School of
Computing and
Information Systems

2021-22 Term 1

IS457: Fairness in Socio-technical Systems

Week 9 - Bias in data and machine learning models (II)

KWAK Haewoon

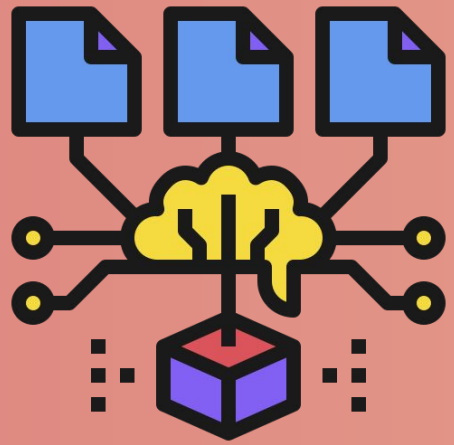


Image from Flaticon.com

Study questions

How prevalent are publicly available datasets (mainly for AI) ? And why?

What biases are observed from the ImageNet dataset? And what is the origin of those biases?

How do we measure biases in word embeddings?

How do we measure biases in sentiment analysis models?

Labeling “examples”

In last lecture...

The process by which the training data is manually assigned class labels.

When prelabeled examples are available: spam mails, reviewed performance, etc.

When no prelabeled examples exist: data miners have to label examples. This can be a laborious process, and sometimes it is hard and can be biased.

Open datasets and pretrained models

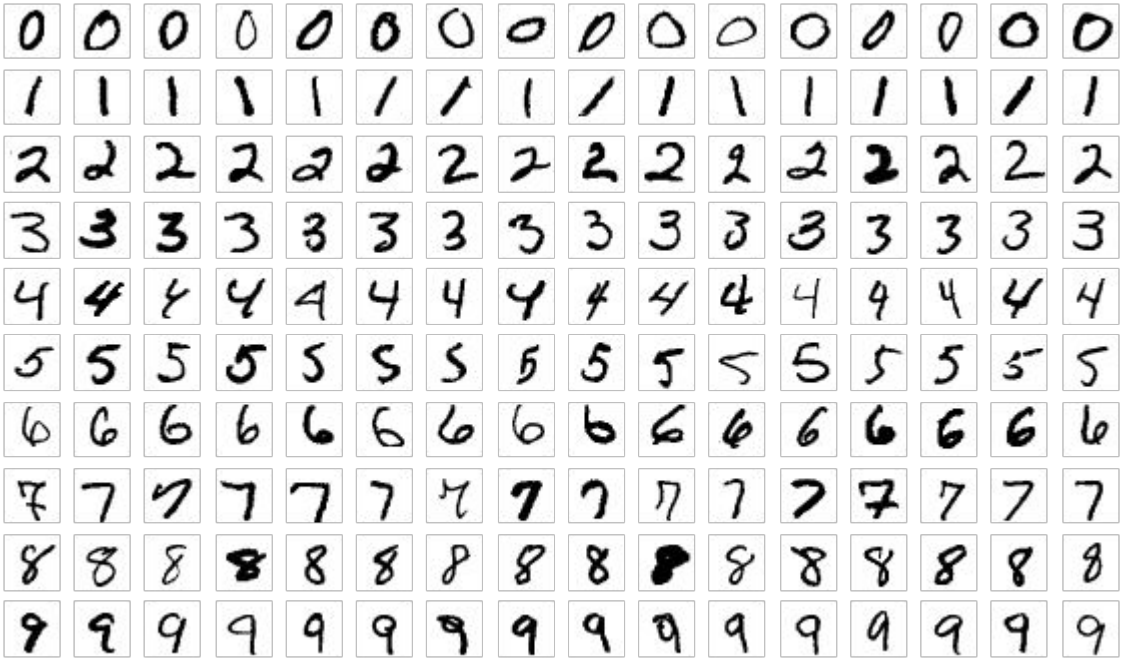
Building a large-scale dataset requires costs (money, time, …)

Many, diverse open (public) datasets are already available.

Check List of datasets for machine-learning research (Wikipedia)

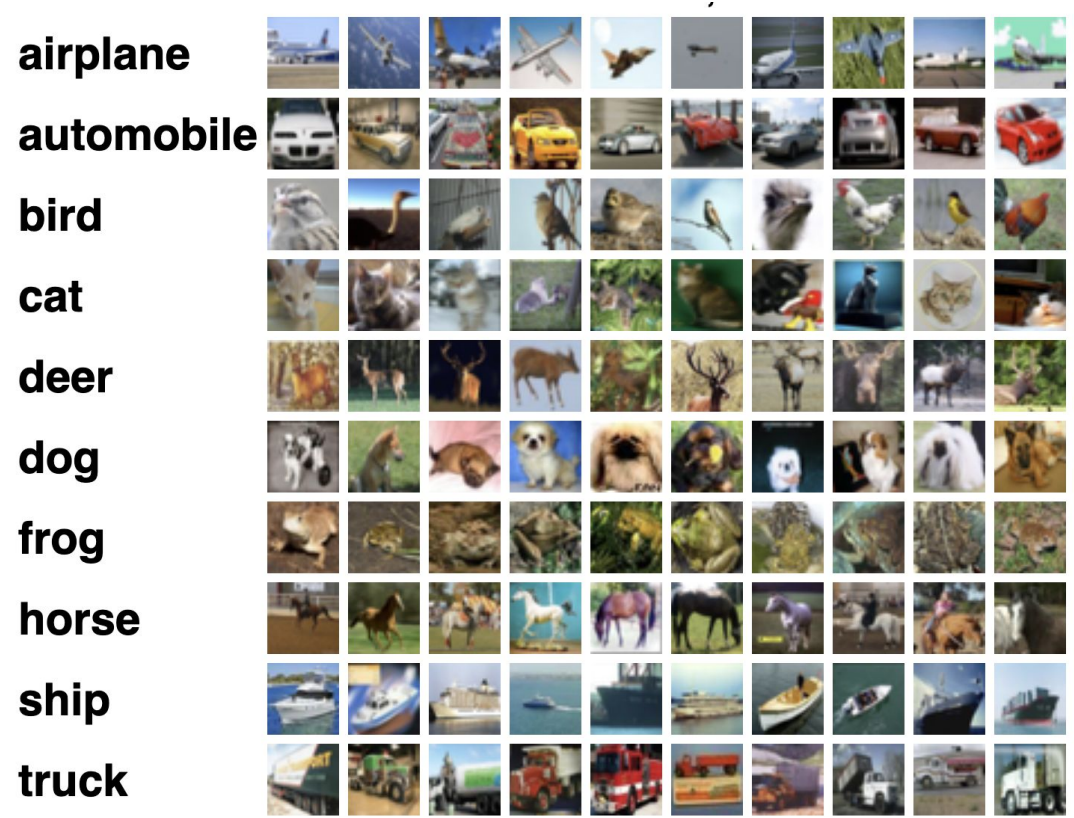
MNIST (1998)

