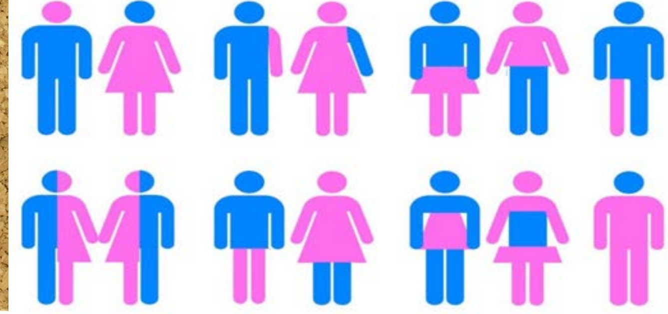



COMPETENT MEN AND WARM WOMEN:

ON THE DETECTION AND ORIGIN OF GENDER STEREOTYPED IMAGE SEARCH RESULTS

JAHNA OTTERBACHER

OPEN UNIVERSITY OF CYPRUS &
RESEARCH CENTRE ON INTERACTIVE MEDIA SMART SYSTEMS AND EMERGING TECHNOLOGIES

NICOSIA, CYPRUS

OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

CY. center for algorithmic transparency

Research Centre on Interactive Media Smart Systems and Emerging Technologies

OVERVIEW

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

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.

