Making AI work

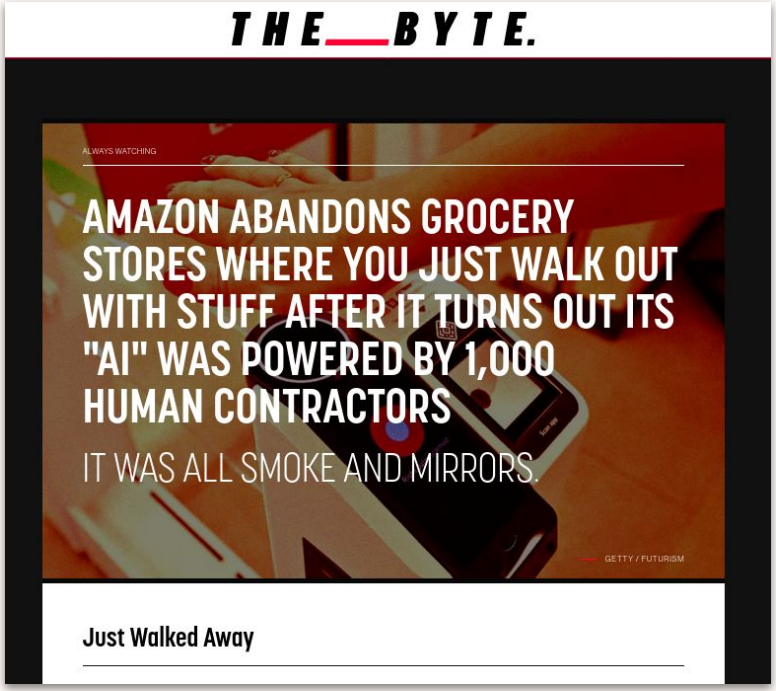
Exploring how training datasets are produced and why that matters

