LAWRENCE CHILLRUD

Ph.D. Candidate

Northwestern University, McCormick School of Engineering
Technological Institute, 2145 Sheridan Rd, Evanston, IL 60208
chili@u.northwestern.edu (845) 652-3798
lawrence-chillrud.github.io github.com/lawrence-chillrud

CV Date of Preparation: 11/2024

Education

9/2022 – Present **Ph.D. in Electrical Engineering** *Evanston, IL*
Northwestern University, McCormick School of Engineering
Advisors: Aggelos Katsaggelos, Ph.D. and Lee Cooper, Ph.D.
Specialization in Signals and Systems

9/2021 – 5/2022 **Post-baccalaureate Scholar** *New York, NY*
Columbia University, School of Professional Studies

9/2016 – 5/2020 **B.A. in Computer Science** *New York, NY*
Columbia University, Columbia College
Specialization in Intelligent Systems

Awards and Honors

9/2024 – Present **National Science Foundation Graduate Research Fellowship**
$159,000 in funding over three years
Awarded to support education of "outstanding graduate students... who have demonstrated their potential for significant research achievements in STEM." Proposal entitled *Closing the gap b/w deep ensembles & Bayesian inference in medical imaging*.

Research

9/2022 – Present **Graduate Student Research Fellow** *Chicago, IL*
Northwestern University, Feinberg School of Medicine
Department of Pathology
Advisor: Lee Cooper, Ph.D.

Working in the Computational and Integrative Pathology Group to develop ML methods to analyze H&E stained histopathology whole slide imaging data. Work includes:

- Close collaboration & weekly presentation of work & experimental results w/pathologists
- Weakly-supervised/multiple instance learning for a) Banff lesion score prediction from renal allograft biopsies and b) survival analysis of a cohort of breast cancer patients
- Investigating DL/CNN digital pathology models using state-of-the-art transformer architectures, self-supervised foundation models
- Working with HistomicsTK to clean, manage & query massive digital pathology datasets
- Experiments investigating a) foundation model sensitivity to types of stains used in whole slide images, and b) efficacy of stain normalization approaches
- Dimension reduction algos. to visualize low dim. embedding spaces of learned features
- Developing novel loss function to minimize false positives at a high true positive rate
- Software dev. on glimr API, a simplified wrapper for hyperparam. search w/Ray Tune

9/2022 – Present **Graduate Student Research Fellow** *Evanston, IL*
Northwestern University, McCormick School of Engineering
Department of Electrical & Computer Engineering
Advisor: Aggelos Katsaggelos, Ph.D.

Working in the Image and Video Processing Lab to develop ML models to predict Meningioma/Glioblastoma biomarker statuses from a patient's mpMRI, entailing:

- Close collaboration & weekly presentation of work & experimental results w/radiologists
- Semi-supervised deep learning w/convolutional networks (e.g. U-Nets, transformers)
- Uncertainty quantification experiments w/Bayesian neural networks & deep ensembles
- Explanatory modeling w/rigorous double nested leave-two-out cross-validation approach to train lasso logistic regression models to learn from limited cohort size
- Texture analysis study using custom-extracted PyRadiomics & CoLIAGe features
- Design & validation of preprocessing pipeline to prepare MRI scans in the cohort for analysis, including but not limited to: scan type cleanup, DICOM to NIfTI conversion, N4 bias field correction, skullstripping, intensity standardization, & image registration
- Other projects include active learning to inform the sampling of collocation points in physics informed neural networks modeling binary star system evolution dynamics

10/2020 – 8/2022 **Senior Programmer** *New York, NY*
Columbia University, Mailman School of Public Health
Department of Environmental Health Sciences
Advisor: Marianthi-Anna Kioumourtzoglou, Sc.D.

