Shayne Longpre

Email: slongpre@media.mit.edu
Web: shaynelongpre.com

INTERESTS

Data-centric AI, language models, & their societal impact. Evaluation, transparency, systemic harms, & policies for responsible AI governance.

EDUCATION

Massachusetts Institute of Technology, Cambridge, Massachusetts — Sept 2021 - Present
Ph.D. Candidate, Media Arts & Sciences
Advisor: Prof. Sandy Pentland

Stanford University, Palo Alto, California — 2012 - 2018
M.S. in Computer Science, Artificial Intelligence
B.A. in Economics, minor in History
Advisors: Chris Manning and Danqi Chen

RESEARCH & INDUSTRY EXPERIENCE

The Data Provenance Initiative — June 2023 - Present
Founder & Lead
A collective of AI researchers passionate about data audits.
Research Advisors: Sara Hooker, Stella Biderman, Sandy Pentland

Cohere 4 AI, Aya Open Science Initiative — March 2023 - Present, (Remote) US
Co-Leading Multilingual Instruction Tuning
Research Advisors: Sara Hooker

Google Brain, Reasoning Team — May 2022 - Sept 2022, (Remote) Canada
Google Student Researcher
Research Advisors: Jason Wei, Barret Zoph, Denny Zhou, & Adam Roberts

BigScience, BLOOM [22] & ROOTS [21] Teams — 2022, (Remote) US
Volunteer Contributor

Apple, Siri & Information Intelligence Team — Feb 2018 - June 2021, Seattle, Washington
Senior Applied Machine Learning Scientist
Research Advisor: Chris Dubois

Stanford NLP Group — 2016 - 2017, Palo Alto, California
Research Assistant
Research Advisors: Chris Manning & Danqi Chen

Salesforce AI Research, previously Metamind — June 2016 - Sept 2016, Palo Alto, California
Deep Learning Research Intern
Research Advisors: Richard Socher & Caiming Xiong

SELECT AI PUBLICATIONS

[1] Global MMLU: Understanding and Addressing Cultural & Linguistic Biases in Multilingual Evaluation
Shivalika Singh, Angelika Romanou, Clémentine Fourrier, David I Adelani, Jian Gang Ngui, Daniel Vila-Suero, Peerat Limkonchotiwat, Kelly Marchisio, Wei Qi Leong, Yosephine Susanto, Raymond Ng, **Shayne Longpre**, Wei-Yin Ko, Madeline Smith, Antoine Bosselut, Alice Oh, Andre FT Martins, Leshem Choshen, Daphne Ippolito, Enzo Ferrante, Marzieh Fadaee, Beyza

Ermis, Sara Hooker
Under Review

[2] The foundation model transparency index v1. 1: May 2024
Rishi Bommasani, Kevin Klyman, Sayash Kapoor, **Shayne Longpre**, Betty Xiong, Nestor Maslej, Percy Liang
Under Review

[3] Bridging the Data Provenance Gap Across Text, Speech and Video
Shayne Longpre, ... (50 authors), Caiming Xiong, Luis Villa, Stella Biderman, Alex Pentland, Sara Hooker, Jad Kabbara
ICLR 2025

[4] The Responsible Foundation Model Development Cheatsheet: A Review of Tools & Resources
Shayne Longpre, Stella Biderman, Alon Albalak, Hailey Schoelkopf, Daniel McDuff, Sayash Kapoor, Kevin Klyman, Kyle Lo, Gabriel Ilharco, Nay San, Maribeth Rauh, Aviya Skowron, Bertie Vidgen, Laura Weidinger, Arvind Narayanan, Victor Sanh, David Adelani, Percy Liang, Rishi Bommasani, Peter Henderson, Sasha Luccioni, Yacine Jernite, Luca Soldaini
TMLR 2025

[5] Consent in Crisis: The Rapid Decline of the AI Data Commons
Shayne Longpre, ... (50 authors), Luis Villa, Stella Biderman, Hanlin Li, Daphne Ippolito, Sara Hooker, Jad Kabbara, Sandy Pentland
NeurIPS 2024

[6] A Systematic Review of NeurIPS Dataset Management Practices
Yiwei Wu, Leah Ajmani, **Shayne Longpre**, Hanlin Li
NeurIPS 2024

