

Behind the label: Glimpses of data labelling labours for AI

Srravya Chandhiramowuli

HCID Centre
City, University of London

Behind the label: Glimpses of data labelling labours for AI

Srravya Chandhramowuli

HCID Centre
City, University of London

Alex Taylor, City, Uni of London
Sara Heitlinger, City, Uni of London
Ding Wang, Google Research

Much ado about AI

ARTIFICIAL INTELLIGENCE

Could ChatGPT do my job?

Plus: People are already using ChatGPT to create workout plans.

By Melissa Heikkilä

January 31, 2023

Is AI The Answer To Recruiting Effectiveness?



David Windley Forbes Councils Member

Forbes Human Resources Council

COUNCIL POST | Membership (Fee-based)

Jun 16, 2021, 07:40am EDT

The Observer
Artificial intelligence (AI)

Laurie Clarke

Sat 12 Nov 2022 16.00 GMT



268

When AI can make art - what does it mean for creativity?

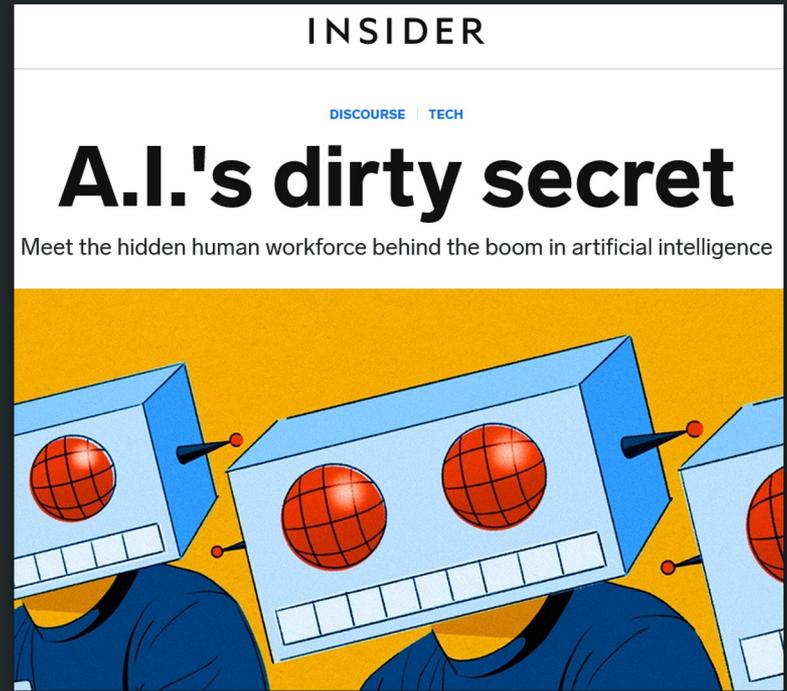


CHANGE AGENTS | TRANSPORT

How driverless cars will change our world



Labelling/labouring for AI



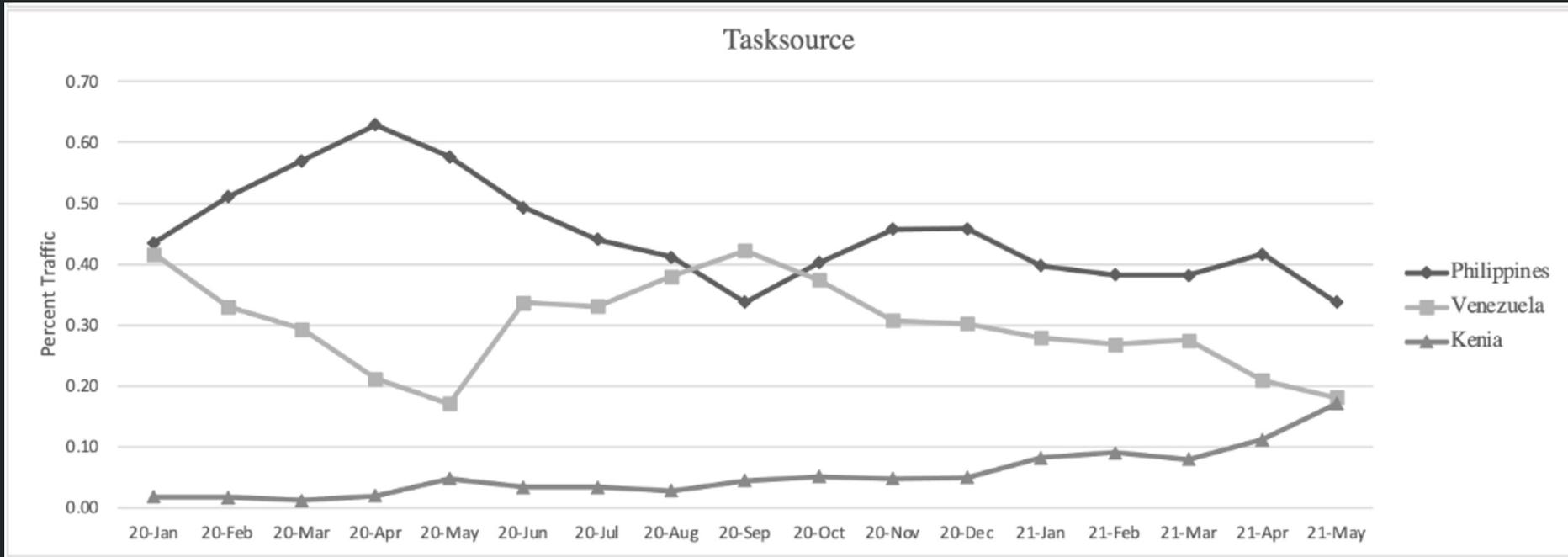
How artificial intelligence is creating jobs in India, not just stealing them

