Subhabrata Majumdar
✉ *zoom.subha@gmail.com*
🌐 *shubhobm.github.io*

Curriculum Vitae

Education

2017 **PhD Statistics**, *University of Minnesota Twin Cities, Minneapolis, MN, USA*

2012 **MS Statistics**, *Indian Statistical Institute, Kolkata, India*

2010 **BS Statistics**, *Indian Statistical Institute, Kolkata, India*

Experience

2023- **Co-founder, Head of AI**, *Vijil*

2023 **Principal Machine Learning Scientist**, *GEICO*

2022-23 **Applied Scientist II**, *Amazon (Twitch)*

2021-22 **Senior Applied Scientist**, *Splunk*

2018-21 **Senior Inventive Scientist**, *AT&T Labs Research*

2017-18 **Postdoctoral Researcher**, *University of Florida*

2016 **Research Intern**, *IBM Research*

Research Interests

AI Security, AI Safety, Trustworthy AI/ML, Representation learning, Statistical machine learning.

Publications

Theory and Methods

2024 D. Rosati, J. Wehner, K. Williams, L. Bartoszcze, D. Atanasov, R. Gonzales, **S. Majumdar**, C. Maple, H. Sajjad, F. Rudzicz. Representation noising effectively prevents harmful fine-tuning on LLMs. *Neural Information Processing Systems (NeurIPS)*, to appear.

2023 F.T. Brito, V.A.E. Farias, C. Flynn, J.C. Machado, **S. Majumdar**, D. Srivastava. Global and local differentially private release of count-weighted graphs. *Proceedings of the ACM on Management of Data (SIGMOD)*, 1 (2), 1-25.

2023 R. Rustamov, **S. Majumdar**. Intrinsic sliced wasserstein distances for comparing collections of probability distributions on manifolds and graphs. *International Conference on Machine Learning (ICML)*, 29388-29415.

2023 V.A.E. Farias, F.T. Brito, C. Flynn, J.C. Machado, **S. Majumdar**, D. Srivastava. Local Dampening: Differential Privacy for Non-numeric Queries via Local Sensitivity. *The VLDB Journal*, 32, 1191–1214.

2022 **S. Majumdar**, S. Chatterjee. Feature selection using e-values. *International Conference on Machine Learning (ICML)*, 14753-14773.

2022 **S. Majumdar**, G. Michailidis. Joint estimation and inference for data integration problems based on multiple multi-layered gaussian graphical models. *Journal of Machine Learning Research*, 23, 1-53.

2022 **S. Majumdar**, S. Chatterjee. On weighted multivariate sign functions. *Journal of Multivariate Analysis*, 105013.

2020 A. Ghosh, **S. Majumdar**. Ultrahigh-dimensional Robust and Efficient Sparse Regression using Non-Concave Penalized Density Power Divergence. *IEEE Transactions on Information Theory*, 66 (12), 7812-7827.

2018 **S. Majumdar**, S. Chatterjee. Non-convex penalized multitask regression using data depth-based penalties. *Stat*, 7, e174.

Applications

2024 M.A. Ayub, **S. Majumdar**. Embedding-based classifiers can detect prompt injection attacks. *Conference on Applied Machine Learning in Information Security (CAMLIS)*.

2023 **S. Majumdar**, S. Basu, M. McGue, S. Chatterjee. Simultaneous selection of multiple important single nucleotide polymorphisms in familial genome wide association studies data. *Scientific Reports*, 13 (1), 8476.

2022 G. Subramaniam, **S. Majumdar**. Network Security Modelling with Distributional Data. *Conference on Applied Machine Learning in Information Security (CAMLIS)*.

2021 N. Derzsy, **S. Majumdar**, R. Malik. An Interpretable Graph-based Mapping of Trustworthy Machine Learning Research. *International Conference on Complex Networks (CompleNet)*.

2019 S.C. Basak, **S. Majumdar**, and others. Computer-Assisted and Data Driven Approaches for Surveillance, Drug Discovery, and Vaccine Design for the Zika Virus. *Pharmaceuticals*, 12, 157.

2019 **S. Majumdar**, S.C. Basak, and others. Finding needles in a haystack: determining key molecular descriptors associated with the blood-brain barrier entry of chemical compounds using machine learning. *Molecular Informatics*, 38, 1800164.

2019 B. Han, **S. Majumdar**, and others. Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates. *Epidemics*, 27, 59-65.

