AI Accountability Reporting

What we will cover in this module

- What makes a good accountability story?
- How to find those stories across the AI accountability framework
- Case studies of good stories

What makes a good accountability story?

Let’s recall our framework for reporting on the AI lifecycle

AI development stages	data +	compute →	AI models →	applications
Related issues	loss of privacy/consent surveillance capitalism labor exploitation erosion of IP	energy/water use mineral extraction emissions geopolitics	embedded abuse automation embedded bias errors/hallucinations	surveillance job loss/degradation discrimination mis/disinformation
Actors	companies governments	companies governments investors	companies talent investors	individuals businesses governments
Impacted people	individuals workers	communities planet citizens	individuals communities businesses	communities consumers democracies workers

Case Study #1

AI development stages

data + compute → AI models → applications

Related issues

surveillance

job loss/degradation

discrimination

mis/disinformation

Stakeholders/ actors

individuals

businesses

governments

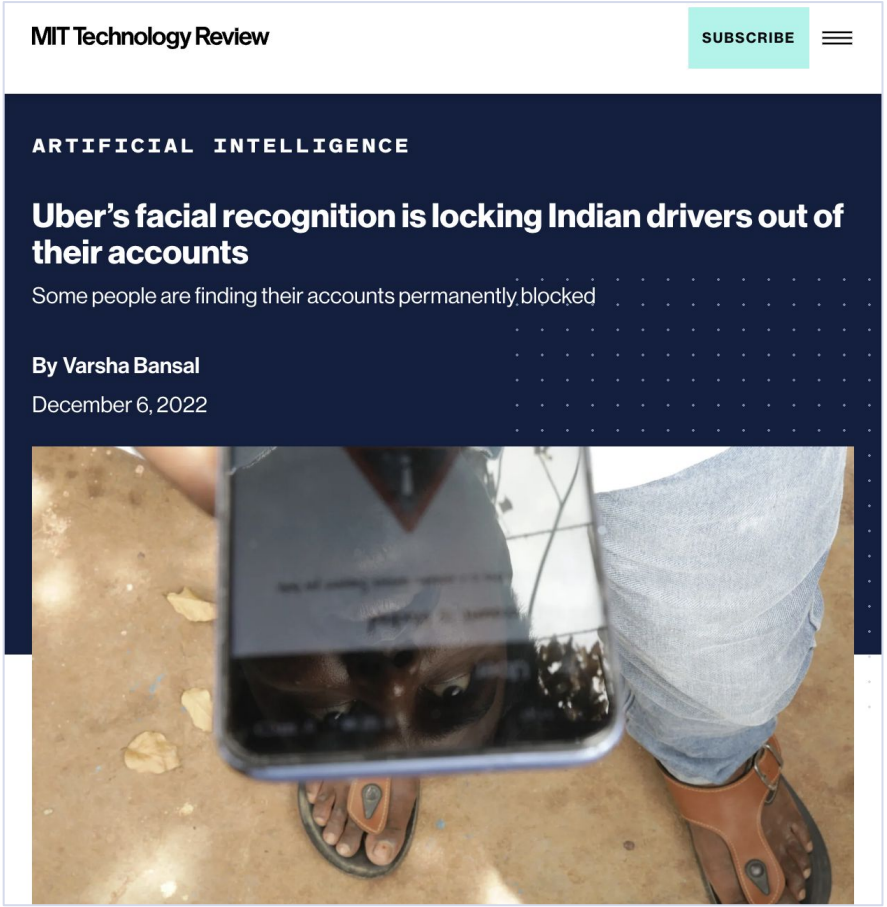
Impacted communities

communities

consumers

democracies

workers



MIT Technology Review

SUBSCRIBE

ARTIFICIAL INTELLIGENCE

Uber's facial recognition is locking Indian drivers out of their accounts

Some people are finding their accounts permanently blocked

By Varsha Bansal

December 6, 2022

Findings

- **Uber drivers in India are being systematically locked out of their accounts because the app’s faulty facial recognition**

- **Drivers struggle through tedious, weeks-long appeals processes that leave them without income they need to pay basic bills**