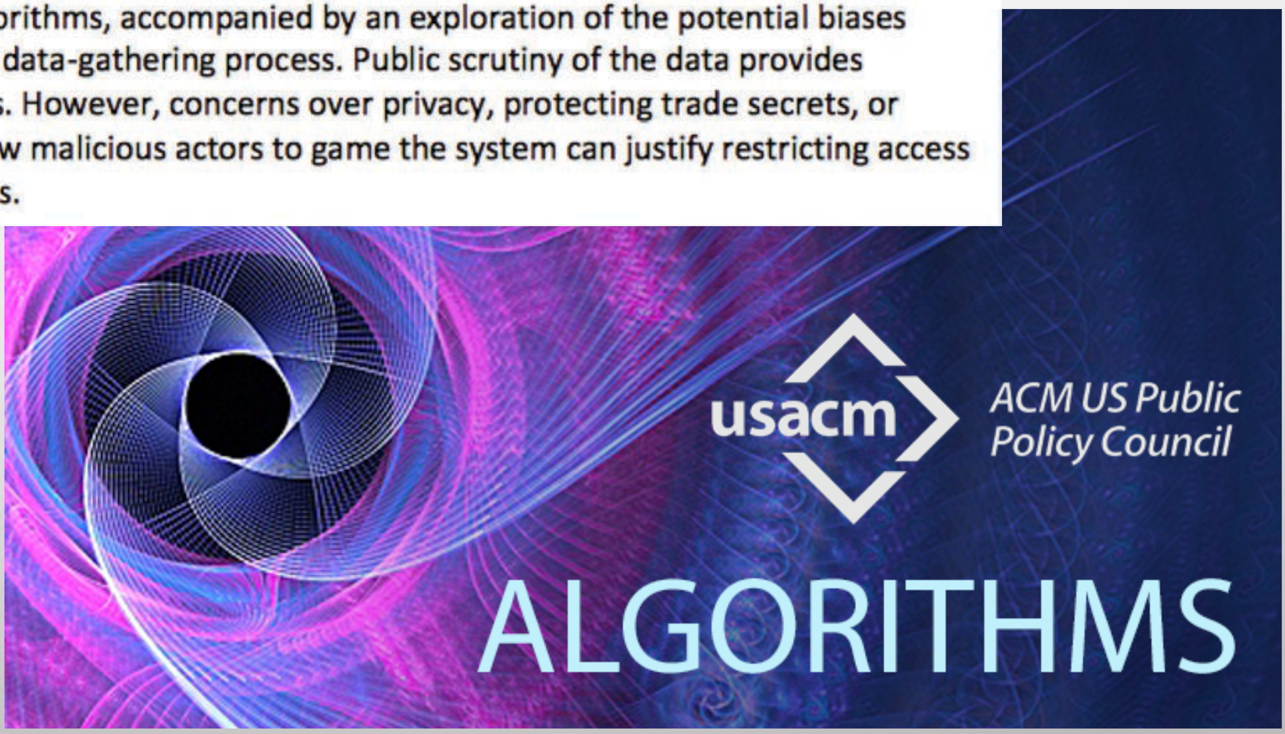
usacm ACM US Public Policy Council

acm Europe Council

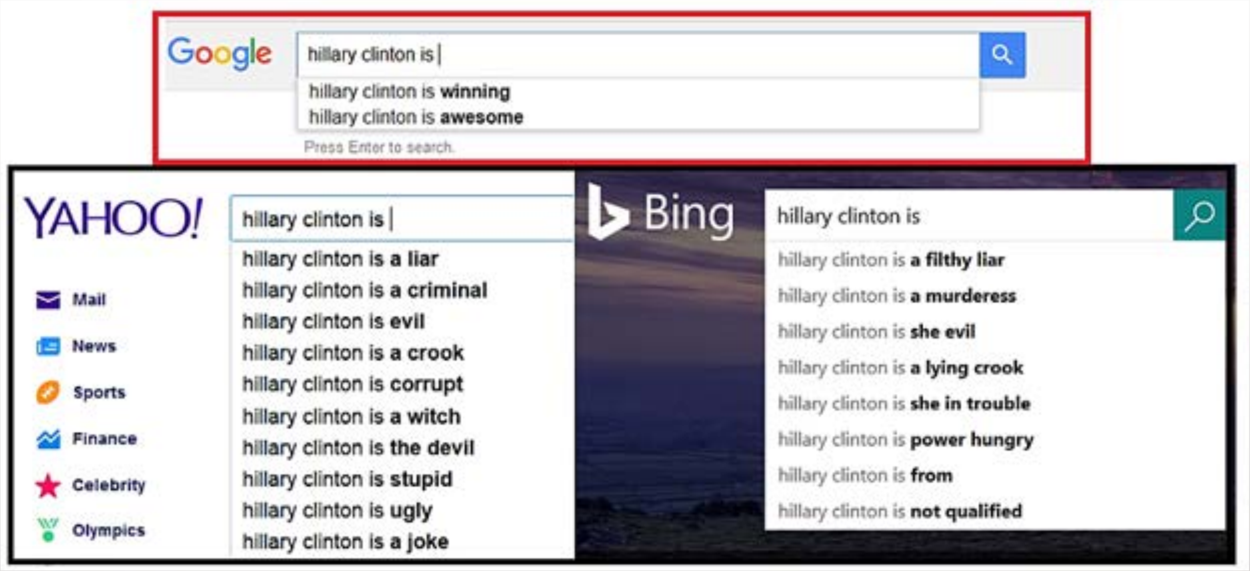
Principles for Algorithmic Transparency and Accountability

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.