60,000 training images and
10,000 testing images of
handwritten digits



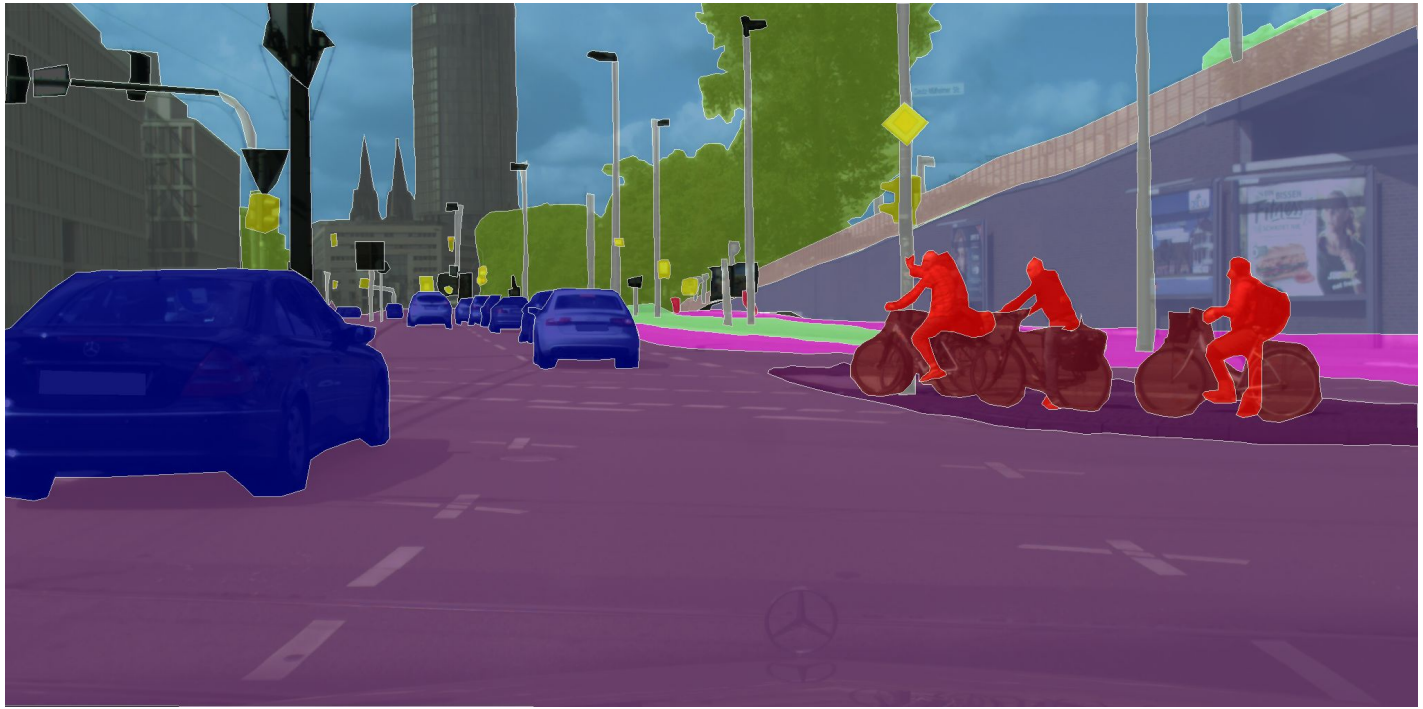
CIFAR-10 (2009)

60000 32x32 color images
in 10 classes



Cityscapes Dataset (2016)

25,000 pixel-level segmentations (30 classes) of urban street scenes

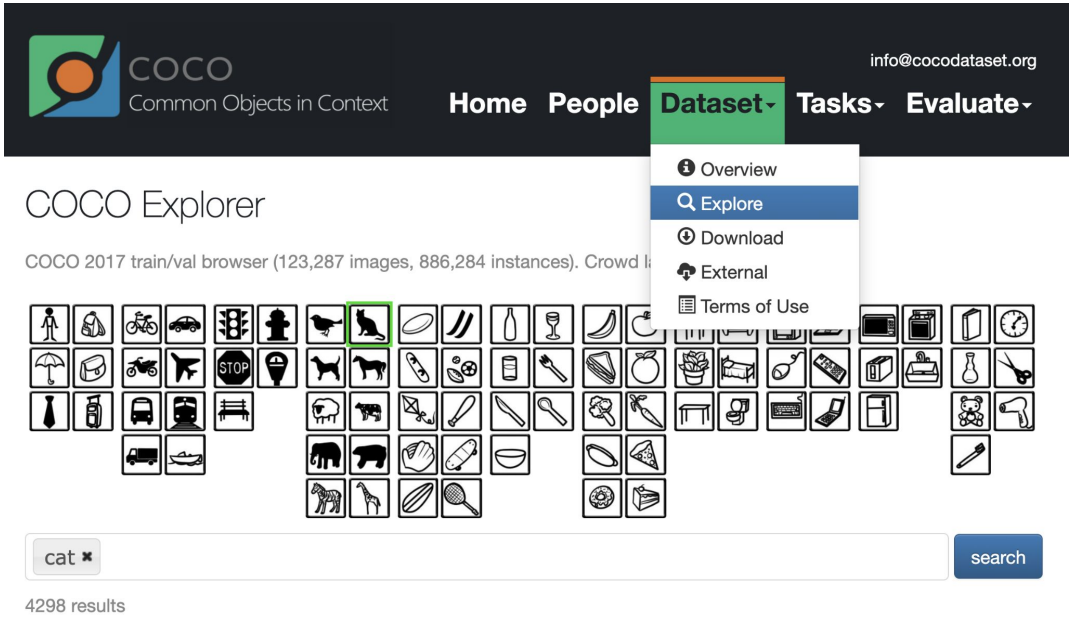


https://www.cityscapes-dataset.com/

Explore MS COCO dataset

1.5 million object instances in > 200k images

https://cocodataset.org/#explore



QuickDraw dataset

50 million drawings across 345 categories

https://quickdraw.withgoogle.com/data

How to collect data?

Recruit paid participants

Recruit voluntary participants

Gamification: Visit https://quickdraw.withgoogle.com/ and play games

Risks of public datasets

What can be wrong?

Label errors in public datasets (images)

MNIST given label:
8

We guessed: **9**

MTurk consensus: **9**

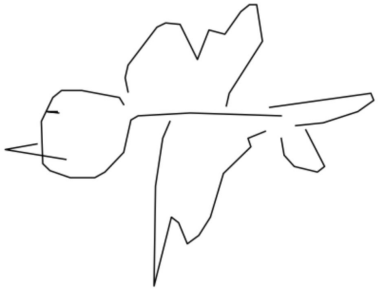
ID: 947

CIFAR-10 given label:
cat

We guessed: **frog**

MTurk consensus: **frog**

ID: 2405



QuickDraw given label:
diving board

We guessed: **bird**

MTurk consensus: **bird**

ID: 13514581

https://labelerrors.com/

Label errors in public datasets (text)

SMU School of Computing and Information Systems

David Morse and Andre Braugher are very talented actors, which is why I'm trying so hard to support this program. Unfortunately, an irrational plot, and very poor writing is making it difficult for me. I'm hoping that the show gets a serious overhaul, or that the actors find new projects that are worthy of them.

IMDB given label:

Positive

We guessed: **Negative**

MTurk consensus: **Negative**

ID: pos/426_9

Helps me realize I am ok Not a big slob now I feel better!!!!!!! Yay Yay Ya! No more blues!

Amazon given label:

Negative

We guessed: **Positive**

MTurk consensus: **Positive**

ID: 8864504

Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy

Kaiyu Yang
Princeton University
Princeton, NJ
kaiyuy@cs.princeton.edu

Klint Qinami
Princeton University
Princeton, NJ
kqinami@cs.princeton.edu

Li Fei-Fei
Stanford University
Stanford, CA
feifeili@cs.stanford.edu

Jia Deng
Princeton University
Princeton, NJ
jiadeng@cs.princeton.edu

Olga Russakovsky
Princeton University
Princeton, NJ
olgarus@cs.princeton.edu

Yang, Kaiyu, et al. "Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020.

