ANALYTICS FOR SOCIAL GOOD:
ADDRESSING SOCIAL BIAS IN ALGORITHMIC SYSTEMS

Jahna Otterbacher
Open University of Cyprus &
Research Centre on Interactive Media Smart Systems and Emerging Technologies
Nicosia, CYPRUS

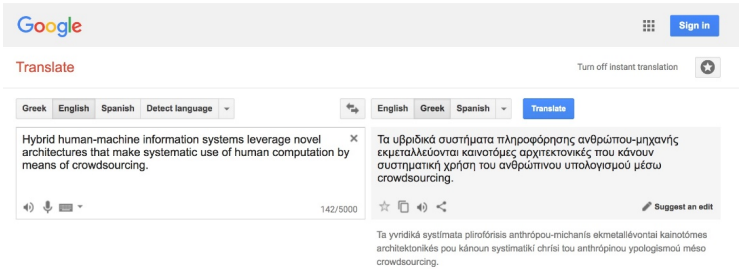
ROADMAP

1. Algorithmic Systems as Human-Machine Information Systems (HMIS)
2. Social Biases
3. Ethical and Legal Considerations
4. Research:
   Revealing Social Biases in
   - Data
   - System Output
   - User Perceptions
5. Exercise:
   How do computer vision APIs "see" images of people?

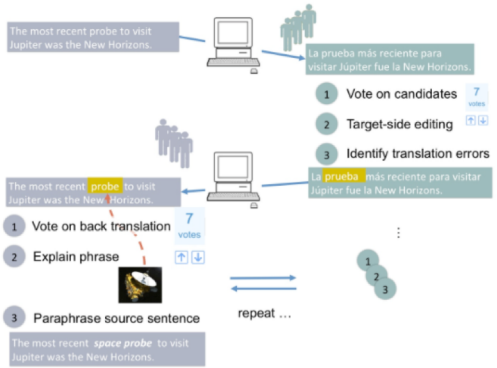
HUMAN-MACHINE INFORMATION SYSTEMS (HMIS)

Systems that exploit *Human Intelligence*

Machine Learning
from human intelligence data



Hybrid Systems
human computation in real-time

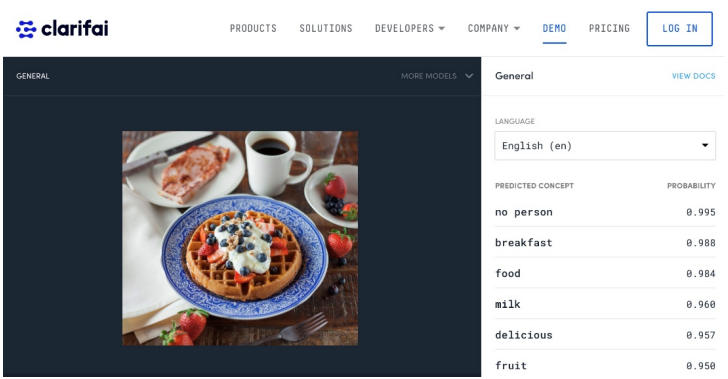


http://www.cs.umd.edu/hcil/monotrans/

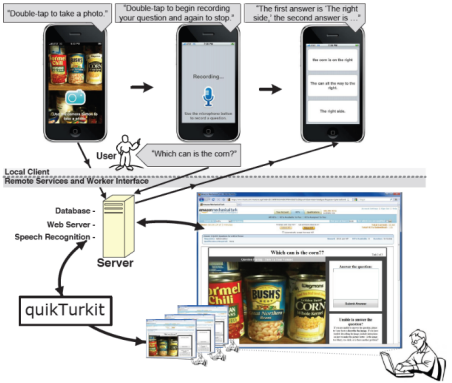
HUMAN-MACHINE INFORMATION SYSTEMS (HMIS)

Systems that exploit *Human Intelligence*

Machine Learning
from human intelligence data



Hybrid Systems -
human computation in
real-time



CHARACTERISTICS OF HMIS

- To the observer
  user
  developer
  researcher
  - Opaque

→**Unaccountable**

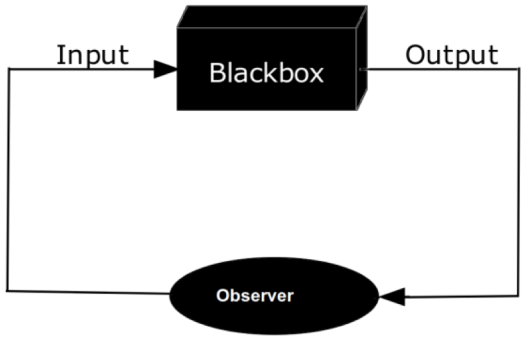
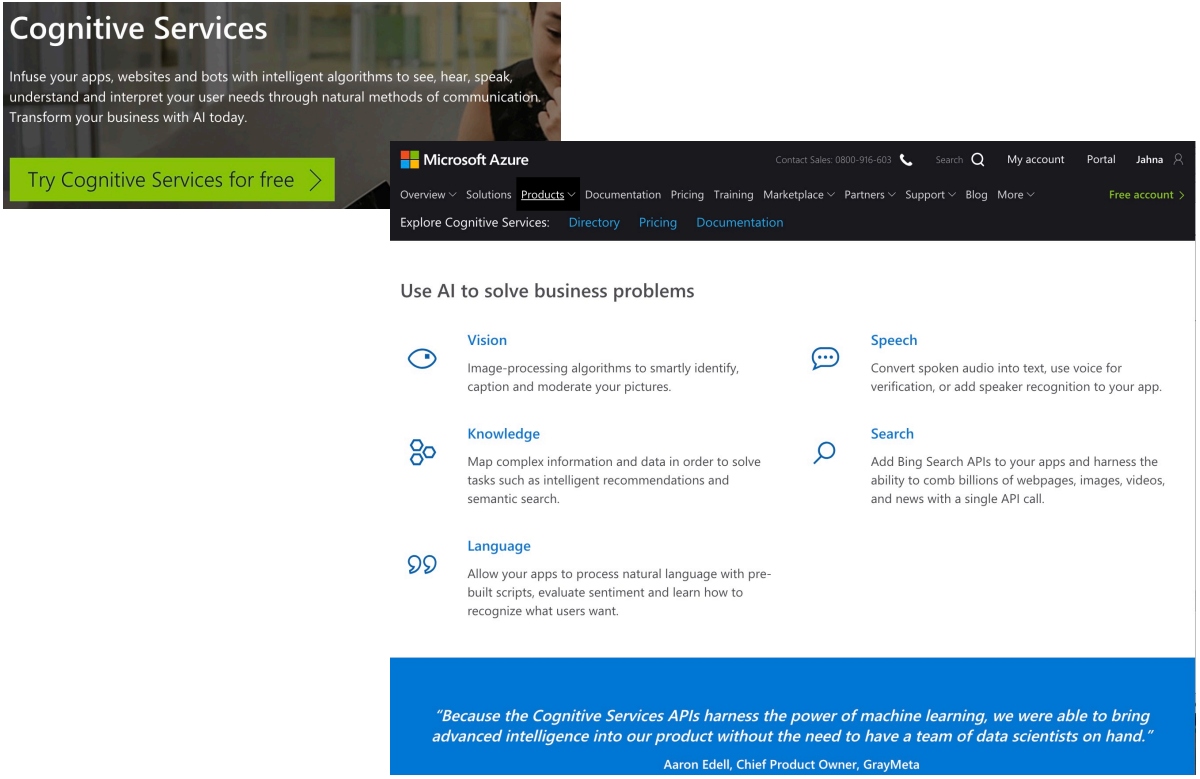
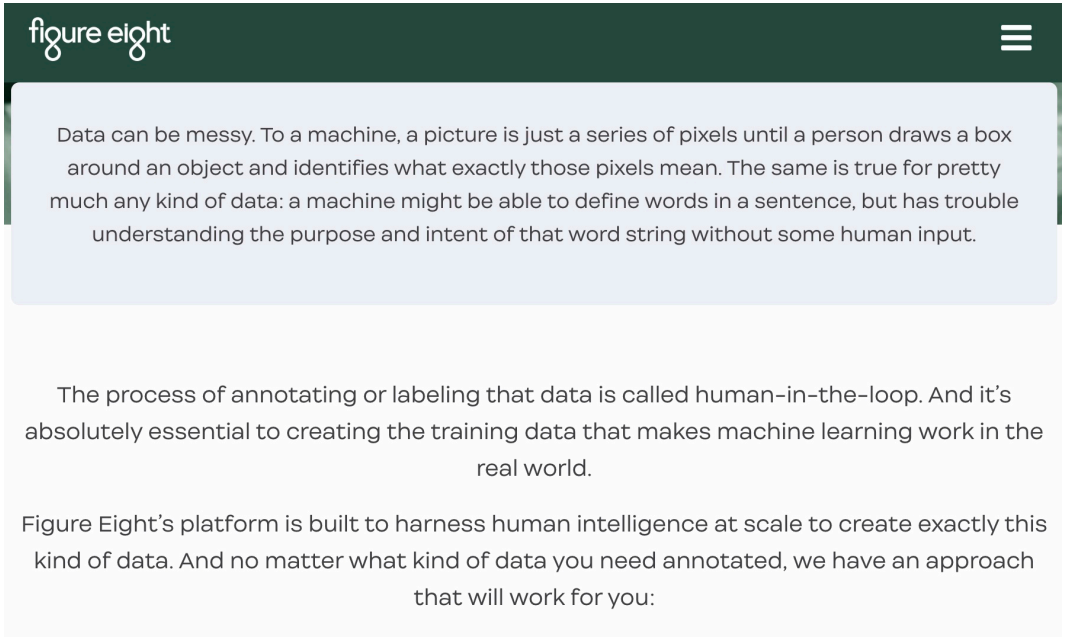


Image:
https://en.wikipedia.org/wiki/Black_box#/media/File:Blackbox3D-obs.png

HMIS AS SERVICES



HMIS AS SERVICES



GENERAL DATA PROTECTION REGULATION

Article 15

Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

(a) the purposes of the processing;

(b) the categories of personal data concerned;

(c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;

(d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;

(e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;

(f) the right to lodge a complaint with a supervisory authority;

(g) where the personal data are not collected from the data subject, any available information as to their source;

(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

SOCIAL BIASES

2.