Worked in makLab to develop interpretable machine learning methods for assessing complex mixtures of environmental exposures. Main project tailored Principal Component Pursuit (PCP), a dimensionality reduction technique from computer vision, for pattern recognition in environmental epidemiology, including development of R software package. Other work included:

- Investigated convex & non-convex approaches to matrix decomposition, dim reduction
- Leveraged Gaussian processes to design faster cross-validated grid searches
- Developed Bayesian non-parametric ensemble model for uncertainty characterization
- Conducted extensive R code reviews for academic papers (reviewed over 9,000 lines)
- Cleaned, visualized & documented public health datasets for various research questions
- Aided in writing, editing of scientific papers & abstracts, presented work at conferences
- Explored methods for source apportionment: PCP, PCA, Autoencoders, Factor Analysis
- Designed, implemented, evaluated environmental/epidemiological health models/analyses

6/2020 – 10/2020 **EHS Research Assistant** *New York, NY*
Columbia University, Mailman School of Public Health
Department of Environmental Health Sciences
Advisor: Marianthi-Anna Kioumourtzoglou, Sc.D.

Worked in makLab to develop interpretable machine learning methods for assessing complex mixtures of environmental exposures:

- Adapted & extended Principal Component Pursuit 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** *New York, NY*
Columbia University
Department of Computer Science
Advisor: Kathleen McKeown, Ph.D.

Worked to develop, train, and validate an automatic fact-checking model for combating misinformation online surrounding COVID-19 & climate-change:

- Researched 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

Peer-Reviewed

7. Wu H, Kalia V, Manz KE, **Chillrud LG**, Dishon NH, Jackson GL, Dye CK, Orvieto R, Aizer A, Levine H, Kioumourtzoglou M-A, Pennell KD, Baccarelli AA, and Machtinger R. Exposome Profiling of Environmental Pollutants in Seminal Plasma and Novel Associations with Semen Parameters. *Environmental Science & Technology*, 2024 58 (31), 13594-13604.

6. Wang G, **Chillrud LG**, Harwood KR, Ananthram A, Subbiah M, McKeown KR. Check-COVID: A Corpus and Task for Fact-Checking COVID-19 Misinformation with Scientific Evidence. *Findings of the Association for Computational Linguistics: ACL 2023*, pages 14114–14127, Toronto, Canada. Association for Computational Linguistics.

5. Cerna-Turoff I, **Chillrud LG**, Rudolph KE, Casey JA. Standards in responsibly sharing cohort data for transparency and reproducibility: response to The Young Lives study. *International Journal of Epidemiology*, 2023 Oct 5;52(5):1666-1669.

4. Tao RH, **Chillrud LG**, Nunez Y, Rowland ST, Boehme AK, Yan J, Goldsmith J, Kioumourtzoglou M-A. Applying Principal Component Pursuit to investigate the association between source-specific fine particulate matter and myocardial infarction hospitalizations in New York City. *Environmental Epidemiology* 7(2):p e243, April 2023.

3. 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. *Environmental Health Perspectives*, 2022.

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. *Environmental Research*. 2022 May 1;207:112229.

Workshop

1. 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.

Presentations

Oral

14. **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. *ENAR 2022*, Houston, TX, March 27-30, 2022.

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.

Invited Talks

12. **Chillrud LG**. Parallel computation in R with the foreach package: A brief introduction. *RClub*, Columbia Mailman School of Public Health, New York, NY, April 7, 2022.

Posters

11. Seesillapachai T, Kunananthaseelan N, **Chillrud LG**, Yousefi S, Cooper LAD. Automated design of deep survival analysis models with SurvivalNet and Glimr. National Cancer Institute Informatics Technology for Cancer Research 2023, Chicago, IL, September 12-13, 2023.

10. **Chillrud LG**, Yan J, Wright J, Goldsmith J, Kioumourtzoglou M-A. Principal Component Pursuit for Source Apportionment from Block Missing Data: An Application to NYC $PM_{2.5}$ Data. *ISEE 2022*, Athens, Greece, September 18-21, 2022.

