

## EDUCATION

---

### Princeton University

*Doctor of Philosophy in Computer Science; GPA: 3.96/4.00*

Princeton, NJ

*January 2021 – June 2026 (expected)*

### Indian Institute of Technology Kanpur

*Bachelor of Technology in Computer Science; GPA: 9.9/10.0*

Kanpur, India

*July 2015 – June 2019*

### École Polytechnique Fédérale de Lausanne

*Exchange Student in Computer Science; GPA: 5.7/6.0*

Lausanne, Switzerland

*August 2017 – May 2018*

## AWARDS AND RECOGNITION

---

### Porter Ogden Jacobus Fellowship

September 2025

*The highest honor that the Princeton Graduate School bestows on an enrolled Ph.D. student*

*One of four awardees across all departments*

### 2026 Siebel Scholar

September 2025

*One of 5 Princeton computer science graduate students selected for an award of \$35,000*

### Mozilla Senior Fellowship in Trustworthy AI

January 2025

*\$130,000 fellowship to advance trustworthy AI through evidence-based tech policy*

### Princeton School of Engineering and Applied Science Award for Excellence

September 2024

### Laurance S. Rockefeller Graduate Prize Fellowship

2024-25

*First computer scientist in 20 years to receive the \$60,000 graduate prize fellowship from*

*Princeton's University Center for Human Values*

### McGraw Center for Teaching and Learning Inaugural Exemplar Mentor Award

May 2024

### Featured in the inaugural list: *TIME100 Most Influential People in AI*

September 2023

### Impact Recognition Award, ACM CSCW

November 2022

### Motorola Gold Medal, IIT Kanpur

June 2019

### Best Paper Award, ACM FAccT

January 2019

### CMMRS 2018, Pre-Doctoral Research School, Max Planck Institute (Saarbrücken)

August 2018

### Bronze Medal, ACM ICPC SWERC, École Normale Supérieure

November 2017

### Academic Excellence Award, IIT Kanpur

July 2016, July 2017

### Outstanding Freshman Award, IIT Kanpur

March 2016

## BOOKS

---

- [1] **AI Snake Oil**  
Arvind Narayanan, **Sayash Kapoor**  
Princeton University Press (2024)  
*Peer-reviewed book on what AI can and can't do. Featured in Nature's list of the 10 best books of 2024, Bloomberg's 49 best books of 2024, and Forbes's 10 must-read tech books of 2024*
- [2] **AI as Normal Technology**  
Arvind Narayanan, **Sayash Kapoor**  
Princeton University Press (in preparation)

## PEER-REVIEWED PUBLICATIONS

---

*Link to Google Scholar · H-Index: 23 · I10-Index: 34 · Citations: 2,653*

- [1] **AI Agents That Matter**  
**Sayash Kapoor\***, Benedikt Stroebel\*, Zachary S. Siegel, Nitya Nadgir, Arvind Narayanan  
Transactions on Machine Learning Research (TMLR 2025)
- [2] **CORE-Bench: Fostering the Credibility of Published Research Through a Computational Reproducibility Agent Benchmark**  
Zachary S. Siegel, **Sayash Kapoor**, Nitya Nadgir, Benedikt Stroebel, Arvind Narayanan  
Transactions on Machine Learning Research (TMLR 2025)
- [3] **Establishing Best Practices for Building Rigorous Agentic Benchmarks**  
Yuxuan Zhu, Tengjun Jin, Yada Pruksachatkun, Andy Zhang, Shu Liu, Sasha Cui, **Sayash Kapoor**, Shayne Longpre, Kevin Meng, Rebecca Weiss, Fazl Barez, Rahul Gupta, Jwala Dhamala, Jacob Merizian, Mario Giulianelli, Harry Coppock, Cozmin Ududec, Jasjeet Sekhon, Jacob Steinhardt, Antony Kellermann, Sarah Schwettmann, Matei Zaharia, Ion Stoica, Percy Liang, Daniel Kang  
Neural Information Processing Systems Datasets and Benchmarks (NeurIPS 2025)
- [4] **The Leaderboard Illusion**  
Shivalika Singh, Yiyang Nan, Alex Wang, Daniel D'Souza, **Sayash Kapoor**, Ahmet Üstün, Sanmi Koyejo, Yuntian Deng, Shayne Longpre, Noah Smith, Beyza Ermis, Marzieh Fadaee, Sara Hooker  
Neural Information Processing Systems Datasets and Benchmarks (NeurIPS 2025)
- [5] **Build Agent Advocates, Not Platform Agents**  
**Sayash Kapoor\***, Noam Kolt\*, Seth Lazar\*  
International Conference on Machine Learning position paper track (ICML 2025)
- [6] **In-House Evaluation Is Not Enough: Towards Robust Third-Party Flaw Disclosure for General-Purpose AI**  
Shayne Longpre, Kevin Klyman, Ruth E Appel, **Sayash Kapoor** et al.  
International Conference on Machine Learning position paper track (ICML 2025)
- [7] **The Reality of AI and Biorisk**  
Aidan Peppin, Anka Reuel, Stephen Casper, Elliot Jones, Andrew Strait, Usman Anwar, Anurag Agrawal, **Sayash Kapoor**, Sanmi Koyejo, Marie Pellat, Rishi Bommasani, Nick Frosst, Sara Hooker  
ACM Conference on Fairness, Accountability, and Transparency (FAccT 2025)
- [8] **The 2024 Foundation Model Transparency Index**  
Rishi Bommasani, Kevin Klyman, **Sayash Kapoor**, Shayne Longpre, Betty Xiong, Nestor Maslej, Percy Liang  
Transactions on Machine Learning Research (TMLR 2025)
- [9] **The 2023 Foundation Model Transparency Index**  
Rishi Bommasani, Kevin Klyman, Shayne Longpre, **Sayash Kapoor**, Nestor Maslej, Daniel Zhang, Percy Liang  
Transactions on Machine Learning Research (TMLR 2025 **Featured certification**)
- [10] **REFORMS: Consensus-based Recommendations for Machine-learning-based Science · Blog post**  
**Sayash Kapoor**, Emily Cantrell, Kenny Peng, Thanh Hien (Hien) Pham, Christopher A. Bail, Odd Erik Gundersen, Jake M. Hofman, Jessica Hullman, Michael A. Lones, Momin M. Malik, Priyanka Nanayakkara, Russell A. Poldrack, Inioluwa Deborah Raji, Michael Roberts, Matthew J. Salganik, Marta Serra-Garcia, Brandon M. Stewart, Gilles Vandewiele, Arvind Narayanan  
Science Advances (2024)