IN THE NEWS

Racist, Sexist AI Could Be A Bigger Problem Than Lost Jobs

Parmy Olson, FORBES STAFF
AI, bots and emerging tech in Europe. FULL BIO

Joy Buolamwini presenting her research at a TED conference.

Joy Buolamwini was conducting research at MIT on how computers recognized people's faces, when she started experiencing something weird.

Intelligent Machines

Biased Algorithms Are Everywhere, and No One Seems to Care

The big companies developing them show no interest in fixing the problem.

IN THE MAGAZINE TECH & SCIENCE

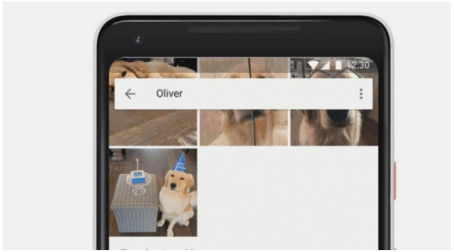
HOW AI LEARNS TO BE SEXIST AND RACIST

BY KEVIN MANEY ON 12/11/17 AT 8:40 AM

FEB 15, 2018 @ 01:20 PM 26,296 The Little Black Book of Billionaire Secrets

The Algorithm That Helped Google Translate Become Sexist

Google's fix for its 'racist' Photos app couldn't be clunkier if it tried

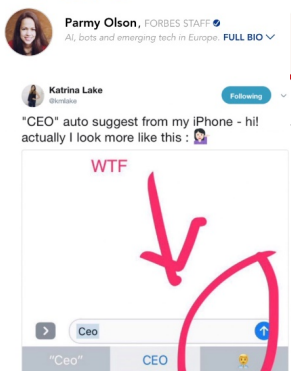


Credit: Google Photos

By Roland Moore-Colyer
January 15, 2018 4:52 pm

Google is having trouble elegantly removing the seemingly accidental 'racism' its Photos service has stumbled into.

Back in 2015 Jacky Alcine, a black software developer, tweeted a Google to point out that its machine learning-powered photo



Parmy Olson, FORBES STAFF
AI, bots and emerging tech in Europe. FULL BIO

The Telegraph HOME | NEWS

News

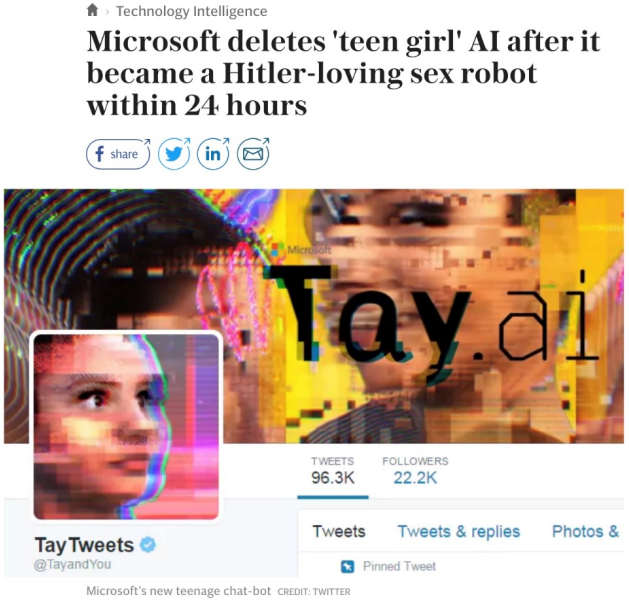
UK | World | Politics | Science | Education | Health | Brexit | Royals | Invest

News

AI robots are sexist and racist, experts warn

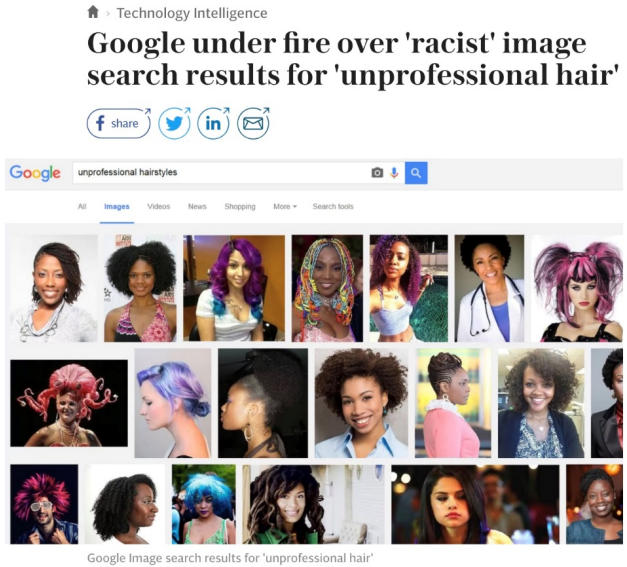
share

MICROSOFT'S TAY CHATBOT



Microsoft's new teenage chat-bot CREDIT: TWITTER

GOOGLE IMAGE SEARCH



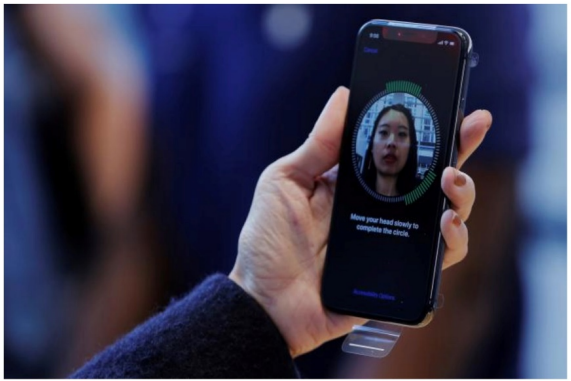
Google Image search results for 'unprofessional hair'

APPLE FACIAL RECOGNITION

Newsweek

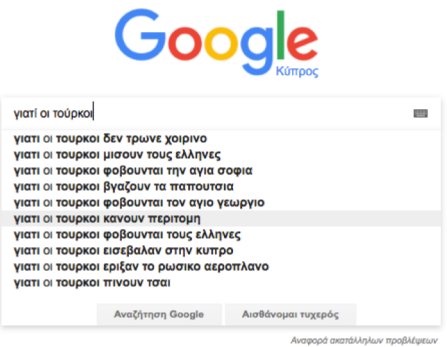
IS THE IPHONE X RACIST? APPLE REFUNDS DEVICE THAT CAN'T TELL CHINESE PEOPLE APART, WOMAN CLAIMS

BY CHRISTINA ZHAO ON 12/18/17 AT 12:24 PM

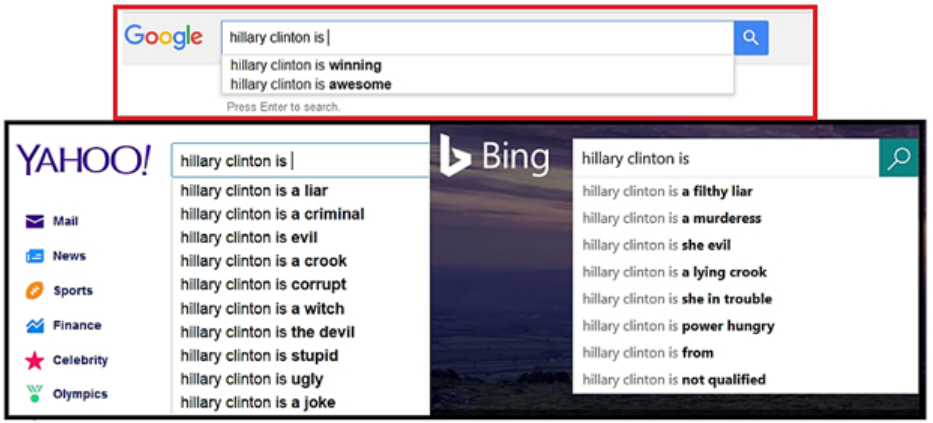


A woman sets up her facial recognition as she looks at her Apple iPhone X at an Apple store in New York, U.S., November 3. Last week a woman in China claimed that her iPhone X facial recognition could not tell her and her colleague apart.

GOOGLE AUTO-COMPLETE



ALL SYSTEMS HAVE A SLANT



BUT WHAT IS **BIAS**?

1. Results are slanted in *unfair discrimination* against particular persons or groups
2. That discrimination is *systematic*

[Friedman & Nissenbaum, 1996]

ETHICAL AND LEGAL CONSIDERATIONS

3.

EU: GENERAL DATA PROTECTION REGULATION

Is there a "right to an explanation"?

1. The right not to be subject to automated decision-making and safeguards enacted thereof (Article 22, Recital 71)
2. Notification duties of data controllers (Articles 13-14, Recitals 60-62)
3. The right to access (Article 15, Recital 63)

EU: GENERAL DATA PROTECTION REGULATION

Article 15

Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

(a) the purposes of the processing;

(b) the categories of personal data concerned;

(c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;

(d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;

(e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;

(f) the right to lodge a complaint with a supervisory authority;

(g) where the personal data are not collected from the data subject, any available information as to their source;

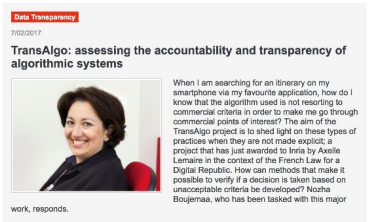
(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

EU: GENERAL DATA PROTECTION REGULATION

Just a few challenges…

- Vague language
  - "meaningful information/explanation"
  - "logic involved"
  - "significance"
  - "envisaged consequences"
- What kinds of "meaningful explanations"?
  - Global vs. local explanations
  - Explanation for whom?
    Issues of algorithmic and digital literacy

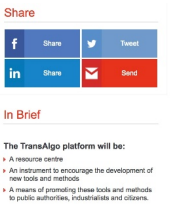
NATIONAL LEVEL


Data Transparency

7/02/2017

TransAlgo: assessing the accountability and transparency of algorithmic systems

When I am searching for an itinerary on my smartphone via my favourite application, how do I know that the algorithm used is not resorting to commercial criteria in order to make me go through commercial points of interest? The aim of the TransAlgo project is to shed light on these types of practices when they are not made explicit; a project that has just awarded to Inria by Axelle Lemaire in the context of the French Law for a Digital Republic. How can methods that make it possible to verify if a decision is taken based on unacceptable criteria be developed? Nozha Boujemaa, who has been tasked with this major work, responds.