Challenges

- No access to Uber’s facial recognition algorithm
- Individual cases of harm, but want to prove that this harm is systematic

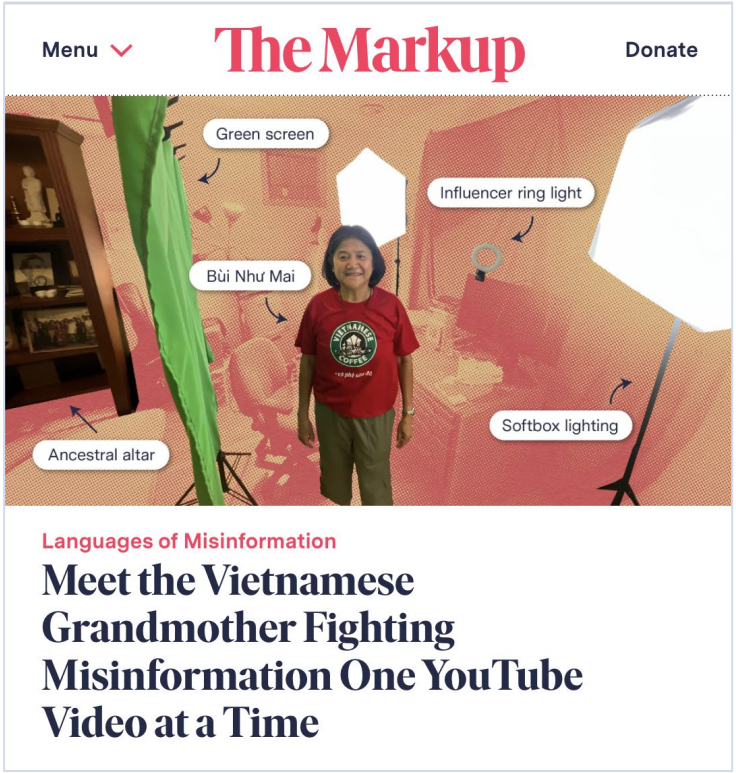
What would your reporting starting point be for this story?