ALL SYSTEMS HAVE A *SLANT*



http://ipullrank.com/dr-epstein-you-dont-understand-how-search-engines-work/

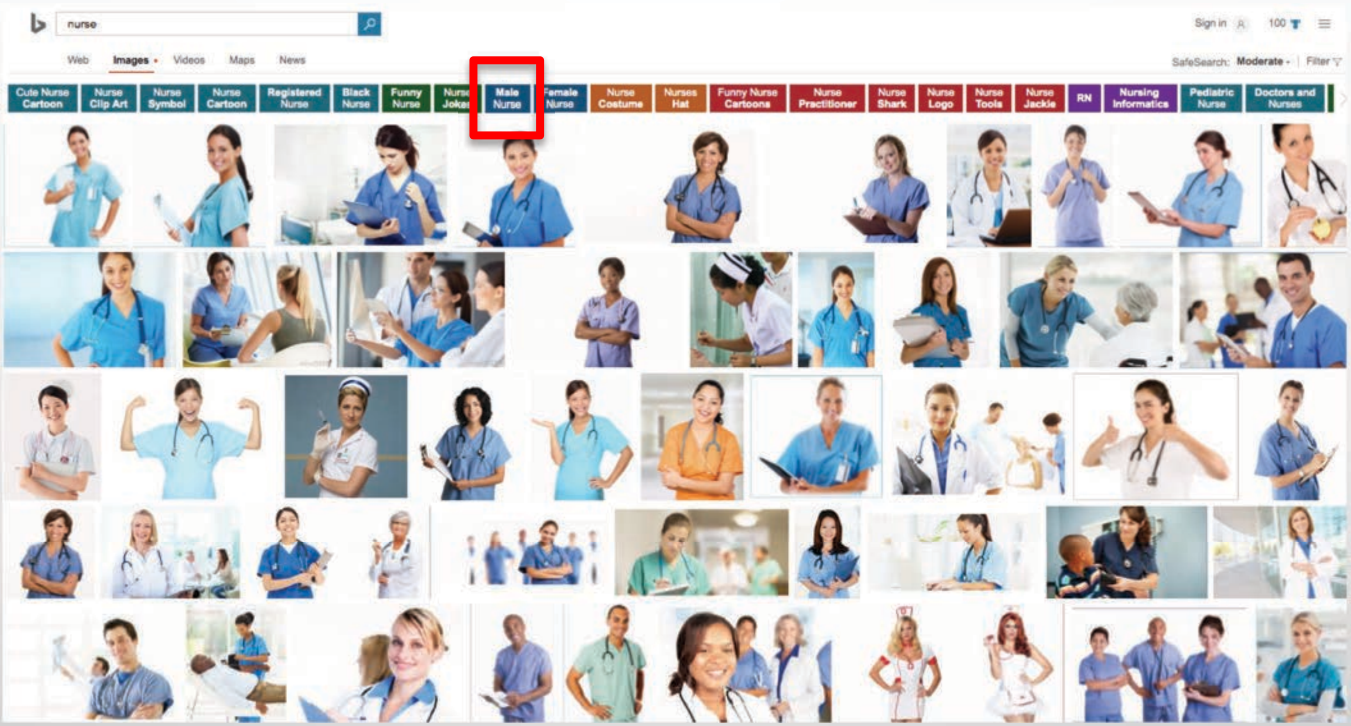
BUT WHAT EXACTLY IS **BIAS**?

1. RESULTS ARE SLANTED IN *UNFAIR DISCRIMINATION* AGAINST PARTICULAR PERSONS OR GROUPS
2. THAT DISCRIMINATION IS *SYSTEMATIC*