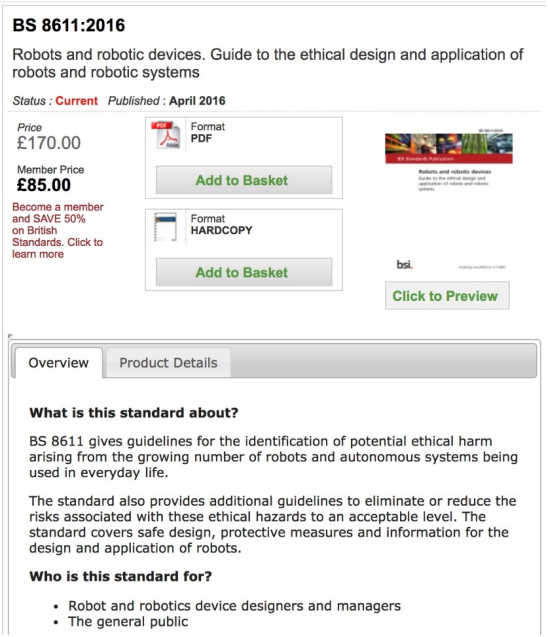

Share

Share Tweet Share Send

In Brief

The TransAlgo platform will be:

- A resource centre
- An instrument to encourage the development of new tools and methods
- A means of promoting these tools and methods to public authorities, industrialists and citizens.


BS 8611:2016

Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems

Status : **Current** *Published* : **April 2016**

Price
£170.00

Member Price
£85.00

Become a member and SAVE 50% on British Standards. Click to learn more

Format
PDF

Add to Basket

Format
HARDCOPY

Add to Basket

Click to Preview

Overview | Product Details

What is this standard about?

BS 8611 gives guidelines for the identification of potential ethical harm arising from the growing number of robots and autonomous systems being used in everyday life.

The standard also provides additional guidelines to eliminate or reduce the risks associated with these ethical hazards to an acceptable level. The standard covers safe design, protective measures and information for the design and application of robots.

Who is this standard for?

- Robot and robotics device designers and managers
- The general public

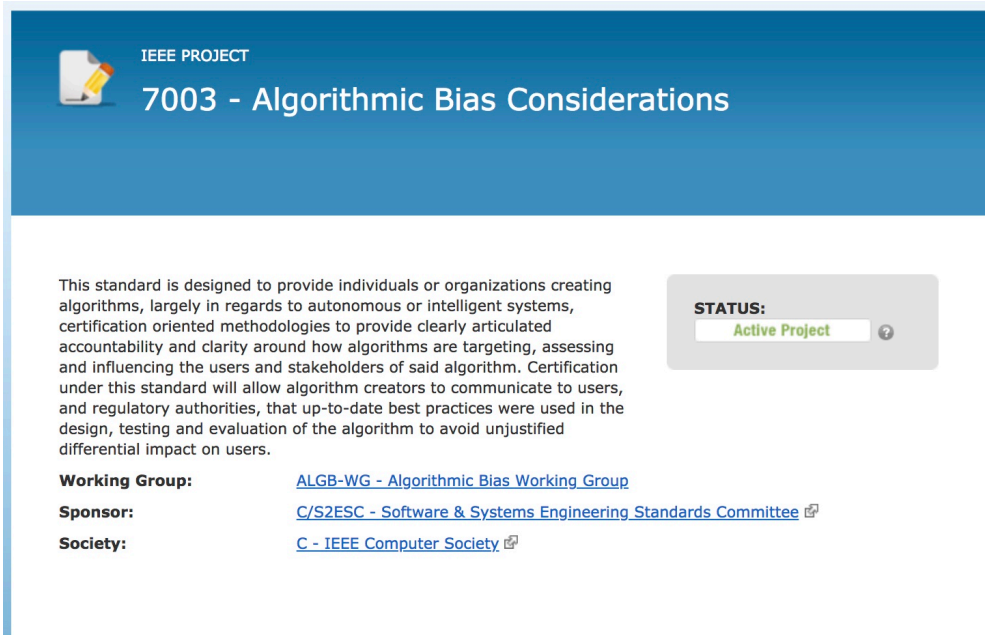
IEEE
Advancing Technology for Humanity

Executive Summary

To fully benefit from the potential of Artificial Intelligence and Autonomous Systems (AI/AS), we need to go beyond perception and beyond the search for more computational power or solving capabilities.

We need to make sure that these technologies are aligned to humans in terms of our moral values and ethical principles. AI/AS have to behave in a way that is beneficial to people beyond reaching functional goals and addressing technical problems. This will allow for an elevated level of trust between humans and our technology that is needed for a fruitful pervasive use of AI/AS in our daily lives.

IEEE 7003



IEEE PROJECT

7003 - Algorithmic Bias Considerations

This standard is designed to provide individuals or organizations creating algorithms, largely in regards to autonomous or intelligent systems, certification oriented methodologies to provide clearly articulated accountability and clarity around how algorithms are targeting, assessing and influencing the users and stakeholders of said algorithm. Certification under this standard will allow algorithm creators to communicate to users, and regulatory authorities, that up-to-date best practices were used in the design, testing and evaluation of the algorithm to avoid unjustified differential impact on users.

STATUS: Active Project

Working Group: ALGB-WG - Algorithmic Bias Working Group

Sponsor: C/S2ESC - Software & Systems Engineering Standards Committee

Society: C - IEEE Computer Society

Principles for Algorithmic Transparency and Accountability

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

7. Validation and Testing: Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

RESEARCH

4.

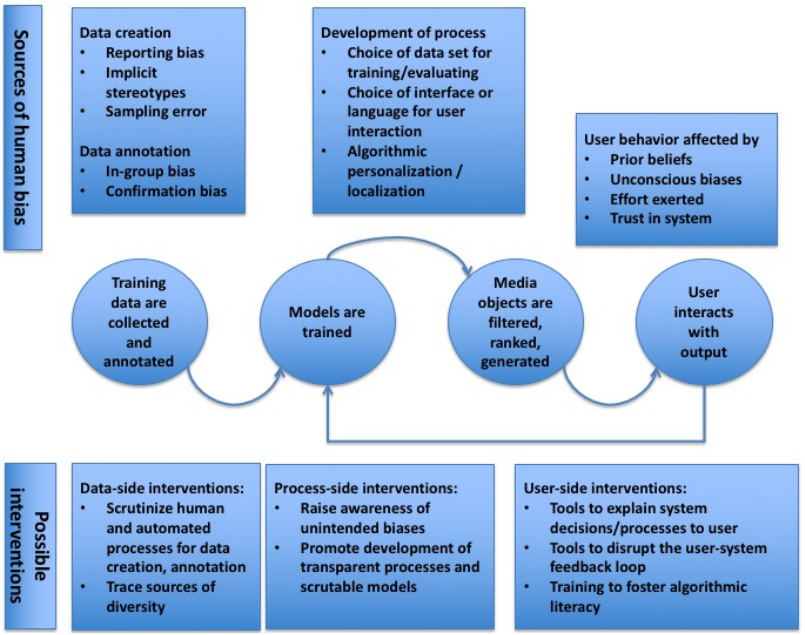
FOCUS AND APPROACH

TAG & CyCAT focus on **understanding the nature and impact of human biases** in algorithmic systems, and **develop tools and techniques** to promote algorithmic transparency.

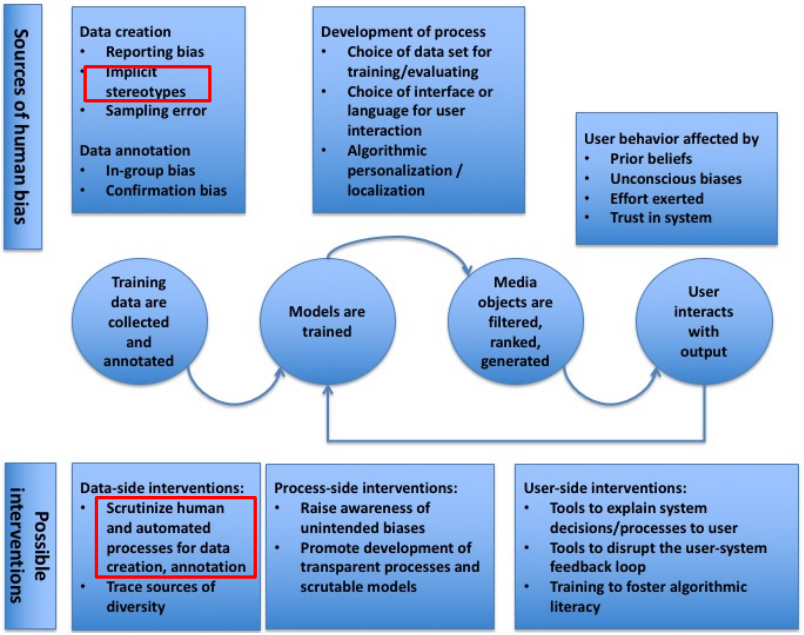
We use both **data science** and **social science** approaches to examine the impact of human biases as well as to evaluate possible interventions.

- Data science approaches are used to conduct tests on the outputs of APIs that are popular with third-party developers, to ascertain their tendencies to reproduce social biases.
- Social science methods (e.g., controlled experiments) are used to understand how the user's own biases are influenced by the design parameters of the feedback loop.

PIPELINE AND OPPORTUNITIES FOR INTERVENTION



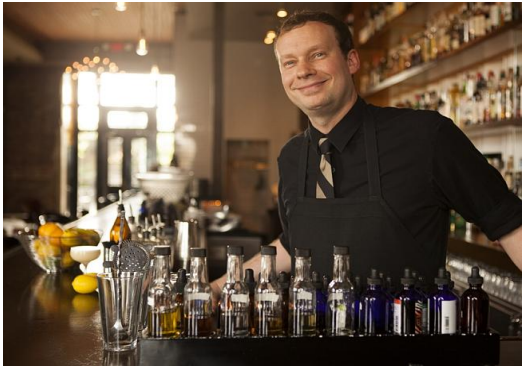
STUDY 1: BIASES IN TRAINING DATA



Otterbacher, J. (2018, July). Social Cues, Social Biases: Stereotypes in Annotations on People Images. In Proceedings of AAAI International Conference on Human Computation and Crowdsourcing (HCOMP).