9. **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*, New York, NY, August 23-26, 2021.

Abstracts

8. Benavides J, Rowland ST, **Chillrud LG**, Gallegos CC, Kumar V, Paisley J, Coull B, Fiore A, Kioumourtzoglou M-A. bneR-A collaborative workflow for air quality modeling and uncertainty characterization using the Bayesian Nonparametric Ensemble. *ISEE 2024*, Santiago, Chile, August 25-28.

7. Wu H, Kalia V, Manz KE, **Chillrud LG**, Dishon NH, Jackson GL, Dye CK, Orvieto R, Aizer A, Levine H, Kioumourtzoglou M-A, Pennell KD, Baccarelli AA, and Machtinger R. Application of Pattern Recognition and Mixture Methods in Exposome Profiling of Seminal Plasma and Associations with Semen Parameters. *ISEE 2024*, Santiago, Chile, August 25-28.

6. Wu H, Kalia V, Manz KE, **Chillrud LG**, Dishon NH, Jackson GL, Orvieto R, Dye CK, Aizer A, Levine H, Kioumourtzoglou M-A, Pennell KD, and Machtinger R, Baccarelli AA. Exposome Profiling Reveals Widespread Environmental Pollutant Exposure in Seminal Plasma and Previously Unknown Associations with Male Fertility. *Fertility and Sterility*, 2023, 120 (4), e304.

5. Benavides J, **Chillrud LG**, DeSerisy M, Cohen J, Goldsmith J, Kioumourtzoglou M-A, Margolis A. Do complex mixtures of prenatal environmental and social exposures explain variation in risk for behavioral symptoms in adolescence? *ISEE 2022*, Athens, Greece, September 18-21.

4. Wu H, Kalia V, Manz KE, **Chillrud LG**, Dishon NH, Orvieto R, Aizer A, Levine H, Kioumourtzoglou M-A, Pennell KD, Machtinger R, Baccarelli AA. Exposomic Analysis of Organic Pollutants in Seminal Plasma and Male Reproductive Parameters. *ISEE 2022*, Athens, Greece, September 18-21.

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. *ISEE 2021*, New York, NY, August 23-26.

2. 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*, New York, NY, August 23-26.

1. 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*, New York, NY, August 23-26.

Technical Skills

Languages: Python, R, LaTeX, C, C++, Java, HTML, CSS, MATLAB, Bash, Zsh

Operating Sys: UNIX, macOS

VC Systems: Git, GitHub

Databases: MongoDB, NoSQL

Certifications: HIPAA, CITI, Human Subjects Protection, Responsible Conduct of Research

ML Libraries: TensorFlow, Keras, PyTorch, Scikit-learn, Hugging Face Transformers, NumPy, SciPy, Pandas, Matplotlib, Seaborn

Dev Tools: Visual Studio Code, iTerm2, Vim, GNU Screen, tmux, RStudio, Jupyter Notebook, Google Cloud Platform, Conda, pip, Homebrew

Relevant Coursework

Courses marked **in bold** indicate Graduate level coursework

Computer Sci: **Machine Learning for Medical Images & Signals**, **Deep Learning for Medical Image Analysis**, **Machine Learning**, **Deep Learning**, **Statistical Pattern Recognition**, **Probabilistic Graphical Models**, **Deep Reinforcement Learning**, **Seminar in Reliable Machine Learning Theory**, **HCI Methods for Health Research**, **Explanation & Reproducibility in Data Driven Science**, **Computational Genomics**, **Natural Language Processing**, **Artificial Intelligence**, **Analysis of Algorithms**, Causal Inference, Computer Science Theory, Advanced Programming in C, Computer Systems, Data Structures & Algorithms

Mathematics: **Random Processes**, **Probability & Statistics**, Foundations of Linear Optimization, Linear Algebra, Calculus I, II, & III, Discrete Mathematics, Number Theory, Cryptography

Other Projects

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 **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. Academic transcripts are available upon request.