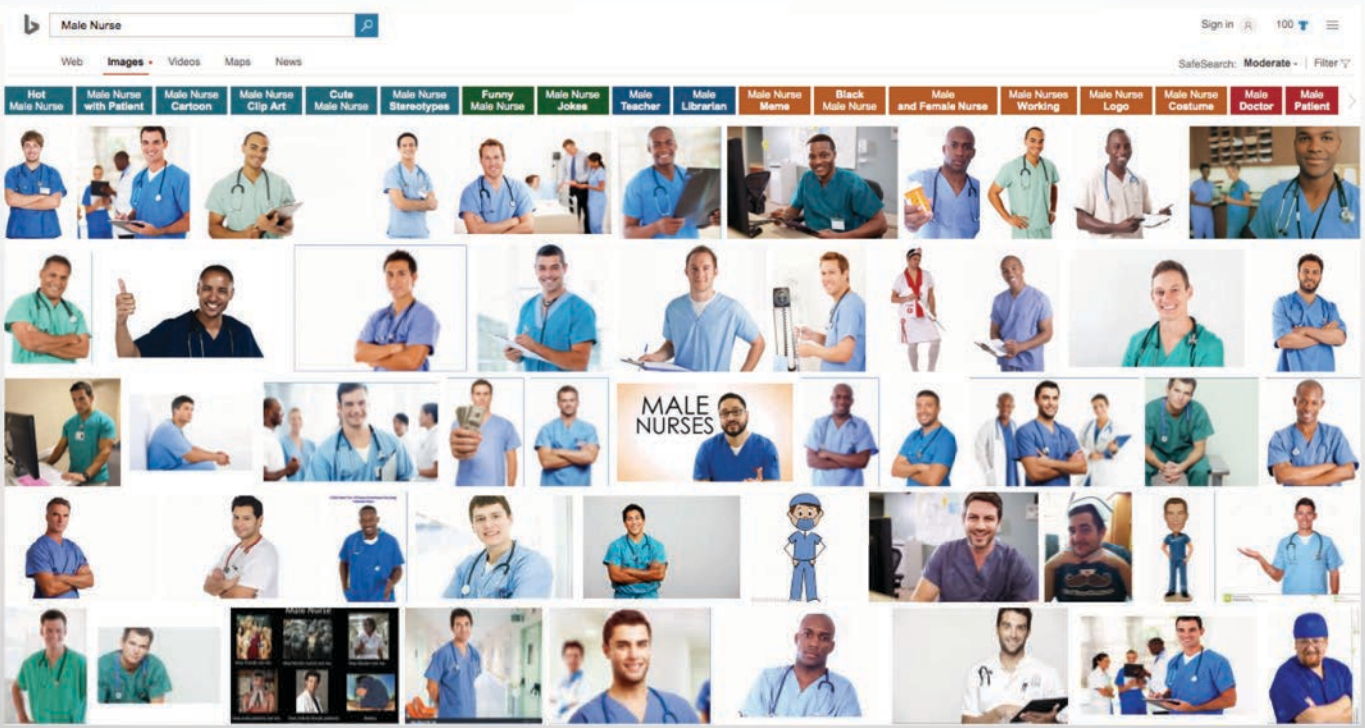
[FRIEDMAN & NISSENBAUM, 1996]

WHO IS A NURSE?

WHO IS A NURSE?

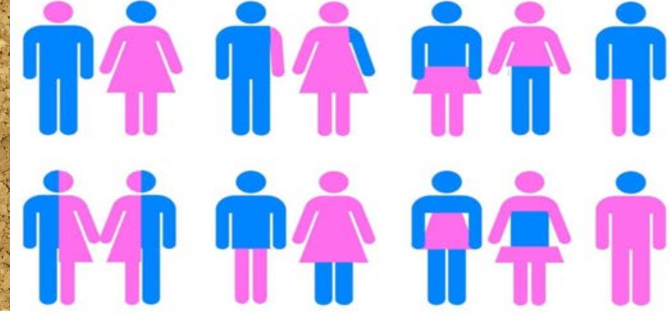


MALE NURSE


Male Nurse
Sign in
100
Web
Images
Videos
Maps
News
SafeSearch: Moderate
Filter
Hot Male Nurse
Male Nurse with Patient
Male Nurse Cartoon
Male Nurse Clip Art
Cute Male Nurse
Male Nurse Stereotypes
Funny Male Nurse
Male Nurse Jokes
Male Teacher
Male Librarian
Male Nurse Meme
Black Male Nurse
Male and Female Nurse
Male Nurses Working
Male Nurse Logo
Male Nurse Costume
Male Doctor
Male Patient
MALE NURSES

TWO KEY QUESTIONS

1. CAN WE **DETECT** SOCIALLY BIASED IMAGE RESULTS AUTOMATICALLY?
   - AWARENESS
2. WHAT MIGHT BE THE UNDERLYING **CAUSE** OF SOCIAL BIAS IN IMAGE SEARCH?
   - DATA PROVENANCE