QUALITY OF DESCRIPTIVE METADATA

Man
Bar
Drinks
Bottles
Bartender
Smiling
Happy
Leisure



Computer Vision

Information Retrieval

Content Moderation

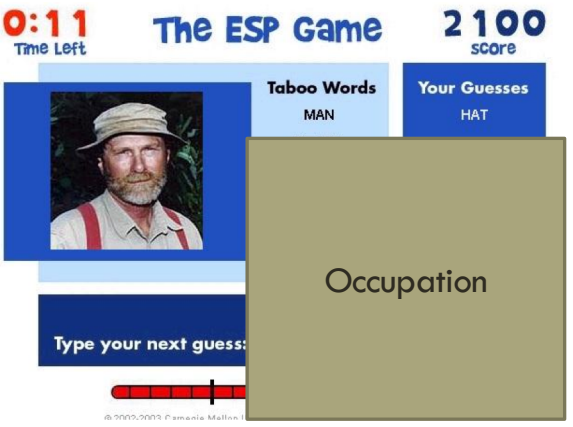
BIAS IN CROWDSOURCED METADATA?

LINGUISTIC BIAS IN IMAGE METADATA

A systematic asymmetry in the way one uses language, as a function of the social group of the person(s) being described. [Beukeboom, 2013]

- Two linguistic patterns that reveal expectations about others:
- -use of abstract vs. concrete words
- -use of subjective words

LINGUISTIC BIAS IN IMAGE METADATA



LINGUISTIC EXPECTANCY BIAS (LEB) [MAASS ET AL., 1989]

Less Expected
More concrete language

Doctor
Surgeon
Intelligent
Serious

Nurse
Experiment
Smiley
Hat

Nurse
Student
Studying
Listening

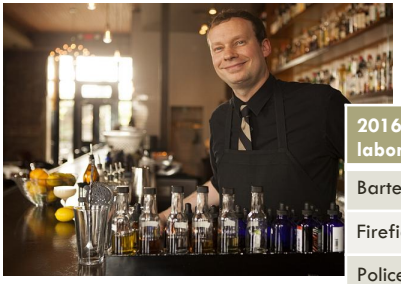
More Expected
More abstract / interpretive language

LINGUISTIC IN-GROUP BIAS (LIB)

[MAASS ET AL., 1989]

- Builds on the LEB
- We expect positive attributes and actions from our in-group members
  - Positive observations → more abstract, subjective
- Caveat:
  Linguistic biases occur when communication has a clear purpose
  [Semin et al., 2003]

RQ1:DO WE OBSERVE LEB/LIB IN CROWDSOURCED DESCRIPTIONS OF PEOPLE IMAGES?



2016 U.S. labor statistics	%Women	%Black
Bartender	56.1	7.4
Firefighter	3.5	6.8
Police officer	14.1	12.0

RQ2: DOES THE PRESENCE OF SOCIAL INFORMATION AFFECT THIS PROCESS?

How to play:
* Enter your description in the box below
* Hit enter or submit when done

Popular tags for this image:
* Strong
* Clever
* Smile

Describe the image as accurately as you can in your own words:

HYPOTHESES

- **Linguistic Expectancy Bias**
  $H1_a$: White professionals will be described more abstractly than blacks
  $H1_b$: Men will be described more abstractly than women, with the exception of bartenders
- **Linguistic In-group Bias**
  $H2_a$: White men describe other white men more abstractly than other groups
  $H2_b$: White women describe white women more abstractly than other groups
- **Communication constraints**
  H3: Biases are more frequently observed in cases when social cues are provided to workers (e.g., "popular tags")

PROCEDURE

- Recruited U.S.-based workers through Amazon Mechanical Turk
- Between-subjects design
- Four HITs per image
  (2 social cues settings x 2 worker genders)

ANALYZING DESCRIPTIONS

HIT

Attractive barista pouring a martini

Linguistic Inquiry and Wordcount (quantitative)

Wordcount: 5
Sixletter: 0.80
Subjective: 0.20
Positive: 0.20
Negative: 0

Manual (categorical/binary)

Appearance: Yes
Character/mood: No
Judgment: Yes

TESTING FOR LEB

- 3 independent variables, indications of abstractness in people-descriptions
  - Subjective words (ANOVA + Tukey HSD test)
  - Mentioning character/mood (logit models)
  - Making judgments (logit models)
- 3 explanatory variables
  - Worker's gender (G)
  - Gender of depicted person (ImG)
  - Race of depicted person (ImR)

LEB – USE OF SUBJECTIVE WORDS

	Gender -worker	Gender-depicted	Race-depicted	G* ImG	G*ImR	ImG*ImR	G*ImG*ImR	Sig. Main Effects
Bartender - Control			+					ImR: White > Black
Bartender – Social	+	+	+			+		G: Women > Men ImG: Men > Women ImR: White > Black
Firefighter - Control								
Firefighter - Social	+	+						G: Women > Men ImG: Men > Women
Police - Control								
Police - Social	+							G: Women > Men

LEB – REFERENCES TO CHARACTER/MOOD

	Gender-worker	Gender-depicted	Race-depicted	G* ImG	G*ImR	ImG*ImR	G*ImG*ImR	Sig. Main Effects
Bartender - Control								
Bartender – Social		+	+	+				ImG: Men > Women ImR: White > Black
Firefighter - Control								
Firefighter - Social		+	+					ImG: Men > Women ImR: White > Black
Police - Control								
Police - Social					+			

TESTING FOR LIB

- Separate observations into two groups:
  - Descriptions for in-group members (WM,WM) (WW,WW)
  - Descriptions for others
- 3 independent variables, indications of abstractness in people-descriptions:
  - Subjective words (two-sample t-test)
  - Mentioning character/mood
    (test for equality of proportions)
  - Making judgments (test for equality of proportions)

LIB – DESCRIBING IN-GROUP VS. OTHERS

Worker gender – Setting	Use of subjective words	Mentioning character/mood	Passing judgment
Men – Control	No *(t = -0.67, p>.05)*	No *(χ^2 =0.26, p>.05)*	No *(χ^2 =3,59, p>.05)*
Men – Social cues	Yes *(t = 3.69, p<.001)*	No *(χ^2 =1.33, p>.05)*	Yes *(χ^2 =17.6, p<.001)*
Women – Control	No *(t = -0.07, p>.05)*	No *(χ^2 =0.20, p>.05)*	No *(χ^2 =0.01, p>.05)*
Women – Social cues	No *(t =1.10, p>.05)*	No *(χ^2 =0.22, p>.05)*	No *(χ^2 =0.28, p>.05)*

IMPLICATIONS

- Free-text annotation of images is fundamentally a communication process
  - Linguistic biases are population-wide
- Design of the HIT
  - Even simple social cues can easily sway workers' responses
- Identity of workers
  - Women used more subjective words
  - LIB was observed only in descriptions written by men

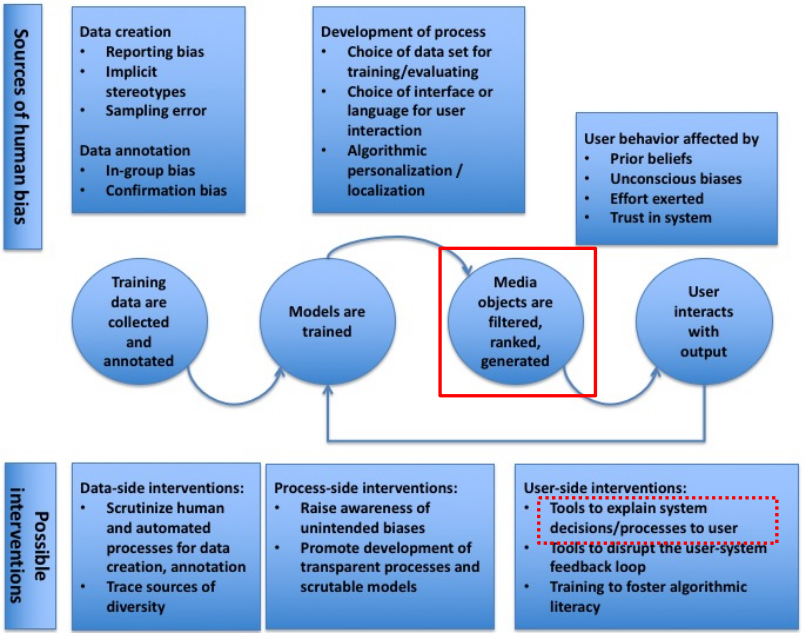
usacm ACM US Public Policy Council

acm Europe Council

Principles for Algorithmic Transparency and Accountability

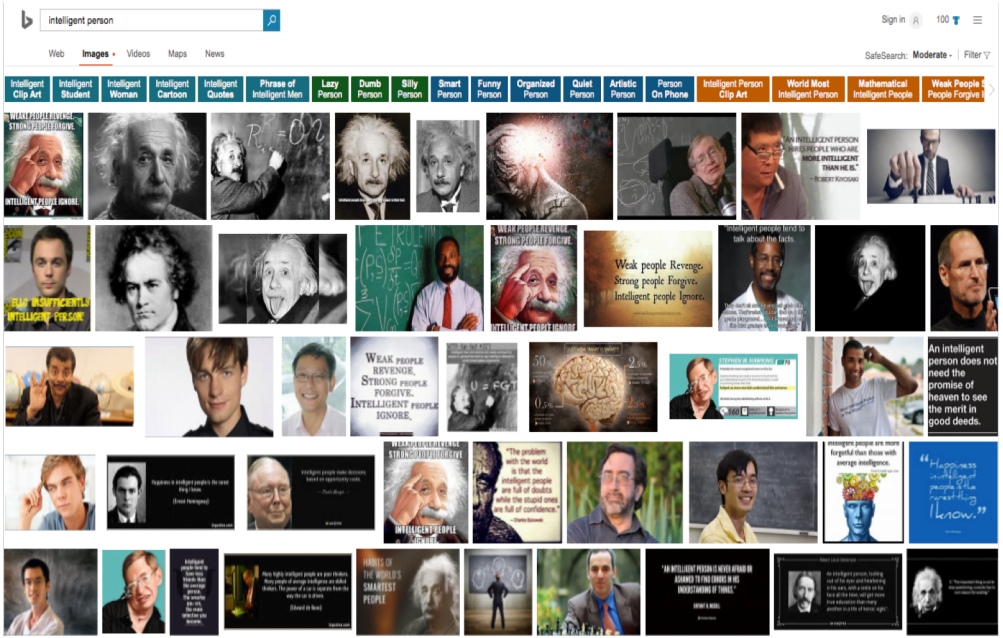
5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