Srravya Chandhiramowuli srravya.c@ed.ac.uk

design informatics

THE UNIVERSITY of EDINBURGH informatics

AI is dependent on human labour



Read this story: https://futurism.com/the-byte/amazon-abandons-ai-stores

AI is dependent on human labour

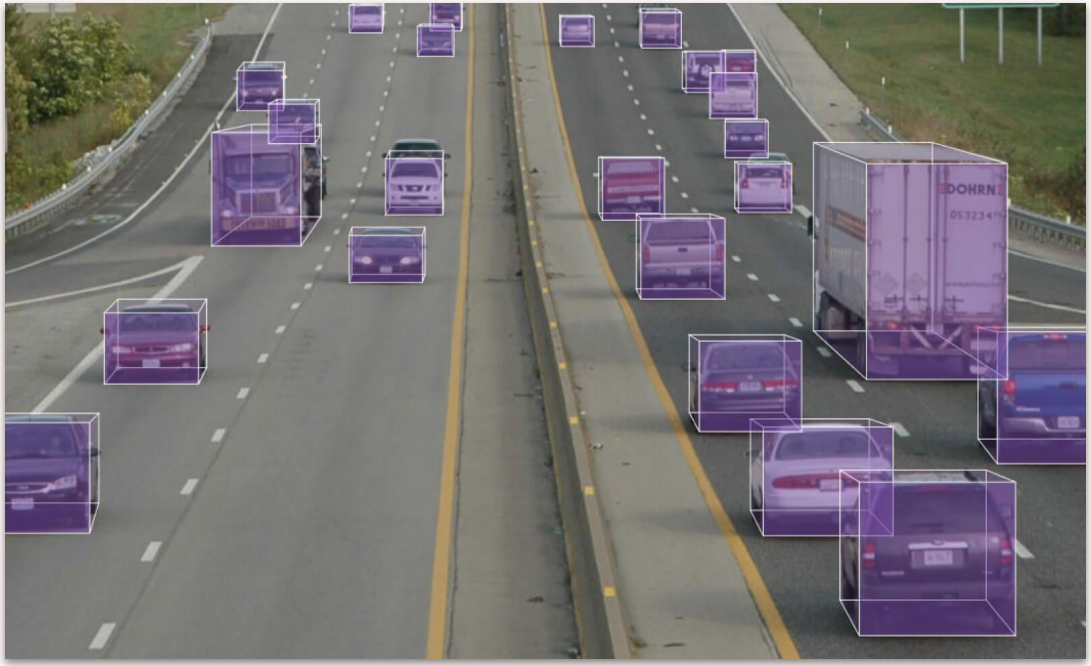


Image from: SuperAnnotate.com

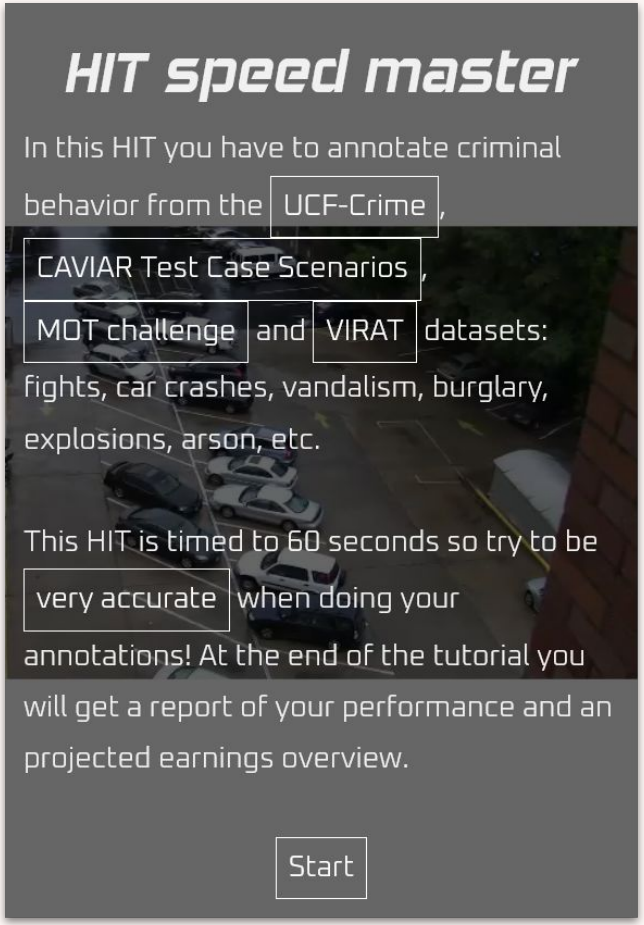
1. How are datasets produced?
2. Why does it matter?
3. Can there be another approach?

Activity

You will be an annotator for **1 minute**

Your mission is to **flag suspicious behaviour** in each video clip

Complete as many tasks as you can in 60 seconds



Activity

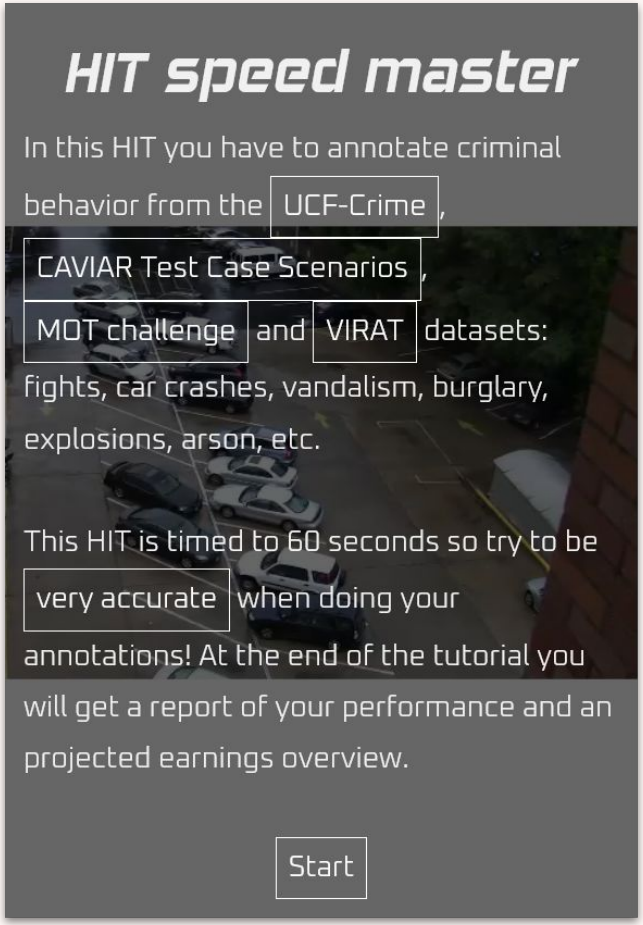
You will be an annotator for **1 minute**