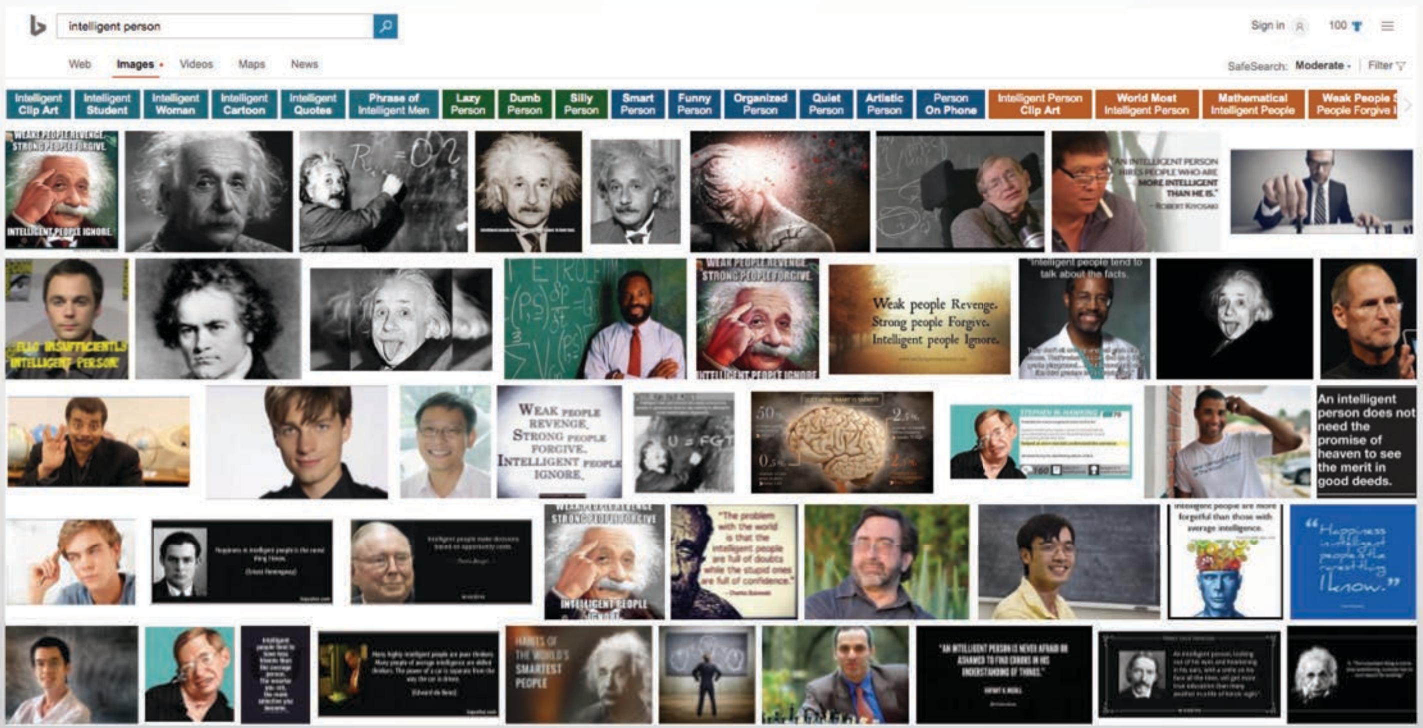

PART I:
CAN WE DETECT SOCIALLY BIASED IMAGE RESULTS AUTOMATICALLY?

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 2017 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS* (PP. 6620-6631). NEW YORK: ACM PRESS.