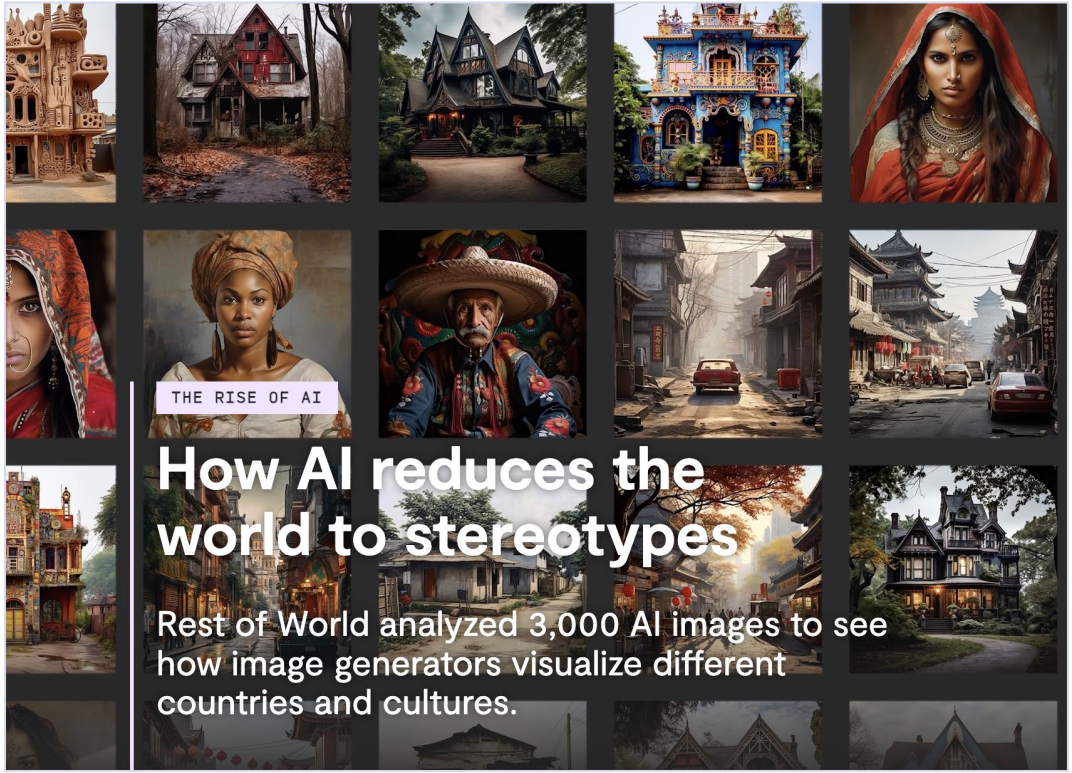
Methodology

- **Work with informal networks of gig workers in India**
- **Survey Uber drivers to see how many report being locked out of their accounts**
- **Find compelling narratives around the appeal procedure**

Uber drivers in India say that problems with the facial recognition they use to access their apps is costing them work. In a survey conducted by MIT Technology Review of 150 Uber drivers in the country, almost half say they have been either temporarily or permanently locked out of their accounts as a result of problems with their selfie. Many suspected that a change in their appearance, such as facial hair, a shaved head, or a haircut, was to blame. Another quarter of them believe it was due to low lighting.

resolved. "We have to keep calling their help line incessantly before they unlock our accounts, constantly telling us that the server is down," said Taqi, with a tone of frustration—but mostly a sense of defeat—in his voice. "It's like their server is always down."





Don’t have access to an AI system?
Build your own dataset!

A powerful example of AI Accountability Reporting...

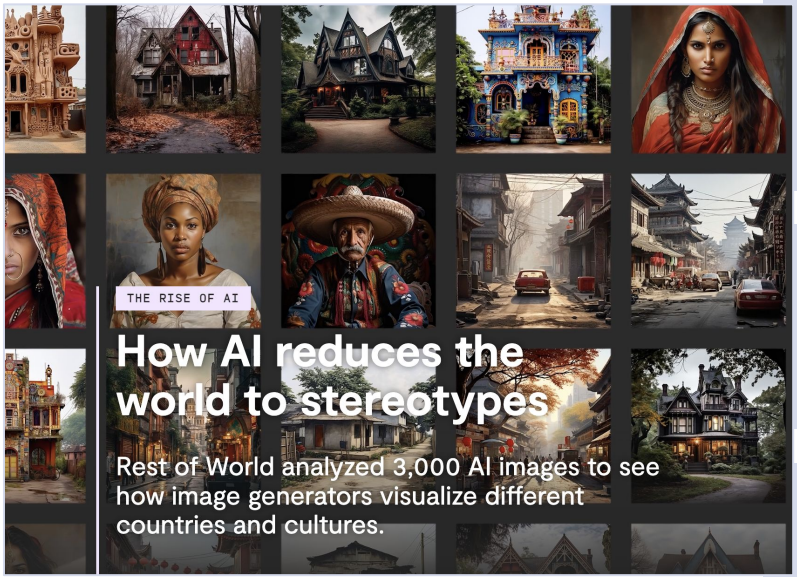
- Unique, ground-up methodology to demonstrate scale and circumvent opacity of the system
- Worked with informal networks of gig workers
- Illustrated consequences with human stories

Case Study #2

AI development stages

data + compute → AI models → applications

Related issues



embedded abuse	surveillance
automation	job loss/degradation
embedded bias	discrimination
errors/hallucinations	mis/disinformation

Actors

companies	individuals
talent	businesses
investors	governments

Impacted communities

individuals	communities
communities	consumers
businesses	democracies
	workers

Findings

- **Midjourney AI reproduces bias and stereotypes**
- **Flattens cultural differences and hierarchies**

A new *Rest of World* analysis shows that generative AI systems have tendencies toward bias, stereotypes, and reductionism when it comes to national identities, too.

