Aaron Schein

www.aaronschein.com

Research Interests

probabilistic modeling • Bayesian methodology • tensor decomposition • causal inference • large language models • complex networks • time series analysis • computational social science • political science • ···

Education

PhD Computer Science	University of Massachusetts Amherst	2019
M.S. Computer Science	University of Massachusetts Amherst	2017
M.A. Linguistics	University of Massachusetts Amherst	2012
B.A. Linguistics	University of Massachusetts Amherst	2011
B.A. Political Science	University of Massachusetts Amherst	2011

Academic Appointments

Assistant Professor	University of Chicago Department of Statistics & Data Science Institute	2022–present
Postdoctoral Fellow	Columbia University Mentors: David Blei, Donald Green	2019–2022

Professional Experience

Visiting Research Scientist	Google DeepMind	2025–present
Senior Research Scientist	Ocurate	2021–2023
Senior Research Scientist	PredictWise	2021–2022
Research Intern	Microsoft Research (New York, NY) *Computational Social Science group*	Summer 2014, 2015
Software Engineering Intern	Google (Mountain View, CA) *Machine Intelligence group*	Summer 2013
Artificial Intelligence Engineer	MITRE Corporation (McLean, VA) *Human Language Technologies group*	2011–2013

Teaching

Instructor	STAT 348 *Modern Methods of Applied Statistics* University of Chicago	Spring 2023–2026
Instructor	DATA 221/231 *Intro to Machine Learning* University of Chicago	Winter 2024–2026
Instructor	STAT 451 *Topics in Advanced Bayesian Methodology* University of Chicago	Fall 2023
Instructor	CICS 191 FYS *Philosophy of A.I. & Theories of Consciousness* University of Massachusetts Amherst	Fall 2016

Preprints and papers under review

- J. Hood and **A. Schein**. Near-Universal Multiplicative Updates for Nonnegative Einsum Factorization. *arXiv preprint arXiv:2602.02759*, 2026.
- J. Lederman and **A. Schein**. Modeling latent underdispersion with discrete order statistics. *arXiv preprint arXiv:2507.09032*, 2025. Under revision at *Journal of the American Statistical Association (JASA)*.

- Y. Wang, **A. Schein**, J. Shou, and D. Blei. The Synthetic Control Method with Many Control Outcomes. 2025. Under revision at *Journal of Machine Learning Research (JMLR)*.
- S. O'Hagan and **A. Schein**. Measurement in the Age of LLMs: An Application to Ideological Scaling. *arXiv preprint arXiv:2312.09203*, 2023.
- A. N. Albert, P. Flaherty, and **A. Schein**. Doubly Non-Central Beta Matrix Factorization for Stable Dimensionality Reduction of Bounded Support Matrix Data. *arXiv preprint arXiv:2410.18425*, 2024. Under revision at *Journal of Machine Learning Research (JMLR)*.

Journal articles

- J. Hood, C. D. Bacco, and **A. Schein**. Broad Spectrum Structure Discovery in Large-Scale Higher Order Networks. *arXiv preprint arXiv:2505.21748*, 2025. To appear in *Nature Communications*.
- E. Dong, **A. Schein**, Y. Wang, and N. Garg. Addressing Discretization-Induced Bias in Demographic Prediction. *PNAS Nexus*, 2025.
- D. P. Green, L. Ong, K. Rutherford, and **A. Schein**. Do billboard advertisements increase voter turnout? a large-scale field experiment. *Quarterly Journal of Political Science*, 19(3), 2024.
- B. J. Radford, Y. Dai, N. Stoehr, **A. Schein**, M. Fernandez, and H. Sajid. Estimating conflict losses and reporting biases. *Proceedings of the National Academy of Sciences (PNAS)*, 120(34), 2023.
- S. He*, **A. Schein***, V. Sarsani, and P. Flaherty. A Bayesian nonparametric model for inferring subclonal populations from structured DNA sequencing data. *Annals of Applied Statistics*. * = Equal contribution.

