- Conducting extensive code reviews for academic papers (reviewed over 9,000 lines).
- Cleaning, visualizing & documenting public health datasets for various research questions.
- Aiding in writing, editing of scientific papers & abstracts. Presenting work at conferences.
- Exploring methods for source apportionment: PCP, PCA, Autoencoders, Factor Analysis.
- Building environmental / epidemiological health models and analyses.

6/2020 – 10/2020 **EHS Research Assistant** Columbia Mailman School of Public Health
Environmental Health Sciences *PI: Dr. Marianthi-Anna Kioumourtzoglou*

- Adapted & extended Principal Component Pursuit (PCP) for environmental mixtures data.
- Benchmarked PCP's computational efficiency, interrogated its mathematical foundations.
- Developed novel, user-friendly R packages for implementation of environmental PCP.
- Designed & ran synthetic & applied experiments to assess PCP's statistical performance.

6/2020 – 10/2020 **NLP Research Assistant** Columbia University
Computer Science Department *PI: Dr. Kathleen McKeown*

- Developed, trained, evaluated fact-checking model for COVID-19 & climate-change.
- Worked on transformer architectures (BERT), few-shot learning, claim detection, named entity recognition, unsupervised data augmentation, transfer learning for fact-checking.
- Built a COVID-19-specific dataset to train and test RoBERTa-based fact-checking model.
- Scraped millions of online news articles for COVID-19 claims & mapped to scientific papers.
- Wrote IRB protocol to receive approval for human annotators to tag fact-checking dataset.
- Implemented & maintained user-friendly annotation interface to facilitate annotations.
- Assisted in writing, editing of scientific paper detailing our novel fact-checking pipeline.

PUBLICATIONS

Papers

1. Makkar A, Liu JZ, Kioumourtzoglou M-A, Coull B, Rowland ST, **Chillrud LG**, Paisley J. Fast Bayesian Nonparametric Ensemble Using Random Fourier Features. Submitted to *AISTATS*, 2022.

2. Gibson EA, Zhang J, Yan J, **Chillrud LG**, Benavides JP, Nunez Y, Herbstman JB, Goldsmith J, Wright J, Kioumourtzoglou M-A. Principal Component Pursuit for Pattern Identification in Environmental Mixtures. Under review at *Environmental Health Perspectives*, 2021.

3. 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. *Environmental Research*, 2021.

4. Wang G, **Chillrud LG**, McKeown KR. Evidence based Automatic Fact-Checking for Climate Change Misinformation. *SocialSens Workshop on The International AAAI Conference on Web and Social Media 2021*.

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.

Poster Presentations

11. **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.

12. **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.

13. 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, Bash, Zsh
Operating Systems:	UNIX, macOS
Version Control Systems:	GitHub
Databases	MongoDB, NoSQL
Certifications	HIPAA, CITI, Human Subjects Protection
Experience Writing & Editing	academic papers, IRB protocols
Machine Learning 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

RELEVANT COURSEWORK

COMPUTER SCIENCE

- Statistical Machine Learning**
- Machine Learning
- Computational Genomics
- Natural Language Processing
- Artificial Intelligence
- Analysis of Algorithms
- Computer Science Theory
- Advanced Programming
- Computer Systems
- Data Structures & Algorithms

MATHEMATICS

- Probability & Statistics*
- Linear Algebra
- Calculus I, II, & III
- Discrete Mathematics
- Number Theory
- Cryptography
- Real Analysis I**
- Bayesian Statistics**
- Analysis & Optimization**

OTHER

- Geochemistry
- Organic Chemistry I
- General Chemistry I & II
- General Chemistry Lab
- Death Valley Geology
- Intro Linguistics

*Taken in the fall 2021 semester; ***might be taken* in the spring 2022 semester; as a student at **Columbia University's School of Professional Studies**, in the *Post-baccalaureate Studies Program*.

OTHER PROJECTS

11/2020 **Scraping Georgia Jails for Georgia Get Out the Vote**
Wrote a Python web-crawler to scrape Georgia's jails for information needed to help register incarcerated voters. Read more here.

10/2020 **RoBERTa for Claim Detection**
Fine-tuned RoBERTa under-the-hood to identify and rank claims worth fact-checking. Implemented with PyTorch and Scikit-learn. Read more here.

5/2020 **Automatic Diagnosis of COVID-19 Chest X-rays with Neural Nets**
Trained a CNN via transfer learning (TensorFlow) to diagnose patient chest x-rays from: COVID-19, no condition, viral-, or bacterial-pneumonia. Read more here.

5/2020 **SARS-CoV-2 Sequence Analysis**
Identified conserved RNA secondary structures across coronavirus spike proteins in a sequence analysis of SARS-CoV-2. Read more here.

REFERENCES

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