Using Midjourney, we chose five prompts, based on the generic concepts of “a person,” “a woman,” “a house,” “a street,” and “a plate of food.” We then adapted them for different countries: China, India, Indonesia, Mexico, and Nigeria. We also included the U.S. in the survey for comparison, given Midjourney (like most of the biggest generative AI companies) is based in the country.


“An Indian person” is almost always an old man with a beard.


“A Mexican person” is usually a man in a sombrero.

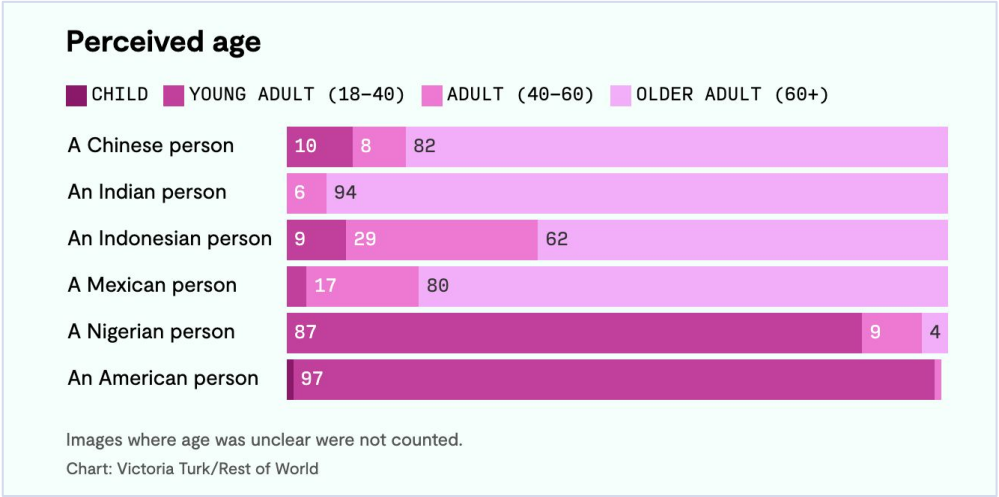
A Nigerian person

Prompt

Woman | Chunky jewellery | Street scene | Age 60+ | Red or orange clothing

Methodology

- **NOT a fancy statistical experiment; two people manually labeling data in a spreadsheet**
- **Analyze 3,000 images with consistent prompts**
- **Breakdown by country**



Hits on a core issue: Much attention is paid to how Generative AI functions in a Western context, while very little attention is paid to the rest of the world

Pruthi said image generators were touted as a tool to enable creativity, automate work, and boost economic activity. But if their outputs fail to represent huge swathes of the global population, those people could miss out on such benefits. It worries him, he said, that companies often based in the U.S. claim to be developing AI for all of humanity, “and they are clearly not a representative sample.”

Why it’s a good example of an AI Accountability Story

- Uses clever, low-tech approach to audit an AI system
- Demonstrates scale
- Situates the tech within a larger societal trend / context

Case Study #3

AI development stages

data + compute → AI models → applications

KAREN NAUNDORF SECURITY SEP 13, 2023 6:00 AM

The Twisted Eye in the Sky Over Buenos Aires

A scandal unfolding in Argentina shows the dangers of implementing facial recognition—even with laws and limits in place.

embedded abuse	surveillance
automation	job loss/degradation
embedded bias	discrimination
errors/hallucinations	mis/disinformation

companies	individuals
talent	businesses
investors	governments

individuals	communities
communities	consumers
businesses	democracies
	workers

Findings

- **AI facial recognition is deployed at scale in Buenos Aires to indiscriminately surveil residents.**

- **Errors in the system have left people wrongfully arrested.**

- **While the system was presented as a way to catch fugitives, in fact authorities have requested data from thousands of innocent citizens, including opposition figures and the vice president.**

What do you think the key reporting questions are for this story?