- [11] **Considerations for governing open foundation models**  
Rishi Bommasani, **Sayash Kapoor**, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaaake, Daniel E. Ho, Arvind Narayanan, Percy Liang  
Science (2024)
- [12] **On the Societal Impact of Open Foundation Models · Blog post**  
**Sayash Kapoor\***, Rishi Bommasani\*, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Peter Cihon, Aspen Hopkins, Kevin Bankston, Stella Biderman, Miranda Bogen, Rumman Chowdhury, Alex Engler, Peter Henderson, Yacine Jernite, Seth Lazar, Stefano Maffulli, Alondra Nelson, Joelle Pineau, Aviya Skowron, Dawn Song, Victor Storch, Daniel Zhang, Daniel E. Ho, Percy Liang, Arvind Narayanan  
International Conference on Machine Learning position paper track (ICML 2024 **Oral**)
- [13] **A Safe Harbor for AI Evaluation and Red Teaming · Blog post**  
Shayne Longpre, **Sayash Kapoor**, Kevin Klyman, Ashwin Ramaswami, Rishi Bommasani, Borhane Blili-Hamelin, Yangsibo Huang, Aviya Skowron, Zheng-Xin Yong, Suhas Kotha, Yi Zeng, Weiyan Shi, Xianjun Yang, Reid Southen, Alexander Robey, Patrick Chao, Diyi Yang, Ruoxi Jia, Daniel Kang, Sandy Pentland, Arvind Narayanan, Percy Liang, Peter Henderson  
International Conference on Machine Learning position paper track (ICML 2024 **Oral**)  
*Our open letter to AI companies calling for a safe harbor was signed by over 350 academics, researchers, and civil society members.*
- [14] **Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy · Blog post**  
Angelina Wang\*, **Sayash Kapoor\***, Solon Barocas, Arvind Narayanan  
ACM Journal on Responsible Computing (JCR 2024)  
*Also presented at: Philosophy, AI, and Society (2023); Data (Re)Makes the World (2023); ACM FAccT (2023)*
- [15] **How large language models can reshape collective intelligence**  
Jason W. Burton, Ezequiel Lopez-Lopez, Shahar Hechtlinger, Zoe Rahwan, Samuel Aeschbach, Michiel A. Bakker, Joshua A. Becker, Aleks Berdichevskaya, Julian Berger, Levin Brinkmann, Lucie Flek, Stefan M. Herzog, Saffron Huang, **Sayash Kapoor**, Arvind Narayanan et al.  
Nature Human Behaviour (2024)
- [16] **The Responsible Foundation Model Development Cheatsheet: A Review of Tools & Resources**  
Shayne Longpre, Stella Biderman, Alon Albalak, Gabriel Ilharco, **Sayash Kapoor**, Kevin Klyman, Kyle Lo, Maribeth Rauh, Nay San, Hailey Schoelkopf, Aviya Skowron, Bertie Vidgen, Laura Weidinger, Arvind Narayanan, Victor Sanh, David Adelani, Percy Liang, Rishi Bommasani, Peter Henderson, Sasha Luccioni, Yacine Jernite, Luca Soldaini  
Transactions on Machine Learning Research (TMLR 2024 **Survey certification**)
- [17] **Foundation Model Transparency Reports · Blog post**  
Rishi Bommasani, Kevin Klyman, Shayne Longpre, Betty Xiong, **Sayash Kapoor**, Nestor Maslej, Arvind Narayanan, Percy Liang  
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES 2024)
- [18] **Leakage and the reproducibility crisis in ML-based science**  
**Sayash Kapoor**, Arvind Narayanan  
Patterns (2023)
- [19] **Weaving Privacy and Power: On the Privacy Practices of Labor Organizers in the U.S. Technology Industry**  
**Sayash Kapoor\***, Matthew Sun\*, Mona Wang\*, Klaudia Jaźwińska\*, Elizabeth Anne Watkins\*  
ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW 2022 **Impact Recognition Award**)
- [20] **The worst of both worlds: A comparative analysis of errors in learning from data in psychology and machine learning**  
Jessica Hullman, **Sayash Kapoor**, Priyanka Nanayakkara, Andrew Gelman, Arvind Narayanan  
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES 2022)
- [21] **Controlling polarization in personalization: an algorithmic framework**  
L. Elisa Celis, **Sayash Kapoor**, Farnood Salehi, and Nisheeth K. Vishnoi  
ACM Conference on Fairness, Accountability, and Transparency (FAccT 2019 **Best Paper Award**)
- [22] **Corruption-tolerant bandit learning**  
**Sayash Kapoor**, Kumar Kshitij Patel, and Purushottam Kar  
Machine Learning (2019)

