Rishi Bommasani
Senior Research Scholar, Stanford Institute for Human-Centered Artificial Intelligence (HAI)

EDUCATION

Stanford University

Ph.D Computer Science, September 2020 – June 2025
Advisors: Percy Liang, Dan Jurafsky
Funding: Stanford Lieberman Fellowship, NSF Graduate Research Fellowship

Cornell University

M.S. Computer Science, August 2019 – May 2020
B.A. Computer Science, B.A. Mathematics, August 2016 – May 2019
Advisor: Claire Cardie

SUMMARY

Research.
Quality: 1 best paper, 10 paper awards/recognitions, 21k citations, h-index: 32
Fellowships: NSF Graduate Research Fellowship, Stanford Lieberman Fellowship
Awards: Stanford Open Innovator Award, Stanford Open Source Software Prize, Cornell Academic Excellence Award
Grants: 5 major grants (total of $2.4M)
Media: New York Times, Nature, Wall Street Journal, Washington Post, Quanta, MIT Tech Review

Policy.
Leadership: California Report on Frontier AI Policy, EU AI Act Code of Practice
Impact: EU AI Act, California SB 53, US federal position on openness, veto of California SB 1047, proposal of US Foundation Model Transparency Act, G7 International AI Code of Conduct
Briefings: 40+ private briefings including the Secretary of Commerce, Director of US CAISI, Director of UK AISI, Netherlands Prime Minister, Luxembourg Prime Minister, UK Secretary of State, Spain Minister of Technology, Singapore Deputy Prime Minister, France Envoy to Prime Minister, UN Secretary-General's Envoy on Technology, Vatican delegation on AI

Leadership, Advising, Teaching, and Service.
Leadership: Co-founder of Stanford CRFM, founder of Stanford AI Policy
Advising: 20 students (7 women, 4 URM) with outcomes at Stanford, Harvard, Princeton, US CAISI, OpenAI, Google, Microsoft, Meta
Teaching: 3-time TA at Stanford, 6-time TA at Cornell (5 teaching awards)
Lectures: Guest lecturer at Stanford, MIT, Berkeley, CMU, Harvey Mudd
Organizer: 1 conference and 5 workshops
University: 1x faculty hiring committee, 4x admissions committee, 2x visit day committee
Scientific Community: Reviewer at Science, Nature, PNAS, NeurIPS, ICML, ICLR, FAccT, TMLR

2025 **Rishi Bommasani** et al. Advancing Science- and Evidence-based AI Policy.
Science.

Alexander Wan et al. The 2025 Foundation Model Transparency Index.
Under review at TMLR.

Rishi Bommasani*, Sarah Bana*, Kathleen A. Creel, Dan Jurafsky, Percy Liang. Hiring Algorithms in Practice: Bias and Homogeneity.

Sayash Kapoor et al. Holistic Agent Leaderboard: The Missing Infrastructure for AI Agent Evaluation.
Under review at ICLR.

Rishi Bommasani. NeurIPS should lead scientific consensus on AI policy.
Neural Information Processing Systems (NeurIPS).

Irene Solaiman et al. Beyond Release: Access Considerations for Generative AI Systems

Jennifer Wang, Kayla Huang, Kevin Klyman, **Rishi Bommasani**. Do Companies Make Good on their Voluntary Commitments to the White House?
AAAI Conference on AI, Ethics, and Society (AIES).

Judy Shen et al. The Disclosure Delusion: Systemic Challenges in AI Data Transparency Policy.

Vyoma Raman et al. Deconstructing Provider and Deployer Obligations for Fairness in General-Purpose AI.
AAAI Conference on AI, Ethics, and Society (AIES).

Xinyu Yang et al. Reliable and Responsible Foundation Models.
Transactions on Machine Learning Research (TMLR). **Outstanding Survey Paper.**

Yoshua Bengio et al. International AI Safety Report
France AI Action Summit.

Laura Weidinger*, Deb Raji* et al. Toward an Evaluation Science for Generative AI Systems
National Academy of Engineering.

Shayne Longpre et al. In-House Evaluation Is Not Enough. Towards Robust Third-Party Evaluation and Flaw Disclosure for General-Purpose AI
International Conference on Machine Learning (ICML). **Spotlight — top 2.5% of 12000 submissions.**

Andy Zhang et al. Language model developers should report train-test overlap.
International Conference on Machine Learning (ICML). **Spotlight — top 2.5% of 12000 submissions.**

Rishi Bommasani*, Kevin Klyman* et al. The 2024 Foundation Model Transparency Index.
Transactions on Machine Learning Research (TMLR).

Rishi Bommasani*, Kevin Klyman* et al. The 2023 Foundation Model Transparency Index.
Transactions on Machine Learning Research (TMLR). **Outstanding Paper.**

2024 **Rishi Bommasani** et al. Considerations for Governing Open Foundation Models.
Science.

