

# NAKUL VERMA

Columbia University ◊ Computer Science Department ◊ New York, USA  
verma@cs.columbia.edu ◊ (212) 853-8471 ◊ <https://www.cs.columbia.edu/~verma>

## EDUCATION

---

**University of California, San Diego**, La Jolla, California, USA

Doctor of Philosophy	Machine Learning	September 2012
Master of Science	Machine Learning	June 2008
Bachelor of Science	Computer Science	December 2004

## EMPLOYMENT

---

**Columbia University, New York City, New York USA** 2017-Present

*Director of MS Data-Science Program*

*Senior Teaching Professor*

- Teaching graduate-level courses in Machine Learning, Algorithms and CS theory.
- Heading research work in understanding and exploiting the intrinsic structure in data (eg. manifold or sparse structure) to design effective machine learning algorithms in missing- or weakly labelled data regimes.

**Janelia Research Campus, Ashburn, Virginia USA** 2013-2017

*Research Specialist*

- Developing biologically plausible statistical models to understand organism behavior from observational data.
- Conduct machine learning research to understand the mathematical structure of behavior.

**Amazon.com Inc., Seattle, Washington USA** 2012-2013

*Research Scientist*

- Developing and analyzing automated models for detecting fraudulent transactions in real time.
- Part of the fraud analytics team; responsible for all live European digital orders.

**Yahoo! Labs, Bangalore, India** Summer 2011

*Visiting Researcher*

- Developed models for improved classification and visualization for hierarchical multi-class data.

**Qualcomm Incorporated, San Diego, California USA** Summer 2008

*Visiting Researcher*

- Applying machine learning and data mining concepts to design provably correct algorithms for phone personalization in a memory constrained environment.
- Part of the Advanced Technologies Group in the Office of the Chief Scientist.

## HONORS AND AWARDS

---

Research & professional service related

- NeurIPS one of the top scoring reviewers, 2019
- ICML reviewer award, 2015, 2020
- Best paper, Wireless Health 2012.

Teaching related

- Janelia Teaching Fellowship, 2015.

Programming related

- Second place in UCSD Programming Contest, 2005.
- Fourth place in Southern California Regional ACM Programming Contest, 2004.
- Third place in UCSD Programming Contest, 2004. (top undergraduate finisher)
- Fourth place in UCSD Datamining Contest, 2004.

University of California, San Diego

- Provost honors, 2001-2004.

## GRANTS AWARDED

---

**SEAS GenAI Education Grant**

2025-26

*Principal Investigator*

- Awarded funding to develop a system to leverage generative AI to enhance student learning.

**Summer@SEAS Research Program**

2025

*Principal Investigator*

- Study and improve the effectiveness of non-linear embedding and visualization techniques in machine learning systems.
- Awarded funding to support undergraduate students for advancing fundamental research.

**Collaboratory Fellows Fund**

2023-26

*Co-Principal Investigator*

- Developing a cross-disciplinary project-based course where teams of students with a diverse skillset work together on creating products or services to address important social problems.
- Received funding to create an interdisciplinary high-impact course.

**Provost Teaching and Learning Grant**

2020, 2021, 2024

*Principal Investigator*

- Investigating data-driven methodologies to improve teaching and learning in tech-heavy classes.
- Awarded funding to develop ML models to learn effective representations of unstructured student solution data and use it to do predictive analytics to improve student performance.

**Alliance Joint Project Grant**

2019-20

*co-Principal Investigator*

- Investigating data-driven methodologies to automatically discover physical laws in mechanics and astrophysics data.

- Awarded seed-funding to conduct high-impact research and foster transatlantic collaborations with researchers.

**Data Science Institute Scholars Grant**

2019, 2020

*Principal Investigator*

- Investigating non-traditional embedding techniques for modern large-scale data.
- Awarded funding twice to support students to work on research projects representation learning.

## INVITED TALKS

---

### **Perils of using t-SNE and friends**

- Theory seminar, Janelia Research, HHMI, 2024.