[7] A Safe Harbor for AI Evaluation and Red Teaming
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
ICML 2024, Oral (1.5%)

[8] On the Societal Impact of Open Foundation Models
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 Storchan, Daniel Zhang, Daniel E Ho, Percy Liang, Arvind Narayanan
ICML 2024, Oral (1.5%)

[9] AI-Powered Autonomous Weapons Risk Geopolitical Instability and Threaten AI Research
Riley Simmons-Edler, Ryan Badman, **Shayne Longpre**, Kanaka Rajan
ICML 2024, Oral (1.5%)

[10] Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?
Shayne Longpre, Robert Mahari, Naana Obeng-Marnu, William Brannon, Tobin South, Katy Gero, Sandy Pentland, Jad Kabbara
ICML 2024, Spotlight

[11] The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI
Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, Deb Roy, Sara Hooker
Nature Machine Intelligence, 2024

[12] A Survey on Data Selection for Language Models
Alon Albalak, Yanai Elazar, Sang Michael Xie, **Shayne Longpre**, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, Colin Raffel, Shiyu Chang, Tatsunori Hashimoto, William Yang Wang
TMLR 2024

[13] Aya model: An instruction finetuned open-access multilingual language model
Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Freddie Vargus, Phil Blunsom, **Shayne Longpre**, Niklas Muennighoff, Marzieh Fadaee, Julia Kreutzer, Sara Hooker
ACL 2024, Best Paper Award

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

[15] Prometheus: Inducing Fine-grained Evaluation Capability in Language Models
Seungone Kim, Jamin Shin, Yejin Cho, Joel Jang, **Shayne Longpre**, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, Minjoon Seo
ICLR 2024

[16] OctoPack: Instruction Tuning Code Large Language Models
Niklas Muennighoff, Qian Liu, Armel Zebaze, Qinkai Zheng, Binyuan Hui, Terry Yue Zhuo, Swayam Singh, Xiangru Tang, Leandro von Werra, **Shayne Longpre**
ICLR 2024

[17] Mixture-of-Experts Meets Instruction Tuning: A Winning Combination for Large Language Models
Sheng Shen, Le Hou, Yanqi Zhou, Nan Du, **Shayne Longpre**, Jason Wei, Hyung Won Chung, Barret Zoph, William Fedus, Xinyun Chen, Tu Vu, Yuexin Wu, Wuyang Chen, Albert Webson, Yunxuan Li, Vincent Zhao, Hongkun Yu, Kurt Keutzer, Trevor Darrell, Denny Zhou
ICLR 2024

[18] A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity
Shayne Longpre, Gregory Yauney, Emily Reif, Katherine Lee, Adam Roberts, Barret Zoph, Denny Zhou, Jason Wei, Kevin Robinson, David Mimno, Daphne Ippolito
NAACL 2024, Outstanding Paper Award

[19] Scaling instruction-finetuned language models
{Hyung Won Chung, Le Hou, **Shayne Longpre**}, ... Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, Jason Wei (35 authors)
JMLR 2024

[20] The Flan Collection: Designing data and methods for effective instruction tuning
Shayne Longpre, Le Hou, Tu Vu, Albert Webson, Hyung Won Chung, Yi Tay, Denny Zhou, Quoc V. Le, Barret Zoph, Jason Wei, Adam Roberts
ICML 2023

[21] The Bigscience Roots Corpus: A 1.6 tb composite multilingual dataset
Hugo Laurençon... **Shayne Longpre**... Margaret Mitchell, Sasha Luccioni, Yacine Jernite (52 authors)
NeurIPS 2022

[22] BLOOM: A 176B-Parameter Open-Access Multilingual Language Model
Teven Le Scao, ... **Shayne Longpre**, ... Matteo Manica (128 authors)
ArXiv, 2022.

