

# Chirag Agarwal

## CONTACT INFORMATION

---

Email: [chiragagarwall12@gmail.com](mailto:chiragagarwall12@gmail.com)

Webpage: [chirag126.github.io](http://chirag126.github.io)

## ACADEMIC & PROFESSIONAL EXPERIENCE

---

### Harvard University

Research Fellow

Advisor: [Dr. Hima Lakkaraju](#)

Boston, MA

2023 – Present

### Adobe

Research Scientist

Noida, IN

2022 – 2023

### Harvard University

Postdoctoral Research Fellow

Advisor: [Dr. Marinka Zitnik](#) and [Dr. Hima Lakkaraju](#)

Boston, MA

2020 – 2022

### Auburn University

Research Assistant

Advisor: [Dr. Anh Nguyen](#)

Auburn, AL

Summer 2019

### Robert Bosch LLC

Computer Vision/Augmented Reality Intern

Sunnyvale, CA

Summer 2018

### Tempus labs Inc.

Imaging Science Intern

Chicago, IL

Spring 2018

### Kitware Inc.

Research and Development Intern

Clifton Park, NY

Summer 2017

### Geisinger Health Systems

Research Intern

Danville, PA

Summer 2016

## EDUCATION

---

### University of Illinois at Chicago

Ph.D. in Electrical and Computer Engineering

Chicago, IL

2020

- Committee: [Dr. Dan Schonfeld](#), [Dr. Bharati Prasad](#), [Dr. Mojtaba Soltanalian](#),  
[Dr. Piotr Gmytrasiewicz](#), [Dr. Anh Nguyen](#)
- Thesis: “Robustness and Explainability of Deep Neural Networks”

### University of Illinois at Chicago

M.S. in Electrical and Computer Engineering

Chicago, IL

2018

## SELECTED HONORS & ACHIEVEMENTS

---

[AINet Fellow](#) by DAAD

2021

AI for Social Good Google Workshop with [Dr. Marinka Zitnik](#) and [Dr. Hima Lakkaraju](#) (US \$10,000)

2021

[Spotlight presentation](#), ICML workshop on Human Interpretability in Machine Learning

2020

2 × Research Proposal accepted by Google Cloud Platform (US \$1,000)	2020
Spotlight paper, IEEE Conference on Image Processing (ICIP)	2019
Finalist for the Deans Scholarship Award at UIC	2019

## RESEARCH ARTICLES

---

### Articles in Peer-Reviewed Journals

49. **C. Agarwal**, O. Queen, H. Lakkaraju, M. Zitnik: Evaluating Explainability for Graph Neural Networks, *Nature Scientific Data*, 2023.
48. H. Honarvar, **C. Agarwal**, S. Somani, A. Vaid, J. Lampert, T. Wanyan, V. Y. Reddy, G. N. Nadkarni, R. Miotto1, M. Zitnik, F. Wang, B. S. Glicksberg: Enhancing convolutional neural network predictions of electrocardiograms with left ventricular dysfunction using a novel sub-waveform representation, *Cardiovascular Digital Health Journal*, 2022.
47. **C. Agarwal**, S. Gupta, M. Y. Najjar, T. E. Weaver, X. J. Zhou, D. Schonfeld, B. Prasad: Deep Learning Analyses of Brain MRI to Identify Sleepiness in Treated Obstructive Sleep Apnea: A Pilot Study, *Journal of Sleep and Vigilance (JSV)*, 2022.
46. B. Prasad\*, **C. Agarwal\***, E. Schonfeld, D. Schonfeld, B. Mokhlesi: Deep learning applied to polysomnography to predict blood pressure in obstructive sleep apnea and obesity hypoventilation: A proof-of-concept study, *Journal of Clinical Sleep Medicine (JCSM)*, 2020.
45. **C. Agarwal**, J. Klobusicky, D. Schonfeld: Convergence of backpropagation with momentum for network architectures with skip connections, *Journal of Computational Mathematics (JCM)*, 2019.
44. E. Cha, Y. Veturi, **C. Agarwal**, M. Arbabshirani, S. Pendergrass: Using Adipose Measures from Electronic Health Record Imaging Based Data for Discovery, *Journal of Obesity*, 2018.