### **Modern supervised learning techniques for neuroscience**

- CAJAL Neuroscience Institute, Champalimaud Foundation, 2023.

### **Hidden potential of non-Euclidean representations for machine learning**

- Bloomberg LP, CTO Datascience Speaker Series, 2019.

### **Distance preserving embeddings for Riemannian manifolds**

- Carnegie Mellon University, Machine Learning Department, 2015.
- IBM Research, Almaden, 2014.
- University of Washington, Math Department, 2013.
- Yahoo Labs, Bangalore, 2012.

### **Sample complexity of metric learning**

- IBM Research, Almaden, 2015.
- Machine Learning seminar, Janelia Research, HHMI, 2015.

### **A tutorial on metric learning with some recent advances**

- Bay Area Machine Learning Group, 2014.

### **An introduction to statistical theory of learning**

- Neurotheory seminar, Janelia Research, HHMI, 2015.

## PATENTS

---

D. Krishnaswamy, R. Daley, N. Verma. Delivery of targeted content related to a learned and predicted future behavior based on spatial, temporal, and user attributed and behavioral constraints. *US Patent Application # 20110282964*, May 2010.

D. Krishnaswamy, N. Verma, V. Bychkovsky. Method and system using keyword vectors and associated metrics for learning and prediction of user correlation of targeted content messages in a mobile environment. *US Patent Application # 20090125462*, May 2009.

D. Krishnaswamy, N. Verma, V. Bychkovsky. Method and system for keyword correlation in a mobile environment. *US Patent Application # 20090125517*, May 2009.

## EDUCATION ACTIVITIES

---

### ❖ *Courses taught at Columbia*

**Machine Learning (COMS 4771)** 2017-2025

- Took ownership of the course. One of the most in-demand courses. Have streamlined the course and made multiple sections available every semester.

**Unsupervised Learning (COMS 4774)** 2018-2025

- Designed a new graduate-level special topics course. This course introduces modern techniques in unsupervised learning.

**Data-Driven Design for Social Innovation (COMS 4995)** 2024

- Designed a new interdisciplinary course that brings students from various departments to work together on creating products or services to address important real-world social problems.

**Analysis of Algorithms (CSOR 4231)** 2024-2025

- Offered the introductory Algorithms design course.

**Computational Linear Algebra (COMS 3251)** 2021-2023

- Revamped the classic Linear Algebra course to include practical applications and specific focus for Computer Science audience.

**Automata and Complexity Theory (COMS 3261)** 2017

- Offered the introductory Computer Science theory course.

### ❖ *Outreach Activities*

**Professor Advisor to the IEEE Student Chapter Columbia** 2023–Present

- Serving as a professor advisor to help undergraduate and graduate students network with people in industry, academia, and among universities for opportunities and connections.

**Guest speaker at Dalton High-School Robotics Club** Spring 2022

- Invited speaker to generate interest among high-school students to study data analysis and machine learning and the importance of fundamental mathematics to such quantitative fields.

**Guest speaker on Columbia Data Science Society Podcast** Summer 2020

- Invited speaker to discuss the successes and challenges of machine learning in modern society, and how young professionals can enhance their datascience and analysis skills.

**Served on Columbia ADI faculty panel** Fall 2018

- Invited for a panel discussion to introduce students to different career paths in academia, help build an understanding of research, and bridge the gap between faculty and students.

**Hour of Code Volunteer for Grade-school Students** 2017

- Taught grade-school students introductory programming skills once a month.

**California Summer School for Math and Science (COSMOS) program** Summer 2006

- Taught high-school students develop creative multi-media applications, and introductory programming skills.