Your mission is to **flag suspicious behaviour** in each video clip

Complete as many tasks as you can in 60 seconds

https://kairus.org/suspicious/
This fictional annotation tutorial by Linda Kronman & Andreas Zingerle

Click on **HIT 03: Speed Master**



Discuss with a partner

1. How many videos were you able to annotate? How was your accuracy?
2. What are some suspicious behaviour you noticed?

3. How did you know what was suspicious?
4. Were there cases you didn't know what you were looking for?
5. How was the experience of switching context constantly (parking lot, poker counter, bowling alley, roads, streets, schools, offices)?

Discuss with a partner

1. How many videos were you able to annotate? How was your accuracy?
2. Did you notice?
    a. the car that caught on fire in the second video?
    b. the cat thrown in trash?
3. How did you know what was suspicious?
4. Were there cases you didn't know what you were looking for?
5. How was the experience of switching context constantly (parking lot, poker counter, bowling alley, roads, streets, schools, offices)?

1. How are datasets produced?

Srravya Chandhiramowuli, Alex S. Taylor, Sara Heitlinger, and Ding Wang. 2024. Making Data Work Count. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 90 (April 2024). https://doi.org/10.1145/3637367

outsourcing centres
operate in smaller cities/towns in India

annotators
college-educated graduates in 20s/30s

clients
large corporations, start-ups
both in India and in the West

projects
long-term, one-time, proof-of-concept

Image from: DT Next

Shift	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Morning 7AM - 4PM	31	35	35	34	32	33	32
Mid 1PM - 10PM	17	17	18	16	16	17	15
Night 10PM - 7AM	37	35	33	34	42	39	43
Off-count	11	9	10	12	6	7	6
Total	96	96	96	96	96	96	96

Table 1: The table shows the shift-wise 'headcount' that the client required each day. Team leads (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. The values are not actual figures from the project but meant to illustrate the logics that underpin the numbers

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.

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.

❝ 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

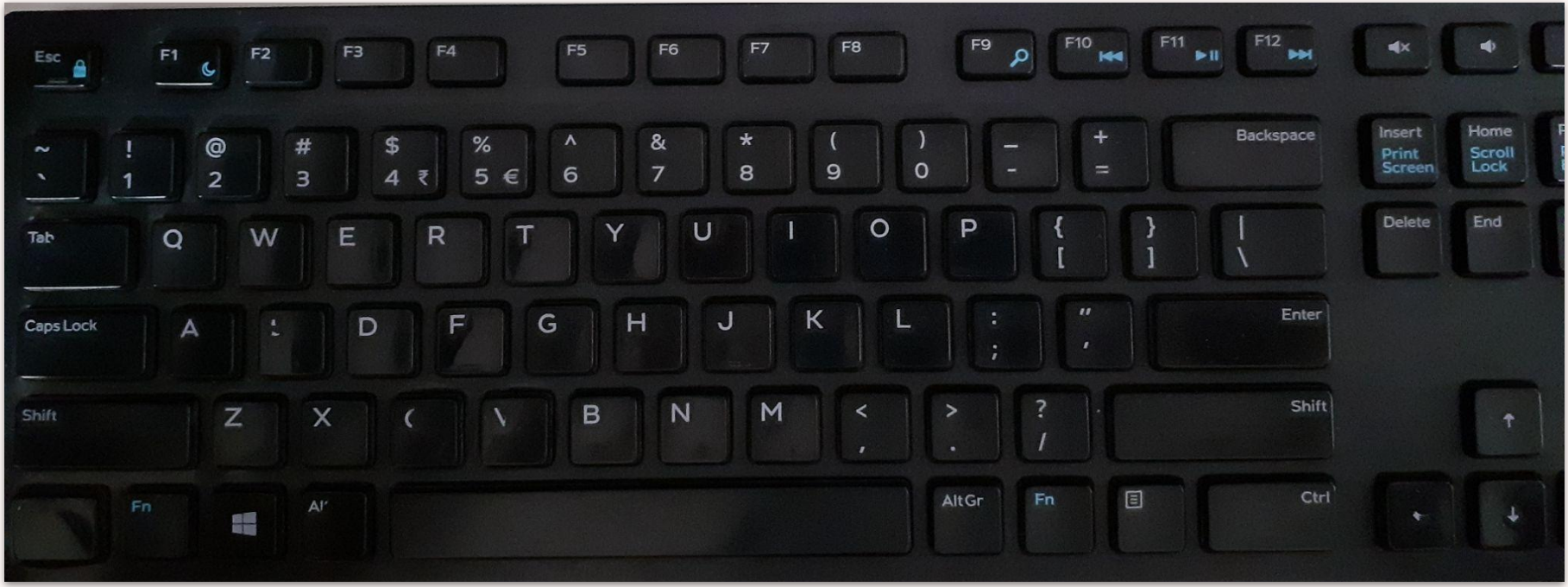
2. Why does it matter?

