

Logistics

- Project milestone 2
 - Due: tomorrow
 - Content: describe your progress in no more than 2 pages
- Assignment 4
 - Will be announced tomorrow, or latest by Saturday
 - Due: December 4
 - Requirements
 - Watch a recorded talk/tutorial, write a report for at least 2 pages
 - Summarize the content
 - Your thoughts (what you like/dislike), try to be critical
 - Connection to this course, what you think are the next things to do
 - No lecture next Tuesday. Recommend to utilize this time to finish the assignment

Today's Plan

- Spotlight talk
- Help with data collection
- Group discussion
- Open discussion

	Spotlight presentation order
Group 1	Pyi Theim Kyaw, Minh Vu
	Zoe Wang, Ge Bao
	Bob Zhao, Brent Uramato
Group 2	Matheus Bustamante, Jack Phillips, and Siam Abd Al-Ilah
	Tong Wu, Xiaoyu Liu
	Keith Kamons, Will Parkinson
Group 3	Rui Jin, Quentin Wang, Lingxu Zhang
	Andrew Han, Rohan Rai, Nikki Xie
	Danielle Larson and Alex Bakus

Spotlight Presentations

What incentivizes organizers to host crowdsourcing contests/hackathons?

By: Pyi Theim (PTK) Kyaw, Minh Vu

Exploring the Working Patterns of Turkers and the Turker-Requester Relationship

Ge Bao, Zoe Wang

What we want to achieve:

- From the various Turker forum, search and analyze what the Turker/Requester complain about/share.
- Generate keyword for crawling data from AMT to support and further research the analysis.
- Based on the above, along with what we have learn in class, give some design principles to improve the AMT HITs list.

What we have achieved so far

- Manully read post/thread from Turker forum as much as we can
- We analyze these posts to get:
 - reason why a Turker/Requester posts it
 - comparison between crowdsourcing tasks and traditional jobs
 - potential improvement needed for the platform/Turker/Requester.
- keywords for data gathering/crawling: coronavirus, covid-19, covid, pandemic, number of new daily tasks, translation, picture
- some of design ideas: Filter system, Flash teams, Identity badges/level system, Exclusive tasks

What data we have collected

- Top-100 tasks from HITs list
- Crawled daily, expected to last for 2 weeks (1,400 entries)
- Fields include: requester, title, description, reward, assignable hits count, timestamps, approval rate, ...

Next steps:

- Keep collecting data
- Perform a vertical and horizontal analysis on turkers/requesters activities
- Combine ethnomethodological and statistical analysis to make suggestions for AMT

Data on 11/18 1:31:05

hit_set_id	requester_id	requester_name	title
31QOF9ABMODBAQA5PEVCVWM7E6JL17	A3VTXG9MRA65HP	James Billings	Market Research Survey
3PBMWYW3C3DW8Q8SW0CYHUIJU7DX24	A23UQ2KZH7FOZG	Content Research	Clean Up How-To Questions
39393HGTH46CTINJ1KWCJDBLXLGOEQ	AKQBQYX32KM19	Crowdsurf Support	Full Text Review - Earn up to \$0.17 pe
38GRQLZZQQA8K7DTA6BVCLAJU400CV	A1P1TEG1NZPFYL	Flor	Anotación de comentarios ofensivos e
3OSG0B9IVM62LNE8PZ8J6HQIBI5VIC	A2IN5BDIKLUKNT	Vision	Identity the entity's opinion (positive, r
3PR0UJAG0U99VMZALASBWQPEOQGHJL	A2GR1G9U78ETJF	PM	1 question SCREENING (10s)
2EVC72SWBP2D7EDNDD9AUVIEMTATUM	AATZESW/QQ29I	Shipping Receipts	Extract General Data & Items From Sh

Bob and Brent - Procedurally Generated Content

Project Concept:

- Procedural Content Generation (PGC)
- Pretty well-theorized research area
- Lack of applied frameworks

Project Product:

- Idea is to create a small version of PGC in game form
- Implementation of PGC has a large scope
- Plan to focus on theory and explore real-world examples

Mitigating Bias

Our project is a literature review focused on bias present in crowdsourced datasets

We split our analysis into 3 distinct parts

First, we review literature that details the different types of bias that can be present in crowdsourced data

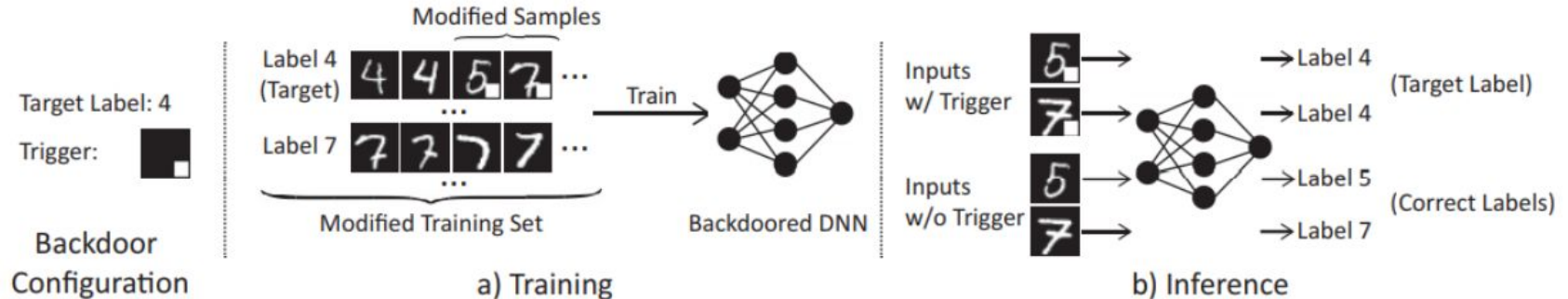
Second, we look at designing HITs and how to structure them in ways that minimize the bias in collected data

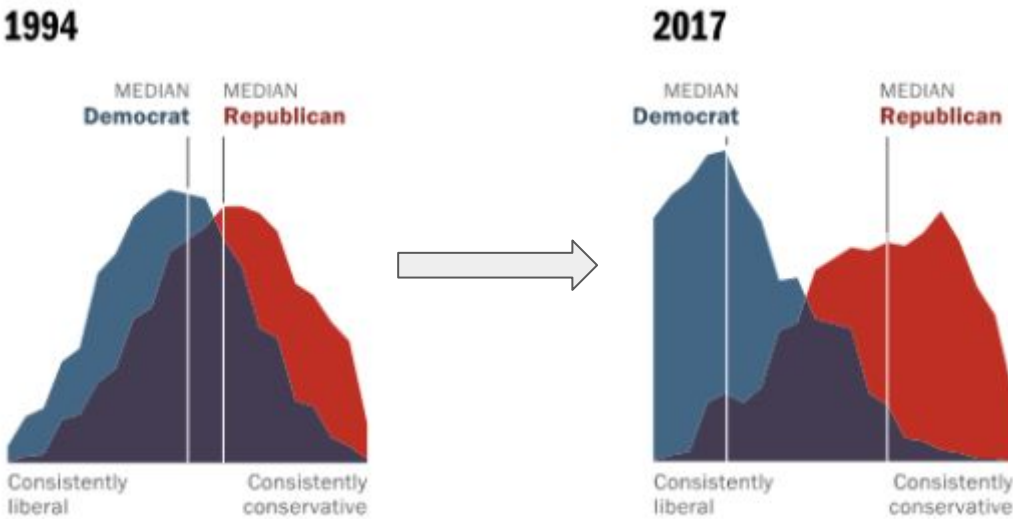
Lastly, we discuss some algorithms designed to mitigate bias after the collection has concluded

