Rishi Bommasani

CONTACT INFORMATION

Department of Computer Science
Stanford University

https://rishibommasani.github.io
nlprishi@stanford.edu

RESEARCH INTERESTS

Societal impact of AI
foundation models, evaluation, inequity, systemic harm, governance, norms, policy, power

EDUCATION

Stanford University

Ph.D. Candidate, Computer Science, September 2020 – Present
Advisors: Percy Liang, Dan Jurafsky
Committee: Percy Liang, Dan Jurafsky, Chris Manning
Funding: NSF GRFP

Cornell University

M.S. Computer Science, August 2019 – May 2020
B.A. Computer Science, August 2016 – May 2019
B.A. Mathematics, August 2016 – May 2019
Thesis: *Generalized Optimal Linear Orders*
Advisor: Claire Cardie
Committee: Claire Cardie, Bobby Kleinberg

PEER-REVIEWED PAPERS

Rishi Bommasani, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, Percy Liang. Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? *Proceedings of the 36th Conference on Neural Information Processing Systems* (NeurIPS), 2022.

Jason Wei, Yi Tay, **Rishi Bommasani** et al. Emergent Abilities of Large Language Models. *Transactions of Machine Learning Research* (TMLR), 2022. **Outstanding Survey Paper.**

Rishi Bommasani, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, Percy Liang. Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? *ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (EAAMO), 2022.

Yacine Jernite et al. Data Governance in the Age of Large-Scale Data-Driven Language Technology. *ACM Conference on Fairness, Accountability, and Transparency* (FAccT), 2022.

Rishi Bommasani and Claire Cardie. Intrinsic Evaluation of Summarization Datasets. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2020.

Rishi Bommasani, Kelly Davis, Claire Cardie. Interpreting Pretrained Contextualized Representations via Reductions to Static Embeddings. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (ACL), 2020.

Rishi Bommasani, Zhiwei Steven Wu, Alexandra Schofield. Towards Private Synthetic Text Generation. *Machine Learning with Guarantees* (NeurIPS Workshop), 2019.

Rishi Bommasani. Long-Distance Dependencies Don't Have to Be Long: Simplifying through Provably (Approximately) Optimal Permutations. *Context and Compositionality in Biological and Artificial Neural Systems* (NeurIPS workshop), 2019.

Rishi Bommasani and Claire Cardie. Towards Understanding Position Embeddings. *BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (ACL workshop), 2019.

Rishi Bommasani. Long-Distance Dependencies Don't Have to Be Long: Simplifying through Provably (Approximately) Optimal Permutations. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (ACL), 2019.

Rishi Bommasani, Arzoo Katiyar, Claire Cardie. SPARSE: Structured Prediction using Argument-Relative Structured Encoding. *Proceedings of the Third Workshop on Structured Prediction for NLP* (NAACL workshop), 2019.

PAPERS UNDER REVIEW

Rishi Bommasani*, ..., Percy Liang*. On the Opportunities and Risks of Foundation Models. *Journal of Machine Learning Research* (JMLR), 2023. **Under Review.**

Percy Liang*, **Rishi Bommasani***, Tony Lee* et al. Holistic Evaluation of Language Models. *Transactions of Machine Learning Research* (TMLR), 2023. **Under Review.**

Rishi Bommasani and Percy Liang. Trustworthy Social Bias Measurement. *61th Annual Meeting of the Association for Computational Linguistics* (ACL), 2023. **Under Review.**

Rishi Bommasani. Evaluation for Change. *61th Annual Meeting of the Association for Computational Linguistics* (ACL), 2023. **Under Review.**

Mina Lee et al. Evaluating Human-Language Model Interaction. *61th Annual Meeting of the Association for Computational Linguistics* (ACL), 2023. **Under Review.**

Deepak Narayanan, Keshav Santhanam, Peter Henderson, **Rishi Bommasani**, Tony Lee, Percy Liang. Evaluating Efficiency-Capability Tradeoffs for Black-Box Autoregressive Language Models. *Machine Learning and Systems* (MLSys), 2023. **Under Review.**

PAPERS IN PREPARATION

Rishi Bommasani, Thomas Liao*, Dilara Soylu*, Kathleen A. Creel, Percy Liang. Ecosystem Graphs: Documenting the Social Footprint of Foundation Models. *To be released on arXiv*, February 2023. **To be submitted to FAccT 2023.**