- [23] **A dashboard for controlling polarization in personalization**  
L. Elisa Celis, **Sayash Kapoor**, Vijay Keswani, Farnood Salehi, and Nisheeth K. Vishnoi  
AI Communications (2019)  
*Also presented at: International Joint Conference on Artificial Intelligence Demos Track (IJCAI 2018)*

---

## PREPRINTS, MANUSCRIPTS, AND COMMENTS

---

- [1] **Holistic Agent Leaderboard: The Missing Infrastructure for AI Agent Evaluation**  
**Sayash Kapoor\***, Benedikt Stroebel\*, Peter Kirgis, Nitya Nadgir, Zachary S Siegel, Boyi Wei, Tianci Xue, Ziru Chen, Felix Chen, Saiteja Utpala, Franck Ndzomga, Dheeraj Oruganty, Sophie Luskin, Kangheng Liu, Botao Yu, Amit Arora, Dongyoon Hahm, Harsh Trivedi, Huan Sun, Juyong Lee, Tengjun Jin, Yifan Mai, Yifei Zhou, Yuxuan Zhu, Rishi Bommasani, Daniel Kang, Dawn Song, Peter Henderson, Yu Su, Percy Liang, Arvind Narayanan  
Preprint (2025)
- [2] **AI as Normal Technology**  
Arvind Narayanan, **Sayash Kapoor**  
Knight First Amendment Institute (2025). Received press coverage in New York Times, Economist, New Yorker etc.
- [3] **Why an overreliance on AI-driven modelling is bad for science**  
Arvind Narayanan, **Sayash Kapoor**  
Nature (2025)
- [4] **International AI Safety Report**  
Yoshua Bengio, ..., **Sayash Kapoor** et al. (2025)  
*A report on the state of advanced AI capabilities and risks written by 100 AI experts*
- [5] **Inference Scaling fLaws: The Limits of LLM Resampling with Imperfect Verifiers**  
Benedikt Stroebel, **Sayash Kapoor**, Arvind Narayanan  
Preprint (2024)
- [6] **Towards a Framework for Openness in Foundation Models: Proceedings from the Columbia Convening on Openness in Artificial Intelligence**  
Adrien Basdevant, Camille François, Victor Storch, Kevin Bankston, Ayah Bdeir, Brian Behlendorf, Merouane Debbah, **Sayash Kapoor**, Yann LeCun, Mark Surman, Helen King-Turvey, Nathan Lambert, Stefano Maffulli, Nik Marda, Govind Shivkumar, Justine Tunney  
Preprint (2024)
- [7] **Promises and pitfalls of artificial intelligence for legal applications · Blog post**  
**Sayash Kapoor**, Peter Henderson, Arvind Narayanan  
Journal of Cross-disciplinary Research in Computational Law (CRCL 2024)
- [8] **The limitations of machine learning models for predicting scientific replicability**  
M. J. Crockett, Xuechunzi Bai, **Sayash Kapoor**, Lisa Messeri, and Arvind Narayanan  
Proceedings of the National Academy of Sciences (PNAS 2023)
- [9] **How to Prepare for the Deluge of Generative AI on Social Media**  
**Sayash Kapoor**, Arvind Narayanan  
Knight First Amendment Institute (2023)