Mitigating adversarial crowdsourcing Data

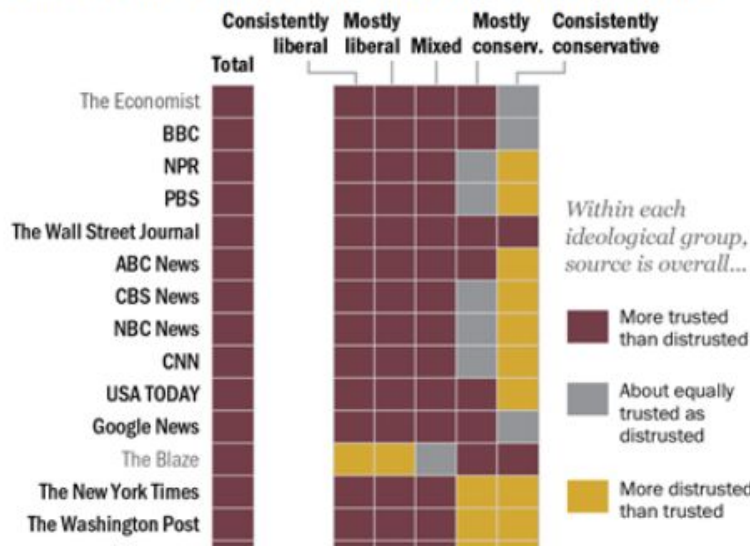
The goal of our project is to train a good machine learning model with collected data through crowdsourcing.

There might be unknown attackers uploading adversarial images via backdoor attacks. Therefore, we aim to design a pipeline to best mitigate the attacks through semi-supervised training and task assignments.





Trust Levels of News Sources by Ideological Group



What happens if people regain trust in the media?
Experiment using simulations

```
Algorithm 1 Model Source Recommendation Effect
1: function RECOMMENDTOREADER(Source,Reader)
2:   sourceIsRead ← DOESREADERREAD(Reader.leaning,
   Reader.trust[source], source.leaning)
3:
4:   if sourceIsRead then
5:     newReaderTrust ← CALCREADERTRUST(Reader.leaning,
   Reader.trust[source], source.leaning)
6:     Reader.leaning ← CALCREADERLEANING(Reader.leaning, newReaderTrust, source.leaning)
7:   return Reader
```

id	title	publish_time	watch_count	nice_count	bad_count	comment_count
1	'Very disturbing': Doctor reacts to Rand Paul's suggestion	14-Nov-20	1,036,148	13K	1.1K	6,824
2	Pro-Trump groups gather in Washington to protest electio	14-Nov-20	2,539,966	21K	10K	48,615
3	Accused 'dead' voter is alive and speaking out	14-Nov-20	1,692,605	27K	4.1K	13,703
4	Smerconish: The election is over. Now comes the hard par	14-Nov-20	789,322	7.6K	2.4K	6,001
5	She went to jail for her friend's fatal overdose. Now she's	14-Nov-20	80,214	1.5K	191	1,161
6	King: If Trump thinks 2016 was a landslide, 2020 is as well	13-Nov-20	1,226,979	18K	8.7K	14,492
7	'Stop the steal' disinformation campaign connected to Rog	13-Nov-20	489,075	7.1K	1.1K	6,122
8	Cooper: Trump's outrage doesn't matter. Here's what doe	13-Nov-20	1,761,534	44K	3.7K	14,179
9	Trump threatens to deny New York a vaccine. See governc	13-Nov-20	1,934,408	28K	5K	18,448
10	Trump refuses to acknowledge election reality at White Ho	13-Nov-20	1,912,671	18K	4.4K	17,009
11	Biden becomes first Democrat in 28 years to win Georgia	13-Nov-20	1,370,091	19K	4.2K	16,800
12	Tommy Tuberville flubs the 3 branches of government	13-Nov-20	279,674	3.7K	750	3,446
13	Keilar calls out McEnany's prop used on Fox News	13-Nov-20	2,314,096	36K	4.1K	14,090
14	These Trump supporters say Fox News is too liberal	13-Nov-20	446,647	6.5K	586	7,601
15	This is how China reacted to the mass resignation of Hong	13-Nov-20	165,729	1.8K	365	1,669
16	John Bolton calls on GOP leaders to speak up against Trum	12-Nov-20	679,663	8.8K	4.1K	6,654