[23] Combining Compressions for Multiplicative Size Scaling on Natural Language Tasks

Rajiv Movva, Jinhao Lei, **Shayne Longpre**, Ajay Gupta, Chris DuBois
COLING 2022

[24] You reap what you sow: On the Challenges of Bias Evaluation Under Multilingual Settings
Zeerak Talat, Aurélie Névéol, Stella Biderman, Miruna Clinciu, Manan Dey, **Shayne Longpre**, Sasha Luccioni, Maraim Masoud, Margaret Mitchell, Dragomir Radev, Shanya Sharma, Arjun Subramonian, Jaesung Tae, Samson Tan, Deepak Tunuguntla, Oskar Van Der Wal
ACL 2022 BigScience Workshop

[25] MIA 2022 Shared Task: Evaluating Cross-lingual Open-Retrieval Question Answering for 16 Diverse Languages
Akari Asai, **Shayne Longpre**, Jungo Kasai, Chia-Hsuan Lee, Rui Zhang, Junjie Hu, Ikuya Yamada, Jonathan H. Clark, Eunsol Choi
NAACL 2022 Multilingual Information Access Workshop

[26] Active Learning Over Multiple Domains in Natural Language Tasks
Shayne Longpre, Julia Reisler, Edward Greg Huang, Yi Lu, Andrew Frank, Nikhil Ramesh, Chris DuBois
NeurIPS 2022 Workshop on Distribution Shifts

[27] Entity-Based Knowledge Conflicts in Question Answering
{**Shayne Longpre**, Kartik Perisetla, Anthony Chen}, Nikhil Ramesh, Chris DuBois, Sameer Singh
EMNLP 2021

[28] MKQA: A Linguistically Diverse Benchmark for Multilingual Open Domain Question Answering
Shayne Longpre, Yi Lu, Joachim Daiber
TACL 2021

[29] Evaluating Entity Disambiguation and the Role of Popularity in Retrieval-Based NLP
Anthony Chen, Pallavi Gudipati, **Shayne Longpre**, Xiao Ling, Sameer Singh
ACL 2021

[30] Evaluating Question Rewriting for Conversational Question Answering
Svitlana Vakulenko, **Shayne Longpre**, Zhucheng Tu, Raviteja Anantha
WSDM 2021

[31] Open-Domain Question Answering Goes Conversational via Question Rewriting
Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, **Shayne Longpre**
NAACL 2021

[32] Pivot Through English: Reliably Answering Multilingual Questions without Document Retrieval
Ivan Montero, **Shayne Longpre**, Ni Lao, Andrew Frank, Christopher DuBois
NAACL 2021 Multilingual Information Access Workshop

[33] On the Transferability of Minimal Prediction Preserving Inputs in Question Answering
Shayne Longpre, Yi Lu, Chris DuBois
NAACL 2021

[34] A Wrong Answer or a Wrong Question? An Intricate Relationship between Question Reformulation and Answer Selection in Conversational Question Answering
Svitlana Vakulenko, **Shayne Longpre**, Zhucheng Tu, Raviteja Anantha
EMNLP 2020 Search-Oriented Conversational AI Workshop Best Paper Award

POLICY WRITINGS

[35] International AI Safety Report
A team of authors, led by Yoshua Bengio. **Shayne Longpre**, as part of the core writing group.

[36] Considerations for governing open foundation models
Rishi Bommasani, Sayash Kapoor, Kevin Klyman, **Shayne Longpre**, Ashwin Ramaswami,

Daniel Zhang, Marietje Schaake, Daniel E Ho, Arvind Narayanan, Percy Liang.
Science 2024
Stanford HAI, Foundation Model Issue Brief Series, 2023.

[37] UK Gov: International Scientific Report on the Safety of Advanced AI — Interim Report (2024)
A team of authors, led by Yoshua Bengio. **Shayne Longpre**, as part of the core writing group.

[38] An Open Letter: A Safe Harbor for Independent AI Evaluation
An open letter signed by 350+ researchers, journalists, and civil society members. Effort led by **Shayne Longpre**, as part of [7], 2024.

[39] A Safe Harbor for AI Evaluation and Red Teaming
Shayne Longpre, Sayash Kapoor, Kevin Klyman, Ashwin Ramaswami, Rishi Bommasani, Arvind Narayanan, Percy Liang, Peter Henderson
Knight First Amendment Institute at Columbia University cross-posted with AI Snake Oil blog, 2024.

[40] Long Comment to the US Copyright Office, Ninth Triennial Proceeding, Class 4
Kevin Klyman, **Shayne Longpre**, Sayash Kapoor, Arvind Narayanan, Aleksandra Korolova, Peter Henderson
Comment to the US Copyright Office, 2024.