Conference proceedings

- C. Wolfram and **A. Schein**. Layers at Similar Depths Generate Similar Activations Across LLM Architectures. In *Conference on Language Models (COLM)*, 2025.
- J. Kim, J. Evans, and **A. Schein**. Linear Representations of Political Perspective Emerge in Large Language Models. In *International Conference on Learning Representations (ICLR)*, 2025. Selected for a full oral presentation (among 5% of accepted papers).
- N. Stoehr, K. Du, V. Snæbjarnarson, R. West, R. Cotterell, and **A. Schein**. Activation Scaling for Steering and Interpreting Language Models. In *Association for Computational Linguistics (ACL)*, 2024.
- K. Du, V. Snæbjarnarson, N. Stoehr, J. C. White, **A. Schein**, and R. Cotterell. Context versus prior knowledge in language models. In *Association for Computational Linguistics (ACL)*, 2024.
- E. Dong, **A. Schein**, Y. Wang, and N. Garg. What to Do about Argmax Bias: An Application to US Voter Registration Data. In *Conference on Fairness, Accountability, and Transparency (FaaCT)*, 2024.
- J. Hood and **A. Schein**. The ALℓ_0CORE Tensor Decomposition for Sparse Count Data. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.
- N. Stoehr, B. J. Radford, R. Cotterell, and **A. Schein**. The Ordered Matrix Dirichlet for state-space models. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.
- N. Stoehr, L. T. Hennigen, J. Valvoda, R. West, R. Cotterell, and **A. Schein**. An ordinal latent variable model of conflict intensity. In *The 61st Annual Meeting Of The Association For Computational Linguistics (ACL)*, 2023.
- N. Stoehr, R. Cotterell, and **A. Schein**. Sentiment as an ordinal latent variable. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2023.

- **A. Schein**, A. Nagulpally, H. M. Wallach, and P. Flaherty. Doubly non-central Beta matrix factorization for DNA methylation data. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021.
- **A. Schein**, K. Vafa, D. Sridhar, V. Veitch, J. Quinn, J. Moffet, D. M. Blei, and D. P. Green. Assessing the effects of friend-to-friend texting on turnout in the 2018 US midterm elections. *The Web Conference (WWW)*, 2021.
- **A. Schein**, S. Linderman, M. Zhou, D. M. Blei, and H. M. Wallach. Poisson-randomized gamma dynamical systems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- **A. Schein**, Z. Wu, A. Schofield, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *International Conference on Machine Learning (ICML)*, 2019.
- **A. Schein**, M. Zhou, and H. M. Wallach. Poisson–gamma dynamical systems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2016. Selected for a full oral presentation (among 8% of accepted papers).
- **A. Schein**, M. Zhou, D. M. Blei, and H. M. Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. *International Conference on Machine Learning (ICML)*, 2016.
- **A. Schein**, J. Paisley, D. M. Blei, and H. M. Wallach. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2015.
- K. Miller, E. Richardson, S. McLeod, J. Finley, and **A. Schein**. International multicultural name matching competition: Design, execution, results, and lessons learned. *International Conference on Language Resources and Evaluation (LREC)*, 2012.

Workshop proceedings and refereed abstracts

- C. Wolfram and **A. Schein**. World Models and Consistent Mistakes in LLMs. In *ICML Workshop on World Models*, 2025.
- J. Jian and **A. Schein**. Bayesian Tensor Decomposition for Uncovering Complex Dependencies in International Trade. In *International Conference on Statistics and Data Science (ICSDS)*, 2025.
- J. Kim, J. Evans, and **A. Schein**. Linear Representations of Political Perspective Emerge in Large Language Models. In *Annual Meeting of the Society for Political Methodology (PolMeth)*, 2025.
- A. Lee, T. Nief, K. Lum, P. Toulis, and **A. Schein**. The Veil of Elections: Assessing the Causal Effect of Homophily on Partisan Segregation. In *Annual Meeting of the Society for Political Methodology (PolMeth)*, 2025.
- J. Kim, J. Evans, and **A. Schein**. Linear Representations of Political Perspective Emerge in Large Language Models. In *International Conference on Computational Social Science (IC2S2)*, 2025.
- J. Lederman and **A. Schein**. Bayesian Analysis of Latent Underdispersion Using Discrete Order Statistics. In *Objective Bayes Methodology Conference (O'Bayes)*, 2025.
- J. Hood, C. D. Bacco, and **A. Schein**. Modeling Higher Order Interactions Among Latent Communities in Hypergraphs. In *SIAM Conference on Applications of Dynamical Systems*, 2025.
- J. Hood and **A. Schein**. Hurdle Conjugate Priors for Scalable Tucker Decomposition. In *ICML Workshop on Structured Probabilistic Inference & Generative Modeling (SPIGM@ICML)*, 2024.
- S. O'Hagan, K. Park, and **A. Schein**. Representations of Ideology in Large Language Models. In *International Conference on Computational Social Science (IC2S2)*, 2024.
- S. O'Hagan and **A. Schein**. Eliciting ideological scales from large language models. *New Directions in Analyzing Text as Data (TADA)*, 2023.
- N. Stoehr, B. J. Radford, R. Cotterell, and **A. Schein**. Escalatory dynamics in international relations. In *International Conference on Computational Social Science (IC2S2)*, 2023.