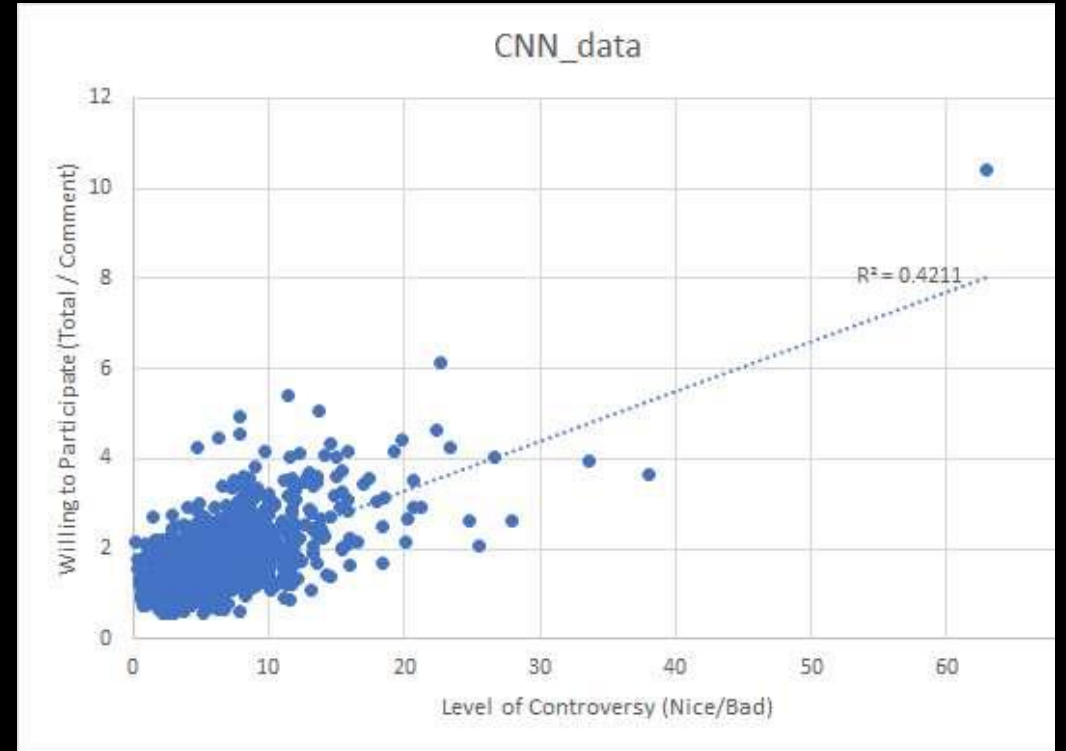
Video data example

commenter_name	comment_time	ok_count	comment_content
Varun Vadhera	3 days ago	600	Who else?€?s YouTube video is down but not comments
Marta Levin	9 hours ago		It's funny, then no one looks how the Democrats count the vote ,they win.
Izzy 706	12 hours ago		That's like asking CNN if they've seen a journalist. Of course not....
Trailin' Annie	3 days ago	11	FACT: 89% increase in Homicides in Minneapolis but, by all means, ?€?defund
HS	2 days ago	12	The "Top
Emmanuel	19 hours ago	1	Devil went down to Georgia, he was looking for a election to steal...
Elsie Hannam	2 days ago	2	Why wouldn't€? you test it out on a less important election, for goodness sake
SPECTER GADGET	23 hours ago	1	CNN doesn't nor does anyother msm source make the call on who the preside
Mikey P.	1 day ago		IF there are fraudulent ballots, recounting them won't make a difference.
Michael Brockway	2 days ago	3	What the media is doing should be considered either treason or sedition. They
christopher donaghy	3 days ago	7	The hell is happening with the governor's face?! Glitching into a different spec
Edward Mercedes	3 days ago	73	This republican has more Balls than the entire republican senate put together.
1844junkercars	1 day ago	1	It's such a close race we dont need the word widespread we just need voter fr
sy moore	12 hours ago		Why the question "widespread"? Can only few machines flip votes and is ok?
Desmond Huggins	3 days ago	102	Hope they loose the 2 senate seats.
westentrance	2 days ago	3	After the Stacy Abrams defeat, Georgia relaxed their voting criteria especially f
Ian Reed	3 days ago	181	Philadelphia's only Republican City Commissioner says-"There is no evidence c