---

## PUBLIC WRITING

---

*In addition to the texts below, I write extensively on the AI as Normal Technology newsletter, which has over 65,000 subscribers.*

- [1] **Worry About Misuse of AI, Not Superintelligence**  
Arvind Narayanan, **Sayash Kapoor**  
WIRED (2024)
- [2] **We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem.**  
**Sayash Kapoor**, Arvind Narayanan  
Knight First Amendment Institute (2024)
- [3] **Is AI too dangerous to release openly?**  
**Sayash Kapoor**, Arvind Narayanan  
Princeton Engineering Magazine (2024)

- [4] **A Safe Harbor for AI Evaluation and Red Teaming**  
Shayne Longpre, **Sayash Kapoor**, Kevin Klyman, et al.  
Knight First Amendment Institute (2024)
- [5] **Does AI Pose an Existential Risk to Humanity? Two Sides Square Off**  
Arvind Narayanan, **Sayash Kapoor**  
The Wall Street Journal, November 2023
- [6] **How to report better on artificial intelligence**  
**Sayash Kapoor**, Hilke Schellmann, Ari Sen  
Columbia Journalism Review (2023)
- [7] **Generative AI companies must publish transparency reports**  
Arvind Narayanan, **Sayash Kapoor**  
Knight First Amendment Institute (2023)
- [8] **A Checklist of Eighteen Pitfalls in AI Journalism**  
**Sayash Kapoor**, Arvind Narayanan  
Reporting on artificial intelligence: a handbook for journalism educators, UNESCO (2023)
- [9] **The LLaMA is out of the bag. Should we expect a tidal wave of disinformation?**  
Arvind Narayanan, **Sayash Kapoor**  
Knight First Amendment Institute (2023)
- [10] **Through the Wire**  
Klaudia Jazwińska, **Sayash Kapoor**, Matthew Sun, Mona Wang  
Logic Mag (2022)
- [11] **The platform as the city**  
Mac Arboleda, Palak Dudani, **Sayash Kapoor**, Lorna Xu  
ACM Interactions Mag (2021)

## POLICY INPUT

---

- [1] **Safeguarding Third-Party AI Research**  
Kevin Klyman, Shayne Longpre, **Sayash Kapoor**, Rishi Bommasani, Percy Liang, Peter Henderson  
*Stanford HAI Policy Brief (2025)*
- [2] **Generative AI Companies: Safe Harbor and Whistleblower Protections**  
**Sayash Kapoor**, Arvind Narayanan  
*Testimony to the New Jersey Assembly Science, Innovation and Technology Committee (2024)*
- [3] **Response to the EU AI Office’s Consultation on the AI Act**  
Varun Nagaraj Rao, Kyler Zhou, **Sayash Kapoor**, Arvind Narayanan  
*Submitted to the EU AI Office (2024)*
- [4] **Princeton Dialogues in AI: Predictive AI**  
Arvind Narayanan, **Sayash Kapoor**, Peter Henderson  
*Senate AI Caucus (2024)*
- [5] **Princeton Dialogues in AI: AI Safety**  
**Sayash Kapoor**, Mihir Kshirsagar  
*Senate AI Caucus and House AI Caucus (2024)*
- [6] **A Safe Harbor For AI Researchers: Promoting Safety And Trustworthiness Through Good-Faith Research**  
Kevin Klyman, **Sayash Kapoor**, Shayne Longpre  
*Federation of American Scientists: Policy memo (2024)*
- [7] **Reducing harm from deepfakes**  
**Sayash Kapoor**, Arvind Narayanan  
*Testimony to the New Jersey Assembly Science, Innovation and Technology Committee (2024)*

