Rishi Bommasani

INFORMATION

Department of Computer Science
Stanford University

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

RESEARCH

I work on the societal impact of AI systems, especially foundation models.

EDUCATION

Stanford University

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

Cornell University

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

PAPERS

Rishi Bommasani, Kevin Klyman, Betty Xiong, Daniel Zhang, Nestor Maslej, Sayash Kapoor, Shayne Longpre, Percy Liang. The Foundation Model Transparency Index. 2023.

Connor Toups*, **Rishi Bommasani***, Kathleen A. Creel, Sarah Bana, Dan Jurafsky, Percy Liang. Homogeneous Outcomes for Individuals from Deployed ML APIs. *Proceedings of Neural Information Processing Systems* (NeurIPS). 2023.

Deepak Narayanan, Keshav Santhanam, Peter Henderson, **Rishi Bommasani**, Tony Lee, Percy Liang. Cheaply Evaluating Inference Efficiency Metrics for Autoregressive Transformer APIs. *Proceedings of Neural Information Processing Systems* (NeurIPS). 2023.

Rishi Bommasani, Dilara Soylu, Thomas I. Liao, Kathleen A. Creel, Percy Liang. Ecosystem Graphs: Documenting the Social Footprint of Foundation Models. *To be submitted to FAccT 2024.*

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

Mina Lee et al. Evaluating Human-Language Model Interaction. *Transactions on Machine Learning Research* (TMLR), 2023.

Rishi Bommasani and Percy Liang. Trustworthy Social Bias Measurement. *To be submitted to FAccT 2024.*

Rishi Bommasani. Evaluation for Change. *Proceedings of the Association for Computational Linguistics* (ACL). 2023.

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

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 Neural Information Processing Systems* (NeurIPS). 2022.

Jason Wei, Yi Tay, **Rishi Bommasani** et al. Emergent Abilities of Large Language Models. *Transactions on 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? *Proceedings of the 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. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency* (FAccT). 2022.

Rishi Bommasani*, ..., Percy Liang*. On the Opportunities and Risks of Foundation Models. *Under review at JMLR*. 2021.

Rishi Bommasani and Claire Cardie. Intrinsic Evaluation of Summarization Datasets. *Proceedings of 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 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. *Association for Computational Linguistics: Student Research Workshop* (ACL SRW). 2019.

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

POLICY

Rishi Bommasani. Drawing Lines: Tiers for Foundation Models. *Policy brief*. 2023.

Caroline Meinhardt et al. By the Numbers: Tracking The AI Executive Order. *Policy brief*. 2023.

Rishi Bommasani et al. Decoding the White House AI Executive Order's Achievements. *Policy brief*. 2023.

Rishi Bommasani, Kevin Klyman, Daniel Zhang, Marietje Schaake, Percy Liang. Do Foundation Model Providers Comply with the Draft EU AI Act? *Policy brief*. 2023.

Rishi Bommasani, Sayash Kapoor, Daniel Zhang, Arvind Narayanan, Percy Liang. Joint Stanford-Princeton response on AI Accountability Policy. *Request from Comment from the US National Telecommunications and Information Administration (NTIA)*. 2023.

Rishi Bommasani, Daniel Zhang, Tony Lee, Percy Liang. Improving Transparency in AI Language Models: A Holistic Evaluation. *Policy brief*. 2023.

IN PREPARATION

Rishi Bommasani. Towards Compromise: A Concrete Two-tier Proposal for Foundation Models in the EU AI Act. *Policy brief*. 2023.

Rishi Bommasani, Sayash Kapoor, Kevin Klyman, Daniel Zhang, Shayne Longpre, Ashwin Ramaswami, Arvind Narayanan, Percy Liang. Governing Open Foundation Models. *Policy brief*. 2023.

Sayash Kapoor*, **Rishi Bommasani*** et al. Open Foundation Models. 2023.

Rishi Bommasani, Sarah Bana, Kathleen A. Creel, Connor Toups, Dan Jurafsky, Percy Liang. Homogeneous Outcomes in Algorithmic Hiring. *To be submitted to PNAS 2024.*

INVITED TALKS

Transparency for Foundation Models: A Lost Cause or a Valiant Fight?. Workshop on **Sociotechnical AI Safety** (Host: Seth Lazar; Discussant: Tyna Eloundou). November 2023.

Foundation Models and the EU AI Act. Roundtable with **UK Secretary of State for Science, Innovation & Technology Michelle Donelan**. November 2023.

Policy for Foundation Models. **Accenture** @ Stanford HAI. November 2023.

Foundation Models and the EU AI Act. **EU Parliament**: Private briefing of the European Parliament (Host: Members of European Parliament Eva Maydell and Dragos Tudorache). November 2023.