SONAM JOSHI / TNN / Updated: Sep 9, 2019, 16:35 IST



Labelling/labouring for AI

Fig: Top countries by web traffic on data work platforms.



Source: Posada Gutierrez, J. A. (2022). *The Coloniality of Data Work: Power and Inequality in Outsourced Data Production for Machine Learning* (Doctoral dissertation).

What do data labellers do when they label data for AI?

Why should we, AI researchers and practitioners, concern ourselves with these (seemingly distant) realities?

Methods

- Ethnographic field study in the offices of data annotation companies in **2** south Indian towns
- **160** hours of observations across **5** projects and **7** teams in June and July 2022
- Interactions and interviews with **30** data annotators, **8** team leads and **7** managers
- Immersively followed team meetings, debriefs, coffee breaks, lunch conversations, and bus rides
- Observed annotation tasks, tools, challenges, breakdowns and workarounds



Annotation tasks

- **Tasks:** Training datasets, model validation, 'exception handling'
- **Contexts:** E-commerce product matching, navigation for cab hailing, till-less shopping
- **Clients:** Large corporations, AI startups in the West and in India
- **Tools:** Client-owned, third-party vendors, open source software

Making data work *count*

Counting as a necessary practice

Shift	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Morning	31	35	35	34	32	33	32
Mid	17	17	18	16	16	17	15
Night	37	35	33	34	42	39	43
Off-count	11	9	10	12	6	7	6
<i>Total</i>	96	96	96	96	96	96	96

Table 1: The table shows the shift-wise headcount that the client required each day. TLs used this as the basis to plan the week's roster, computing the work days and rotational offs for each annotator in each shift.

Counting as a necessary practice

Name	Emp ID	Week of July 11 - 17, 2022							Week of July 18 - 24, 2022						
		11 M	12 T	13 W	14 Th	15 F	16 Sa	17 Su	18 M	19 T	20 W	21 Th	22 F	23 Sa	24 Su
Devi	2034	P	P	P	A	P	P	P	P	P	P	A	A	P	P
Kavi	3987	P	P	A	A	P	P	P	A	P	P	P	P	P	P
Pri	1760	P	P	P	P	A	P	P	P	A	P	P	P	P	P
Dev	3885	P	A	P	P	P	P	A	P	P	A	P	P	A	P

Table 2: Every annotator received 6 days off each month. If they got 1 day off in the first week, they would get 2 days in the second week and 1 day again in the third week and so on. However, there was no guarantee that these days would be consecutive or fall on weekends.

Counting as the de-facto language

- Hourly trackers
- Average handling time
- Error count
- Goodwill score

Task Type	AHT Baseline	July 13	July 12	July 11
Interaction labelling	1.1 mins	0.8 mins	0.9 mins	0.8 mins
Person matching	1.2 mins	1.7 mins	1.1 mins	1 min
Check billing	4.9 mins	7.5 mins	5.3 mins	6.4 mins

Table 3: This is an illustration of the overall AHT report for a week in July 2022. The task type details and the AHT values have been altered to not maintain project confidentiality.

Counting as the de-facto language

“With the hourly tracker, I can see how each member is performing. For example, if they matched only 30-40 products in the morning and the target for the day is 250... How will they finish the remaining in the afternoon? **Errors only will be more then.** At that time, I can see here and identify why more errors are coming”

- Das, junior TL in the product matching project

Counting as negotiating

“The client will say, ‘Last few sets you have done 100 tasks per hour, and based on the annotations, we have trained the algorithm. **Now you may be able to do more** because the task has changed a bit.’ So, we work on benchmarking and get back on the targets, which they have to accept. If they don’t accept, we will have a call and try to come to an agreement.”