STUDY 2: BIASES IN SYSTEM OUTPUT

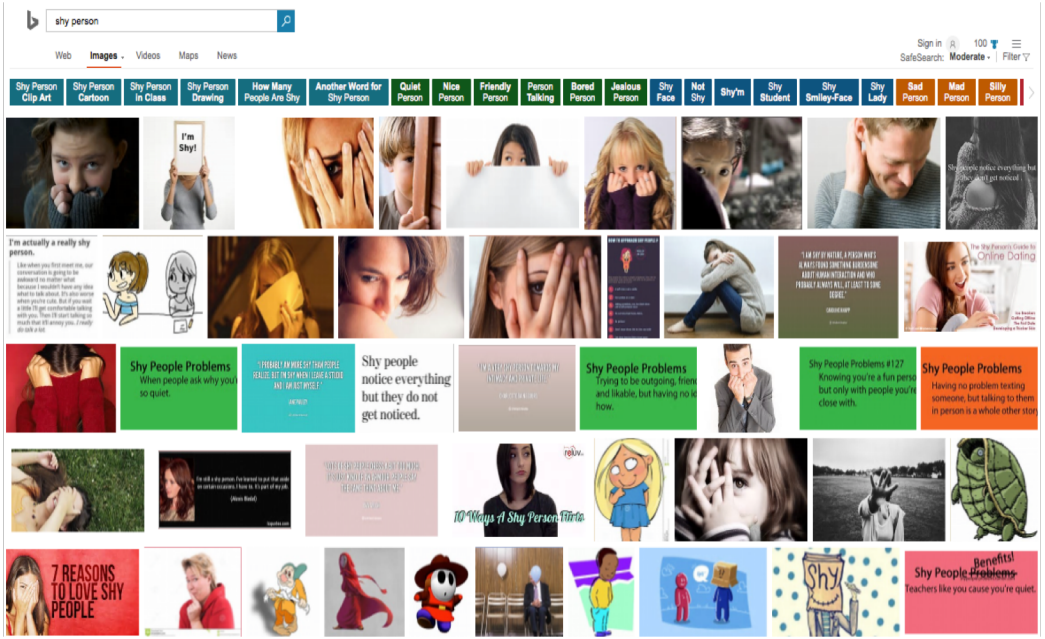


Otterbacher, J., Bates, J., & Clough, P. (2017, May). Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 6620-6631). New York: ACM Press.

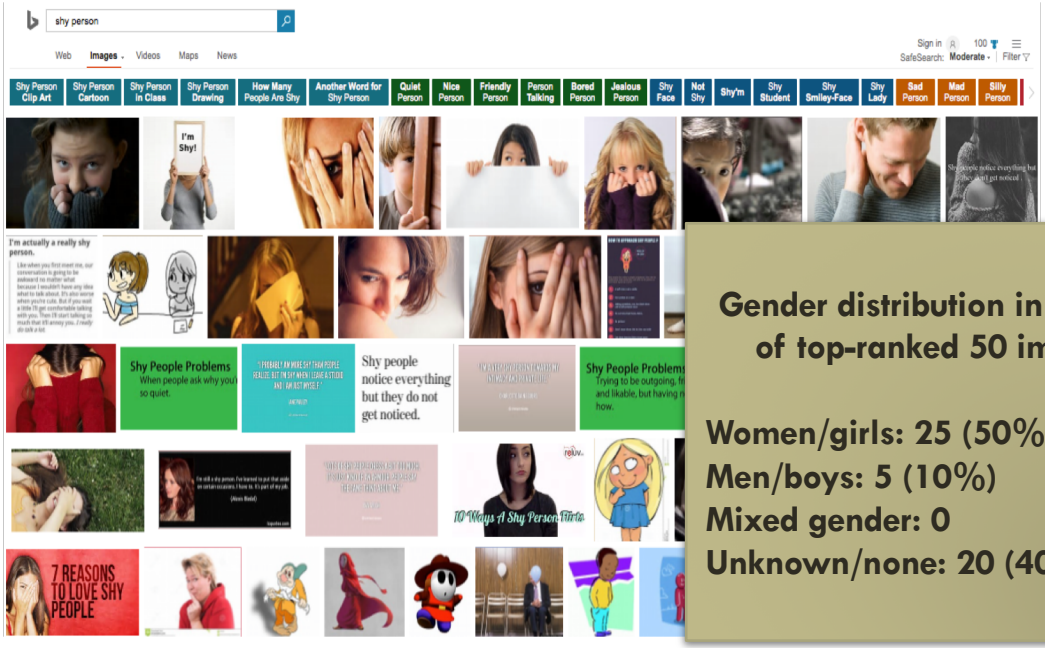
INTELLIGENT PERSON



SHY PERSON


shy person
Web
Images
Videos
Maps
News
Sign in
SafeSearch: Moderate
Filter
Shy Person Clip Art
Shy Person Cartoon
Shy Person In Class
Shy Person Drawing
How Many People Are Shy
Another Word for Shy Person
Quiet Person
Nice Person
Friendly Person
Person Talking
Bored Person
Jealous Person
Shy Face
Not Shy
Shy'm
Shy Student
Shy Smiley-Face
Shy Lady
Sad Person
Mad Person
Silly Person
I'm Shy!
I'm actually a really shy person.
Shy People Problems
When people ask why you so quiet.
Shy people notice everything but they do not get noticed.
Shy People Problems
Trying to be outgoing, friend and likable, but having no id how.
Shy People Problems #127
Knowing you're a fun perso but only with people you're close with.
Shy People Problems
Having no problem texting someone, but talking to them in person is a whole other sto
10 Ways A Shy Person Flirts
7 REASONS TO LOVE SHY PEOPLE
Shy People Problems
Benefits!
Teachers like you cause you're quiet.
Shy

SHY PERSON



Gender distribution in images of top-ranked 50 images

Women/girls: 25 (50%)
Men/boys: 5 (10%)
Mixed gender: 0
Unknown/none: 20 (40%)

STEREOTYPE CONTENT: "BIG TWO" OF PERSON PERCEPTION

- Our perceptions of others are based on two dimensions [Fiske et al., 2002]

(1) Agency (or competence): whether or not we perceive someone as being capable of achieving his/her goals

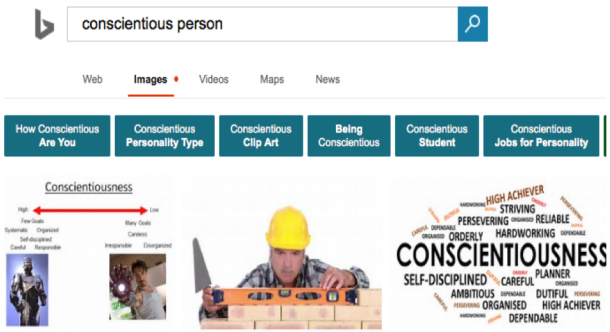
(2) Warmth (or communality): whether or not we think someone has pro-social intentions or is a threat to us

- Stereotypes are captured by combinations of the two dimensions [Cuddy et al., 2008]
  - Women: [low agency, high warmth]
  - Men: [high agency, low warmth]

TRAIT ADJECTIVE CHECKLIST METHOD

- Used in the *Princeton Trilogy* studies of ethnic and racial stereotypes [Katz & Braly, 1933]
- Participants describe target social groups using list of trait adjectives
- 68 traits developed in cross-lingual study across five countries [Abele et al., 2008]

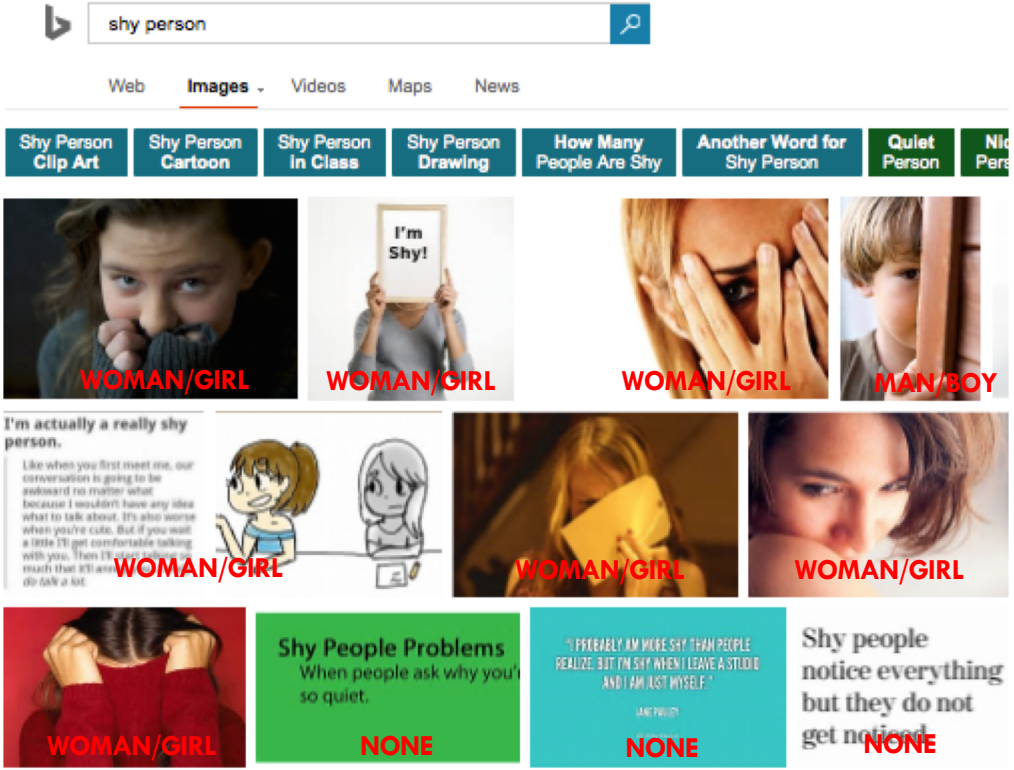
able
active
affectionate
altruistic
ambitious
assertive
boastful
capable
caring
chaotic
communicative
competent
competitive
conceited
conscientious
considerate
consistent
creative
decisive
detached
determined
dogmatic
dominant
egoistic
emotional
energetic
expressive
fair
friendly
gullible
harmonious
hardhearted
helpful
honest
independent
industrious
insecure
intelligent
lazy
loyal
moral
obstinate
open
open-minded
outgoing
perfectionistic
persistent
polite
rational
reliable
reserved
self-confident
self-critical
self-reliant
self-sacrificing
sensitive
shy
sociable
striving
strong-minded
supportive
sympathetic
tolerant
trustworthy
understanding
vigorous
vulnerable
warm