- [8] **Response to Request for Comment on Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights**  
Alondra Nelson, Arvind Narayanan, Caroline Meinhardt, Daniel E. Ho, Daniel Zhang, Dawn Song, Inioluwa Deborah Raji, Kevin Klyman, Marietje Schaake, Mihir Kshirsagar, Percy Liang, Peter Henderson, Rishi Bommasani, Rohini Kosoglu, Rumman Chowdhury, **Sayash Kapoor**, Seth Lazar, Shayne Longpre, Stefano Maffulli, Stella Biderman, Victor Storchan  
*Submitted to the National Telecommunications and Information Administration (2024)*
- [9] **Comment to the Copyright Office in Support of a Safe Harbor Exemption for Generative AI Research**  
Kevin Klyman, Shayne Longpre, **Sayash Kapoor**, Arvind Narayanan, Aleksandra Korolova, Peter Henderson  
*Submitted to the U.S. Copyright Office (2024)*
- [10] **Beyond the AI hype**  
**Sayash Kapoor**, Arvind Narayanan  
*Government of Canada's Federal Foresight Network (2024)*
- [11] **Intro to AI/ML for Regulators**  
**Sayash Kapoor**, Mihir Kshirsagar  
*Consumer Finance Protection Bureau (2024)*
- [12] **How to Prepare for the Deluge of Generative AI on Social Media**  
**Sayash Kapoor**, Arvind Narayanan  
*Federal Trade Commission Division of Advertising Practices Tech Speaker Series (2023)*
- [13] **Considerations for governing open foundation models · Blog post**  
Rishi Bommasani, **Sayash Kapoor**, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaake, Daniel E. Ho, Arvind Narayanan, Percy Liang  
*Stanford HAI Issue Brief (2023)*
- [14] **The urgent need for accountability in predictive AI**  
Arvind Narayanan, **Sayash Kapoor**  
*Congressional Forum (2023)*
- [15] **Three Ideas for Regulating Generative AI · Blog post**  
Rishi Bommasani, **Sayash Kapoor**, Daniel Zhang, Arvind Narayanan, Percy Liang  
*Submitted to the National Telecommunications and Information Administration (2023)*
- [16] **CITP Comments on AI Accountability · Blog post**  
Archana Ahlawat, Justin Curl, **Sayash Kapoor**, Aleksandra Korolova, Mihir Kshirsagar, Surya Mattu, Jakob Mökander, Arvind Narayanan, Matthew J. Salganik  
*Submitted to the National Telecommunications and Information Administration (2023)*
- [17] **Calling for Investing in Equitable AI Research in Nation's Strategic Plan · Blog post**  
Solon Barocas, **Sayash Kapoor**, Mihir Kshirsagar, Arvind Narayanan  
*Submitted to the White House Office of Science and Technology Policy (2022)*
- [18] **National AI Research Infrastructure Needs to Support Independent Evaluation of Performance Claims · Blog post**  
**Sayash Kapoor**, Mihir Kshirsagar, Arvind Narayanan  
*Submitted to the White House Office of Science and Technology Policy and National Science Foundation*

## WORK EXPERIENCE

---

### Facebook

*Software Engineer, Integrity*

London, UK

*July 2019 - December 2020*

Developed machine learning models to combat Covid-19 misinformation and non-consensual intimate imagery across Facebook and Instagram. Interned from May – August 2018; developed machine learning models to detect and remove child sexual abuse material from the platform.

## SERVICE AND WORKSHOPS

---

### **Workshop organizer**

The Future of Third-Party AI Evaluation (Stanford & Princeton)  
*Over 400 registrations. Video recordings seen over 2,000 times.*

Workshop on Useful and Reliable AI Agents (Princeton)  
*Over 500 registrations. Video recordings seen over 3,300 times.*

Responsible and open foundation models (Princeton & Stanford)  
*Over 900 registrations. Video recordings seen over 3,200 times.*

The Reproducibility Crisis in ML-based Science (Princeton)  
*Over 1,700 registrations. Video recordings seen over 6,500 times.*

Resistance AI (NeurIPS 2020)

### **Program committee member**

AIES 2022, FAccT 2022, FAccT 2023, AIES 2024, FAccT 2024

### **Reviewer**

Nature, Science Advances, PLoS ONE, JMLR, Patterns, ICML 2022, ICML 2025

### **Mentor**

Princeton Computer Science Pre-Application Graduate Mentoring Program

## TEACHING

---

### **Open Source AI: Benefits and Risks**

Wilson Center Advanced AI Labs. Four courses taught in Fall 2024 and Spring 2025

### **Princeton Policy Precepts: AI's impact on labor**

Princeton University School of Public and International Affairs, Fall 2025

### **Princeton Policy Precepts: Predictive AI**

Princeton University School of Public and International Affairs, Fall 2024

### **CS 5382: Practical Principles for Designing Fair Algorithms**

Cornell University. Guest Lecturer. Spring 2024.

### **COS 350: Ethics of Computing**

Princeton University. Preceptor and teaching assistant. Fall 2023.

### **COS 324: Introduction to Machine Learning**

Princeton University. Preceptor and teaching assistant. Spring 2023.

### **PHI 543: Machine Learning: A Practical Introduction for Humanists and Social Scientists**

Princeton University. Guest Lecturer. Fall 2023.

### **SOC 306: Machine Learning with Social Data: Opportunities and Challenges**

Princeton University. Guest Lecturer. Spring 2022, Spring 2023.