[41] Discit Ergo Est: Training Data Provenance and Fair Use
Robert Mahari, **Shayne Longpre**
Network Law Review, 2024.

[42] Request for Comment on Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights
Researchers from Stanford HAI, CRFM, RegLab, and other institutions
Stanford HAI, Comment to National Telecommunications and Information Administration, 2024.

[43] Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies
Shayne Longpre, Marcus Storm, Rishi Shah
MIT Science Policy Review, Volume III, 2022.

[44] Invigorating Competition in Social Networking: An Interoperability Remedy
Cristian Santesteban, **Shayne Longpre**
Competition Policy International, 2021.

[45] How Big Data Confers Market Power to Big Tech: Leveraging the Perspective of Data Science
Cristian Santesteban, **Shayne Longpre**
The Antitrust Bulletin, 2020.

AWARDS & FUNDRAISING

ACL 2024 Best Paper Award [13] 2024
NAACL 2024 Outstanding Paper Award [18] 2024
Federation of American Scientists AI Legislative Proposal Award 2024
Awarded to top "New Legislative Proposals To Deploy Artificial Intelligence Strategically"
Mozilla Data Futures Lab, Infrastructure Grant Award, 2023 2023
Awarded for the Data Provenance Initiative, accompanied by $25,000 research grant
MIT Generative AI Impact Award, 2023 2023
Awarded for the Data Provenance Initiative, accompanied by $70,000 research grant.
ICLR Highlighted Reviewer (3%) 2022
EMNLP 2020, *Search-Oriented Conversational AI Workshop* **Best Paper Award** 2020
EMNLP 2019, *MRQA Workshop Shared Task* **2nd place** 2019
TREC 2019, *Conversational Assistance Track (CAsT) Shared Task* **1st place ("A Team")** 2019

TEACHING EXPERIENCE

Instructor, MAS.S68 Generative AI: Evaluation and New Research Methods, MIT 2023

Instructed research seminar on large language models, and the landscape of socio-political concerns with their adoption.

Instructor, AI4ALL, Stanford University 2017

Instructed course on Computer Vision fundamentals to young women in STEM.

Teaching Assistant, Natural Language Processing w/ Deep Learning (CS224N), Stanford 2017

Teaching Assistant, Computer Vision with Deep Learning (CS231N), Stanford 2017

SERVICE

Leadership & Organization

Workshop Organizer Instruction Following & Finetuning Workshop (ITIF), NeurIPS 2023 2023

Model Training Co-Lead Cohere For AI Aya Initiative 2023 - Present

MozFest Facilitator Bringing Light to Shadow Data 2023

Workshop Co-Lead Organizer Defining Transparency Workshop, Brown University 2022

(*hosted w/ the Algorithmic Transparency Institute and Brown's Information Futures Lab*)

Workshop Organizer Multilingual Information Access Workshop (MIA), NAACL 2022 2022

Shared-task Organizer MIA 2022 Shared task, NAACL 2022 2022

Academic Service

Reviewer (2024): ARR, ICML, NeurIPS, AAAI, ICLR, COLM

Area Chair (2023): EMNLP, *Large Language Models and the Future of NLP Track*

ACL Professional Conduct Committee 2021 - 2023

Trained volunteer, responding to complaints from ACL's Anti-Harassment Policy.

Reviewer (2023): ARR, NeurIPS, FAccT, ICML, ICLR, EMNLP

Reviewer (2022): ARR, NeurIPS, ICLR, EMNLP, various workshops

ADVISING

Volunteer at MIT **Students Offering Support Program** 2022-2024

Advising underrepresented students in MIT graduate applications.

