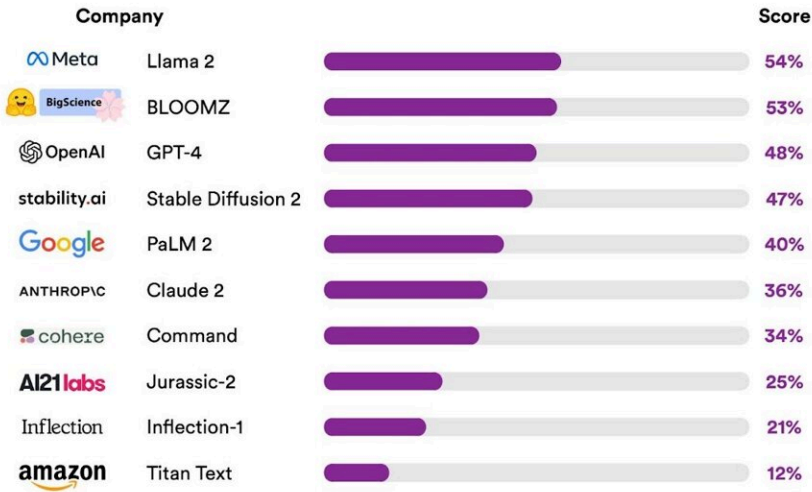


New index rates transparency of ten foundation model companies, and finds them lacking

October 29 2023, by Katharine Miller

Foundation Model Transparency Index Total Scores, 2023
Source: 2023 Foundation Model Transparency Index



Credit: Stanford University

Companies in the foundation model space are becoming less transparent, says Rishi Bommasani, Society Lead at the Center for Research on Foundation Models (CRFM), within Stanford HAI. For example, OpenAI, which has the word "open" right in its name, has clearly stated that it will not be transparent about most aspects of its flagship model,

GPT-4.

Less [transparency](#) makes it harder for other businesses to know if they can safely build applications that rely on commercial foundation models; for academics to rely on commercial foundation models for research; for policymakers to design meaningful policies to rein in this powerful technology; and for consumers to understand model limitations or seek redress for harms caused.

To assess transparency, Bommasani and CRFM Director Percy Liang brought together a multidisciplinary team from Stanford, MIT, and Princeton to design a scoring system called the [Foundation Model Transparency Index](#). The FMTI evaluates 100 different aspects of transparency, from how a company builds a foundation model, how it works, and how it is used downstream.

When the team scored 10 major foundation model companies using their 100-point index, they found plenty of room for improvement: The highest scores, which ranged from 47 to 54, aren't worth crowing about, while the lowest score bottoms out at 12. "This is a pretty clear indication of how these companies compare to their competitors, and we hope will motivate them to improve their transparency," Bommasani says.

Another hope is that the FMTI will guide policymakers toward effective regulation of foundation models. "For many policymakers in the EU as well as in the U.S., the U.K., China, Canada, the G7, and a wide range of other governments, transparency is a major policy priority," Bommasani says.

The [index](#), accompanied by an extensive [100-page paper](#) on the methodology and results, makes available all of the data on the 100 indicators of transparency, the protocol use for scoring, and the

developers' scores along with justifications. The paper has also been posted to the *arXiv* preprint server.

Why transparency matters

A lack of transparency has long been a problem for consumers of digital technologies, Bommasani notes. We've seen deceptive ads and pricing across the internet, unclear wage practices in ride-sharing, dark patterns tricking users into unknowing purchases, and myriad transparency issues around content moderation that have led to a vast ecosystem of mis- and disinformation on [social media](#). As transparency around commercial FMs wanes, we face similar sorts of threats to consumer protection, he says.

In addition, transparency around commercial foundation models matters for advancing AI policy initiatives and ensuring that upstream and downstream users in industry and academia have the information they need to work with these models and make informed decisions, Liang says.

Foundation models are an increasing focus of AI research and adjacent scientific fields, including in the social sciences, says Shayne Longpre, a Ph.D. candidate at MIT: "As AI technologies rapidly evolve and are rapidly adopted across industries, it is particularly important for journalists and scientists to understand their designs, and in particular the raw ingredients, or data, that powers them."

To policymakers, transparency is a precondition for other policy efforts. Foundation models raise substantive questions involving intellectual property, labor practices, energy use, and bias, Bommasani says. "If you don't have transparency, regulators can't even pose the right questions, let alone take action in these areas."

And then there's the public. As the end-users of AI systems, Bommasani says, they need to know what foundation models these systems depend on, how to report harms caused by a system, and how to seek redress.

Creating the FMTI

To build the FMTI, Bommasani and his colleagues developed [100 different indicators of transparency](#). These criteria are derived from the AI literature as well as from the social media arena, which has a more mature set of practices around consumer protection.

About a third of the indicators relate to how foundation model developers build their models, including information about training data, the labor used to create it, and the computational resources involved. Another third is concerned with the model itself, including its capabilities, trustworthiness, risks, and mitigation of those risks. And the final third involves how the models are being used downstream, including disclosure of company policies around model distribution, user data protection and model behavior, and whether the company provides opportunities for feedback or redress by affected individuals.