Foundation Models and the EU AI Act. Private briefing of **ITRE and INTA delegations of EU Parliament** @ Stanford HAI. November 2023.

Foundation Models and the EU AI Act. Private briefing of **German AI and Data delegation** @ Stanford HAI. October 2023.

Foundation Models and the EU AI Act. Private briefing of **Italian Vice-President of the Chamber of Deputies Anna Ascani** @ Stanford HAI. October 2023.

Evaluation of Foundation Models. Spotlight at Data Nutrition Project Workshop (Host: Kasia Chmielinski). September 2023.

Foundation Models and the EU AI Act. Private briefing of **Spanish Secretary for Digital Affairs Carme Artigas** @ Stanford. September 2023.

The societal impact of foundation models. **Nonprofit Management Institute** Conference. September 2023.

Foundation models. **PwC** @ Stanford HAI. August 2023.

Foundation models. **Banco Itau** @ Stanford HAI. August 2023.

Governance of foundation models. **Congressional Boot Camp on Artificial Intelligence** @ Stanford HAI. August 2023.

Foundation Models and the EU AI Act. **EU Parliament**: Private briefing at the European Parliament in Brussels, Belgium (Host: Members of European Parliament Eva Maydell and Dragos Tudorache). July 2023.

The Societal Impact of Foundation Models. **United Nations**: Invited Talk at the Conference on Large Generative Models in Geneva, Switzerland (Host: Philipp Hacker and Sarah Hammer). July 2023.

Language Models, Law and Policy. Roundtable Panelist at **NYU** School of Law (Host: Paul Friedl

and Gabriel Nicholas). July 2023.

Foundation Models in Academia and Industry. Panelist at **ICML Workshop** on Efficient Systems for Foundation Models. July 2023.

Making Foundation Models Transparent. **NIST** AI Metrology Colloquium Series (Host: Afzal Godil). June 2023.

Ethics of Language Models. Invited Speaker at **Northeastern** AI and Data Ethics Summer School (Host: Kathleen A. Creel). June 2023.

Societal Impact of AI. **Princeton** CS + Center for Information Technology Policy (Host: Arvind Narayanan). May 2023.

Holistic Evaluation of Foundation Models. **Google** @ Stanford HAI. March 2023.

The Center for Research on Foundation Models. **Barclays** @ Stanford HAI. March 2023.

The Center for Research on Foundation Models. **Axa** @ Stanford HAI. February 2023.

Holistically Evaluating Language Models on the Path to Evaluating Foundation Models. Guest lecture at **MIT** MAS.S68 Generative AI for Constructive Communication (Host: Deb Roy, Shayne Longpre, Hang Jiang, Hope Schroeder). February 2023.

Evaluation in AI. **ML Collective** Seminar (Host: Rosanne Liu). February 2023.

Evaluation in AI. **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 and Zachary Bamberger). January 2023.

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

Holistic Evaluation of Language Models. Guest lecture at **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]
After: Corporate Development at D.E. Shaw.
Dilara Soylu [M.S. CS, Stanford University, 2022]
After: Ph.D. CS at Stanford University.
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.
Jonathan Xue [High School Student, 2023]
Joseph Kihang'a [B.A. French, Cornell University, 2018]
Julie Phan [B.S. CS, Cornell University, 2020]
Kevin Klyman [M.S. International Policy, Stanford University, 2025]
After: J.D. Law at Harvard University.
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 Renewal, Stanford HAI (lead PI: Percy Liang). 2023

Amount: $1,000,000
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

Human-Centered Natural Language Processing (Stanford CS 329X) Spring 2023
Graduate Teaching Assistant
Foundation 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
Society Lead at the The Center for Research on Foundation Models. 2021–Present.
New Horizons in Generative AI: Science, Creativity, and Society (w/ Percy Liang, Surya Ganguli).

2023.
Workshop on Responsible and Open Foundation Models (w/ Sayash Kapoor, Percy Liang, Arvind Narayanan). 2023.
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, 2024.
Stanford Computer Science PhD Visit Day Organizer (AI). 2021, 2022.
Stanford Computer Science PhD Student-Applicant Support Program Reviewer. 2021, 2022.

Research Service
Area Chair at EMNLP
Reviewer at ACL, EMNLP, NAACL, AACL, ARR, COLING, various workshops
Organizer of CRAFT Session on Language Models and Society: Bridging Research and Policy (w/ Ioana Baldini, Stefania Druga, Mihaela Vorvoreanu). *FAccT 2023*.
Organizer of Special Interest Group on Foundation Models in Healthcare
(w/ Anja Thieme, Marzyeh Ghassemi, Tariq Osman Andersen, Ewa Luger, Aditya Nori). *CHI 2023*.

LAST UPDATED *November 25, 2023*