Niklas Muennighoff, Peking University → Now Stanford CS PhD. Published [5][13][16]. 2023-2025

Campbell Lund, Wellesley CS → Now Edinburgh University AI Ethics MS. Published [5]. 2023-2024

An My Dinh, MIT Undergrad. Published [5]. 2023 - 2024

Minnie Liang, MIT Undergrad. Published [5]. 2023 - 2024

Seungone Kim, KAIST AI MS → Now CMU LTI CS PhD student. Published [15]. 2022 - 2023

Rajiv Movva, MIT CS → Now Cornell Tech CS PhD. Published [23]. 2021 - 2022

Xuhui Zhou, UW MS → Now LTI-CMU CS PhD. 2021 - 2022

Erik Jones, Stanford MS → Now UC Berkeley CS PhD. 2021

Ivan Montero, UW CS MS student → Now Apple Machine Learning. Published [32]. 2020 - 2021

INVITED TALKS & PANELS

Talks & Lectures for the Multimodal Data Provenance [3]

Twelve Labs — Multimodal Weekly (Host: James Le) 2024

Talks & Lectures for the Foundation Model Development Cheatsheet [4]

Linux Foundation AI & Data Seminar Series (Host: Anni Lai) 2024

Talks & Lectures for the Consent in Crisis [5]

BBC Radio 4 — Will AI Eat Itself? — Podcast Guest 2024

Risks of AI in the Military — Panel Moderator 2024

Women in AI & Robotics Reading Group (Host: Cleo Norris) 2024

UC Berkeley Responsible AI Workshop (Host: Genevieve Smith) 2024

MIT Sloan Initiative on the Digital Economy Guest Speaker (Host: Sinan Aral) 2024

AI Tinkerer Paper Club (Host: Human Feedback Foundation) 2024
Mozilla AI Salon (Host: Abeba Birhane) 2024
PLAMADISO – Platforms, Markets, and the Digital Society (Host: Volker Stocker) 2024
Creative Commons – Workshop on Preference Signals (Host: Anna Tumadóttir) 2024
MozFest Data Futures Lab Showcase (Host: Mozilla) 2024
Stanford HAI (Host: Digital Economy Lab) 2024
Sony AI (Host: Wiebke Hutiri) 2024
MIT CIO Symposium on AI (Host: Irving Wladawsky-Berger) 2024

Talks & Lectures for A Safe harbor for AI Evaluation & Red Teaming [7]
Harvard Kempner Center — Reading Group (Host: Ryan Badman) 2024
Alignment Workshop Lightning Talk 2024

Talks & Lectures for the Data Provenance Initiative [11]
MIT Imagination in Action (Host: MIT Media Lab) 2024
USC NLG Seminar Series (Host: Justin Cho) 2024
UT Austin Data Ethics course, Guest Lecture (Host: Hanlin Li) 2024
MLCommons Croissant Group 2024
Mozilla Data Futures Lab Speaker Series 2024
Ethical Commerce Alliance (Host: Nina Müller) 2023
Harvard Library Innovation Lab (Host: Greg Leppert) 2023
MIT Algorithmic Alignment Group (Host: Dylan Hadfield-Menell) 2023

Talks & Lectures for the A Pretrainer's Guide [18]
Microsoft Research India (Host: Sanchit Ahuja) 2024
Salesforce AI (Host: Caiming Xiong) 2024
Cohere For AI (Host: Sara Hooker) 2024
Mosaic ML (Host: Jonathan Frankle) 2023
Harvard & MIT: Policymaking for AI Series, Invited Talk & Panel (Host: Getting Plurality Research Network) 2023
University of Washington (Hosts: Akari Asai, Sewon Min) 2023
Allen Institute of AI (AI2) (Host: Maria Antoniak) 2023

Talks & Lectures for Effective Instruction Tuning [20][19]
Instituto Superior Técnico Seminar Series (Host: Nuno Guerreiro) 2023
Amazon Data-centric AI Seminar Series (Host: Li Lihong) 2023
Databricks Seminar Series (Host: Mike Conover) 2023
Apple Applied ML Reading Group (Host: Michael Tu) 2023
Kailua Labs AI Seminar Series (Host: Pablo Mendes) 2023
Oracle ML Seminar Series (Host: Ari Kobren) 2023
Google Research (Host: Denny Zhou) 2022

Other Talks & Lectures
Truth & Trust Online 2022 *Evaluating Transparency in Online Social Platforms* 2022
Panel moderator at NAACL 2022, *MIA Workshop* 2022
UC Irvine Reading Group *Knowledge Conflicts in QA* [27] (Host: Sameer Singh) 2021
Question Answering Evaluation Panel moderator at EMNLP 2021, *SCAI Workshop* 2021