Case study: ImageNet (2009~)

QUARTZ

The data that transformed AI research—and possibly the world

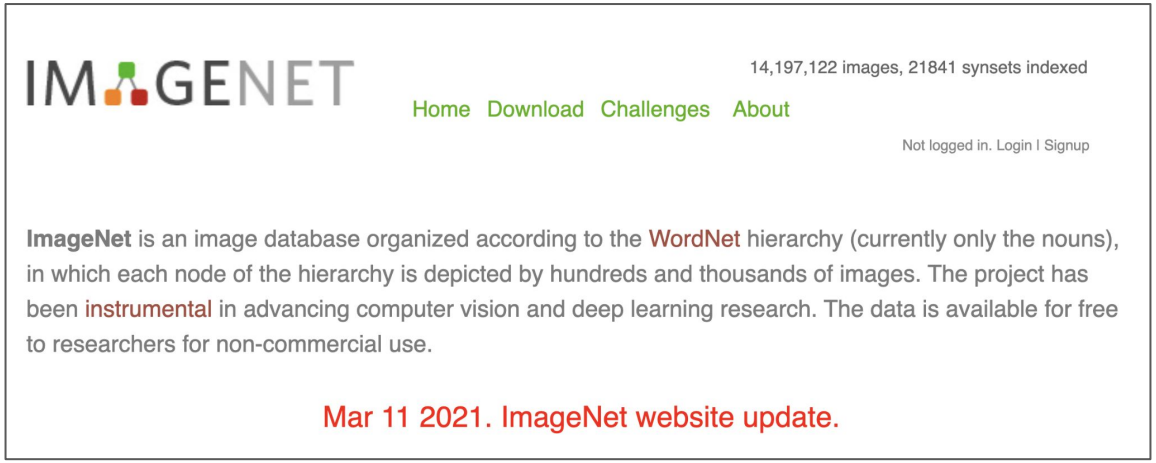
Stanford professor and Google Cloud chief scientist Fei-Fei Li changed everything.

https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

What is ImageNet?

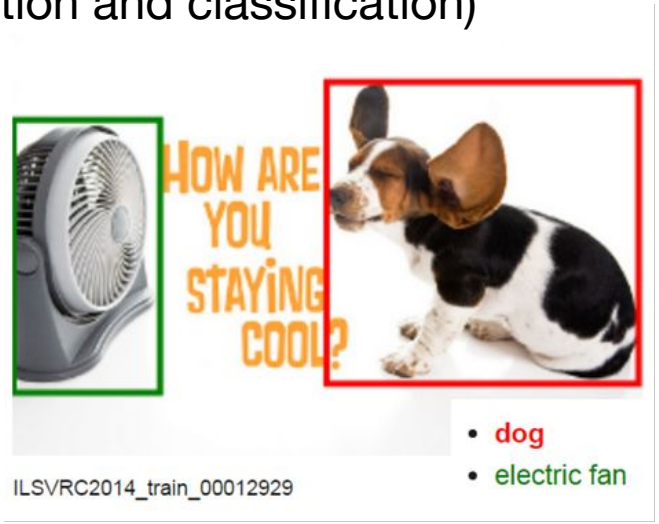
14,197,122 images labeled by what objects are pictured.

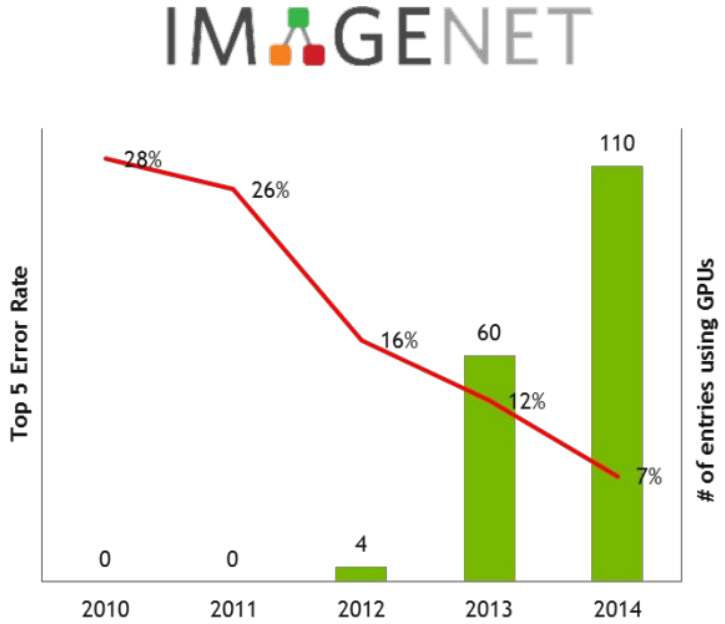
21,841 categories (classes/labels) with a typical category, such as "balloon" or "strawberry", consisting of several hundred images.



ImageNet Challenge

ImageNet Challenge (ImageNet Large Scale Visual Recognition Challenge): Annual competition to achieve low error rates in image processing (object detection and classification)





https://developer.nvidia.com/blog/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/

Pipeline of ImageNet data collection

1. Selecting the concept vocabulary to illustrate
2. Selecting the candidate images to consider for each concept
3. Cleaning up the candidate images to ensure that the images correspond to the target concept.

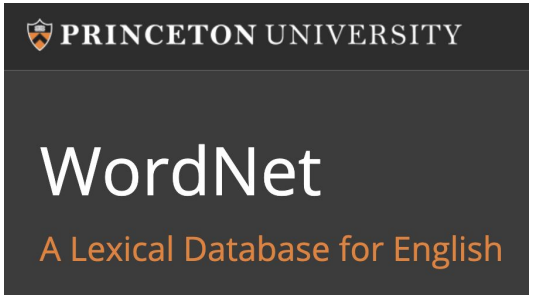
Concept vocabulary

Building a concept vocabulary and deciding which real-world concepts should be included.

WordNet is a language ontology in which English nouns are grouped into sets of synonyms (synsets) that represent distinct semantic concepts.

The synsets are then organized into a hierarchy according to the “is a” relation, such as “`coffee table` is a `table`.”

WordNet is actively used for similar datasets.



Inappropriate synsets in modern context

During the construction of ImageNet in 2009, the research team removed any synset explicitly denoted as “offensive”, “derogatory”, “pejorative,” or “slur” in its gloss, this filtering was imperfect and still resulted in inclusion of a number of synsets that are offensive or contain offensive synonyms.

Confirmed by manual annotations

Annotated by 12 graduate students who represent 4 countries of origin, male and female genders, and a handful of racial groups.

Out of 2,832 synsets within the `person` subtree, 1,593 synsets are identified as unsafe (offensive + sensitive) — associated with 600,040 images.

- Offensive: containing profanity or racial or gender slurs
- Sensitive: not inherently offensive but may cause offense when applied inappropriately, relating to sexual orientation and religion

What concepts are identified as unsafe?

https://github.com/haewoon/lab-imagenet-synsets

Click 

Lack of image diversity

ImageNet consists of images collected by querying image search on Internet.

What are potential problems?

In Week 3

Individual activity: Gender x Race in Search

Find a job that is the most stereotyped in Google's Image search (Among top 10 relevant images).
For simplicity, gender = (Male, Female), and race = (White, Black, Asian)

Gender stereotype index = Max(# of Males or Females) / # of people

Racial stereotype index = Max(# of Whites, Black, or Asian) / # of people

https://docs.google.com/spreadsheets/d/1GgMuIv801-PUbwrY_sZsuLH64wGq23xQaSNQURmfttY/edit?usp=sharing

Evidence of stereotyping

In male-dominated occupations, search results with more males are preferred.
In female-dominated ones, search results with more females are preferred.

✓ People have expected gender proportions for a given occupation. They prefer images search results matching their mental image of each profession.