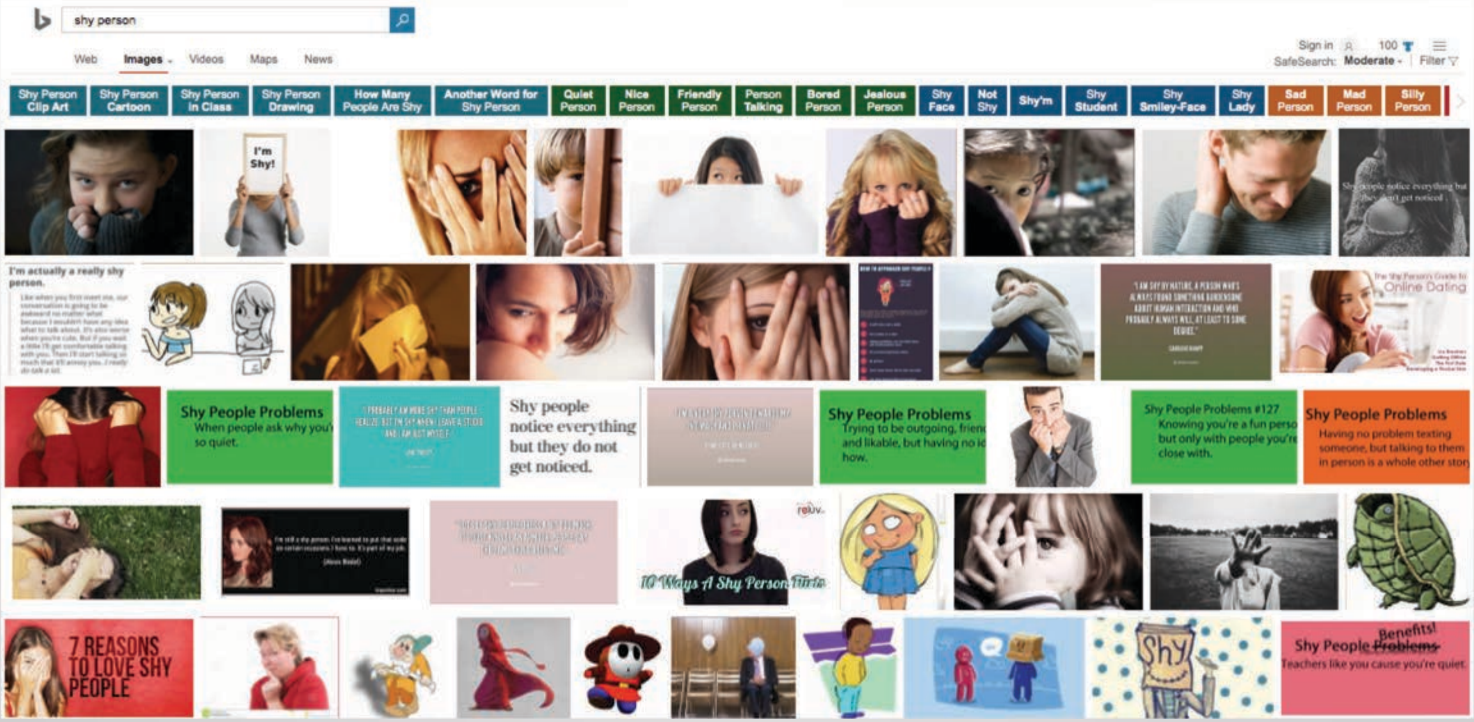
OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

The University Of Sheffield.