PRESS & MEDIA **Select Press for Multimodal Data Provenance [3]** 2024
MIT Technology Review. *This is where the data to build AI comes from*

The Globe and Mail. *AI-generated video has come a long way. Can you spot the difference between real and fake?*

Select Press for Consent in Crisis [5] 2024

The New York Times. *The Data That Powers A.I. Is Disappearing Fast*
Vox. *It's practically impossible to run a big AI company ethically*
Yahoo!Finance. *Data to train AI models is becoming restricted. Here's why.*
404 Media. *The Backlash Against AI Scraping Is Real and Measurable*
404 Media. *Anthropic AI Scraper Hits iFixit's Website a Million Times in a Day*
The StackOverflow Podcast. *The Stack Overflow Podcast: The Data Provenance Initiative*
MIT Technology Review. *AI that Feeds on a Diet of AI Garbage Ends up Spitting out Nonsense*
IEEE Spectrum. *AI Has Created a Battle Over Web Crawling*
Wired. *A New Group Is Trying to Make AI Data Licensing Ethical*
Le Monde. *A cause des intelligences artificielles, le Web se ferme de plus en plus*
Variety. *Generative AI & Licensing: A Special Report*
Nature Press. *The AI revolution is running out of data. What can researchers do?*
Riff Reporter. *Immer weniger aktuelle Daten für das Training: Künstliche Intelligenz in der Kris*
The Observer. *AI Companies Are Running Out of Training Data: Study*
Futurism. *Crisis Looms as AI Companies Rapidly Losing Access to Training Data*
Start Magazine. *L'intelligenza artificiale è a corto di dati?*
Mozilla. *AI Training Can Undermine the Open Web. This Team Is Thinking Through Solutions*

Select Press for the Geopolitical risks of Autonomous Weaponry [9] 2024

Harvard Medical School News. *The Risks of Artificial Intelligence in Weapons Design*

Select Press for A Safe Harbor for AI Evaluation & Red Teaming [7] [38] 2024

The Washington Post. *Top AI researchers say OpenAI, Meta hinder independent evaluations*
VentureBeat. *Experts call for 'safe harbor' so researchers, journalists & artists can evaluate AI*
404 Media. *It May Soon Be Legal to Jailbreak AI to Expose How it Works*
Vox. *How would we even know if AI went rogue?*
Decipher. *The Emerging Ecosystem Dedicated to AI Accountability*

Select Press for the Aya Model [13] 2023

Cohere For AI. *The Journey of Aya - Accelerating Multilingual AI Through Open Science*

Select Press for Data Provenance Initiative [11] 2023

The Washington Post. *AI researchers uncover ethical, legal risks to using popular data sets*
VentureBeat. *MIT, Cohere for AI, others launch platform to track and filter audited AI datasets*
IEEE Spectrum. *Public AI Training Datasets Are Rife With Licensing Errors*
MIT News. *Study: Transparency is often lacking in datasets used to train large language models*
Reuters. *Legal transparency in AI finance: facing the accountability dilemma in digital decision-making*
TechCircle. *MIT, Cohere for AI, others launch platform to enhance transparency in AI data*
Cohere Blog. *Data Provenance Explorer Launches to Tackle Data Transparency Crisis*

Select Press for Foundation Model Transparency Index [14] 2023

Stanford HAI Blog. *Introducing The Foundation Model Transparency Index*
The New York Times. *Maybe we will finally learn more about how AI works*
The Atlantic. *We Don't Actually Know If AI Is Taking Over Everything*
Bloomberg. *Klobuchar Says AI Regulation Still Possible Before End of Year*
The Information. *How Transparent is your model?*
VentureBeat. *How transparent are AI models? Stanford researchers found out.*

The Verge. *The world's biggest AI models aren't very transparent, Stanford study says*
Reuters. *Stanford researchers issue AI transparency report, urge tech companies to reveal more*
Fast Company. *Why everyone seems to disagree on how to define Artificial General Intelligence*

Hacker News Our course on Generative AI trending 2023
Google AI Blog The Flan Collection: Advancing open source methods for instruction tuning 2023
PaLM 2 Technical Report Flan Collection & Methods cited several times as key components. 2023
DAIR.AI Top ML Papers of the Week 2023

LAST UPDATED *August 2024.*