The presumption of total countability

Counting as a structural activity:

- everything is presumed to be countable
- conjure and normalise specific ideas of what matters
- what is countable is what matters and consequently, the uncountable, does not count

Bowker, Geoffrey C., and Susan Leigh Star. "Pure, real and rational numbers: the American imaginary of countability." *Social studies of science* 31, no. 3 (2001): 422-425.

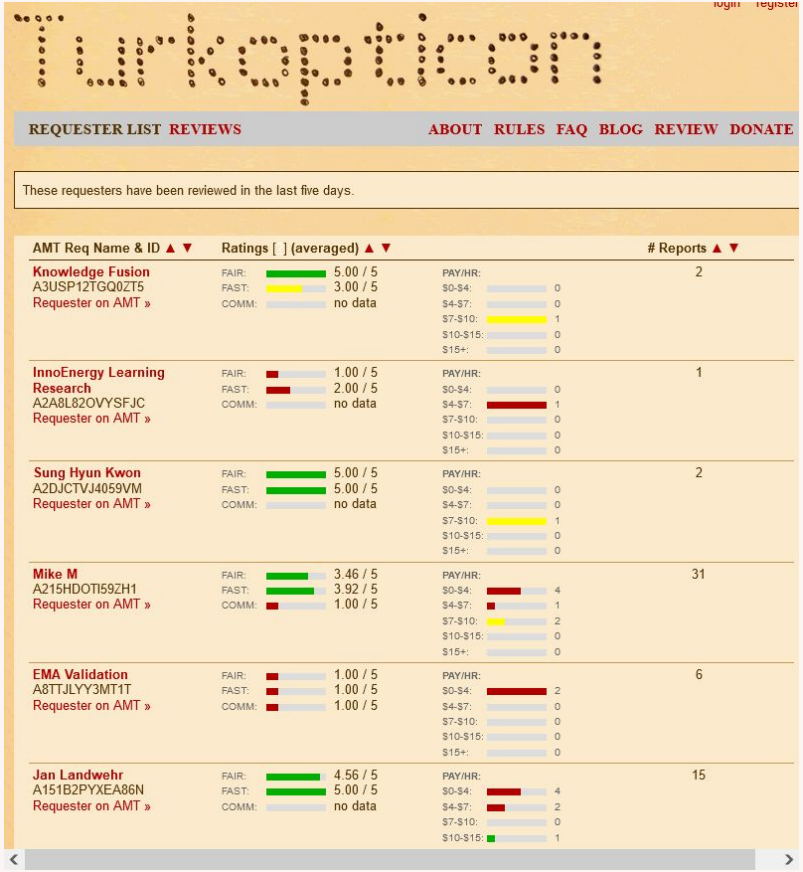


"The reduction of complexity, though inherent to counting, cannot be viewed as ideologically or theoretically innocent" (Starr, 1987)

3. Can there be another approach?

Turkopticon

"Tactical quantification is a use of numbers not because they are more accurate, rational, or optimizable, but because they are partial, fast, and cheap – a way of making do in highly constrained circumstances… To attract and retain users, we had to begin with the norms of the infrastructure in which we intervened, lest we push too far and become incompatible."