Methodology

- **Saw an opportunity in an ongoing scandal to go deeper**
- **Track down people affected by the system**
- **Isolate a key character (a critical judge investigating the facial recognition system) and use them as an entry point to find out more about the system**

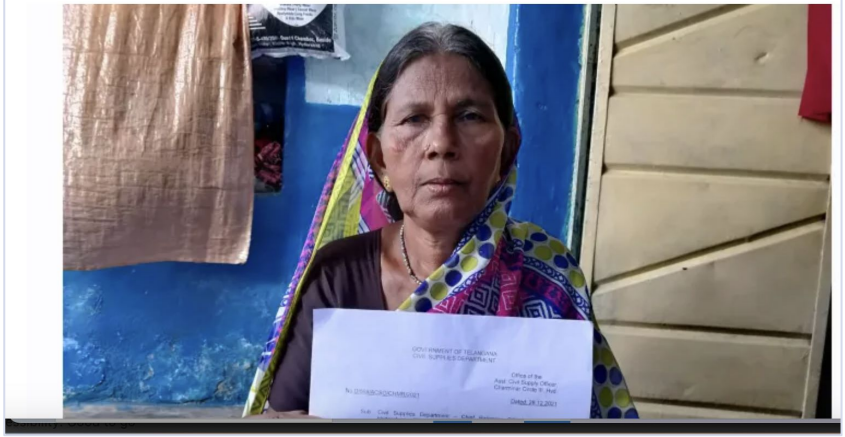
Gallardo says. He sits in his office on the boulevard Avenida de Mayo, just a few hundred meters from the pink government building. The office is spacious and flooded with light; a photo of the judge with Pope Francis hangs prominently on the wall. Historical penal codes stand on a bookshelf. “Only Argentina’s around 40,000 fugitives from justice may be searched for with the system,” he says. “But the number of personal data requested by the city was almost 10 million. The government could never explain why so much data was requested that did not belong to fugitives.”

ALJAZEERA News Israel War on Gaza Features Opinion Video

Economy | Poverty and Development

How an algorithm denied food to thousands of poor in India's Telangana

It adopted AI in welfare schemes to weed out ineligible ones, but has wrongfully removed thousands of legitimate ones.



investico onderzoeksjournalisten

To research About us Support us

'I want someone to say I'm not a fraud'

Duo's discriminatory fraud checks

Dear reader,

With successful journalistic collaboration, the research result also reaches exactly the group that concerns the news. For example, in the autumn of 2022 we started investigating a 'lead' that came from the Higher Education Press Agency: how does student financing organization DUO search for fraudsters? After a thorough reconstruction, we came across an exceptional number of lawsuits in which alleged fraudsters had a migration background. That was the moment to ask NOSop3, with a large and diverse viewing audience, on board. Together with them, we called more than 80 lawyers who showed us that in 97% of their lawsuits an accused student had a migration background.

Lawyers and people in the legal system can be a great way to access human stories and technical details!

Why it’s a powerful example of good AI Accountability Reporting

- **Captures and centers human stories**
- **Uses the legal system to establish damning facts about how the system works**
- **Situates findings within the larger context of growing surveillance states in Latin America**