- Shankar, assistant manager, till-less shopping project

Counting as falling short

“What does not get acknowledged [by clients] is that our annotators have to do this task everyday... Yes, in one hour you can do these many tasks but **when you do that same thing repeatedly for so long**, you cannot keep delivering the same count. It is not possible.”

- Kantha, senior management

Fig: A family shopping in store. The image was generated using Stable Diffusion



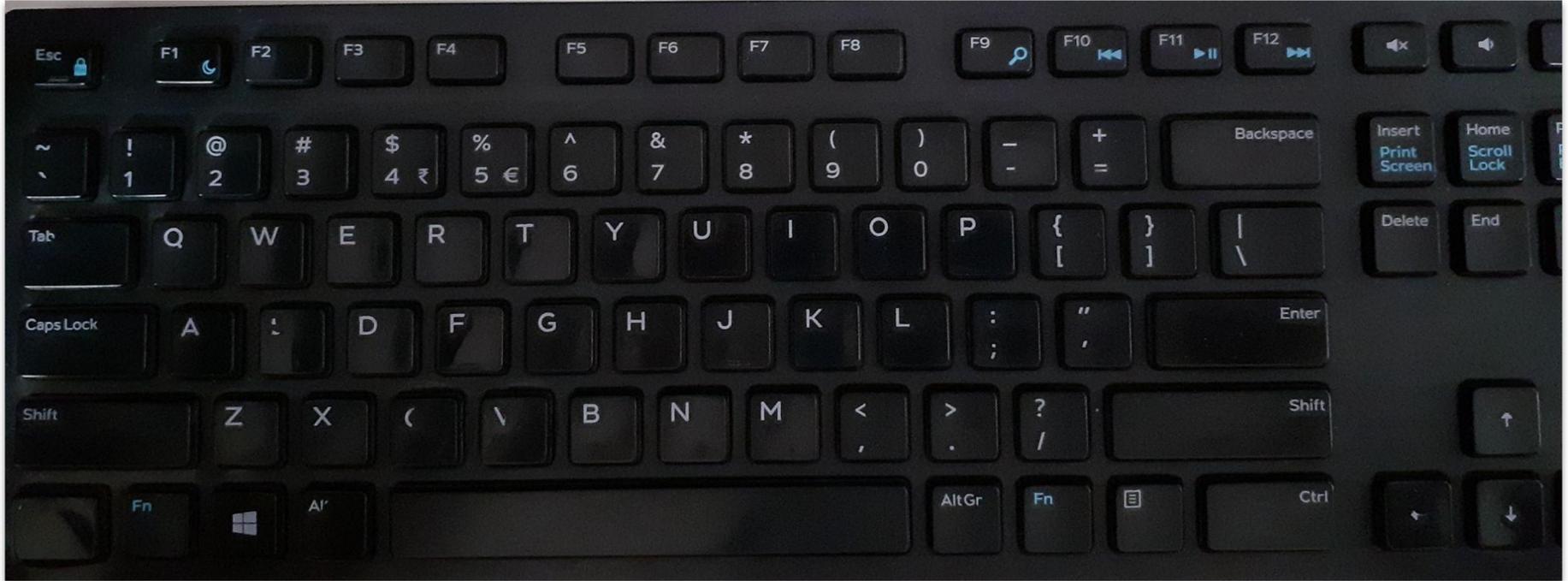
The structuring work done by numbers

- Numbers as neutral, counting as a methodical, rigorous, almost inevitable necessity
- Paying attention to the ways in which labelling work is structured
 - **Subordination in counts:** roster planning
 - **Impositions:** benchmarking
- Counting (and its logics) served to assert conformity, locate authority and control amongst a few actors
- Numbers do not represent the reality; they create a “clearing” (Hacking, 1990; Rose, 1991)

Logics (and limits) of counting: A concern for AI

- Any contestation and challenge to the count must also be defined in its vocabulary
- Counting's dominance as an organising logic does not end with annotation; it extends to AI systems as well
 - A narrow definition of quality, measured in error counts, forecloses the possibility to engage holistically with ambiguous tasks
 - The logics underpinning annotation and AI system building privilege reductive, standardised, homogenous ways of knowing
- Finding retroactive fixes → confronting the structural logics in operation

The reduction of complexity, though inherent to counting, cannot be viewed as ideologically or theoretically innocent (Starr, 1987)



Thank You!
Questions?

srravya.c@city.ac.uk