## SELECTED PRESS

---

### **What if artificial intelligence is just a "normal" technology?**

The Economist, September 2025

### **How ChatGPT surprised me**

The New York Times, August 2025

### **A.I. May Be Just Kind of Ordinary**

The New York Times, August 2025

### **Silicon Valley Is Drifting Out of Touch With the Rest of America**

The New York Times, August 2025

**AI isn't magic. It's just weird**

Prospect Magazine, June 2025

**How to build a better AI benchmark**

MIT Technology Review, May 2025

**Experts Explain: AI as 'normal' technology**

The Indian Express, May 2025

**We need to start thinking of AI as "normal"**

MIT Technology Review, April 2025

**What works in AI, and what's just hype**

The Hindu, February 2025

**Artificial intelligence dominates annual convention**

Al Jazeera, January 2025

**'AI Snake Oil': A conversation with Princeton AI experts**

Princeton Engineering, January 2025

**AI Snake Oil with Sayash Kapoor**

BCG Henderson Institute, December 2024

**The Year of the AI Election Wasn't Quite What Everyone Expected**

WIRED, December 2024

**AI Snake Oil: What Artificial Intelligence Can and Cannot Do**

Harvard Gazette, October 2024

**Seeing the Forest Through the A.I. Trees**

Air Mail, October 2024

**Popping the AI Hyperbole Bubble**

The Deal, October 2024

**Why AI isn't as clever – or as dangerous – as we think**

The Telegraph, October 2024

**Ray Kurzweil Still Says He Will Merge With A.I.**

The New York Times, October 2024

**AI Snake Oil: Exposing The Truth Behind Overhyped Claims**

NDTV, October 2024

**AI Snake Oil (excerpt)**

Stanford Social Innovation Review, October 2024

**Generative AI Hype Feels Inescapable. Tackle It Head On With Education**

WIRED, September 2024

**Professor Arvind Narayanan and Sayash Kapoor Explain AI**

Princeton Alumni Weekly, September 2024

**Snake Oil: Don't believe the artificial intelligence hype**

Financial Review, September 2024

**A new book tackles AI hype – and how to spot it**

Science News, September 2024

**Arvind Narayanan and Sayash Kapoor on AI Snake Oil**

Princeton University Press, September 2024

**Princeton SPIA AI Experts Separate Hype from Substance in New Book**

Princeton SPIA, September 2024

## **AI Snake Oil: Separating Hype from Reality**

Tech Policy Press, September 2024

## **In the Age of A.I., What Makes People Unique?**

The New Yorker, August 2024

## **'AI Snake Oil' Sorts Promise from Hype**

Practical Ecommerce, August 2024

## **Chatbots Are Primed to Warp Reality**

The Atlantic, August 2024

## **Science has an AI problem. This group says they can fix it.**

UC San Diego Today, May 2024

## **Experts call for legal 'safe harbor' so researchers, journalists and artists can evaluate AI tools**

VentureBeat, March 2024

## **Top AI researchers say OpenAI, Meta and more hinder independent evaluations**

Washington Post, March 2024

## **Researchers, legal experts want AI firms to open up for safety checks**

Computer World, March 2024

## **Stanford study outlines risks and benefits of open AI models**

Axios, March 2024

## **A Mistral chills European regulators**

Politico, March 2024

## **What are LLMs, and how are they used in generative AI?**

Computer World, February 2024

## **Princeton University's 'AI Snake Oil' authors say generative AI hype has 'spiraled out of control'**

VentureBeat, August 2023

## **Computer Science Researchers Call Out AI Hype as 'Snake Oil'**

Princeton Alumni Weekly, December 2023

## **OpenAI's ChatGPT turns one year old; what it did (and didn't do)**

Computer World, November 2023

## **Artificial intelligence is not a silver bullet**

NPR, December 2023

## **AI's Spicy-Mayo Problem**

The Atlantic, November 2023

## **AI Is Becoming More Powerful—but Also More Secretive**

WIRED, October 2023

## **How Does AI 'Think'? We Are Only Starting to Understand That**

The Wall Street Journal, October 2023

## **The world's biggest AI models aren't very transparent**

The Verge, October 2023

## **Maybe We Will Finally Learn More About How A.I. Works**

The New York Times, October 2023

## **Klobuchar Says AI Regulation Still Possible Before End of Year**

Bloomberg, October 2023

## **Why everyone seems to disagree on how to define artificial general intelligence**

Fast Company, October 2023

**OpenAI Is Human After All: Sharing Is Caring, Researchers Tell Model Developers**

The Information, October 2023

**How transparent are AI models? Stanford researchers found out**

VentureBeat, October 2023

**Newsletter helped us dissect fake claims about AI in real-time**

The Indian Express, September 2023

**Prominent AI fairness advocates among Princeton AI luminaries**

The Daily Princetonian, September 2023

**OpenAI Worries About What Its Chatbot Will Say About People's Faces**

The New York Times, July 2023