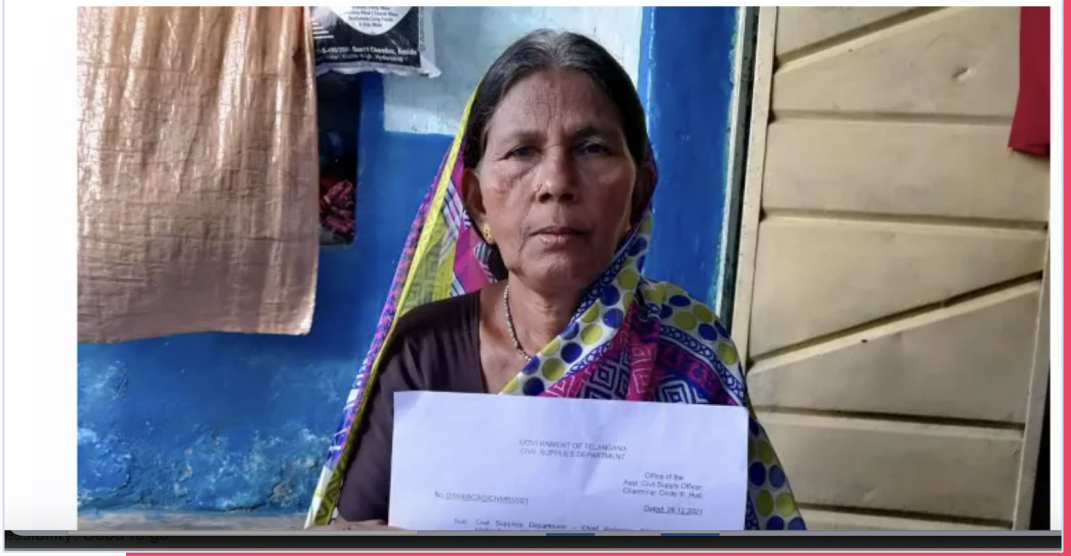
Case Study #3

Case Study #2

Economy | Poverty and Development

How an algorithm denied food to thousands of poor in India's Telangana

It adopted AI in welfare schemes to weed out ineligible ones, but has wrongfully removed thousands of legitimate ones.



AI development stages

data + compute → AI models → applications

Related issues

embedded abuse	surveillance
automation	job loss/degradation
embedded bias	discrimination
errors/hallucinations	mis/disinformation

Actors

companies	individuals
talent	businesses
investors	governments

Impacted communities

individuals	communities
communities	consumers
businesses	democracies
	workers

ALJAZEERA News Israel War on Gaza Features Opinion Video

Economy | Poverty and Development

How an algorithm denied food to thousands of poor in India's Telangana

It adopted AI in welfare schemes to weed out ineligible ones, but has wrongfully removed thousands of legitimate ones.

Findings

- **An AI system deployed to determine food benefits eligibility excluded thousands of people — with profound consequences.**

- **People wrongly thrown off welfare had to prove their own innocence.**

Bee is not the only victim of such digital snafus. From 2014 to 2019, Telangana cancelled more than 1.86 million existing food security cards and rejected 142,086 fresh applications without any notice.

The government initially claimed that these were all fraudulent claimants of subsidy and that being able to “weed out” the ineligible beneficiaries had saved it large sums of money.

But our investigation reveals that several thousands of these exclusions were done wrongfully, owing to faulty data and bad algorithmic decisions by Samagra Vedika.

Once excluded, the onus is on the removed beneficiaries to prove to government agencies that they were entitled to the subsidised food.

Even when they did so, officials often favoured the decision of the algorithm.

What would your reporting starting point be for this story?

Good AI accountability stories DO NOT require fancy technical data work.

Methodology

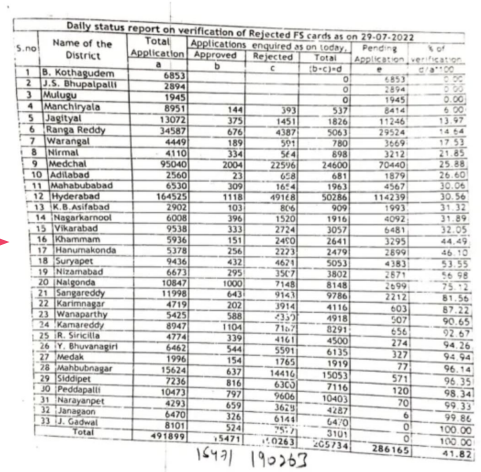
- **Find people most affected by working together with lawyers / civil society**
- **Obtain statistics that show scale of harm**
- **Put findings against company and government claims of fairness and efficiency**

The state IT department denied our requests under the Right to Information Act to share the source code and the formats of the data used by Samagra Vedika to make decisions, saying the company had “rights over” them. Posidex Technologies denied our requests for an interview.

Over a dozen interviews with state officials, activists and those excluded from welfare schemes, as well as perusal of a range of documents including bidding records, gave us a glimpse of how Samagra Vedika uses algorithms to triangulate a person’s identity in multiple government databases – as many as 30 – and combines all the information to build a person’s profile.