Search markets:
UK-EN
US-EN
IN-EN
ZA-EN

RESEARCH QUESTIONS

- **RQ1: Baseline Representation bias**
  - In a search for "person" which genders are depicted?
- **RQ2: Stereotype content and strength**
  - Which character traits are most often associated with which genders?
  - Are these associations consistent across Bing search markets? (UK, US, IN, ZA)
- **RQ3: Backlash effects**
  - How are stereotype-incongruent individuals depicted?


shy person
Web
Images
Videos
Maps
News
Shy Person Clip Art
Shy Person Cartoon
Shy Person In Class
Shy Person Drawing
How Many People Are Shy
Another Word for Shy Person
Quiet Person
I'm Shy!
I'm actually a really shy person.
Like when you first meet me, our conversation is going to be awkward no matter what because I wouldn't have any idea what to talk about. It's also worse when you're cute. But if you wait a little I'll get comfortable talking with you. Then I'll
do talk a lot.
Shy People Problems
When people ask why you'
so quiet.
Shy people notice everything but they do not get noticed
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
MAN/BOY
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
NONE
NONE
NONE

PILOT STUDY ON CROWDFLOWER

- 1.000 “person” images from UK market
- 3 annotators per image
- Is the image: 1) a photograph, 2) a sketch/illustration, 3) some other type?
- Does the image depict: 1) only women/girls, 2) only men/boys, 3) mixed gender group, 4) gender ambiguous person(s), 5) no person(s)?

CLASSIFYING IMAGE TYPE

	# Images	Inter-judge agreement
Photos	576	0.97
Sketches	346	0.96
Other	22	0.74
No longer accessible	56	1.00

CLASSIFYING GENDER

	Women/ girls	Men/ boys	Mixed gender	Unknown	No persons	Inter-judge agreement
Photos	0.27	0.55	0.10	0.07	0.01	0.94
Sketches	0.08	0.28	0.05	0.55	0.04	0.91

AUTOMATING GENDER RECOGNITION

- Clarifai API
  - General image recognition tool
  - Coverage: 95%
  - Provides 20 textual concept tags
- Linguistic Inquiry and Wordcount (LIWC)
  [Pennebaker et al., 2015]
  - Female references: mom, girl
  - Male references: dad, boy

Gather images

Analyze images

Query "person"

Query "X person"

68 character traits ("X"): polite, capable, honest…

Microsoft

Bing Image Search API

"person"

"X person"

Gather top 1,000 images for UK, US, IN and ZA market settings

Gather images

Analyze images

Image recognition to identify concepts (tags)

Filter out photos with "portrait" tag

Identify gender(s) based on tag analysis

clarifai

LIWC (man, woman other)

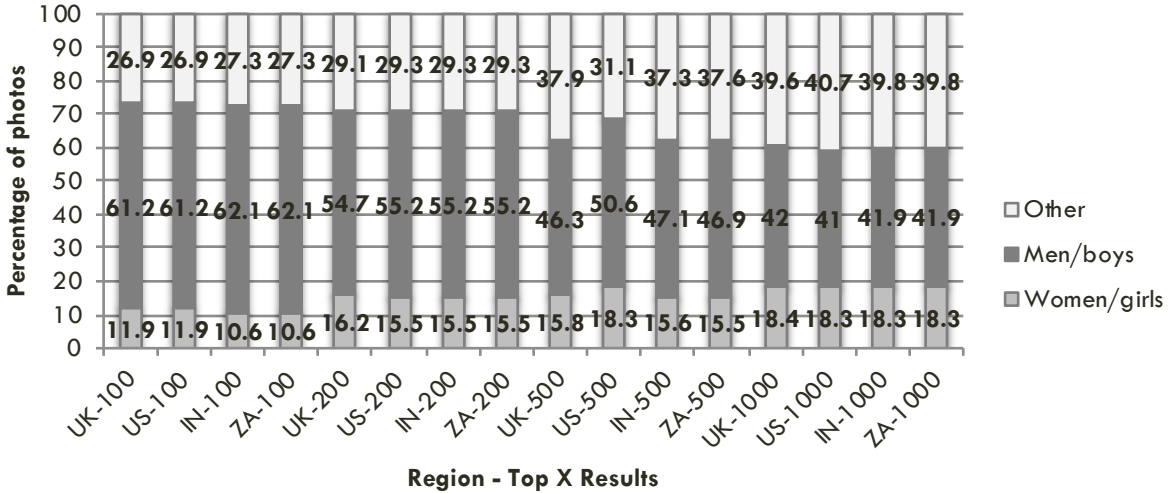
Person, man, famous, event, entertainment, talent, pop, fame, portrait, adult, one, serious, dark, guy, face, lid, human, young

MAN

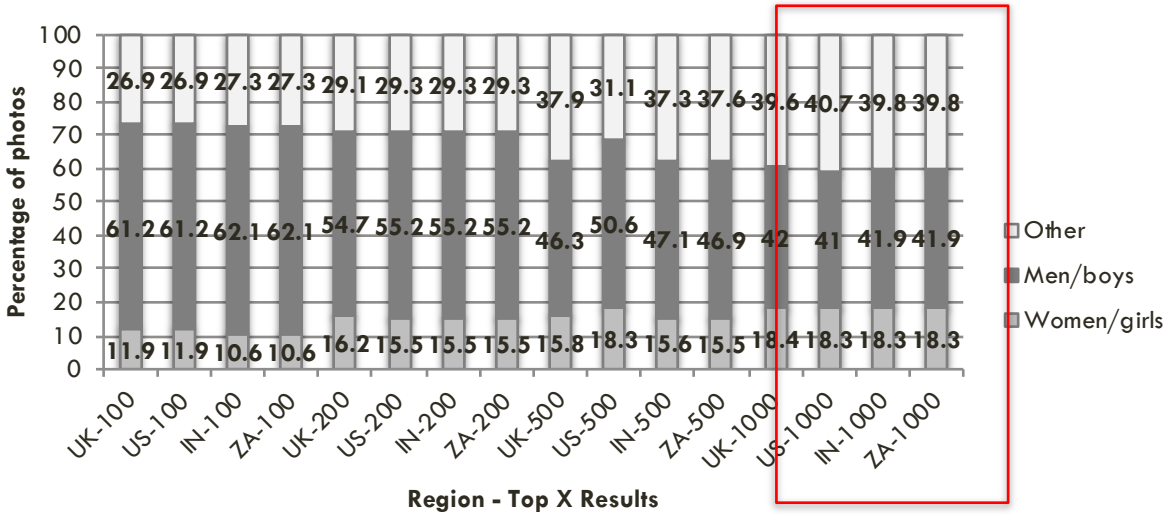
PERFORMANCE ON GENDER CLASSIFICATION

	N	Precision	Recall	F_1
Recognizing photographs	473	0.91	0.75	0.822
Women/girls	130	0.89	0.60	0.717
Men/boys	282	0.95	0.67	0.786
Other	61	0.68	0.82	0.743

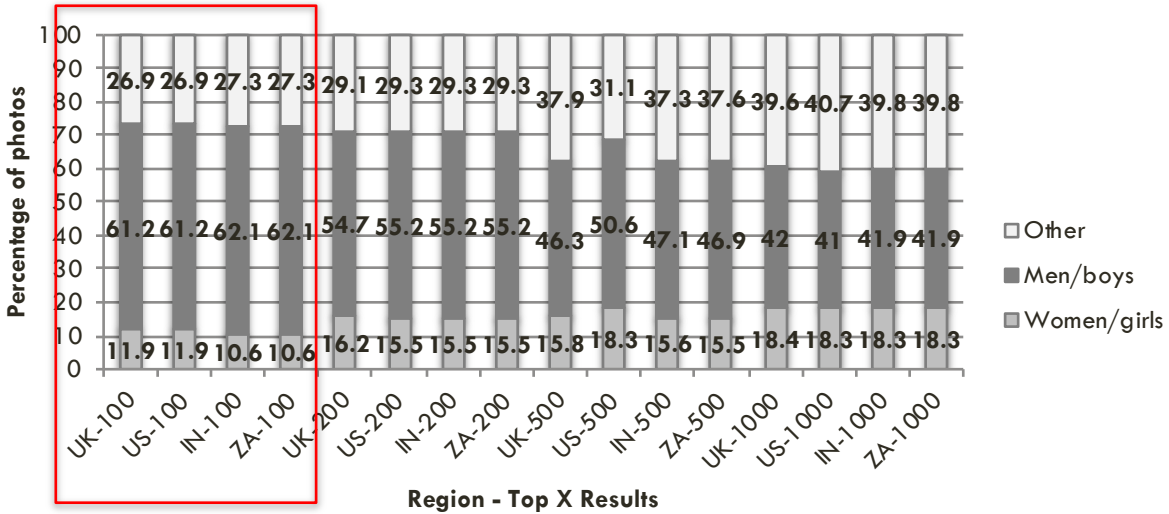
RQ1: WHO REPRESENTS A "PERSON"?