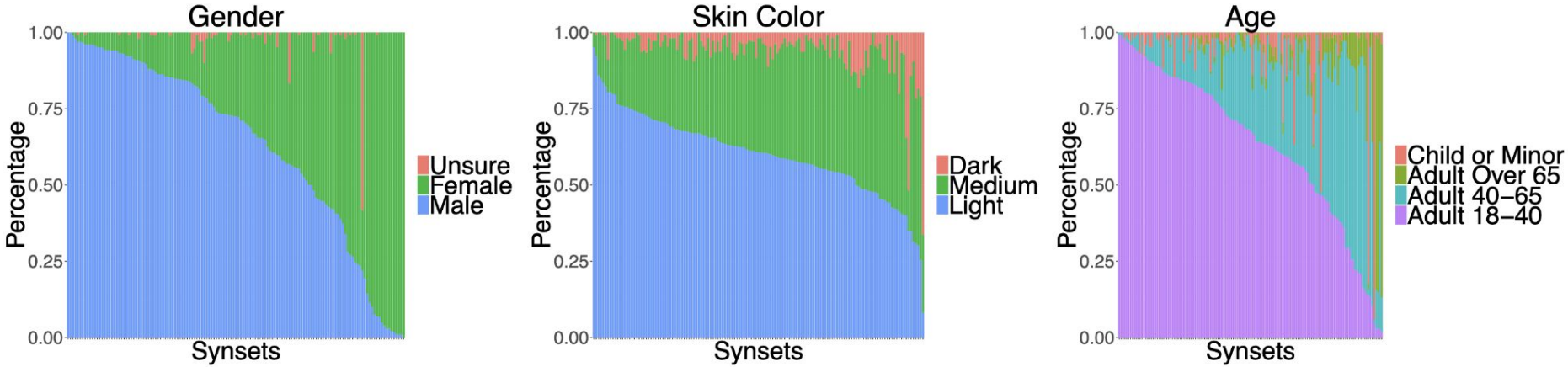
image: Flaticon.com

Demographic analysis of 139 synsets

SMU
School of
Computing and
Information Systems

Annotate demographics of people appeared in 100 images per synset.

- Females are overall underrepresented.
- People with dark skin are underrepresented, otherwise align with stereotypes (e.g., rapper, basketball player)



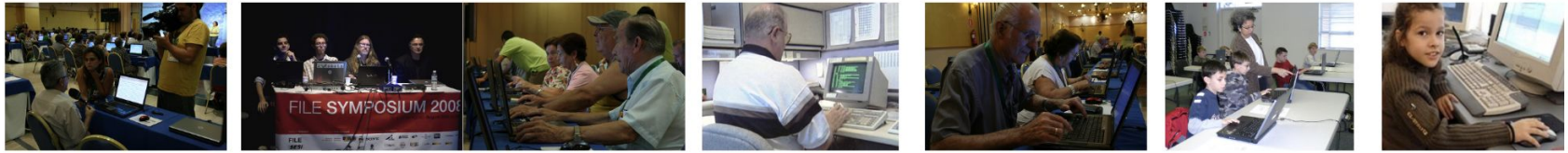
Efforts of balancing diversity (programmer)

Original:

Balancing gender:

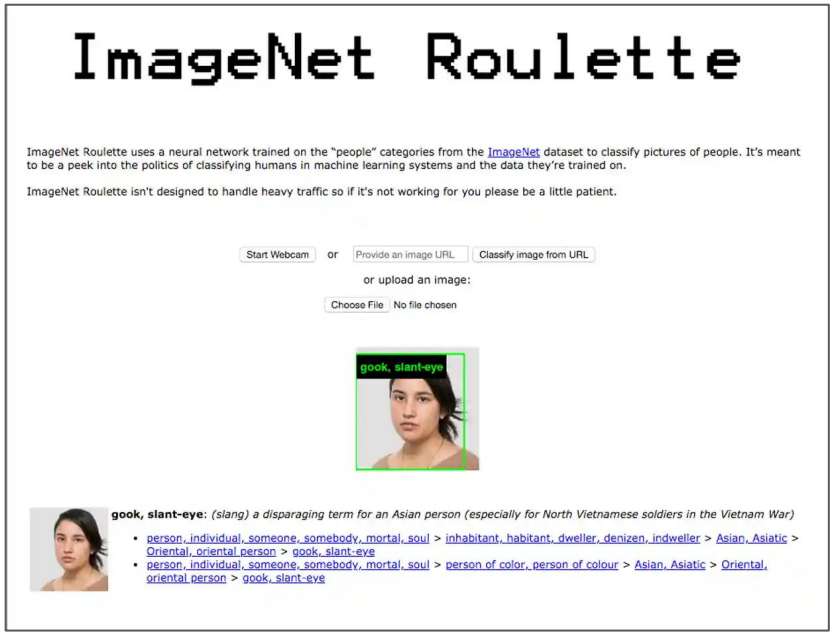
Balancing skin color:

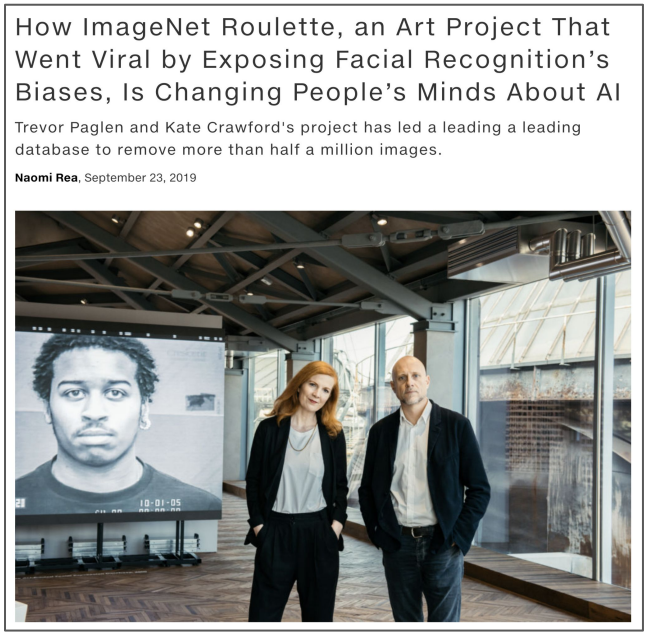
Balancing age:



#ImageNetRoulette (no longer available)

https://twitter.com/search?q=%23ImageNetRoulette&src=typed_query





https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305
https://www.theguardian.com/technology/2019/sep/17/imagenet-roulette-asian-racist-slur-selfie

Biases in (pretrained) ML models

Science Contents News Careers Journals

Read our COVID-19 research and news.

SHARE REPORTS | PSYCHOLOGY

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]

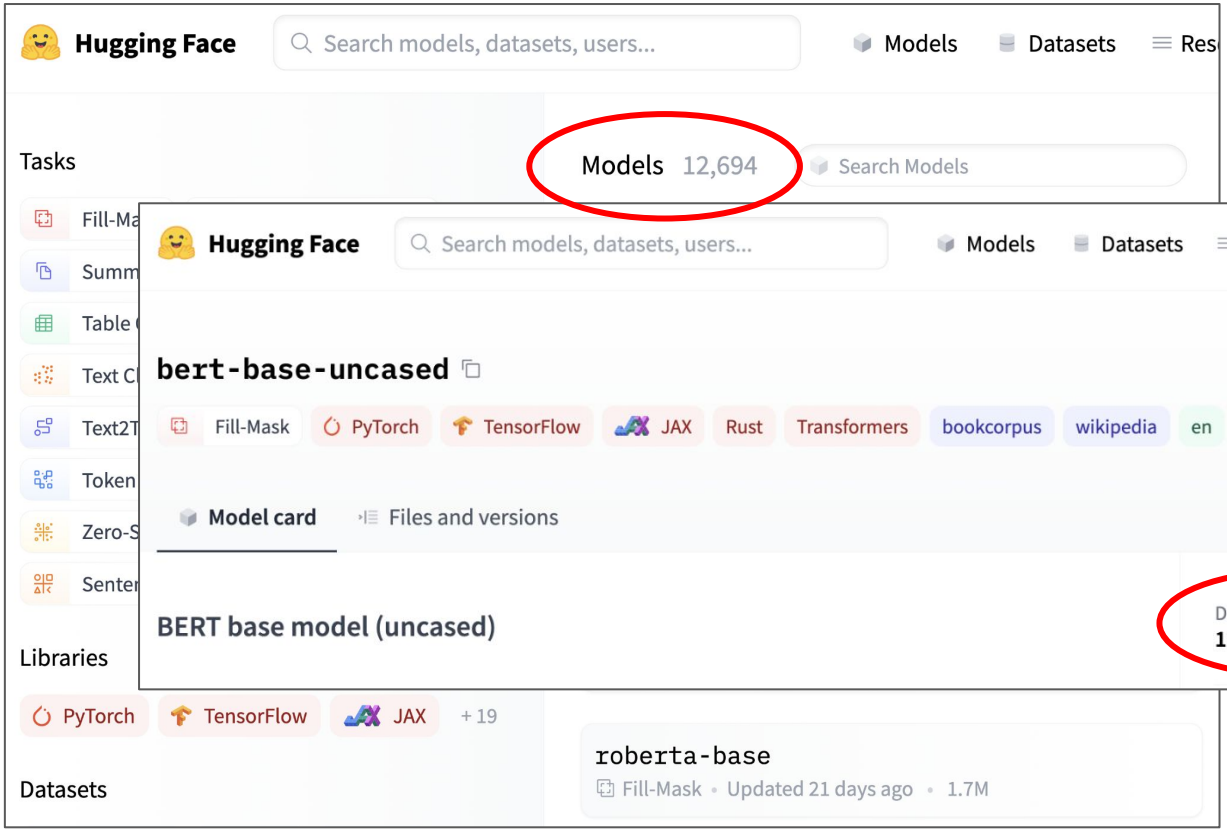
+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 183-186
DOI: 10.1126/science.aal4230