Sayash Kapoor*, **Rishi Bommasani*** et al. On the Societal Impact of Open Foundation Models.
International Conference on Machine Learning (ICML). **Oral — top 1.5% of 10000 submissions.**

Shayne Longpre et al. A Safe Harbor for AI Evaluation and Red Teaming.
International Conference on Machine Learning (ICML). **Oral — top 1.5% of 10000 submissions.**

Shayne Longpre*, Stella Biderman* et al. The Responsible Foundation Model Development Cheatsheet: A Review of Tools & Resources.
Transactions on Machine Learning Research (TMLR). **Outstanding Survey Paper.**

Yoshua Bengio et al. International Scientific Report on the Safety of Advanced AI (Interim Report)
AI Seoul Summit.

Aidan Peppin et al. The Reality of AI and Biorisk

Risto Uuk et al. Effective Mitigations for Systemic Risks from General-Purpose AI

Rishi Bommasani et al. Foundation Model Transparency Reports.
AAAI Conference on AI, Ethics, and Society (AIES). **Oral**.

Rishi Bommasani, Dilara Soylu, Thomas I. Liao, Kathleen A. Creel, Percy Liang. Ecosystem Graphs: Documenting the Social Footprint of Foundation Models.
AAAI Conference on AI, Ethics, and Society (AIES).

Rishi Bommasani and Percy Liang. Trustworthy Social Bias Measurement.
AAAI Conference on AI, Ethics, and Society (AIES).

2023

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

Neel Guha*, Christie Lawrence* et al. AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing.
George Washington Law Review.

Connor Toups*, **Rishi Bommasani*** et al. Homogeneous Outcomes for Individuals from Deployed ML APIs.
Neural Information Processing Systems (NeurIPS).

Deepak Narayanan et al. Cheaply Evaluating Inference Efficiency Metrics for Autoregressive Transformer APIs.
Neural Information Processing Systems (NeurIPS).

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

2022

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

Rishi Bommasani. Evaluation for Change. *Association for Computational Linguistics* (ACL).

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

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

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

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

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

Rishi Bommasani, Kelly Davis, Claire Cardie. Interpreting Pretrained Contextualized Representations via Reductions to Static Embeddings. *Association for Computational Linguistics* (ACL).

POLICY **Rishi Bommasani***, Scott Singer* et al. California Frontier AI Policy Report. *Request from California governor Gavin Newsom*. 2025.

Nuria Oliver, **Rishi Bommasani** et al. EU AI Act Code of Practice (Transparency). 2025.

Conrad Stosz et al. Minimum Operating Conditions for Independent Third Party AI Evaluations. 2025.

Caroline Meinhardt et al. Stanford Response on US AI Action Plan. *Request for Comment: White House Office of Science and Technology Policy*. 2025.

Kevin Klyman et al. Safeguarding Third-Party AI Research. 2025.

Rishi Bommasani*, Alexander Wan* et al. Stanford Response on Managing Misuse Risk for Dual-Use Foundation Models. *Request for Comment: US Artificial Intelligence Safety Institute*. 2024.

Rishi Bommasani*, Sayash Kapoor*, Kevin Klyman* et al. Stanford-Princeton led Response on Open Foundation Models. *Request for Comment: US National Telecommunications and Information Administration (NTIA)*. 2024.

Alondra Nelson et al. AI Policy and Governance Working Group Response on Open Foundation Models. *Request for Comment: US National Telecommunications and Information Administration (NTIA)*. 2024.

Rishi Bommasani et al. Considerations for Governing Open Foundation Models. 2023.

Rishi Bommasani, Tatsunori Hashimoto, Daniel E. Ho, Marietje Schaake, Percy Liang. Towards compromise: A concrete two-tier proposal for foundation models in the EU AI Act. 2023.

Rishi Bommasani. Drawing Lines: Tiers for Foundation Models. 2023.

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

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

Arvind Narayanan, Sayash Kapoor, **Rishi Bommasani**. What the executive order means for openness in AI. 2023.

Rishi Bommasani, Kevin Klyman, Daniel Zhang, Marietje Schaake, Percy Liang. Do Foundation Model Providers Comply with the Draft EU AI Act? *Trust & Safety Research Conference*. 2023.

Rishi Bommasani, Sayash Kapoor, Daniel Zhang, Arvind Narayanan, Percy Liang. Stanford-Princeton response on AI Accountability.
Request for Comment: 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. 2023.

INVITED TALKS

AI Innovation. Private roundtable with **Director of the White House Office of Science and Technology Policy, Arati Prabhakar**. August 2024.

AI Policy for Foundation Models. **Stanford Congressional Boot Camp on Artificial Intelligence**. August 2024.

Foundation Model Transparency Index. Private briefing of **UK Competition and Markets Authority**. July 2024.

Panel on Foundation Models for Space. **IEEE** Space Mission Challenges for Information Technology (SMC-IT) and Space Computing Conference. July 2024.

Policy for Foundation Models. **Accenture** @ Stanford HAI. July 2024.

Panel on Scoping Evaluations. **National Academies of Sciences, Engineering, and Medicine** (Host: William Isaac, Mona Sloane, Ben Shneiderman). June 2024.

Policy for Foundation Models. Private roundtable with **French President's Special Envoy for the AI Action Summit, Anne Bouverot**. June 2024.