- **A. Schein**, D. M. Blei, and D. P. Green. Assessing the effects of friend-to-friend texting on turnout in the 2020 US presidential election. *Conference on Digital Experimentation (CODE@MIT)*, 2021.
- **A. Schein**, D. M. Blei, and D. P. Green. A pair of large-scale digital field experiments reveal large effects of friend-to-friend texting on voter turnout in the 2018 and 2020 US elections. *Conference on Politics and Computational Social Science (PaCSS)*, 2021.
- A. Nagulpally, **A. Schein**, H. M. Wallach, and P. Flaherty. Matrix factorization for DNA methylation data. In *Great Lakes Bioinformatics Conference (GLBIO)*, 2021.
- **A. Schein**, M. Zhou, D. M. Blei, and H. M. Wallach. An experimental study of friend-to-friend GOTV text messages in the 2018 midterm elections. In *International Conference on Computational Social Science (IC2S2)*, 2020. Winner of the Best Presentation award.
- **A. Schein**, Z. Wu, A. Schofield, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *International Conference on Computational and Methodological Statistics*, 2019.
- A. Schofield, **A. Schein**, Z. Wu, M. Zhou, and H. M. Wallach. Toward practical and locally private inference of topic models. *New Directions in Analyzing Text as Data (TADA)*, 2018.
- A. Schofield, **A. Schein**, Z. Wu, M. Zhou, and H. M. Wallach. A variational inference approach for locally private inference of Poisson factorization models. *Advances in Approximate Bayesian Inference*, 2018.
- **A. Schein**, Z. Wu, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *Advances in Approximate Bayesian Inference*, 2017.
- B. Kim, **A. Schein**, B. Desmarais, and H. M. Wallach. A network model for dynamic textual communications with application to government email corpora. *New Directions in Analyzing Text as Data (TADA)*, 2017.
- B. Kim, **A. Schein**, B. Desmarais, and H. M. Wallach. A network model for dynamic textual communications with application to government email corpora. *Political Networks Conference*, 2017.
- **A. Schein**, P. Flaherty, M. Zhou, D. Sheldon, and H. M. Wallach. Beta Tucker decomposition for DNA methylation data. *NeurIPS Workshop on "Computational Biology"*, 2016. Selected for oral presentation (among 20% of accepted papers).
- **A. Schein**, M. Zhou, D. M. Blei, and H. M. Wallach. Modeling international relations with Bayesian Poisson Tucker decomposition. In *International Conference on Computational Social Science (IC2S2)*, 2016.
- **A. Schein**, M. Zhou, J. Paisley, D. M. Blei, and H. M. Wallach. Dynamic Bayesian Poisson tensor factorization. In *Tenth Conference on Bayesian Nonparametrics (BNP)*, 2015.
- **A. Schein**, M. Zhou, D. M. Blei, and H. M. Wallach. Modeling topic-partitioned assortativity and disassortativity in dyadic event data. *NeurIPS Workshop on "Networks in the Social and Information Sciences"*, 2015. Winner of Best Student Poster award (prize: $400).
- **A. Schein**, M. Zhou, J. Paisley, D. M. Blei, and H. M. Wallach. Dynamic Bayesian Poisson tensor factorization. In *Tenth Conference on Bayesian Nonparametrics (BNP)*, 2015.
- **A. Schein**, J. Paisley, D. M. Blei, and H. M. Wallach. Inferring polyadic events with Poisson tensor factorization. *NeurIPS Workshop on "From Graphs to Rich Data"*, 2014.
- **A. Schein**, J. Moore, and H. M. Wallach. Inferring multilateral relations from dynamic pairwise interactions. *NeurIPS Workshop on "Frontiers of Network Analysis"*, 2013.

Newspaper articles

- **A. Schein**. Joe Biden's 'virtual ground game' gamble seems to have paid off. *Financial Times Opinion*, November 8, 2020.