❖ *Courses taught at UC San Diego*

**Teaching Assistant and Course Tutor**

2003–2007

- Advanced Data Structures – WI04, SP05, SP06, FA06
- Compiler Construction – WI05, FA05, WI06, FA06, WI07
- Design and Analysis of Algorithms – SP05
- Discrete Mathematics – SP04
- Intro to Computer Science and OOP: Java – WI03, SP03, FA03, FA04, WI07

**PROFESSIONAL SERVICE**

---

Reviewer and program committee member (multiple years):

- Nature Methods
- Journal of Machine Learning Research (JMLR)
- Machine Learning Journal (MLJ)
- IEEE Transactions on Information Theory (T-IT)
- Discrete and Computational Geometry (DCG)
- Journal of the Royal Statistical Society (JRSS)
- Transactions on Machine Learning Research (TMLR)
- Conference on Learning Theory (COLT)
- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Artificial Intelligence and Statistics (AISTATS)
- Association for the Advancement of Artificial Intelligence (AAAI)
- Neural Computation
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Signal Processing (TSP)
- IEEE Signal Processing Letters (SPL)
- International Journal of Computer Vision (IJCV)
- EURASIP Journal of Wireless Communications and Networking (JWCN)
- ACM SIGKDD Conference on Knowledge Data Discovery and Data Mining (KDD)
- IEEE International Conference on Data Mining (ICDM) High Dimensional Mining.
- Asian Conference on Machine Learning (ACML)
- International Conference on Distributed Computing and Networking (ICDCN)

Grant reviewer for:

- Columbia SEAS Interdisciplinary Research Seed (SIRS) Funding, and SEAS Translational Acceleration Research (STAR) Funding program, 2025
- Columbia University's Research Initiatives in Science & Engg. (RISE) competition, 2022
- Columbia University's Data Science and Health Initiative (DASHI) program, 2022
- Israel Science Foundation, 2022

## PUBLICATIONS

---

(2400+ citations; *h*-index: 20)

41. **MathEL: Autonomous self-improving AI for IMO combinatorics.**  
I. Drori, G. Longhitano, M. Mao, M. Bafna, N. Verma, A. Shporer, A. Amit, M. Udell.  
*In preparation*, 2026.
40. **t-SNE Exaggerates Clusters, Provably**  
N. Bergam, S. Snoeck, N. Verma.  
*International Conference on Learning Representations (ICLR)*, 2026.
39. **Compressibility Barriers to Neighborhood-Preserving Data Visualizations.**  
S. Snoeck, N. Bergam, N. Verma.  
*Algorithmic Learning Theory (ALT)*, 2026.
38. **Diverse Inference and Verification for Advanced Reasoning.**  
I. Drori, G. Longhitano, M. Mao, S. Hyun, Y. Zhang, S. Park, Z. Meeks, X. Zhang, B. Segev, H. Yong, N. Verma, A. Shporer, A. Amit, M. Udell.  
*Computing Research Repository (CoRR) abs/2502.09955*, 2025.
37. **Using Deep Autoregressive Models as Causal Inference Engines.**  
D. Im, K. Zhang, N. Verma, K. Cho.  
*Transactions on Machine Learning Research (TMLR)*, 2025.
36. **LogicLearner: Guided Practice for Proofs in Propositional Logic.**  
A. Inamdar, N. Verma and A. Salleb-Aouissi.  
*Under Review at International Journal of Artificial Intelligence in Education (IJAIE)*, 2024.
35. **Contrastive Loss is All You Need to Recover Analogies as Parallel Lines.**  
E. Ri, F. Lee, N. Verma.  
*Association for Computational Linguistics (ACL) workshop on Representation Learning*, 2023.
34. **A Dataset for Learning University STEM Courses at Scale and Generating Questions at a Human Level.**  
I. Drori, S. Zhang, Z. Chin, R. Shuttleworth, A. Lu, L. Chen, B. Birbo, M. He, P. Lantigua, S. Tran, G. Hunter, B. Feng, N. Cheng, R. Wang, Y. Hicke, S. Surbehera, A. Raghavan, A. Siemenn, N. Singh, J. Lynch, A. Shporer, N. Verma, T. Buonassisi, A. Solar-Lezama.  
*Educational Advances in Artificial Intelligence (EAAI)*, 2023.
33. **Improving Model Training via Self-learned Label Representations.**  
X. Yu, N. Verma.  
*Computing Research Repository (CoRR) abs/2209.04528*, 2022.
32. **A Neural Network Solves, Explains, and Generates University Math Problems by Program Synthesis and Few-shot Learning at Human Level.**  
I. Drori, S. Zhang, R. Shuttleworth, L. Tang, A. Lu, E. Ke, K. Liu, L. Chen, S. Tran, N. Cheng, R. Wang, N. Singh, T. Patti, J. Lynch, A. Shporer, N. Verma, E. Wu, and G. Strang.  
*Proceedings of National Academy of Sciences (PNAS)*, 2022.