Panel on Human-Assistant Interaction. **Google DeepMind & Stanford** workshop on the Ethical and Societal Implications of Agentic AI Systems (Host: Leif Wenar, Iason Gabriel). June 2024.

Foundation Model Transparency Index. Private briefing of **EU AI Office**. June 2024.

Homogeneous Outcomes in Algorithmic Hiring. CEO Seminar at **Department of Labor**. June 2024.

AI Policy for Foundation Models. **AI Policy Network Speaker Series** (Host: Hodan Omaar). May 2024.

Moderator for **California Generative AI Summit** (w/ Deb Raji). May 2024.

AI Policy for Foundation Models. Tech Policy Roundtable for the **European Consulates-General** (Host: Nate Persily). April 2024.

Transparency Requirements for AI Act Implementation. Private briefing of **EU AI Office**. April 2024.

Transparency and Openness for Foundation Models. Private briefing of **UN High Level Advisory Board**. April 2024.

Transparency Requirements for AI Act Implementation. Private briefing of **EU AI Office**. April 2024.

Moderator for AI Safety Institute session at **Stanford AI Policy Workshop** (Speakers: Elizabeth Kelly, Oliver Illot, Akiko Murakami). April 2024.

Open Foundation Models. **Partnership on AI & GitHub Workshop**. April 2024.

Open Foundation Models. Private convening by **Carnegie Endowment for International Peace** @ Rockefeller Foundation, Italy (Host: Mariano-Florentino Cuéllar). April 2024.

Open Foundation Models. Private roundtable with **US Secretary of Commerce Gina Raimondo**. March 2024.

Open Foundation Models. Private convening by **Institute of Advance Study** @ Shangri La Museum of Islamic Art, Hawaii (Host: Alondra Nelson). March 2024.

Procurement Standards for AI. Private briefing of **US Department of Homeland Security Chief Information Office, Eric Hysen** @ Stanford HAI. March 2024.

Policy for foundation models. **Stanford Cyber Policy Seminar** (Host: Nathan Persily). February 2024.

Foundation Models and AI Policy. Private roundtable with **Netherlands Prime Minister Mark Rutte**. December 2023.

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. Private 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. 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. **Stanford Congressional Boot Camp on Artificial Intelligence**. 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). **Stanford Congressional Boot Camp on Artificial Intelligence**. 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 Koc (Cambridge undergrad; now Research Engineer at Google)
Albert Tsao (Cornell undergrad; now Software Engineer at Workday)
Alexander Wan (Berkeley undergrad; actively working with me)
Anna Huang (Cornell undergrad; now Software Engineer at Microsoft)
Connor Toups (Stanford masters; now Corporate Development at D.E. Shaw)
Dilara Soylu (Stanford masters; now Stanford PhD)
Gokul Dharan (Stanford masters; now Software Engineer at Zipline)
Han Qiao (Cornell undergrad; now Software Engineer at Meta)
Jennifer Wang (Brown undergrad; now Stanford PhD)
Jonathan Xue (high school student; now undergrad at JHU)
Joseph Kihang'a (Cornell undergrad)
Julie Phan (Cornell undergrad; now Software Engineer at GEM)
Kevin Klyman (Stanford masters; now Senior Adviser at US CAISI and Harvard JD)
Nathan Kim (Stanford undergrad; now Software Engineer at Unify)
Ryan Chi (Stanford undergrad; now Member of Technical Staff at OpenAI)
Sabhya Chhabria (Cornell undergrad; now Princeton masters)
Siddharth Sharma (high school student; now Stanford undergrad)
Virginia Adams (Applied scientist at Nvidia; now Stanford masters)
Wenyi Guo (Cornell undergrad; now Cornell masters)
Ye Jiang (Cornell undergrad; now Cornell masters)

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, Spring 2018, Fall 2017
Teaching Assistant

Service

Leadership

Society Lead at the Stanford Center for Research on Foundation Models. 2021–Present.
Organizer of Stanford AI Policy Working Group. 2023–Present
The Future of Third-Party AI Evaluation (w/ Shayne Longpre, Kevin Klyman, Sayash Kapoor, Michelle Sahar, Rumman Chowdhury, Arvind Narayanan, Percy Liang). 2024.
Closed-Door Workshop on Governing Open Foundation Models (w/ Daniel Zhang, Percy Liang, Daniel E. Ho). 2024.
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

Stanford Computer Science Faculty Hiring Committee (Chair: Kunle Olukotun). 2024
Stanford Computer Science PhD Admissions Committee (Chair: Karen Liu). 2022, 2023, 2024, 2025.
Stanford Computer Science PhD Visit Day Organizer (AI). 2021, 2022.
Stanford Computer Science PhD Student-Applicant Support Program Reviewer. 2021, 2022.
Stanford HAI 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.

Service to the Field

NSF Grant Reviewer
Area Chair at ACL, EMNLP
Reviewer at Science, NeurIPS, ICML, ACL, EMNLP, FAccT, COLM, various workshops
Organizer of AI Policy Social (w/ Hoda Heidari, Judy Shen, Daniel Zhang). *ICML 2024*.
Organizer of AI Policy Social (w/ Divyansh Kaushik, Judy Shen). *ICML 2023*.
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 *December 22, 2025*