### Articles in Peer-Reviewed Conference Proceedings

43. M. Llodes, D. Ganguly, S. Bhatia, **C. Agarwal**: Explain like I am BM25: Interpreting a Dense Model's Ranked-List with a Sparse Approximation, *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2023.
42. A. Seth, M. Hemani, **C. Agarwal**: DeAR: Debiasing Vision-Language Models with Additive Residuals, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
41. S. Deshmukh, A. Dasgupta, B. Krishnamurthy, N. Jiang, **C. Agarwal**, J. Subramanian, G. Theocharous: Trajectory-based Explainability Framework for Offline RL, *International Conference on Learning Representations (ICLR)*, 2023.
40. J. Cheng, G. Dasoulas, H. He, **C. Agarwal**, M. Zitnik: GNNDlete: A General Unlearning Strategy for Graph Neural Networks, *International Conference on Learning Representations (ICLR)*, 2023.
39. V. Giunchiglia, C. V. Shukla, G. Gonzalez, **C. Agarwal**: Towards Training GNNs using Explanation Directed Message Passing, *Proceedings of the First Learning on Graphs Conference (LoG)*, 2022.
38. **C. Agarwal**, E. Saxena, S. Krishna, M. Pawelczyk, N. Johnson, I. Puri, M. Zitnik, H. Lakkaraju : OpenXAI: Towards a Transparent Evaluation of Model Explanations, *Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
37. **C. Agarwal**, D. D'Souza, S. Hooker: Estimating Example Difficulty using Variance of Gradients, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
36. **C. Agarwal**, M. Zitnik, H. Lakkaraju: Probing GNN Explainers: A Rigorous Theoretical and Empirical Analysis of GNN Explanation Methods, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

35. M. Pawelczyk, **C. Agarwal**, S. Joshi, S. Upadhyay, H. Lakkaraju: Exploring Counterfactual Explanations Through the Lens of Adversarial Examples: A Theoretical and Empirical Analysis, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.
34. **C. Agarwal**, H. Lakkaraju, M. Zitnik: Towards a Unified Framework for Fair and Stable Graph Representation Learning, *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021.
33. S. Agarwal, S. Jabbari, **C. Agarwal**, S. Upadhyay, Z. S. Wu, H. Lakkaraju: Towards the Unification and Robustness of Perturbation and Gradient Based Explanations, *International Conference on Machine Learning (ICML)*, 2021.
32. **C. Agarwal\***, S. Khobahi\*, D. Schonfeld, M. Soltanian: CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images, *SPIE Medical Imaging*, 2021.
31. **C. Agarwal**, A. Nguyen: Explaining image classifiers by removing input features using generative models, *Asian Conference on Computer Vision (ACCV)*, 2020.
30. N. Bansal\*, **C. Agarwal\***, A. Nguyen\*: SAM: The Sensitivity of Interpretability Methods to Hyperparameters, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.  
**Oral presentation (Top 5%).**
29. **C. Agarwal**, S. Khobahi, A. Bose, M. Soltanian, D. Schonfeld: Deep-URL: A Model-Aware Approach To Blind Deconvolution Based On Deep Unfolded Richardson-Lucy Network, *IEEE Conference on Image Processing (ICIP)*, 2020.
28. **C. Agarwal**, A. Nguyen, D. Schonfeld: Improving Robustness to Adversarial Examples by Encouraging Discriminative Features, *IEEE Conference on Image Processing (ICIP)*, 2019.  
**Spotlight presentation (Top 10%).**
27. M. Aloraini, M. Sharifzadeh, **C. Agarwal**, D. Schonfeld: Statistical Sequential Analysis for Object-based Video Forgery Detection, *Electronic Imaging*, 2019.
26. N. Khobragade\*, **C. Agarwal\***: Multi-class segmentation of neuronal electron microscopy images using deep learning, *SPIE Medical Imaging*, 2018.
25. **C. Agarwal**, M. Sharifzadeh, D. Schonfeld: CrossEncoders: A complex neural network compression framework, *IS&T International Symposium on Electronic Imaging*, 2018.
24. M. Sharifzadeh, **C. Agarwal**, M. Aloraini, D. Schonfeld: Convolutional neural network steganalysis's application to steganography, *IEEE Visual Communications and Image Processing (VCIP)*, 2017.
23. **C. Agarwal**, A.H. Dallal, M.R. Arbabshirani, A. Patel, G. Moore: Unsupervised quantification of abdominal fat from CT images using Greedy Snakes, *SPIE Medical Imaging*, 2017.
22. A.H. Dallal, **C. Agarwal**, M.R. Arbabshirani, A. Patel, G. Moore: Automatic estimation of heart boundaries and cardiothoracic ratio from chest X-ray images, *SPIE Medical Imaging*, 2017.
21. M.R. Arbabshirani, A.H. Dallal, **C. Agarwal**, A. Patel, G. Moore: Accurate segmentation of lung fields on chest radiographs using deep convolutional networks, *SPIE Medical Imaging*, 2017.
20. **C. Agarwal**, A. Bose, S. Maiti, N. Islam, S.K. Sarkar: Enhanced data hiding method using DWT based on Saliency model, *IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, 2013.
19. S. Maiti, **C. Agarwal**, A. Bose, S.K. Sarkar: Robust data hiding technique in wavelet domain using saliency map, *International Journal of Advances in Engineering and Technology*, 2013.
18. N. Islam S. Maiti, A. Bose, **C. Agarwal**, S. K. Sarkar: An Improved Method of Pre-Filter Based Image Watermarking in DWT Domain, *International Journal of Computer Science and Technology*, 2013.