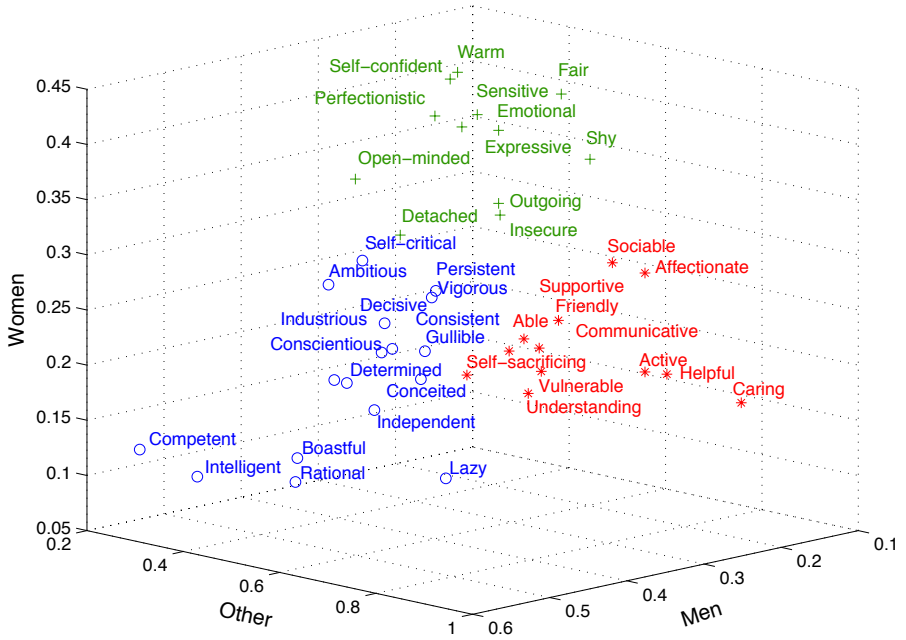
RQ1: WHO REPRESENTS A "PERSON"?



RQ1: WHO REPRESENTS A "PERSON"?



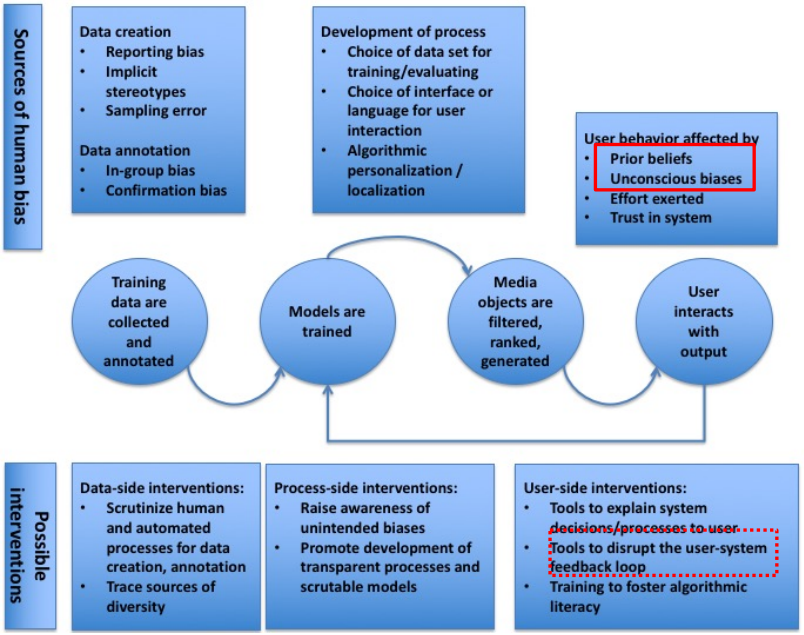
RQ2: WHICH TRAITS ARE GENDERED? (UK)



GENDERING OF TRAITS ACROSS ALL FOUR REGIONS

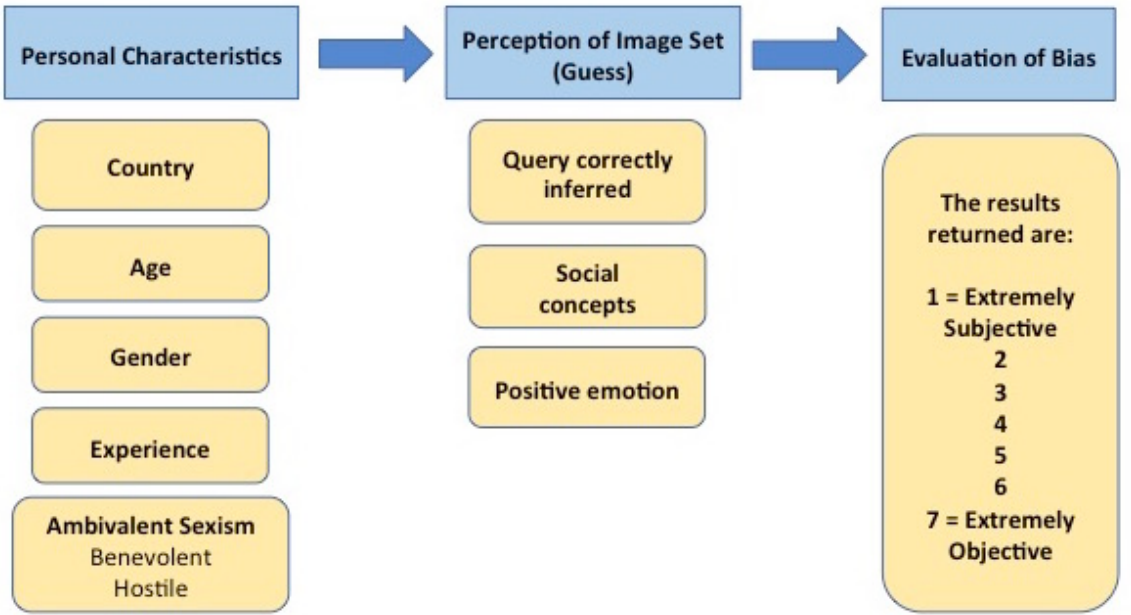
Men/boys: ambitious, boastful, competent, conceited, conscientious, consistent, decisive, determined, gullible, independent, industrious, intelligent, lazy, persistent, rational, self-critical, vigorous
Women/girls: detached, emotional, expressive, fair, insecure, open-minded, outgoing, perfectionistic, self-confident, sensitive, shy, warm
Gender-neutral: able, active, affectionate, caring, communicative, competitive, friendly, helpful, self-sacrificing, sociable, supportive, understanding, vulnerable

STUDY 3: USER PERCEPTIONS OF BIAS



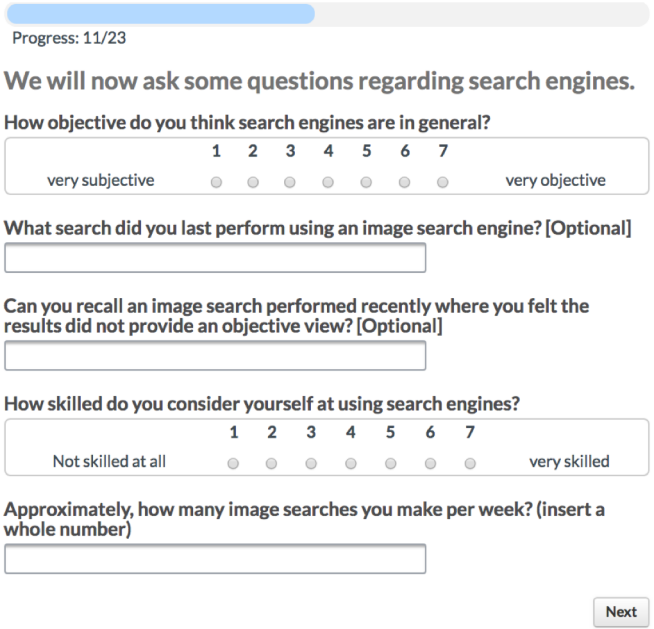
Otterbacher, J., Checco, A., Demartini, G. & Clough, P. (2018, July) Investigating User Perception of Bias in Image Search: The Role of Sexism. In the Proceedings of the 2018 ACM SIGIR Conference, Ann Arbor MI USA. New York: ACM Press.

CONCEPTUAL MODEL



EXPERIMENT ON FIGURE EIGHT: STEP 1

STEP 2


Progress: 11/23
We will now ask some questions regarding search engines.
How objective do you think search engines are in general?
1 2 3 4 5 6 7
very subjective
very objective
What search did you last perform using an image search engine? [Optional]
Can you recall an image search performed recently where you felt the results did not provide an objective view? [Optional]
How skilled do you consider yourself at using search engines?
1 2 3 4 5 6 7
Not skilled at all
very skilled
Approximately, how many image searches you make per week? (insert a whole number)
Next

STEP 3

Progress: 12/23

How objective do you think the results of the search engine are for the query 'working person'?

	1	2	3	4	5	6	7	
very subjective	○	○	○	○	○	○	○	very objective

Can you explain why in your words?

STEP 4

We obtained these images from a search engine by using the query: **working person** , while you described the images with the keywords: **Working hard** . Please answer to some questions about the difference between the two keywords

How similar are the keywords "working person" and "Working hard"?