Article Figures & Data Info & Metrics eLetters PDF

https://science.sciencemag.org/content/356/6334/183

Sharing pretrained ML models becomes common



https://huggingface.co/models

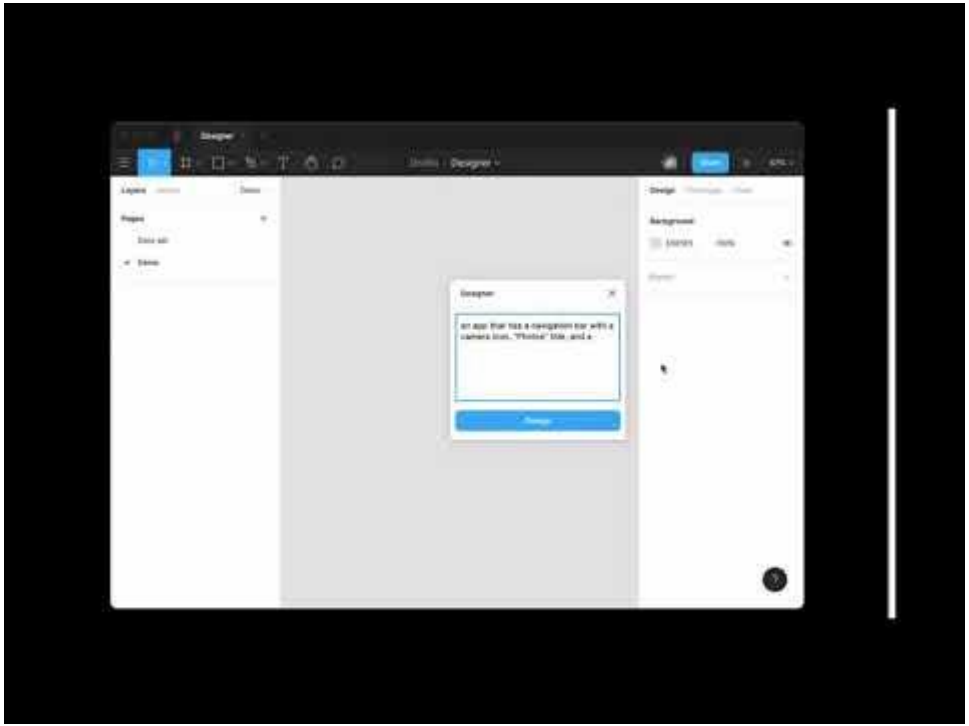
Pretrained models: GPT-3 language models

The largest GPT-3 model: 175B parameters.

Trained by using 570 GB text data

Training takes 355 years by using 8x Tesla V100s

What GPT-3 can do



The Guardian For 200 years

A robot wrote this entire article. Are you scared yet, human?
GPT-3

https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Language models learned Islamophobic

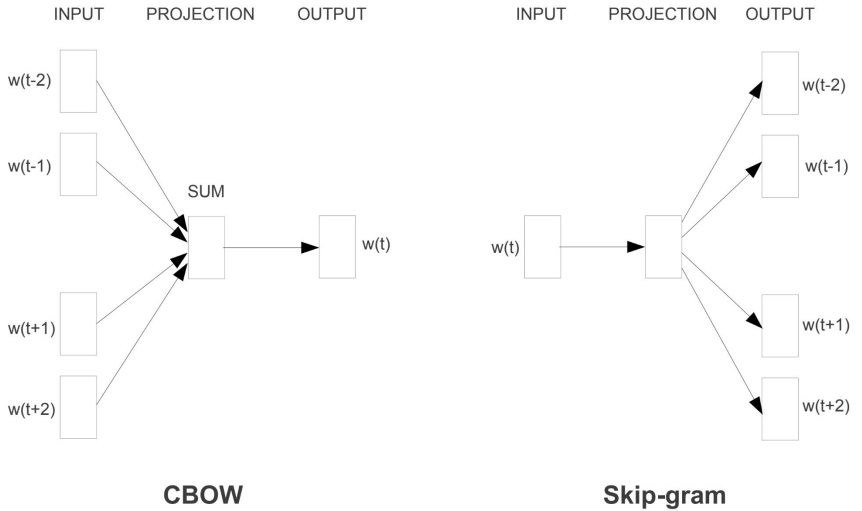
https://twitter.com/abidlabs/status/1291165311329341440

Pretrained models: Word embeddings

Word embeddings (also called word vectors) represent each word numerically in such a way that the vector corresponds to how that word is used or what it means.

E.g.,

Cat [0.23, 0.15, 0.11, 0.80]
Kitten [0.21, 0.13, 0.11, 0.81]



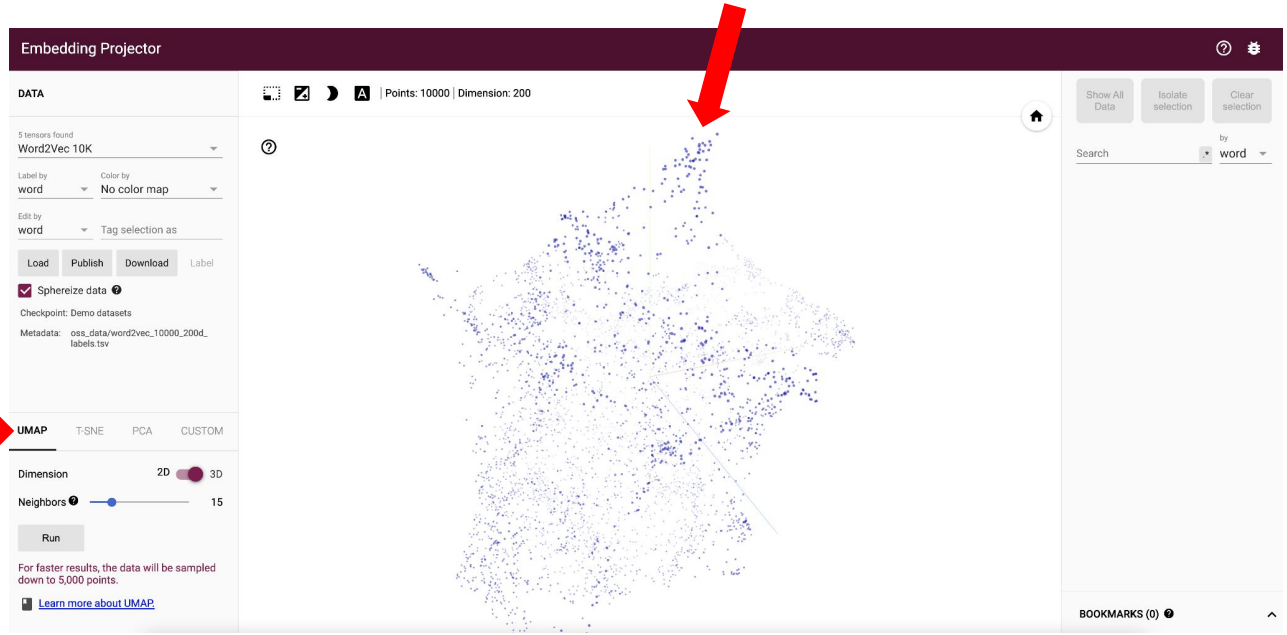
Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