Rishi Bommasani*, Connor Toups*, Shibani Santurkar, Kathleen A. Creel, Sarah Bana, Dan Jurafsky, Percy Liang. Homogeneous Outcomes for Individuals from Deployed ML APIs. *To be released on arXiv*, April 2023. **To be submitted to NeurIPS 2023.**

Rishi Bommasani, Sarah Bana, Kathleen A. Creel, Shibani Santurkar, Inioluwa Deborah Raji, Dan Jurafsky, Percy Liang. Homogeneous Outcomes in Hiring. *To be released on arXiv*, May 2023. **To be submitted to PNAS 2023.**

WRITINGS

Rishi Bommasani, Percy Liang, Tony Lee. Holistic Evaluation of Language Models. *Annals of the New York Academy of Sciences (NYAS)*, February 2023. **In preparation.**

Rishi Bommasani, Daniel Zhang, Tony Lee, Percy Liang. Holistic Evaluation of Language Models. *Policy brief*, January 2023. **In preparation.**

Rishi Bommasani. Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? *Stanford AI and CRFM blogs*, January 2023. **In preparation.**

Rishi Bommasani, Percy Liang, Tony Lee. Language Models are Changing AI: The Need for Holistic Evaluation. *Stanford CRFM and HAI blogs*, November 2022.

Jason Wei and **Rishi Bommasani**. Examining Emergent Abilities in Large Language Models. *Stanford HAI blog*, September 2022.

Percy Liang, **Rishi Bommasani**, Kathleen A. Creel, Rob Reich. The Time Is Now to Develop Community Norms for the Release of Foundation Models. *Stanford CRFM and Stanford HAI blogs; op-ed in Protocol*, September 2022.

Rishi Bommasani and Percy Liang. Reflections on Foundation Models. *Stanford CRFM, Stanford HAI, and the Gradient blogs*, October 2021.

Siddharth Karamcheti*, Laurel Orr*, et al. Mistral — A Journey towards Reproducible Language Model Training. *Stanford CRFM blog*, August 2021.

Invited Talks

Evaluation in AI. Princeton CS + Center for Information Technology Policy (Host: Arvind Narayanan). 2023.

Trustworthy Social Bias Measurement. ML Collective Seminar (Host: Rosanne Liu). February 2023.

Holistic Evaluation of Language Models. ENS Paris Cognitive Machine Learning Seminar (Host: Emmanuel Dupoux, Tú Anh Nguyen, Mathieu Rita). February 2023.

Evaluation in AI. Technion - Israel Institute of Technology (Host: Yonatan Belinkov, Zachary Bamberger). January 2023.

Holistic Evaluation of Language Models. Partnership on AI (Host: Madhulika Srikumar, Elissa Redmiles). January 2023.

Holistic Evaluation of Language Models. CMU CS 15-884 Theoretical and Empirical Foundations of Modern Machine Learning (Host: Aditi Raghunathan). December 2022.

Foundation models. AI Seminar at Ohio State University (Host: Yu Su). October 2022.

Foundation models: Below the surface. Adani Group (Gautam Adani) @ Stanford HAI. October 2022.

Holistic evaluation of language models. Stanford Computing and Society (Host: Roshni Sahoo). October 2022.

Holistic evaluation of language models. National AI Advisory Council @ Stanford HAI. October 2022.

Foundation models: Below the surface. MunichRe @ Stanford HAI. September 2022.

Foundation models: Below the surface. Sony (company-wide; host: Yuki Mitsufuji). August 2022.

Fireside chat on foundation models (w/ Percy Liang). Congressional Boot Camp on Artificial Intelligence @ Stanford HAI. August 2022.

Systemic harms: Picking on the same person. Fairness Lunch at Stanford (Host: Omer Reingold and Judy Shen). May 2022. Holistic evaluation of language models. Google @ Stanford HAI. April 2022.

Foundation models: Below the surface. Wells Fargo (company-wide; host: Angelina Yang). December 2021.

Foundation models: Below the surface. Stanford Vision Lab (Host: Fei-Fei Li and Shyamal Buch). November 2021.

Foundation models: Below the surface. MIT NLP Seminar (Host: Jacob Andreas and Belinda Li). October 2021.

Foundation models: Below the surface. Facebook AI Research (Host: Myle Ott). October 2021.