Invited Talks

Invited speaker	Research seminar	Mila	Dec 2025
Invited speaker	QHMESS Workshop	University of Chicago	Nov 2025
Invited speaker	Statistics department seminar	University of Wisconsin	Oct 2025
Invited speaker	Research seminar	Google DeepMind	Jun 2025
Invited speaker	Statistics Meets Tensors Workshop	IMSI Institute	May 2025
Invited speaker	Quantitative Methods Workshop	University of Chicago	May 2025
Invited speaker	Marshall School statistics seminar	Univ. Southern California	Apr 2025
Invited speaker	Statistics department seminar	Notre Dame	Mar 2025
Invited speaker	CS department seminar	ETH Zurich	Feb 2025
Invited speaker	Statistics department seminar	Duke University	Nov 2024
Invited speaker	School of Information seminar	University of Michigan	Oct 2024
Invited speaker	Quantitative Methods Workshop	University of Chicago	May 2024
Invited speaker	Linguistics department seminar	University of Chicago	Apr 2024
Keynote speaker	Conference of Social Computing	Tsinghua University	Sep 2023
Invited speaker	Empirical Inference seminar	Max Planck Institute	Mar 2023
Invited speaker	Research Seminar Series	Max Planck Institute	Jun 2022
Invited speaker	ICSA Applied Stat Symposium	University of Michigan	Jun 2023
Invited speaker	Cyber Toaster Summer School	Los Alamos National Lab	May 2023
Invited speaker	Statistics department seminar	University of Waterloo	Oct 2023
Invited speaker	Quantitative Methods Workshop	University of Chicago	Jan 2023
Invited speaker	ICBINB Workshop	NeurIPS	Dec 2024
Invited speaker	Marketing Quant Seminar	Columbia University	Mar 2021
Invited speaker	CS department seminar	Georgia Tech	Jul 2018
Invited speaker	Research seminar	Google Brain (Mountain View)	Jun 2018
Invited speaker	Research seminar	Google Brain (Mountain View)	Jan 2017
Invited speaker	Machine Learning & Friends Lunch	UMass Amherst	Sep 2013
Invited speaker	Data Mining Symposium	MITRE Corporation	Aug 2012
Invited speaker	Human Language Tech Symposium	MITRE Corporation	Aug 2012

Academic Community Service

Workshop organizing

Co-founder	"I Can't Believe It's Not Better" (ICBINB) Initiative	2015–present
Co-organizer	NeurIPS Workshop "I (Still) Can't Believe It's Not Better" (ICBINB)	2021
Co-organizer	NeurIPS Workshop "I Can't Believe It's Not Better" (ICBINB)	2020
Co-organizer	NeurIPS Workshop "Practical Bayesian Nonparametrics"	2016
Co-organizer	NeurIPS Workshop "Bayesian Nonparametrics: The Next Generation"	2016

Reviewing for journals

- **Statistics:** Annals of Applied Statistics (AoAS) • Bayesian Analysis • Journal of the American Statistical Association (JASA) • Journal of the Royal Statistical Society Series B (JRSS-B) • Journal of the Royal Statistical Society Series C (JRSS-C) • Statistics and Computing
- **Machine learning & signal processing:** IEEE Transactions on Signal Processing • Journal of

Machine Learning Research (JMLR) • Physical Review E • Transactions of Machine Learning Research (TMLR)

- **Social & political science:** American Journal of Political Science (AJPS) • Political Analysis • Sociological Methods and Research
- **Other applications:** PloS Biology • PloS One

Conference reviewing and chairing

Area chair	Neural Information Processing Systems (NeurIPS)	2022–2025
Area chair	Artificial Intelligence and Statistics (AISTATS)	2022–2025
Area chair	Conference on Uncertainty in Artificial Intelligence (UAI)	2022–2024
Reviewer	Conference on Artificial Intelligence (AAAI)	2015
Reviewer	Artificial Intelligence and Statistics (AISTATS)	2015–2021
Reviewer	International Conference on Machine Learning (ICML)	2015–2021
Reviewer	International Joint Conference on AI (IJCAI)	2015, 2016
Reviewer	Neural Information Processing Systems (NeurIPS)	2015–2021
Reviewer	Conference on Uncertainty in Artificial Intelligence (UAI)	2015–2021