2018 **S. Majumdar**, S.C. Basak. Beware of external validation! – A Comparative Study of Several Validation Techniques used in QSAR Modelling. *Current Computer Aided Drug Design*, 14, 284–291.

2018 **S. Majumdar**, S.C. Basak, and others. Mathematical structural descriptors and mutagenicity assessment: A study with congeneric and diverse data sets. *SAR and QSAR in Environmental Research*, 29, 579–590.

2016 **S. Majumdar**, S.C. Basak. Exploring intrinsic dimensionality of chemical spaces for robust QSAR model development: A comparison of several statistical approaches. *Current Computer Aided Drug Design*, 12, 294–301.

2015 S.C. Basak, **S. Majumdar**. Prediction of Mutagenicity of Chemicals from Their Calculated Molecular Descriptors: A Case Study with Structurally Homogeneous versus Diverse Datasets. *Current Computer Aided Drug Design*, 11, 117–123.

2015 E. Potash, J. Brew, A. Loewi, **S. Majumdar**, A. Reece, J. Walsh, E. Rozier, E. Jorgenson, R. Mansour, and R. Ghani. Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. *Proceedings of KDD*, 2039–2047.

2013 **S. Majumdar**, S.C. Basak, G.D. Grunwald. Adapting Interrelated Two-Way Clustering Method for Quantitative Structure-Activity Relationship (QSAR) Modeling of Mutagenicity/Non-Mutagenicity of a Diverse Set of Chemicals. *Current Computer Aided Drug Design*, 9, 463–471.

Preprints

2024 H. Raj, V. Gupta, D. Rosati, **S. Majumdar**. Improving Consistency in Large Language Models through Chain of Guidance. *Under review at Transactions of Machine Learning Research.*

2024 D. Rosati, G. Edkins, H. Raj, D. Atanasov, **S. Majumdar**, J. Rajendran, F. Rudzicz, H. Sajjad. Defending against Reverse Preference Attacks is Difficult. *Under review at 3rd IEEE Conference on Secure and Trustworthy Machine Learning (SaTML).*

2024 L. Derczynski, E. Galinkin, J. Martin, **S. Majumdar**, N. Inie. garak: A Framework for Security Probing Large Language Models. *arXiv:2406.11036.*

Books

2023 Y. Pruksachatkun, M. Mcateer, **S. Majumdar**. *Practicing Trustworthy Machine Learning: Consistent, Transparent, and Fair AI Pipelines.* O'Reilly Media.

Book Chapters

2024 **S. Majumdar**. *Standards for LLM Security.* In: Large Language Models in Cybersecurity, Springer, 225–231.

2024 **S. Majumdar**, T. Vogelslang. *Towards Safe LLMs Integration.* In: Large Language Models in Cybersecurity, Springer, 243–247.

2019 **S. Majumdar**. *Data-driven Strategies to Model and Mitigate the Threat of Zika.* In: Zika virus: Basic biology, surveillance, vaccine design and anti-Zika drug discovery: Computer-assisted strategies to combat the menace, Nova Science Publishers, Inc., 129-152.

2015 S.C. Basak, **S. Majumdar**. *Current Landscape of Hierarchical QSAR Modeling and its Applications: Some Comments on the Importance of Mathematical Descriptors as well as Rigorous Statistical Methods of Model Building and Validation.* In: Advances in Mathematical Chemistry and Applications: Vol. 1, Elsevier and Bentham e-Books, 251-281.

2015 U. Mukherjee, **S. Majumdar**, S. Chatterjee, *Fast and Robust Supervised Learning in High Dimensions Using the Geometry of the Data.* In: Advances in Data Mining: Applications and Theoretical Aspects, ser. Lecture Notes in Computer Science, 9165, 109–123.

Refereed Workshops

2022 H. Raj, D. Rosati, **S. Majumdar**. Measuring Reliability of Large Language Models through Semantic Consistency. *NeurIPS 2022 ML Safety Workshop* **(Best paper award)**.

2022 C. Flynn, A. Guha, **S. Majumdar**, D. Srivastava, Z. Zhou. Towards Algorithmic Fairness in Space-Time: Filling in Black Holes. *NeurIPS 2022 Workshop on Trustworthy and Socially Responsible Machine Learning.*

2022 **S. Majumdar**, C. Flynn, R. Mitra. Detecting bias in the presence of spatial autocorrelation. *NeurIPS 2021 Algorithmic Fairness through the Lens of Causality and Robustness workshop.*

2021 C. Last, P. Pramanik, N. Saini, A.S. Majety, D.-H. Kim, M. García-Herranz, **S. Majumdar**. Towards an Open Global Air Quality Monitoring Platform to Assess Children's Exposure to Air Pollutants in the Light of COVID-19 Lockdowns. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.*

Patents

Please find a list of 20+ filed and granted patents here.