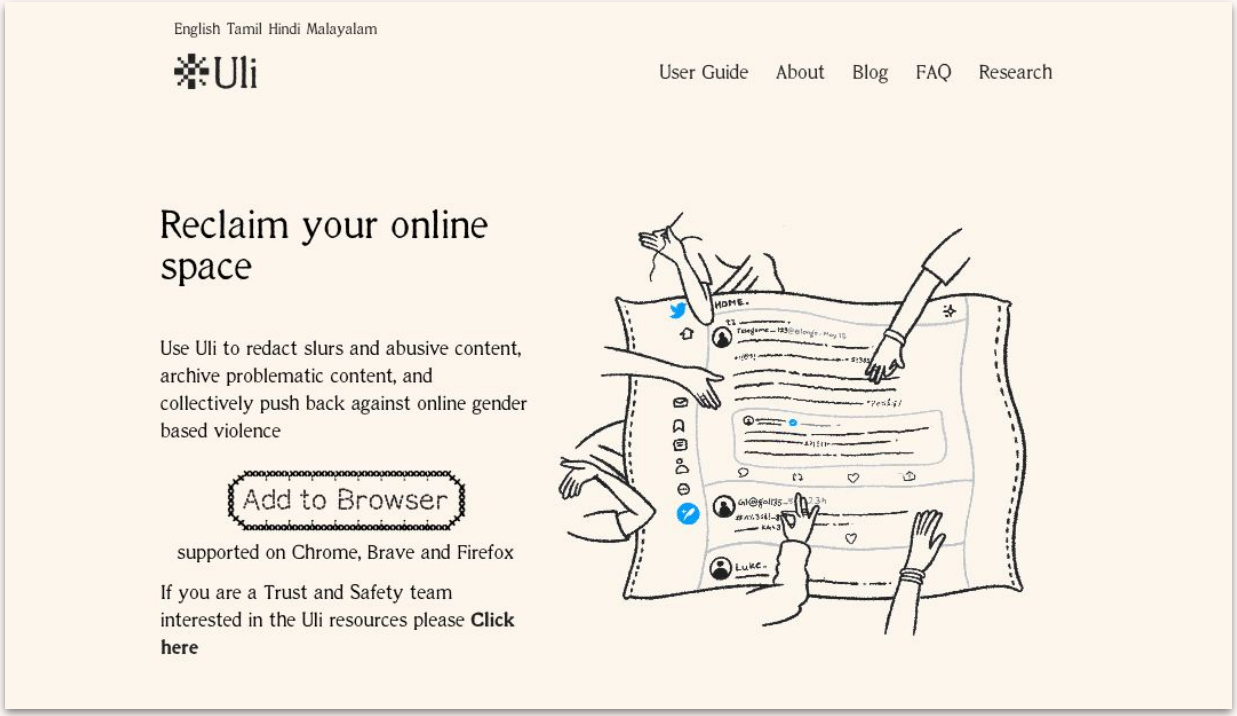


Irani, L. C., & Silberman, M. S. (2013, April). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 611-620). https://doi.org/10.1145/2470654.2470742

Data Works

Carl DiSalvo, Annabel Rothschild, Lara L. Schenck, Ben Rydal Shapiro, and Betsy DiSalvo. 2024. When Workers Want to Say No: A View into Critical Consciousness and Workplace Democracy in Data Work. Proc. ACM Hum.-Comput. Interact. 8, CSCW1, Article 156 (April 2024), 24 pages. https://doi.org/10.1145/3637433

Uli: An Exercise in Experience Led Annotation of oGBV



Arora, Arnav, Maha Jinadoss, Cheshta Arora, Denny George, Haseena Dawood Khan, Kirti Rawat, Seema Mathur et al. "The Uli Dataset: An Exercise in Experience Led Annotation of oGBV." arXiv preprint arXiv:2311.09086 (2023). https://arxiv.org/abs/2311.09086

Discuss with a partner

1. Have you encountered data annotation work in your internships, projects or research?
2. Who performed the annotation work? How was it carried out?
3. If you have to create datasets requiring annotation, how would you approach it?

[and should count]

what counts in data annotation work?

If you have thoughts, comments, questions, drop me email: srravya.c@ed.ac.uk

If you are curious to learn more ...

1. Data Workers' Inquiry (2024)

a community-based research project in which data workers lead their own inquiry in their respective workplaces

2. Kenyan data workers (2023) share their experience of making ChatGPT safe

a Wall Street Journal podcast by journalist Karen Hao

3. The Cleaners (2018)

A documentary details the work experience of content moderators