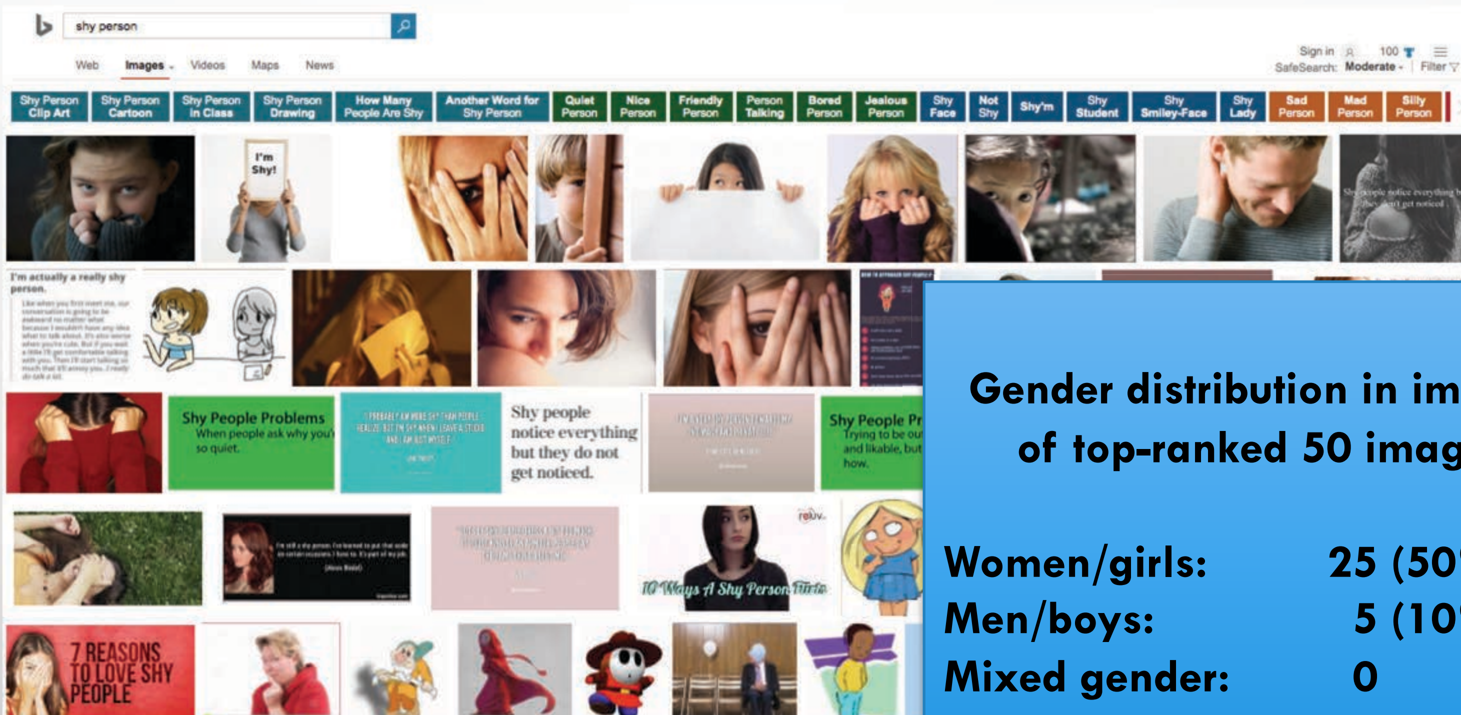
INTELLIGENT PERSON



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're so quiet.
Shy people notice everything but they do not get noticed.
Shy People Problems
Trying to be outgoing, friendly and likable, but having no idea how.
Shy People Problems #127
Knowing you're a fun person 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 story
10 Ways A Shy Person Flirts
7 REASONS TO LOVE SHY PEOPLE
Shy
Benefits!
Shy People Problems
Teachers like you cause you're quiet.

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

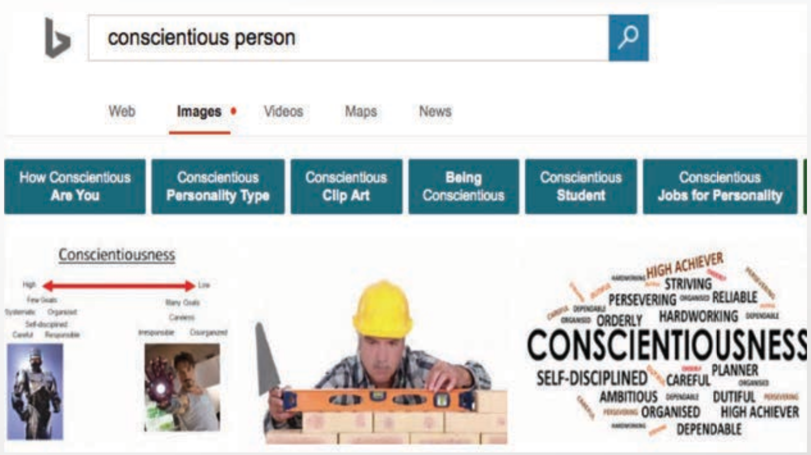
(2) WARMTH (OR COMMUNALITY): WHETHER OR NOT WE THINK SOMEONE HAS PRO-SOCIAL INTENTIONS OR IS A THREAT TO US

- 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]