**GPT-4: Is the AI behind ChatGPT getting worse?**

New Scientist, July 2023

**Tips for Investigating Algorithm Harm and Avoiding AI Hype**

Global Investigative Journalism Network, July 2023

**Six tips for better coding with ChatGPT**

Nature News, June 2023

**The White House AI R&D Strategy Offers a Good Start. Here's How to Make It Better**

Tech Policy Press, May 2023

**The AI backlash is here. It's focused on the wrong things**

Washington Post, April 2023

**What is needed instead of an AI moratorium**

Tagesspiegel Background, March 2023

**Here are 5 reasons people are dunking on that call for a 6-month A.I. development pause**

Fortune, March 2023

**Sloppy Use of Machine Learning Is Causing a 'Reproducibility Crisis' in Science**

WIRED, August 2022

**Could Machine Learning Fuel a Reproducibility Crisis in Science?**

Nature, July 2022

SELECTED TALKS

---

**Scaling Laws podcast**

Lawfare, August 2025

**Interpreting India podcast**

Carnegie India, August 2025

**The reproducibility crisis in ML-based science**

ML reproducibility workshop. Workshop. August 2025.

**AI Snake Oil**

AmCham Lab Global. Invited talk. August 2025.

**AI Snake Oil**

Princeton University Press Author Talk. Book Talk. June 2025.

**AI agents and the law**

Vista Institute for AI Policy. Invited talk. June 2025.

**AI as Normal Technology**

Montclair. Invited talk. June 2025.

**AI as Normal Technology**

World Innovation, Technology and Services Alliance (WITSA). Invited talk. June 2025.

**AI as Normal Technology**

American Enterprise Institute. Invited talk. May 2025.

**AI Snake Oil**

Science, Technology, and Policy fellows, Washington DC. Invited talk. May 2025.

**AI and National Security**

Department of Defense. Panel. May 2025.

**AI as Normal Technology**

Two Sigma. Invited talk. May 2025.

**AI Snake Oil**

Washington D.C. Book Talk. May 2025.

**AI as Normal Technology**

AI Risk Institute (ARI). Panel. May 2025.

**AI as Normal Technology**

RAND. Panel. May 2025.

**Building and evaluating AI Agents That Matter**

AI Safety Institute (AISI). Invited talk. April 2025.

**AI Snake Oil**

NASA Goddard. Invited talk. April 2025.

**AI Snake Oil**

NPR. Interview. April 2025.

**AI Snake Oil**

Zelen Symposium. Invited talk. April 2025.

**AI Snake Oil**

Baltimore Library. Book Talk. April 2025.

**AI's Impact on Science, Law, and Society**

Berkman Klein Center (BKC), Harvard. Invited talk. April 2025.

**Open source AI: Risks, Benefits, Interventions**

Wilson Center. Policy course. March 2025.

**AI Snake Oil**

Section School. Book Talk. March 2025.

**AI Snake Oil**

Loyola University. Invited talk. March 2025.

**AI Snake Oil**

Schmidt Sciences. Invited talk. March 2025.

**Princeton GradFutures science and policy careers**

Princeton University. Panel. March 2025.

**AI Snake Oil**

Oxford Internet Institute (OII). Invited talk. March 2025.

**Building and evaluating AI Agents That Matter**

Snowflake. February 2025.

**Building and evaluating AI Agents That Matter**

Keynote, AI Engineer Conference. February 2025.

### **Open source AI: Risks, Benefits, Interventions**

Wilson Center course for federal staffers in the U.S. government, February 2025.

### **Can AI automate science?**

University of Illinois Generative AI and Future of Research Speaker Series. Invited talk. February 2025.

### **AI Snake Oil**

Keynote at Brussels Winter Academy and AI and Law, February 2025.

### **Is Generative AI a Threat to Democracy?**

GETTING-Plurality, January 2025.

### **AI Snake Oil**

Sony AI Ethics Talk, January 2025.

### **AI Snake Oil**

9<sup>th</sup> Harris Miller Book Talk, Washington D.C. January 2025.

### **AI Snake Oil**

The Prompt Podcast, Denmark. Podcast. November 2024.

### **On the Societal Impact of Open Foundation Models**

University of Rochester AI Policy and Regulation Workshop. Invited talk. November 2024.

### **AI Snake Oil**

Fidelity. Book Talk. November 2024.

### **AI Snake Oil**

Princeton University GradFutures Responsible AI course. Book Talk. November 2024.

### **Promises and Pitfalls of AI in law**

Law and Technology Centre, HKU. Invited talk. October 2024.

### **AI Snake Oil**

AirBnB. Book Talk. October 2024.

### **Is AI-generated disinformation a threat to democracy?**

Global Summit on the Future of Free Speech. Invited talk. October 2024.

### **AI Snake Oil**

Princeton Public Library. Book Talk. October 2024.

### **Types of AI and AI Snake Oil**

AAAS Center for Scientific Evidence in Public Issues. Policy seminar. October 2024.

### **Open source AI and its policy implications**

Wilson Center (Executive staffers). Policy course. October 2024.

### **AI Agents That Matter**

Weaviate Podcast. Podcast. October 2024.