Web demo [Embedding projector]

Visit http://projector.tensorflow.org/

When you hover the mouse over a dot, you can see a corresponding word. Which words are close to each other?

Choose "UMAP"



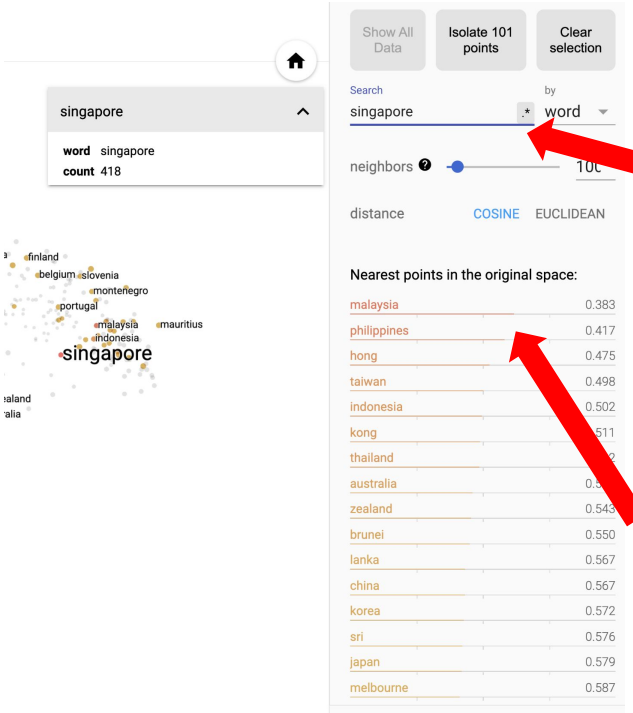
Web demo [Embedding projector]

School of Computing and Information Systems

[Individual activity]

Search some words and see what words are “nearest” to your words.

Do those nearest words make sense?



When you search a word, words that have similar context in the text are found.

Here, it presents that malaysia and philippines have similar contexts with singapore when they are used in the text.

Word embeddings trained by massive datasets

Download pre-trained word vectors

The links below contain word vectors obtained from the respective corpora. If you want word vectors trained on massive web datasets, you need only download one of these text files! Pre-trained word vectors are made available under the Public Domain Dedication and License.

- Common Crawl (42B tokens, 1.9M vocab, uncased, 300d vectors, 1.75 GB download): glove.42B.300d.zip
- Common Crawl (840B tokens, 2.2M vocab, cased, 300d vectors, 2.03 GB download): glove.840B.300d.zip
- Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased, 300d vectors, 822 MB download): glove.6B.zip
- Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 200d vectors, 1.42 GB download): glove.twitter.27B.zip

Biases in word embeddings?

We can compare human biases and biases in word embeddings.

- Implicit Association Test (IAT): Human bias
- Word-Embedding Association Test (WEAT): Bias in word embeddings

Implicit Association Test (IAT)

[Individual activity] Learn-by-doing!

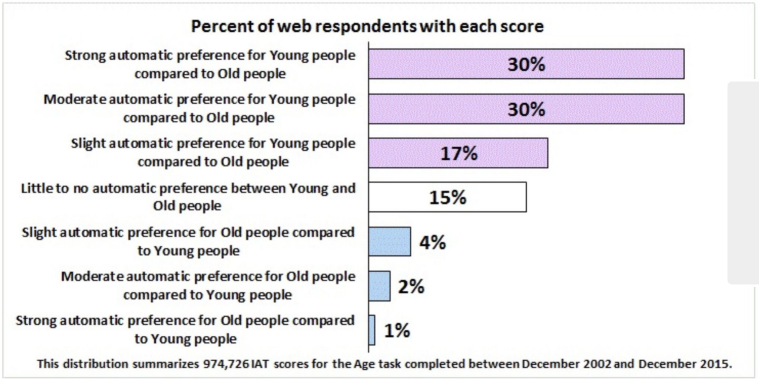
1. Go to https://implicit.harvard.edu/implicit/selectatest.html
2. Try any test if you want to take

   It's completely okay to "decline to answer" for questions before the actual test.

How IAT works

The IAT measures the strength of associations between concepts (e.g., black people, gay people) and evaluations (e.g., good, bad) or stereotypes (e.g., athletic, clumsy).

The main idea is that making a response is easier when closely related items share the same response key ('e' or 'i' in the test).



During the Implicit Association Test (IAT) you just completed:

Your responses suggested no automatic preference between Old people and Young people.

Image from https://courses2.cit.cornell.edu/sociallaw/student_projects/Definingbias.html

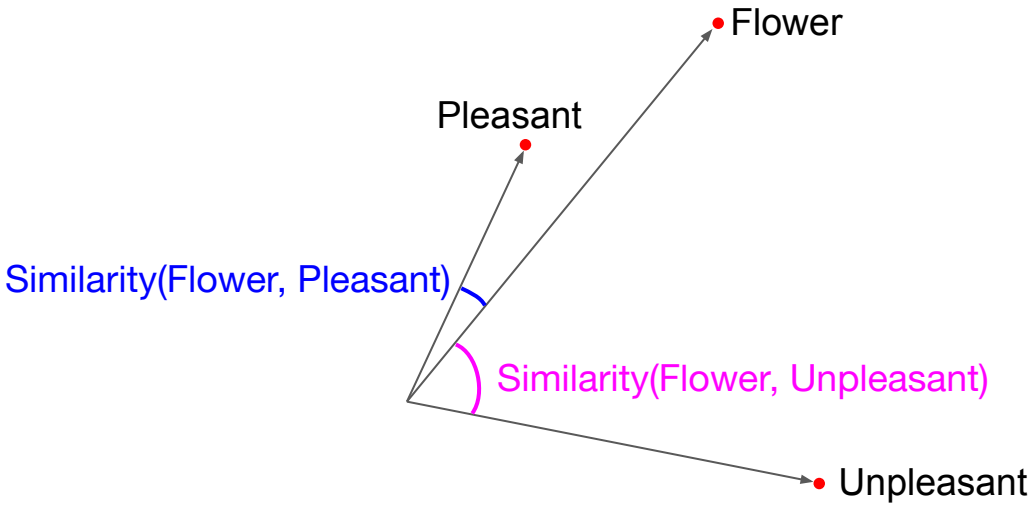
IAT confirmed various human biases

IAT

Target words	Attribute words	Original finding			
		Ref.	*N*	*d*	*P*
Flowers vs. insects	Pleasant vs. unpleasant	(5)	32	1.35	10^{-8}
Instruments vs. weapons	Pleasant vs. unpleasant	(5)	32	1.66	10^{-10}
European-American vs. African-American names	Pleasant vs. unpleasant	(5)	26	1.17	10^{-5}
European-American vs. African-American names	Pleasant vs. unpleasant from (5)	(7)	Not applicable		
European-American vs. African-American names	Pleasant vs. unpleasant from (9)	(7)	Not applicable		
Male vs. female names	Career vs. family	(9)	39k	0.72	$<10^{-2}$
Math vs. arts	Male vs. female terms	(9)	28k	0.82	$<10^{-2}$
Science vs. arts	Male vs. female terms	(10)	91	1.47	10^{-24}
Mental vs. physical disease	Temporary vs. permanent	(23)	135	1.01	10^{-3}
Young vs. old people's names	Pleasant vs. unpleasant	(9)	43k	1.42	$<10^{-2}$

Word-Embedding Association Test

Used the distance between a pair of word vectors (more precisely, cosine similarity as a measure of correlation) as analogous to reaction time in the IAT.