University & Departmental Service

Core faculty	Center for CSS	Univ. of Chicago	Fall 2024–present
Mentor	Stats consulting	Univ. of Chicago	Fall 2024, Winter 2025
Member	Stats prelim committee	Univ. of Chicago	2023, 2024, 2025
Member	Stats curriculum committee	Univ. of Chicago	2022–present
Member	Stats faculty search	Univ. of Chicago	2023
Member	Stats PhD admissions	Univ. of Chicago	2022
Member	DSI curriculum committee	Univ. of Chicago	2022–present
Co-organizer	DSI Seminar Series	Univ. of Chicago	2024
Member	DSI postdoc search	Univ. of Chicago	2022, 2023
Member	DSI "Rising Stars" committee	Univ. of Chicago	2022, 2023
Member	DSI PhD admissions	Univ. of Chicago	2023, 2024
Member	CS PhD admissions	Univ. of Chicago	2022, 2023
Co-organizer	DSI Lecture Series	Columbia Univ.	2017–2020

Funding

Crankstart Foundation "Bayesian Small-Area Estimation for Structural Democracy" 2025
Award amount: $40,000 (12 months).

MacArthur Foundation "Translating Research to Practice: Responsible ML" 2023–2024
Award amount: $60,000 (12 months). Co-PI: Kristian Lum

DSI Data & Democracy Award, "Posterior Inference from Noisy Census Data" 2023–2024
Award amount: $50,000 (12 months). Co-PIs: Aloni Cohen, Moon Duchin

DSI Data & Democracy Award, "Data of Democracy" 2023–2024
Award amount: $50,000 (12 months). Co-PIs: Raul Castro Fernandez

Advising & Mentorship

Current advisees

Postdoctoral researchers:

- Patricia Chiril, Data Science Institute, University of Chicago
- Jie Jian, Data Science Institute, University of Chicago

Doctoral students:

- Laura DeFalco, Department of Statistics, University of Chicago
- Ankur Garg, Department of Statistics, University of Chicago
- Sulagna Ghosh, Department of Statistics, University of Chicago
- John Hood, Department of Statistics, University of Chicago
- Jimmy Lederman, Department of Statistics, University of Chicago
- Amber Lee, Department of Statistics, University of Chicago
- Nihar Mauskar, Department of Computer Science, University of Chicago
- Chris Wolfram, Department of Computer Science, University of Chicago

Masters students:

- Darin Keng, Department of Statistics, University of Chicago
- Pippa Lin, Department of Statistics, University of Chicago
- Arnav Rastogi, Department of Statistics, University of Chicago
- Xiaolong Wang, Department of Statistics, University of Chicago
- Keyao Wu, Department of Statistics, University of Chicago
- Haolin Yang, Department of Statistics, University of Chicago
- Peter Zhang, Computational Social Science, University of Chicago

Former advisees

Doctoral students:

- Niklas Stoehr, Institute for Machine Learning, ETH Zurich 2021-2024
  (Now: Research Scientist, Google DeepMind)

Masters students:

- Ian Joffe, Department of Statistics, University of Chicago 2024-2025
- Yilong Chen, Department of Statistics, University of Chicago 2024-2025
  (Now: PhD student, Department of Statistics, Northwestern University)
- Yu Sen Lui, Department of Statistics, University of Chicago 2024-2025
- Qizhong Zhang, Department of Statistics, University of Chicago 2023-2024
  (Now: PhD student, Department of Statistics, Stanford University)
- Chenfeng Li, Department of Statistics, University of Chicago 2023-2024
- Austin Koort, Department of Statistics, University of Chicago 2023-2024
- Jungho Lee, Department of Statistics, University of Chicago 2022-2023
  (Now: PhD student, Department of Statistics, Carnegie Mellon University)
- Seung Chul (Eric) Lee, Department of Statistics, University of Chicago 2022-2023
  (Now: PhD student, McCombs School of Business, University of Texas at Austin)
- Boxuan Zhou, Department of Statistics, University of Chicago 2022-2023

(Now: Research professional, Booth School of Business, University of Chicago)

Undergraduate students:

- Anqi Qu, Department of Economics, University of Chicago 2024-2025
  (Now: PhD student, Department of Statistics, Oxford University)
- Joshua Shou, Department of Statistics, University of Chicago 2022-2023

Doctoral committee member

- Alexander Hoyle, Department of Computer Science, University of Maryland August 2024
- Anjali Albert, Department of Mathematics and Statistics, UMass Amherst June 2024