## Preprints and Workshop Articles

17. **C. Agarwal**: Intriguing Properties of Visual-Language Model Explanations, *Preliminary version presented at RTML Workshop, ICLR 2023*.
16. S. Krishna, **C. Agarwal**, H. Lakkaraju: On the Impact of Adversarially Robust Models on Algorithmic Recourse, *Preliminary version presented at Trustworthy and Socially Responsible Machine Learning Workshop, NeurIPS 2022*.
15. S. Deshmukh, A. Dasgupta, B. Krishnamurthy, N. Jiang, **C. Agarwal**, G. Theodorou, J. Subramanian: Trajectory-based Explainability Framework for Offline RL, *Preliminary version presented at Offline RL Workshop, NeurIPS 2022*.
14. **C. Agarwal**, O. Queen, M. Zitnik: An Explainable AI Library for Benchmarking Graph Explainers, *Preliminary version presented at Graph Learning Benchmarks Workshop, WWW, 2022*.
13. **C. Agarwal**, N. Johnson, M. Pawelczyk, S. Krishna, E. Saxena, M. Zitnik, H. Lakkaraju: Rethinking Stability for Attribution-based Explanations, *Preliminary version presented at PAIR<sup>2</sup> Struct Workshop, ICLR, 2022*.  
**Oral Presentation.**
12. **C. Agarwal**, M. Zitnik, H. Lakkaraju: Towards a Unified Framework for Fair and Stable Graph Representation Learning, *Preliminary version presented at Socially Responsible Machine Learning Workshop, ICML, 2021*.
11. **C. Agarwal**, H. Lakkaraju, M. Zitnik: Towards a Rigorous Theoretical Analysis and Evaluation of GNN Explanations, *Preliminary version presented at Theoretic Foundation, Criticism, and Application Trend of Explainable AI Workshop, ICML, 2021*.
10. M. Pawelczyk, S. Joshi, **C. Agarwal**, S. Upadhyay, H. Lakkaraju: On the Connections between Counterfactual Explanations and Adversarial Examples, *Preliminary version presented at Theoretic Foundation, Criticism, and Application Trend of Explainable AI Workshop, ICML, 2021*.
9. D. D'Souza, Z. Nussbaum, **C. Agarwal**, S. Hooker: A Tale Of Two Long Tails, *Preliminary version presented at Uncertainty & Robustness in Deep Learning Workshop, ICML, 2021*.
8. H. Honarvar, **C. Agarwal**, S. Somani, A. Vaid, J. Lampert, T. Wanyan, V. Y. Reddy, G. N. Nadkarni, R. Miotto1, M. Zitnik, F. Wang, B. S. Glicksberg: A novel representation of electrocardiogram waveforms for enhancing deep learning predictions, *Preliminary version presented at Interpretable Machine Learning in Healthcare Workshop, ICML, 2021*.
7. **C. Agarwal\***, S. Hooker\*: Estimating Example Difficulty using Variance of Gradients, *Preliminary version presented at Human Interpretability in Machine Learning Workshop, ICML, 2020*.
6. **C. Agarwal\***, P. Chen\*, A. Nguyen: Intriguing generalization and simplicity of adversarially trained neural networks, *Preliminary version presented at Human Interpretability in Machine Learning Workshop, ICML, 2020*.  
**Spotlight Presentation.**
5. **C. Agarwal**, B. Dong, D. Schonfeld, A. Hoogs: An explainable adversarial robustness metric for deep learning neural networks, 2018.
4. M. Sharifzadeh, **C. Agarwal**, M. Salarian, D. Schonfeld: A new parallel message-distribution technique for cost-based steganography, 2017.