Foundation Model Transparency Index Scores by Major Dimensions of Transparency, 2023

Source: 2023 Foundation Model Transparency Index

	Meta	BigScience	OpenAI	stability.ai	Google	ANTHROPIC	cohere	AI21labs	Inflection	amazon	Average
	Llama 2	BLOOMZ	GPT-4	Stable Diffusion 2	PaLM 2	Claude 2	Command	Jurassic-2	Inflection-1	Titan Text	
Data	40%	60%	20%	40%	20%	0%	20%	0%	0%	0%	20%
Labor	29%	86%	14%	14%	0%	29%	0%	0%	0%	0%	17%
Compute	57%	14%	14%	57%	14%	0%	14%	0%	0%	0%	17%
Methods	75%	100%	50%	100%	75%	75%	0%	0%	0%	0%	48%
Model Basics	100%	100%	50%	83%	67%	67%	50%	33%	50%	33%	63%
Model Access	100%	100%	67%	100%	33%	33%	67%	33%	0%	33%	57%
Capabilities	60%	80%	100%	40%	80%	80%	60%	60%	40%	20%	62%
Risks	57%	0%	57%	14%	29%	29%	29%	29%	0%	0%	24%
Mitigations	60%	0%	60%	0%	40%	40%	20%	0%	20%	20%	26%
Distribution	71%	71%	57%	71%	71%	57%	57%	43%	43%	43%	59%
Usage Policy	40%	20%	80%	40%	60%	60%	40%	20%	60%	20%	44%
Feedback	33%	33%	33%	33%	33%	33%	33%	33%	33%	0%	30%
Impact	14%	14%	14%	14%	14%	0%	14%	14%	14%	0%	11%
Average	57%	52%	47%	47%	41%	39%	31%	20%	20%	13%	

Credit: Stanford University

The indicators are designed to circumvent some of the traditional tradeoffs between transparency and other values, such as privacy, security, competitive advantage, or concerns about misuse by bad actors, Bommasani says. "Our intent is to create an index where most indicators are not in contention with competitive interests; by looking at precise matters, the tension between transparency and competition is largely obviated," he says. "Nor should disclosure risk facilitating misuse by other actors in the ecosystem." In fact, for some indicators, a point is awarded if the company doesn't disclose the requested information but does justify why it's not disclosed.

The index intentionally does not focus on rating corporate responsibility. Bommasani says. If a company discloses that training their models requires a lot of energy, that they don't pay their workers a living wage, or that their downstream users are doing something harmful, the

company will still get an FMTI point for those disclosures.

Although more responsible conduct by foundation model companies is the goal, transparency is a first step in that direction, Bommasani says. By surfacing all the facts, the FMTI sets the conditions that will allow a regulator or a lawmaker to decide what needs to change. "As researchers, we play an instrumental role in enabling other actors with greater teeth in the ecosystem to enact substantive policy changes."

The scores

To rate the top model creators, the research team used a structured search protocol to collect publicly available information about each company's leading foundation model. This included reviewing the companies' websites as well as performing a set of reproducible Google searches for each company. "In our view, if this rigorous process didn't find information about an indicator, then the company hasn't been transparent about it," says Kevin Klyman, a Stanford MA student in international policy and a lead co-author of the study.

After the team came up with a first draft of the FMTI ratings, they gave the companies an opportunity to respond. The team then reviewed the company rebuttals and made modifications when appropriate.

Bommasani and his colleagues have now released the scores for 10 companies working in the foundation model space. As shown in the accompanying graphic, Meta achieved the highest score—an unimpressive 54 out of 100.

"We shouldn't think of Meta as the goalpost with everyone trying to get to where Meta is," Bommasani says. "We should think of everyone trying to get to 80, 90, or possibly 100."

And there's reason to believe that is possible: Of the 100 indicators, at least one company got a point for 82 of them.

Maybe more important are the indicators where almost every company did poorly. For example, no [company](#) provides information about how many users depend on their model or statistics on the geographies or market sectors that use their model. Most companies also do not disclose the extent to which copyrighted material is used as training data. Nor do the companies disclose their labor practices, which can [be highly problematic](#).

"In our view, companies should begin sharing these types of critical information about their technologies with the public," Klyman says.

As the foundation model market matures and solidifies, and companies perhaps make progress toward greater transparency, it will be important to keep the FMTI up to date, Bommasani says. To make that easier, the team is asking companies to disclose information for each of the FMTI's indices in a single place, which will earn them an FMTI point. "It will be much better if we only have to verify the information rather than search it out," Bommasani says.

The FMTI's potential impact

Nine of the 10 companies the team evaluated have [voluntarily committed to the Biden-Harris administration](#) an intention to manage the risks posed by AI. Bommasani hopes the newly released FMTI will motivate these companies to follow through on those pledges by increasing transparency.

He also hopes the FMTI will help inform policy making by world governments. Case in point: The European Union is currently working to pass the AI Act. The European Parliament's position as it enters

negotiations requires disclosure of some of the indicators covered by the FMTI, but not all.

By highlighting where companies are falling short, Bommasani hopes the FMTI will help focus the EU's approach to the next draft. "I think this will give them a lot of clarity about the lay of the land, what is good and bad about the status quo, and what they could potentially change with legislation and regulation."

More information: The Foundation Model Transparency Index.
crfm.stanford.edu/fmti/fmti.pdf

Rishi Bommasani et al, The Foundation Model Transparency Index, *arXiv* (2023). [DOI: 10.48550/arxiv.2310.12941](https://doi.org/10.48550/arxiv.2310.12941)

Provided by Stanford University

Citation: New index rates transparency of ten foundation model companies, and finds them lacking (2023, October 29) retrieved 10 September 2024 from
<https://techxplore.com/news/2023-10-index-transparency-ten-foundation-companies.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.