31. **How I Stopped Worrying About Training Data Bugs and Started Complaining.**  
L. Flokas, W. Wu, J. Wang, N. Verma, E. Wu.  
*Proceedings of the Sixth Workshop on Data Management for End-To-End Machine Learning (DEEM)*, 2022.
30. **Solving Probability and Statistics Problems by Program Synthesis at Human Level and Predicting Solvability.**  
L. Tang, E. Ke, N. Singh, B. Feng, D. Austin, N. Verma, I. Drori.  
*International Conference on Artificial Intelligence in Education (AIED)*, 2022.
29. **Complaint-Driven Training Data Debugging at Interactive Speeds.**  
L. Flokas, W. Wu, Y. Liu, J. Wang, N. Verma and E. Wu.  
*International Conference on Data Management (ACM SIGMOD)*, 2022.
28. **An Analysis of Document Graph Construction Methods for AMR Summarization.**  
F. Lee, C. Kedzie, N. Verma and K. McKeown.  
*Computing Research Repository (CoRR) abs/2111.13993*, 2021.
27. **Solving Linear Algebra by Program Synthesis.**  
I. Drori and N. Verma.  
*Computing Research Repository (CoRR) abs/2111.08171*, 2021.
26. **Automated Symbolic Law Discovery: A Computer Vision Approach.**  
H. Xing, A. Salleb-Aouissi, N. Verma.  
*Association for the Advancement of Artificial Intelligence (AAAI)*, 2021.
25. **Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics.**  
B. Cowgill, F. Dell’Acqua, A. Chaintreau, N. Verma and S. Deng.  
*ACM Conference on Economics and Computation (EC)*, 2020.
24. **Cortical Pattern Generation During Dexterous Movement is Input-driven.**  
B. Sauerbrei, J. Guo, J. Cohen, M. Mischiati, W. Guo, M. Kabra, N. Verma, B. Mensh, K. Branson, A. Hantman.  
*Nature*, 2019.
23. **Noise-Tolerant Fair Classification.**  
A. Lamy, Z. Zhong, A. Menon, N. Verma.  
*Neural Information Processing Systems (NeurIPS)*, 2019.
22. **Metric Learning on Non-Euclidean Representations.**  
M. Aalto, N. Verma.  
*Computing Research Repository (CoRR) abs/1902.01738*, 2019.
21. **Model-Agnostic Meta-Learning using Runge-Kutta Methods.**  
D. Im, Y. Jiang, N. Verma.  
*Computing Research Repository (CoRR) abs/1910.07368*, 2019.
20. **Meta-Learning to Cluster.**  
Y. Jiang, N. Verma.  
*Computing Research Repository (CoRR) abs/1910.14134*, 2019.