Talks

Keynotes

June 2023	CVPR Workshop on Fair, Data Efficient and Trusted Computer Vision, Vancouver, Canada
Sep 2022	8th Indo-US Workshop on Mathematical Chemistry, virtual
Aug 2022	Faculty Development Programme, Saranathan College of Engineering, Trichy, India
July 2022	NAACL Workshop on Trustworthy Natural Language Processing, Seattle, WA

Panels

Sep 2024	OctoAI Builders Roundtable: Secure GenAI for Enterprises, virtual
Oct 2022	ML:Integrity Conference, virtual
Feb 2020	National Institute of Statistical Sciences (NISS) Industry Career Fair, virtual

Invited Talks

Jan 2025	International Conference on Data Management, Analytics & Innovation, Kolkata, India
Apr 2024	LinkedIn, Bellevue, WA
Aug 2023	O'Reilly Expert Webinar, virtual
Nov 2022	Open Data Science Conference West, San Francisco, CA
Apr 2022	University of Washington RAISE lab, Seattle, WA
Dec 2020	Data Science Salon, virtual
Nov 2020	(Lecture series) Dept. of Mathematics and Statistics at Indian Institute of Technology, Kanpur, India
Mar 2019	Women in Machine Learning and Data Science meetup, New York, NY
May 2018	International Indian Statistical Association (IISA) Conference, Gainesville, FL
May 2018	Savvysherpa, Inc., Minneapolis, MN
Dec 2017	IISA Conference, Hyderabad, India
Dec 2017	Indian Statistical Institute, Kolkata, India
Aug 2016	(Student paper) IISA Conference, Corvallis, OR

Awards

2016–17	University of Minnesota Interdisciplinary Doctoral Fellowship
2016–17	University of Minnesota School of Statistics Martin Award
2016	IISA Conference student travel award
2015	5th International Workshop on Climate Informatics travel award
2014–16	University of Minnesota School of Statistics travel award
2012	Debesh-Kamal Scholarship, Ramakrishna Mission Institute of Culture, Kolkata, India
2008–12	KVPY national fellowship, Department of Science and Technology, Govt. of India
2005–08	National scholar, National Council of Educational Research and Training, Govt. of India

Advising and Mentorship

2024-	Aditya Karan, PhD student at UIUC / internship mentor, research advisor
2024-	Md. Ahsan Ayub, Postdoc at Vanderbilt University / research advisor

2022- Harsh Raj, incoming MS student at Northeastern Univ / internship mentor, research advisor

2022- Domenic Rosati, PhD student at Dalhousie University / research advisor

2020 Christina Last, MS at Massachusetts Institute of Technology / internship mentor

2020 Prithviraj Pramanik, AQAI / internship mentor

2019–21 Felipe Brito, Universidade Federal do Ceara, Brazil / research advisor

2019–21 Victor Farias, Universidade Federal do Ceara, Brazil / research advisor

Teaching

As Teaching Assistant at School of Statistics, Univ. of Minnesota

Fall 2014 STAT 8051 - Advanced Regression Techniques

Spring 2014 STAT 3022 - Data Analysis

Fall 2013 STAT 5021 - Statistical Analysis

STAT 5031 - Statistical Methods for Quality Improvement

Spring 2013 STAT 5303 - Designing Experiments

STAT 5401 - Applied Multivariate Methods

Fall 2012 STAT 3011 - Introduction to Statistical Analysis

Service

Reviewing

Journals IEEE Transactions on Information Theory, Statistica Sinica, Sankhya B, Scientific Reports, Biometrics, R Journal, Applied Computing and Informatics, Current Computer-Aided Drug Design, Australasian Medical Journal

Conferences AAAI, AISTATS, CAMLIS, ICML, IAAI, NeurIPS, PAKDD

Workshops TrustNLP at NAACL, ML-RSA at NeuRIPS, CHI Extended Abstracts

Organizing

2025 Program Committee member, IISA Conference

2024–25 Co-secretary and IT Committee Chair, IISA

2021 Organizing team, Trustworthy ML Symposium

2017–18 Session Chair, IISA Conferences