### **AI Snake Oil**

Adam Conover's Factually! Podcast. Book Podcast. October 2024.

### **AI Snake Oil**

AI & Social Sciences Seminar, Paris. Book Talk. September 2024.

### **AI Snake Oil**

Eric Topol's Ground Truths Podcast. Book Podcast. September 2024.

### **AI Snake Oil**

City Lights. Book Talk. September 2024.

### **The threat of existential risk from AI**

Machine Learning Street Talk Podcast. Podcast. August 2024.

**A Safe Harbor for AI Evaluation and Red Teaming**

Federation of American Scientists. Congressional briefing. July 2024.

**AI agents that matter**

Meta (Core Applied Sciences). Invited talk. May 2024.

**AI and disinformation**

Dutch Ministry of Interior and Kingdom Relations workshop. Invited talk. May 2024.

**On the Societal Impact of Open Foundation Models**

Toronto AI Safety group. Invited talk. May 2024.

**Understanding and Unlocking AI's Economic Potential**

World Bank Measuring Development 2024. Panel. May 2024.

**Princeton Dialogues in AI**

Senate AI Caucus. April 2024.

**Princeton Dialogues in AI**

House AI Caucus. April 2024.

**On the Societal Impact of Open Foundation Models**

Stanford RegLab. Invited talk. April 2024.

**On the Societal Impact of Open Foundation Models**

Mechanism Design For Social Good Speaker Series. Invited talk. April 2024.

**On the Societal Impact of Open Foundation Models**

World Innovation, Technology and Services Alliance. Invited talk. March 2024.

**Assessing the risks of open models**

This Week in Machine Learning. Podcast. March 2024.

**On the Societal Impact of Open Foundation Models**

Tech Policy Press. Podcast. March 2024.

**On the Societal Impact of Open Foundation Models**

Safe Mode. Podcast. March 2024.

**Intro to AI/ML for Regulators**

Consumer Finance Protection Bureau. Invited talk. March 2024.

**On the Societal Impact of Open Foundation Models**

Princeton Alignment Reading Group. Invited talk. February 2024.

**Against Predictive Optimization**

Cornell University. Guest lecture. February 2024.

**Understanding AI Hype**

Symphony AI. Invited talk. February 2024.

**Against Predictive Optimization**

Stanford University Fairness Lunch Speaker Series. Invited talk. February 2024.

**On the Societal Impact of Open Foundation Models**

Stanford Workshop on Governance of Open Foundation Models. Panel. February 2024.

**Beyond the AI hype**

Government of Canada's Federal Foresight Network. Panel. March 2024.

**How to Prepare for the Deluge of Generative AI on Social Media**

Federal Trade Commission. Invited talk. December 2023.

**Launch of NTIA's Public Consultation Process on Widely Available AI Foundation Model Weights**

Center for Democracy and Technology. Panel. December 2023.

**Data Governance in the Age of AI**  
Washington D.C. Panel. December 2023.

**National Association of Attorneys General**  
Washington D.C. Panel. November 2023.

**AI and its hazards for science**  
ScienceWriters Conference, University of Colorado, Boulder. Invited talk. October 2023.

**How to detect AI hype**  
Princeton University Press. Invited talk. October 2023.

**Tigers on Strike**  
Princeton University. Panel. September 2023.

**Responsible and Open Foundation Models**  
Princeton-Stanford. Workshop organizer and panel moderator. September 2023.

**Improving Reproducibility, Trustworthiness and Fairness in Machine Learning**  
ICIAM Minisymposium, Tokyo. Invited talk. August 2023.

**Investigating algorithmic harm: Best practices and hard-learned lessons**  
Investigative Reporters and Editors, Orlando. Panel. June 2023.

**Against Predictive Optimization**  
ACM FAccT, Chicago. Paper talk. June 2023.

**CITP Digital Investigators Conference**  
Princeton University. Invited talk. May 2023.

**Critical voices on AI**  
Birkbeck Institute of Data Analytics. Invited talk. May 2023.

**Co-opting AI: Language**  
New York University. Invited talk. April 2023.

**Royal Society, UK Reproducibility Network (UKRN)**  
Panel. April 2023.

**Data (Re)Makes the World**  
Yale Law School. Panel. April 2023.

**Yale Quantum Institute**  
Yale University. Invited talk. March 2023.

**AI for Libraries, Archives, and Museums**  
Keynote. November 2022.

**Institute of Data Science and Artificial Intelligence seminar**  
University of Exeter. Invited talk. November 2022.

**Data Science Institute seminar**  
Lawrence Livermore National Lab. Invited talk. October 2022.

**5th Annual conference of the Massive Analysis and Quality Control Society**  
FDA headquarters. Invited talk. September 2022.

**Workshop on The Reproducibility Crisis in ML-based Science**  
Princeton University. Opening talk. July 2022.