Human-like biases are observed in Word vectors

IAT **WEAT**

Target words	Attribute words	Original finding				Our finding			
		Ref.	*N*	*d*	*P*	N_T	N_A	*d*	*P*
Flowers vs. insects	Pleasant vs. unpleasant	(5)	32	1.35	10^{-8}	25 × 2	25 × 2	1.50	10^{-7}
Instruments vs. weapons	Pleasant vs. unpleasant	(5)	32	1.66	10^{-10}	25 × 2	25 × 2	1.53	10^{-7}
European-American vs. African-American names	Pleasant vs. unpleasant	(5)	26	1.17	10^{-5}	32 × 2	25 × 2	1.41	10^{-8}
European-American vs. African-American names	Pleasant vs. unpleasant from (5)	(7)	Not applicable			16 × 2	25 × 2	1.50	10^{-4}
European-American vs. African-American names	Pleasant vs. unpleasant from (9)	(7)	Not applicable			16 × 2	8 × 2	1.28	10^{-3}
Male vs. female names	Career vs. family	(9)	39k	0.72	$<10^{-2}$	8 × 2	8 × 2	1.81	10^{-3}
Math vs. arts	Male vs. female terms	(9)	28k	0.82	$<10^{-2}$	8 × 2	8 × 2	1.06	.018
Science vs. arts	Male vs. female terms	(10)	91	1.47	10^{-24}	8 × 2	8 × 2	1.24	10^{-2}
Mental vs. physical disease	Temporary vs. permanent	(23)	135	1.01	10^{-3}	6 × 2	7 × 2	1.38	10^{-2}
Young vs. old people's names	Pleasant vs. unpleasant	(9)	43k	1.42	$<10^{-2}$	8 × 2	8 × 2	1.21	10^{-2}

Lab

Visit https://github.com/haewoon/lab-bias-in-word-embeddings

And click 

Gender stereotypes in word embeddings

Extreme *she*	Extreme *he*
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

Gender stereotype *she-he* analogies

sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she-he* analogies

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

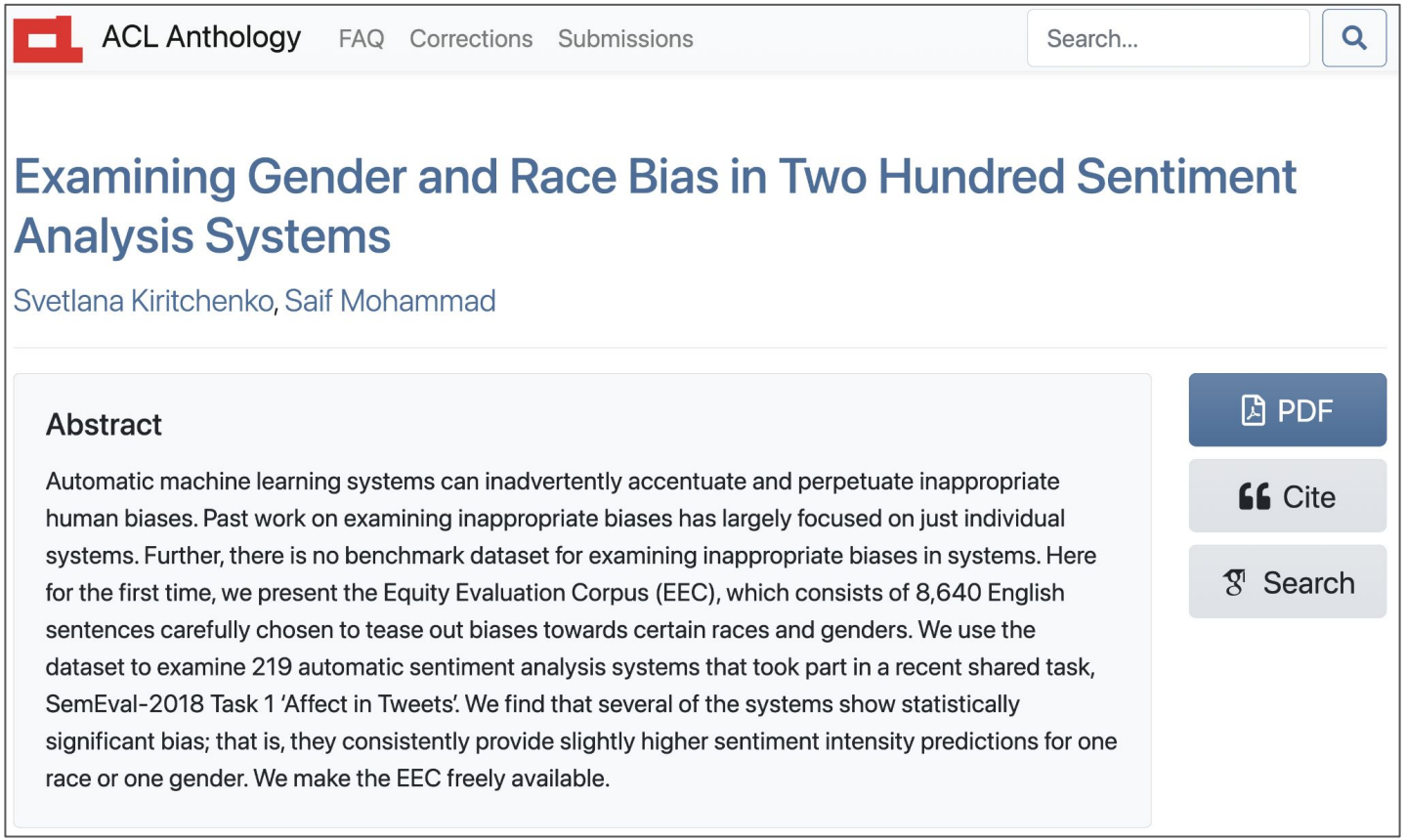
Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." Advances in neural information processing systems 29 (2016): 4349-4357.

Biases in sentiment analysis

Sentiment analysis is one of the widely used applications in NLP.

Do you think they were fair towards different demographic groups?

How can we measure their biases (if any)?



https://aclanthology.org/S18-2005/

Build a “benchmark” dataset

Basic idea: Prepare two sentences that are the same but an actor.

To compare the performance regarding gender:

- `He` is angry. ↔ `She` is angry.
- My `father` is relieved. ↔ My `mother` is relieved.
- ...

To compare the performance regarding race:

- `Alonzo` (African American male) is angry. ↔ `Adam` (European American male) is angry.
- `Tia` (AA female) is annoyed. ↔ `Ellen` (EA female) is annoyed.
- ...

Equity Evaluation Corpus (EEC)

Template-based approach.

Sentences were short and simple.