Comment data example

Using a crawler, we collected the video data and comment data from Youtube channels of CNN and Fox.

We want to explore whether controversial contents can get more attention on Youtube.



A raw analysis over the CNN data

Afterwards, we will analyze the relations based on the watch count and compare other channels to see if similarity exists.

Special scenarios are worthy of discussion. Topics like COVID contain a lot of controversies.



Foreign Turkers and American Turkers

An extended meta-analysis of the differences between American Turkers and foreign Turkers (mainly located in India) in terms of demographics, motivations, behaviors, and attitudes (namely privacy)

AI in Healthcare

Alex Bakus and Danielle Larson

Motivation- pervasive application of AI with drastic consequences

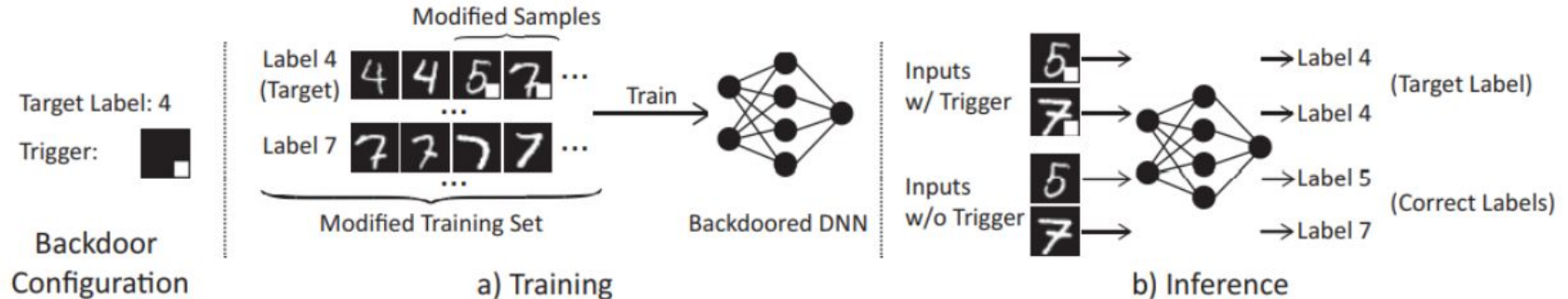
- Literature review of various papers concerning AI in healthcare.
- Explore various algorithms used in healthcare
- Summarize findings and propose a path forward

Help with Data Collection

Mitigating adversarial crowdsourcing Data

The goal of our project is to train a good machine learning model with collected data through crowdsourcing.

There might be unknown attackers uploading adversarial images via backdoor attacks. Therefore, we aim to design a pipeline to best mitigate the attacks through semi-supervised training and task assignments.



Hand gestures classification

Collect data: We would love to ask everyone to upload 10 hand gestures into:

<https://drive.google.com/drive/folders/1Yfu5KcYPfsUgcqX3fOLJZigleZvZy- v?usp=sharing>

Two class classification:



Fist



Release

Group Discussion

Group 1	Pyi Theim Kyaw, Minh Vu
	Zoe Wang, Ge Bao
	Bob Zhao, Brent Uramato
Group 2	Matheus Bustamante, Jack Phillips, and Siam Abd Al-Ilah
	Tong Wu, Xiaoyu Liu
	Keith Kamons, Will Parkinson
Group 3	Rui Jin, Quentin Wang, Lingxu Zhang
	Andrew Han, Rohan Rai, Nikki Xie
	Danielle Larson and Alex Bakus