ADVISING

Agnieszka (Aga) Koc [B.A. CS, Cambridge University, 2019]
After: Engineer at Google.
Albert Tsao [B.S. CS, Cornell University, 2020]
After: M.S. CS at Cornell University.
Aman Achpal [Engineer at Microsoft]
Anna (Wei-An) Huang [B.S. CS, Cornell University, 2021]
After: Engineer at Microsoft.
Connor Toups [M.S. CS, Stanford University, 2023]
Dilara Soylu [M.S. CS, Stanford University, 2022]
Gokul Dharan [M.S. CS, Stanford University, 2022]
After: Engineer at Zipline.
Han (Quintessa) Qiao [B.S. CS, Cornell University, 2022]
After: Engineer at Meta.
Joseph Kihang'a [B.A. French, Cornell University, 2018]
Julie Phan [B.S. CS, Cornell University, 2020]
Nathan Kim [B.S. CS, Stanford University, 2024]
Ryan Chi [B.S and M.S. CS, Stanford University, 2024]
Sabhya Chhabria [B.A. CS, Cornell University, 2022]
After: M.S. CS at Princeton University.
Siddharth Sharma [High School Student, 2021]
After: B.S. CS at Stanford University.
Virginia Adams [Senior Applied Scientist at NVIDIA]
After: M.S. CS at Stanford University.
Wenyi Guo [B.S. CS, Cornell University, 2022]
After: M.Eng. CS at Cornell University.
Ye Jiang [M.Eng. CS, Cornell University, 2020]

FUNDING

Foundation Models: Integrating Technical Advances, Social Responsibility, and Applications.
Hoffman-Yee Grant, Stanford HAI (lead PI: Percy Liang). 2022
Amount: $500,000
Holistic Benchmarking of Language Models, Google (PI: Percy Liang). 2022.
Amount: $105,000
Outcome Homogenization and Algorithmic Monoculture, Stanford HAI (PI: Percy Liang). 2022.
Amount: $50,000 Microsoft Azure Credits
Social Bias Acquisition of Language Models, Stanford HAI (PI: Percy Liang). 2021.
Amount: $15,000 Microsoft Azure Credits
Research on Foundation Models, Google (PI: Percy Liang). 2021.
Amount: $15,000
Graduate Research Fellowship, National Science Foundation. 2020.
Amount: $138,000
NeurIPS Travel Grant. 2019.
Amount: $900
ACL Student Scholarship. 2019.
Amount: $2,300
Mozilla Research Travel Grant. 2019
Amount: $3,500

Awards

Computer Science Prize for Academic Excellence and Leadership, Cornell University
Outstanding Teaching Assistant Award (6x), Cornell University
Phi Beta Kappa
Dean's List, Cornell University
Magna Cum Laude with Distinction in all Subjects, Cornell University

Teaching Experience

Large Language Models (Stanford CS 324) Winter 2022
Graduate Teaching Assistant
Natural Language Understanding (Stanford CS 224U) Spring 2021
Graduate Teaching Assistant
Natural Language Processing (Cornell CS 5740) Spring 2020
Graduate Teaching Assistant
Natural Language Processing (Cornell CS 4740) Fall 2019
Instructor, Graduate Teaching Assistant
Honors Discrete Mathematics (Cornell CS 2802) Spring 2019
Head Teaching Assistant
Natural Language Processing (Cornell CS 4740) Fall 2018
Head Teaching Assistant
Discrete Mathematics (Cornell CS 2800) Fall 2018
Teaching Assistant
Discrete Mathematics (Cornell CS 2800) Spring 2018
Teaching Assistant
Discrete Mathematics (Cornell CS 2800) Fall 2017
Teaching Assistant

Service

Leadership and Organization
The Center for Research on Foundation Models (director: Percy Liang). Ongoing.
The First Workshop on Foundation Models (w/ Percy Liang). 2022.

University Service
AI Audit Challenge (w/ Marietje Schaake, Daniel Zhang). 2022.
Stanford Ethics and Society Review (Chair: Michael Bernstein). 2022.
Stanford NLP Retreat Organizer (w/ Tianyi Zhang, John Hewitt, Tatsu Hashimoto). 2022.
Stanford Computer Science PhD Admissions Committee (Chair: Karen Liu). 2022, 2023.
Stanford Computer Science PhD Visit Day Organizer (AI). 2021, 2022.
Stanford Computer Science PhD Student-Applicant Support Program Reviewer. 2021, 2022.

Reviewing
ACL, EMNLP, NAACL, AACL, ARR, COLING, various workshops

Last Updated

January 28, 2023