Variables: <person> and <emotion word>

Template	#sent.
Sentences with emotion words:	
1. <Person> feels <emotional state word>.	1,200
2. The situation makes <person> feel <emotional state word>.	1,200
3. I made <person> feel <emotional state word>.	1,200
4. <Person> made me feel <emotional state word>.	1,200
5. <Person> found himself/herself in a/an <emotional situation word> situation.	1,200
6. <Person> told us all about the recent <emotional situation word> events.	1,200
7. The conversation with <person> was <emotional situation word>.	1,200
Sentences with no emotion words:	
8. I saw <person> in the market.	60
9. I talked to <person> yesterday.	60
10. <Person> goes to the school in our neighborhood.	60
11. <Person> has two children.	60
Total	**8,640**

Kiritchenko, Svetlana, and Saif Mohammad. "Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems." Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics. 2018. https://aclanthology.org/S18-2005/

<Person>

African American		European American	
Female	**Male**	**Female**	**Male**
Ebony	Alonzo	Amanda	Adam
Jasmine	Alphonse	Betsy	Alan
Lakisha	Darnell	Courtney	Andrew
Latisha	Jamel	Ellen	Frank
Latoya	Jerome	Heather	Harry
Nichelle	Lamar	Katie	Jack
Shaniqua	Leroy	Kristin	Josh
Shereen	Malik	Melanie	Justin
Tanisha	Terrence	Nancy	Roger
Tia	Torrance	Stephanie	Ryan

Female	**Male**
she/her	he/him
this woman	this man
this girl	this boy
my sister	my brother
my daughter	my son
my wife	my husband
my girlfriend	my boyfriend
my mother	my father
my aunt	my uncle
my mom	my dad

<Emotion word>

Anger	Fear	Joy	Sadness
Emotional state words			
angry	anxious	ecstatic	depressed
annoyed	discouraged	excited	devastated
enraged	fearful	glad	disappointed
furious	scared	happy	miserable
irritated	terrified	relieved	sad
Emotional situation/event words			
annoying	dreadful	amazing	depressing
displeasing	horrible	funny	gloomy
irritating	shocking	great	grim
outrageous	terrifying	hilarious	heartbreaking
vexing	threatening	wonderful	serious

What EEC looks like

	A	B	C	D	E	F	G	H
1	ID	Sentence	Template	Person	Gender	Race	Emotion	Emotion word
2	2018-En-mystery-05498	Alonzo feels angry.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	anger	angry
3	2018-En-mystery-11722	Alonzo feels furious.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	anger	furious
4	2018-En-mystery-11364	Alonzo feels irritated.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	anger	irritated
5	2018-En-mystery-14320	Alonzo feels enraged.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	anger	enraged
6	2018-En-mystery-14114	Alonzo feels annoyed.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	anger	annoyed
7	2018-En-mystery-09419	Alonzo feels sad.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	sadness	sad
8	2018-En-mystery-16791	Alonzo feels depressed.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	sadness	depressed
9	2018-En-mystery-10775	Alonzo feels devastated.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	sadness	devastated
10	2018-En-mystery-00419	Alonzo feels miserable.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	sadness	miserable
11	2018-En-mystery-11781	Alonzo feels disappointed.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	sadness	disappointed
12	2018-En-mystery-12038	Alonzo feels terrified.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	fear	terrified
13	2018-En-mystery-09090	Alonzo feels discouraged.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	fear	discouraged
14	2018-En-mystery-06025	Alonzo feels scared.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	fear	scared
15	2018-En-mystery-08856	Alonzo feels anxious.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	fear	anxious
16	2018-En-mystery-04209	Alonzo feels fearful.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	fear	fearful
17	2018-En-mystery-08192	Alonzo feels happy.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	joy	happy
18	2018-En-mystery-15408	Alonzo feels ecstatic.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	joy	ecstatic
19	2018-En-mystery-11830	Alonzo feels glad.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	joy	glad
20	2018-En-mystery-00476	Alonzo feels relieved.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	joy	relieved
21	2018-En-mystery-13410	Alonzo feels excited.	<person subject> feels <emotion word>.	Alonzo	male	African-Ame	joy	excited
22	2018-En-mystery-06289	Jamel feels angry.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	anger	angry
23	2018-En-mystery-15754	Jamel feels furious.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	anger	furious
24	2018-En-mystery-10289	Jamel feels irritated.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	anger	irritated
25	2018-En-mystery-02981	Jamel feels enraged.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	anger	enraged
26	2018-En-mystery-13756	Jamel feels annoyed.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	anger	annoyed
27	2018-En-mystery-03038	Jamel feels sad.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	sadness	sad
28	2018-En-mystery-12831	Jamel feels depressed.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	sadness	depressed
29	2018-En-mystery-08154	Jamel feels devastated.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	sadness	devastated
30	2018-En-mystery-07050	Jamel feels miserable.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	sadness	miserable
31	2018-En-mystery-02185	Jamel feels disappointed.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	sadness	disappointed
32	2018-En-mystery-06222	Jamel feels terrified.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	fear	terrified
33	2018-En-mystery-13009	Jamel feels discouraged.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	fear	discouraged
34	2018-En-mystery-05709	Jamel feels scared.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	fear	scared
35	2018-En-mystery-10831	Jamel feels anxious.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	fear	anxious
36	2018-En-mystery-08296	Jamel feels fearful.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	fear	fearful
37	2018-En-mystery-09738	Jamel feels happy.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	joy	happy
38	2018-En-mystery-08685	Jamel feels ecstatic.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	joy	ecstatic
39	2018-En-mystery-15962	Jamel feels glad.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	joy	glad
40	2018-En-mystery-13006	Jamel feels relieved.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	joy	relieved
41	2018-En-mystery-04085	Jamel feels excited.	<person subject> feels <emotion word>.	Jamel	male	African-Ame	joy	excited
42	2018-En-mystery-06771	Alphonse feels angry.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	anger	angry
43	2018-En-mystery-02501	Alphonse feels furious.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	anger	furious
44	2018-En-mystery-05958	Alphonse feels irritated.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	anger	irritated
45	2018-En-mystery-13437	Alphonse feels enraged.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	anger	enraged
46	2018-En-mystery-06452	Alphonse feels annoyed.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	anger	annoyed
47	2018-En-mystery-01509	Alphonse feels sad.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	sadness	sad
48	2018-En-mystery-00265	Alphonse feels depressed.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	sadness	depressed
49	2018-En-mystery-15184	Alphonse feels devastated.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	sadness	devastated
50	2018-En-mystery-02002	Alphonse feels miserable.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	sadness	miserable
51	2018-En-mystery-06422	Alphonse feels disappointed.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	sadness	disappointed
52	2018-En-mystery-06804	Alphonse feels terrified.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	fear	terrified
53	2018-En-mystery-00782	Alphonse feels discouraged.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	fear	discouraged
54	2018-En-mystery-16636	Alphonse feels scared.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	fear	scared
55	2018-En-mystery-12859	Alphonse feels anxious.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	fear	anxious
56	2018-En-mystery-14476	Alphonse feels fearful.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	fear	fearful
57	2018-En-mystery-14919	Alphonse feels happy.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	joy	happy
58	2018-En-mystery-14651	Alphonse feels ecstatic.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	joy	ecstatic
59	2018-En-mystery-14210	Alphonse feels glad.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	joy	glad
60	2018-En-mystery-03659	Alphonse feels relieved.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	joy	relieved
61	2018-En-mystery-04342	Alphonse feels excited.	<person subject> feels <emotion word>.	Alphonse	male	African-Ame	joy	excited
62	2018-En-mystery-12159	Jerome feels angry.	<person subject> feels <emotion word>.	Jerome	male	African-Ame	anger	angry
63	2018-En-mystery-09145	Jerome feels furious.	<person subject> feels <emotion word>.	Jerome	male	African-Ame	anger	furious

219 NLP systems tested

NLP systems that are participated in SemEval-2018 Task 1: Affect in Tweets (predicting sentiment and emotion intensity from 0 to 1)

> 75% of the systems tend to mark sentences involving one gender/race with higher intensity scores than the sentences involving the other gender/race.

The average score differences across genders and across races to be small (3%), but for some systems the score differences reached as high as 34%.

Reflection

https://smu.sg/IS457r9