19. **Stochastic Neighbor Embeddings Under  $f$ -Divergences.**  
D. Im, N. Verma, K. Branson.  
*Computing Research Repository (CoRR) abs/1811.01247*, 2018.
18. **Time-accuracy Tradeoffs in Kernel Prediction: Controlling Prediction Quality.**  
S. Kpotufe, N. Verma.  
*Journal of Machine Learning Research (JMLR)*, 2017.
17. **Sample Complexity of Learning Mahalanobis Distance Metrics.**  
N. Verma, K. Branson.  
*Neural Information Processing Systems (NIPS)*, 2015.
16. **Distance Preserving Embeddings for General  $n$ -Dimensional Manifolds.**  
N. Verma.  
*Journal of Machine Learning Research (JMLR)*, 2013.
15. **Efficient Energy Management and Data Recovery in Sensor Networks Using Latent Variables Based Tensor Factorization.**  
B. Milosevic, J. Yang, N. Verma, S. Tilak, P. Zappi, E. Farella, L. Benini, T. Rosing.  
*Sixteenth IEEE International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, 2013.
14. **Mobile Technologies and the Exposome: Continuous Assessment of Environmental Exposures Critical to Health.**  
K. Patrick, B. Griswold, G. Norman, F. Raab, J. Kerr, B. Demchak, I. Krueger, S. Dasgupta, N. Nikzad, C. Ziftci, N. Verma, P. Zappi, E. Bales, N. Quick.  
*Lecture Notes on Wireless Healthcare Research*, pp. 25–32, 2013.
13. **Learning from Data with Low Intrinsic Dimension.**  
N. Verma.  
Ph.D. thesis, *Dept. of Computer Science and Engineering, UC San Diego*, 2012.
12. **Distance Preserving Embeddings for Low-Dimensional Manifolds.**  
N. Verma.  
*Conference on Learning Theory (COLT)*, 2012.
11. **Learning Hierarchical Similarity Metrics.**  
N. Verma, D. Mahajan, S. Sellamanickam, V. Nair.  
*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
10. **CitiSense: Improving Geospatial Environmental Assessment of Air Quality Using a Wireless Personal Exposure Monitoring System.**  
N. Nikzad, N. Verma, C. Ziftci, E. Bales, N. Quick, P. Zappi, K. Patrick, S. Dasgupta, I. Krueger, T. Rosing, W. Griswold.  
*Wireless Health*, 2012. (best paper)
9. **Citizensense: Mobile Air Quality Sensing for Individuals and Communities.**  
C. Ziftci, N. Nikzad, N. Verma, P. Zappi, E. Bales, I. Krueger, W. Griswold.  
*Conference on Systems, Programming, and Applications: Software for Humanity*, 2012.
8. **A Note on Random Projections for Preserving Paths on a Manifold.**  
N. Verma.  
*UC San Diego Tech. Report CS2011-0971*, 2011.

7. **Latent Variables Based Data Estimation for Sensing Applications.**  
N. Verma, P. Zappi, T. Rosing.  
*IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2011.
6. **Multiple Instance Learning with Manifold Bags.**  
B. Babenko, N. Verma, P. Dollar, S. Belongie.  
*Twenty-Eighth International Conference on Machine Learning (ICML)*, 2011.
5. **CitiSense - Adaptive Services for Community-Driven Behavioral and Environmental Monitoring to Induce Change.**  
N. Nikzad, C. Ziftci, P. Zappi, N. Quick, P. Aghera, N. Verma, B. Demchak, K. Patrick, H. Shacham, T. Rosing, I. Krueger, W. Griswold, S. Dasgupta.  
*UC San Diego Tech. Report CS2011-0961*, 2011.
4. **Which Spatial Trees are Adaptive to Intrinsic Dimension?**  
N. Verma, S. Kpotufe, S. Dasgupta.  
*Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2009.
3. **Mathematical Advances in Manifold Learning.**  
N. Verma.  
*UC San Diego Tech. Report*, 2008.
2. **Learning the Structure of Manifolds Using Random Projections.**  
Y. Freund, S. Dasgupta, M. Kabra, N. Verma.  
*Twenty-First Conference on Neural Information Processing Systems (NIPS)*, 2007.
1. **A Concentration Theorem for Projections.**  
S. Dasgupta, D. Hsu, N. Verma.  
*Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.