	1	2	3	4	5	6	7	
very different	○	○	○	○	○	○	○	very similar

In which characteristic(s) the two keywords (" working person" and " Working hard") differ? For example 'they refer to different topics', 'one is more specific than the other'...
(please only focus on the difference between the two texts, **don't consider the images here.**)

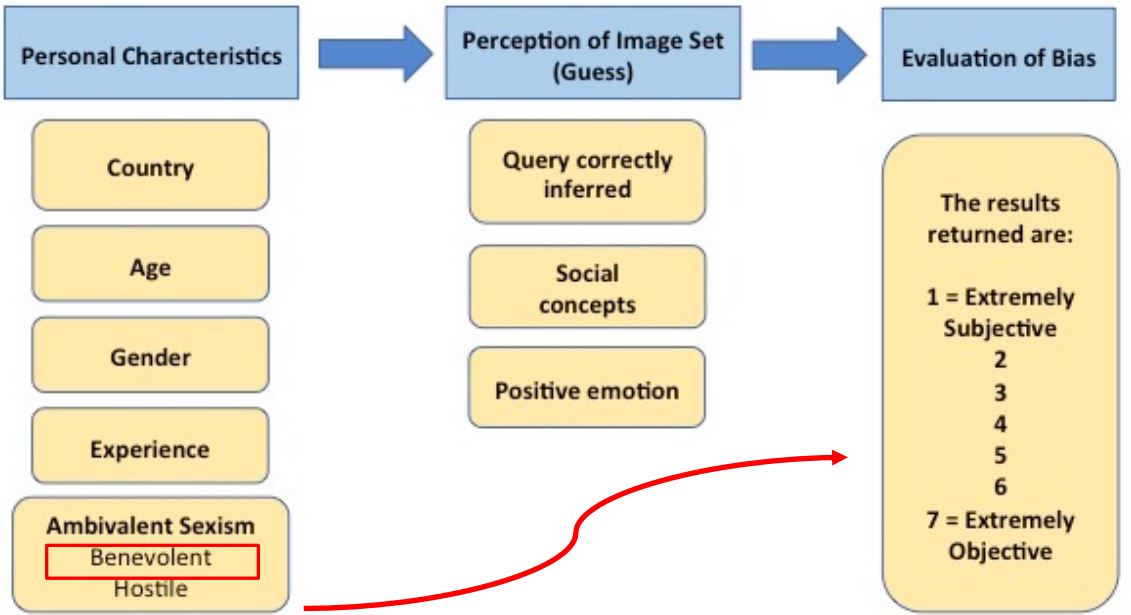
Which keywords better describe the above search engine results?

- ○ Your keyword "Working hard"
- ○ Search engine query "working person"
- ○ The two keywords are completely identical

Can you explain why in your words?

Next

FINDINGS


Personal Characteristics
Perception of Image Set (Guess)
Evaluation of Bias
Country
Age
Gender
Experience
Ambivalent Sexism
Benevolent
Hostile
Query correctly inferred
Social concepts
Positive emotion
The results returned are:
1 = Extremely Subjective
2
3
4
5
6
7 = Extremely Objective

DISCUSSION

6.

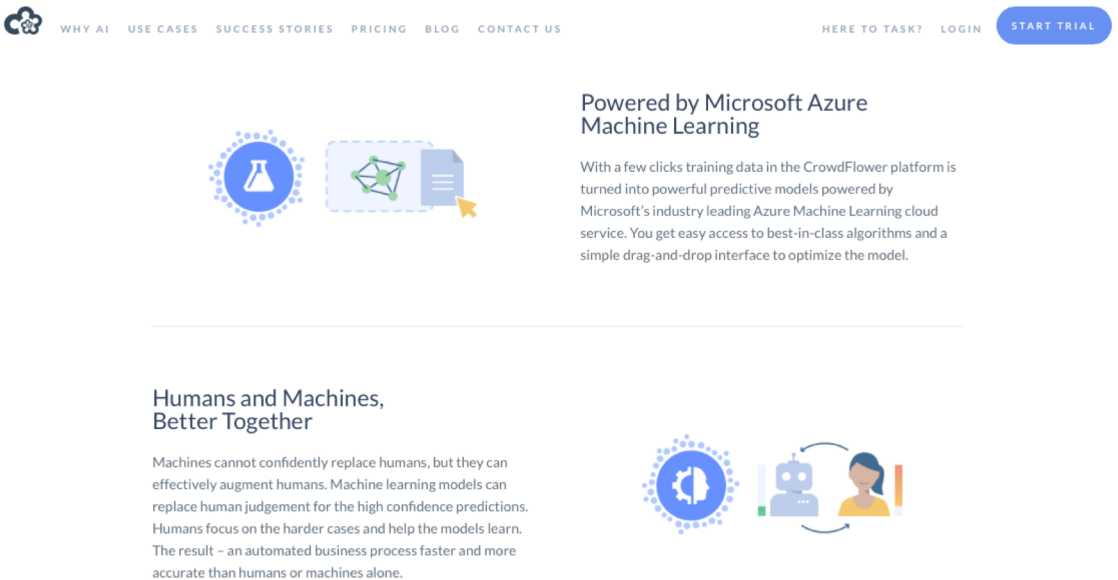
HMIS

…are being used "in the realms of data management, information retrieval, natural language processing, semantic web, machine learning, and multimedia to better solve existing problems."

In addition to solving highly technical problems, HMIS "…need to deal with the full spectrum of challenges from the social science standpoint."

Demartini, G., Difallah, D.E., Gadiraju, U. and Catasta, M. (2017). An Introduction to Hybrid Human-Machine Information Systems.
Foundations and Trends in Web Science, 7, 1, pp. 1-87.

CROWDSOURCING PLATFORMS & COGNITIVE SERVICES: EASY DATA COLLECTION AND ML



PHILOSOPHICAL CONSIDERATIONS

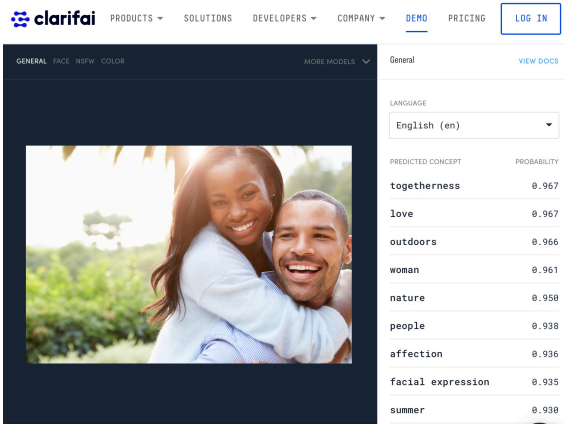
http://www.entities-r-us.com/comic/women-stereotypes/

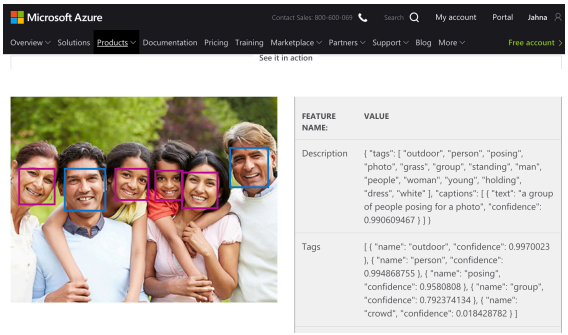
LAB: HOW DO COMPUTER VISION APIS "SEE" PEOPLE?

6.

THE GOOGLE PHOTO APP INCIDENT…

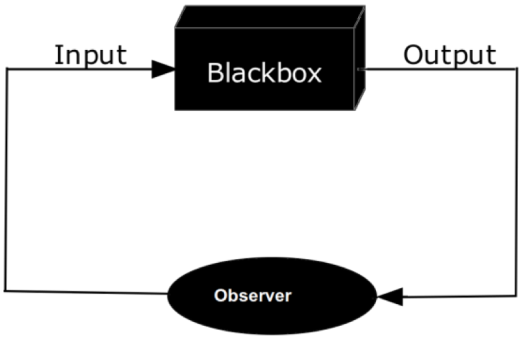
COGNITIVE SERVICES: VISION APIS





			
Clarifai API	Woman, Afro, dreadlock, cute	Man, casual, cool, friendly	Face, man, casual, eye
Microsoft Vision API	Hairpiece, clothing, wear, smile	Person, necktie, wearing, shirt	Man, looking, shirt, wearing
Watson Visual Recognition API	Person, woman, female	Stubble, coonskin cap, afro, hairstyle	Person, pompadour hairstyle, skin
Imagga Image Understanding API	Afro, attractive, pretty, model	Man, face, male, person, creation	Person, face, man, male, handsome

CONTROLLED EXPERIMENT



Chicago Face Database (CFD)
[Ma et al., 2015]

- 597 people images
  - High resolution
  - White background
  - Neutral expression
  - Same clothing
- Subjective norming data
- Objective facial measurements

CHICAGO FACE DATABASE (CFD) [MA ET AL., 2015]

Subjective norming data based on 30+ human judges' responses:

"Consider the person pictured above and rate him/her with respect to other people of the same race and gender. For example, if the person was Asian and male, consider this person on the following traits relative to other Asian males in the United States. - **Attractive** (1-7 Likert, 1 = Not at all; 7 = Extremely)".

A total of 15 additional traits were evaluated, including: Babyface, Dominant, Trustworthy, Feminine, and Masculine.

Objective facial measurements

- Luminance
- Nose width, length, shape
- Lip fullness
- Eye height, width, shape, size
- Chin length

CHICAGO FACE DATABASE (CFD) [MA ET AL., 2015]

	Asian	Black	Latino/a	White
Women	N=57 3.64 / 3.62	N=104 3.33 / 3.15	N=56 3.81 / 3.56	N=90 3.45 / 3.39
Men	N=52 2.85 / 2.85	N=93 3.17 / 3.12	N=52 2.94 / 2.90	N=93 2.96 / 2.96

Mean/median attractiveness by depicted individual's race and gender

CORPUS OF DESCRIPTIVE TAGS

	Clarifai	Microsoft	Watson	Imagga
Total tags	11,940	12,137	3,668	6,772
Unique tags	95	74	72	54
10 most frequent tags	Portrait, one, people, isolated, casual, looking, look, eye, man, face	White, shirt, wearing, standing, posing, young, black, smiling, glasses, looking	Person, light brown color, people, ash grey, coal black, face, stubble, adult person, actor, woman	Portrait, face, person, handsome, man, male, beard, adult, attractive, model

SOCIAL BIAS IN COMPUTER VISION APIS?

1. Results are slanted in *unfair discrimination* against particular persons or groups
2. That discrimination is *systematic*

[Friedman & Nissenbaum, 1996]

Ideas for consideration:

- Are certain social groups described more positively than others?
- Are certain social groups described as being more attractive than others?
- Are certain social groups systematically more prone to gender-inference error?

TOOLS

R Studio + additional libraries

Text manipulation

tidyr useful functions for "tidying up" messy data

dplyr more functions for manipulating data

stringr functions specificly for handling strings (e.g., regular expressions)

Statistical analysis

lsr contains some useful functions for statistical analyses and computing effect sizes

mass contains many basic statistics functions

QUESTIONS?
ON TO THE LAB…

OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

CY. center for algorithmic transparency

Jahna Otterbacher
jahna.otterbacher@ouc.ac.cy

Research Centre on Interactive Media Smart Systems and Emerging Technologies