## Patents

3. S. Deshmukh, A. Dasgupta, **C. Agarwal**, B. Krishnamurthy, G. Theodorou, J. Subramanian.: Novel Trajectory-based Explainability Framework for RL-based Decision Making. Internal Reference: P11853-US.
2. M. Hemani, A. Seth, **C. Agarwal**: Debiasing vision-language models with additive residual learning. Internal Reference: P11919-US.
1. T. Menta, A. Patil, S. Jandial, Balaji K, **C. Agarwal**, M. Sarkar: HASTE: A Novel Method and Apparatus to Estimate Transferability using Hard Subsets. Internal Reference: P11683-US.

## TEACHING EXPERIENCE

---

<b>Guest Lecture</b> at Harvard University <i>Course on Interpretability and Explainability in Machine Learning</i>	Spring 2021, 2023
<b>Teaching Assistant</b> University of Illinois at Chicago <i>Pattern Recognition, Image Analysis &amp; Computer Vision, Digital Signal Processing, Neural Networks.</i>	Spring, Fall 2014 - 2020

## TUTORIALS

---

<a href="#">Explainable ML in the Wild</a> : When Not to Trust Your Explanations	FACCT 2021
----------------------------------------------------------------------------------	------------

## INVITED TALKS

---

<a href="#">Computer Vision Talks</a>	2023
<a href="#">TrustML Young Scientists Seminars</a> at RIKEN-AIP, Japan	2022
Adobe Research: XAI: Challenges and Solutions	2022
<a href="#">CAI Summer School</a> at IIIT-Delhi	2022
<a href="#">LOGML Summer School</a>	2022
Guest Lecture in Interpretability & Explainability course at Harvard	2021
<a href="#">2d3d.ai</a>	2021
<a href="#">W&amp;B</a> - Weights & Biases Salon	2020

## COMMUNITY SERVICE

---

<b>Open Collaboration Initiatives:</b> <a href="#">TrustworthyML Initiative</a> and <a href="#">MLCollective</a>	2021-Present
<b>External Ph.D. Examiner:</b> Jessica Rumbelow - University of St. Andrews	2023
<b>Program Committee for Workshops:</b> <a href="#">XAI4CV</a> - Explainable AI for Computer Vision (XAI4CV) Workshop	CVPR, 2023
<a href="#">SRML</a> - Workshop on Socially Responsible Machine Learning	ICLR, 2022
<a href="#">AdvML</a> - New Frontiers in Adversarial Machine Learning	ICML, 2022
<a href="#">SRML</a> - Workshop on Socially Responsible Machine Learning	ICML, 2021
<a href="#">SeSML</a> - Workshop on Security and Safety in Machine Learning Systems	ICLR, 2021
<a href="#">AROW</a> - Workshop on Adversarial Robustness in the Real World	ECCV, 2020-2021
<a href="#">WHI</a> - Workshop on Human Interpretability in Machine Learning	ICML, 2020
<b>Program Committee for Conferences:</b> AISTATS - International Conference on Artificial Intelligence and Statistics	2023
AAAI - AAAI International Conference on Artificial Intelligence	2023
LOG - Learning on Graphs Conference	2022
FACCT - ACM Conference on Fairness, Accountability, and Transparency	2022-2023
ICLR - International Conference on Learning Representations	2022-2023
NeurIPS - Advances in Neural Information Processing Systems	2021-2023
KDD - ACM SIGKDD Conference on Knowledge Discovery and Data Mining	2021-2023

ICML - International Conference on Machine Learning	2021-2023
WACV - IEEE/CVF Winter Conference on Applications of Computer Vision	2023
CVPR - IEEE/CVF Conference on Computer Vision and Pattern Recognition	2023
ICCV - IEEE/CVF International Conference on Computer Vision	2023
ACL - ACL Rolling Review	2023
<b>Journal Reviewing:</b>	
TMLR - The Transactions on Machine Learning Research	2022-2023
TMI - IEEE Transactions on Medical Imaging	2022