Building connections between scale and harm

Bismillah Bee with the old 'Below Poverty Line' card of the family issued in 2006 [Courtesy of The Reporters'



Daily status report on verification of Rejected FS cards as on 29-07-2022

S.no	Name of the District	Total Application	Applications enquired as on today,			Pending	% of
			Approved	Rejected	Total	Application	verification
		a	b	c	(b+c)=d	e	d*100/a
1	B. Kothagudem	6853			0	6853	0.00
2	J.S. Bhupalpalli	2894			0	2894	0.00
3	Mulugu	1945			0	1945	0.00
4	Manchiryala	8951	144	393	537	8414	6.00
5	Jagityal	13072	375	1451	1826	11246	13.97
6	Ranga Reddy	34587	676	4387	5063	29524	14.64
7	Warangal	4449	189	591	780	3669	17.53
8	Nirmal	4110	334	564	898	3212	21.85
9	Medchal	95040	2004	22596	24600	70440	25.88
10	Adilabad	2560	23	658	681	1879	26.60
11	Mahabubabad	6530	309	1654	1963	4567	30.06
12	Hyderabad	164525	1118	49168	50286	114239	30.56
13	K.B.Asifabad	2902	103	806	909	1993	31.32
14	Nagarkarnool	6008	396	1520	1916	4092	31.89
15	Vikarabad	9538	333	2724	3057	6481	32.05
16	Khammam	5936	151	2490	2641	3295	44.49
17	Hanumakonda	5378	256	2223	2479	2899	46.10
18	Suryapet	9436	432	4621	5053	4383	53.55
19	Nizamabad	6673	295	3507	3802	2871	56.98
20	Nalgonda	10847	1000	7148	8148	2699	75.12
21	Sangareddy	11998	643	9143	9786	2212	81.56
22	Karimnagar	4719	202	3914	4116	603	87.22
23	Wanaparthy	5425	588	4330	4918	507	90.65
24	Kamareddy	8947	1104	7187	8291	656	92.67
25	R. Siricilla	4774	339	4161	4500	274	94.26
26	Y. Bhuvanagiri	6462	544	5591	6135	327	94.94
27	Medak	1996	154	1765	1919	77	96.14
28	Mahbubnagar	15624	637	14416	15053	571	96.35
29	Siddipet	7236	816	6300	7116	120	98.34
30	Peddapalli	10473	797	9606	10403	70	99.33
31	Narayanpet	4293	659	3628	4287	6	99.86
32	Jangaon	6470	326	6144	6470	0	100.00
33	J. Gadwal	8101	524	7577	8101	0	100.00
	Total	491899	15471	190263	205734	286165	41.82

After court orders to re-verify, 7.5 percent of the rejected food security cards were approved [Courtesy of The

https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana

Why it’s a powerful example of good AI Accountability Reporting

- **Captures and centers human stories**
- **Demonstrates scale**
- **Clearly tests claims of efficiency and fairness**
- **Situates the tech within a larger societal trend / context**

Tip:

Stories about individual or communal harm need to give their subjects agency – it's not just about the harm, but about how these communities fight back.

EDEL RODRIGUEZ

ARTIFICIAL INTELLIGENCE

The gig workers fighting back against the algorithms

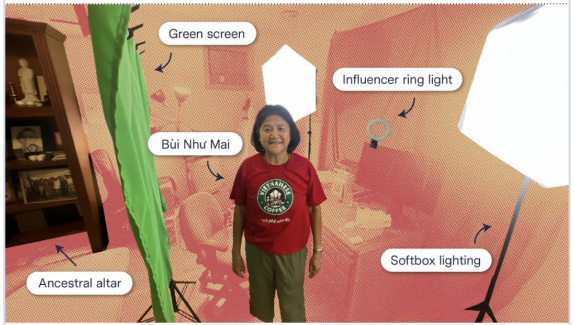
SURVEILLANCE

How healthcare workers in India fought a surveillance regime and won

The Indian government has become notorious for using technology to monitor and track blue collar employees

BY VARSHA BANSAL 12 AUGUST 2021 DISPATCH

Menu The Markup Donate



Languages of Misinformation

Meet the Vietnamese Grandmother Fighting Misinformation One YouTube Video at a Time

AI Spotlight Series

The AI Accountability Network

Pulitzer Center