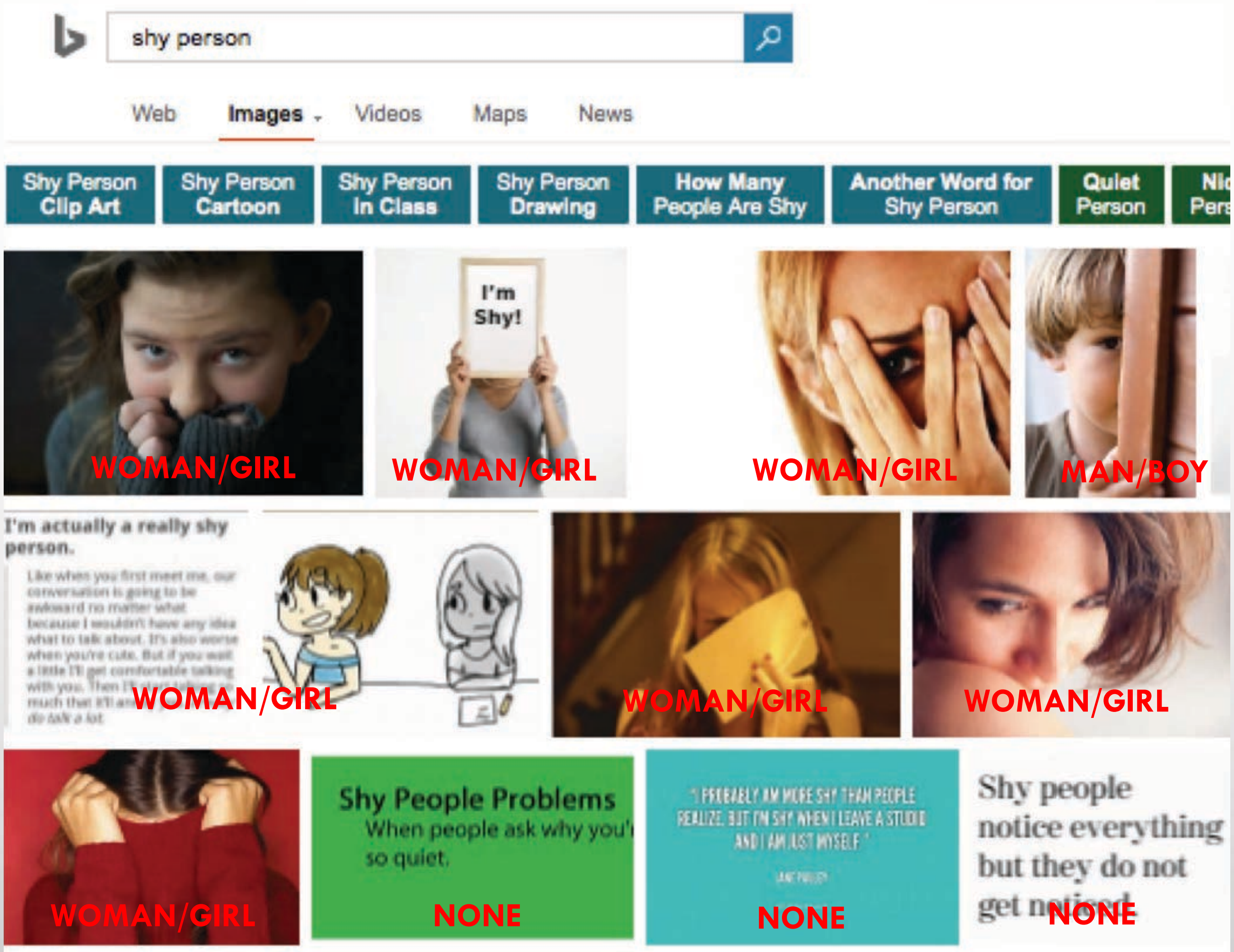
able
active
affectionate
altruistic
ambitious
assertive
boastful
capable
caring
chaotic
communicative
competent
competitive
conceited
conscientious
considerate
consistent
creative
decisive
detached
determined
dogmatic
dominant
egoistic
emotional
energetic
expressive
fair
friendly
gullible
harmonious
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!
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
MAN/BOY
I'm actually a really shy person.
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
WOMAN/GIRL
Shy People Problems
When people ask why you'
so quiet.
NONE
"I PROBABLY AM MORE SHY THAN PEOPLE REALIZE. BUT I'M SHY WHEN I LEAVE A STUDIO AND I AM JUST MYSELF."
NONE
Shy people notice everything but they do not get n
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)

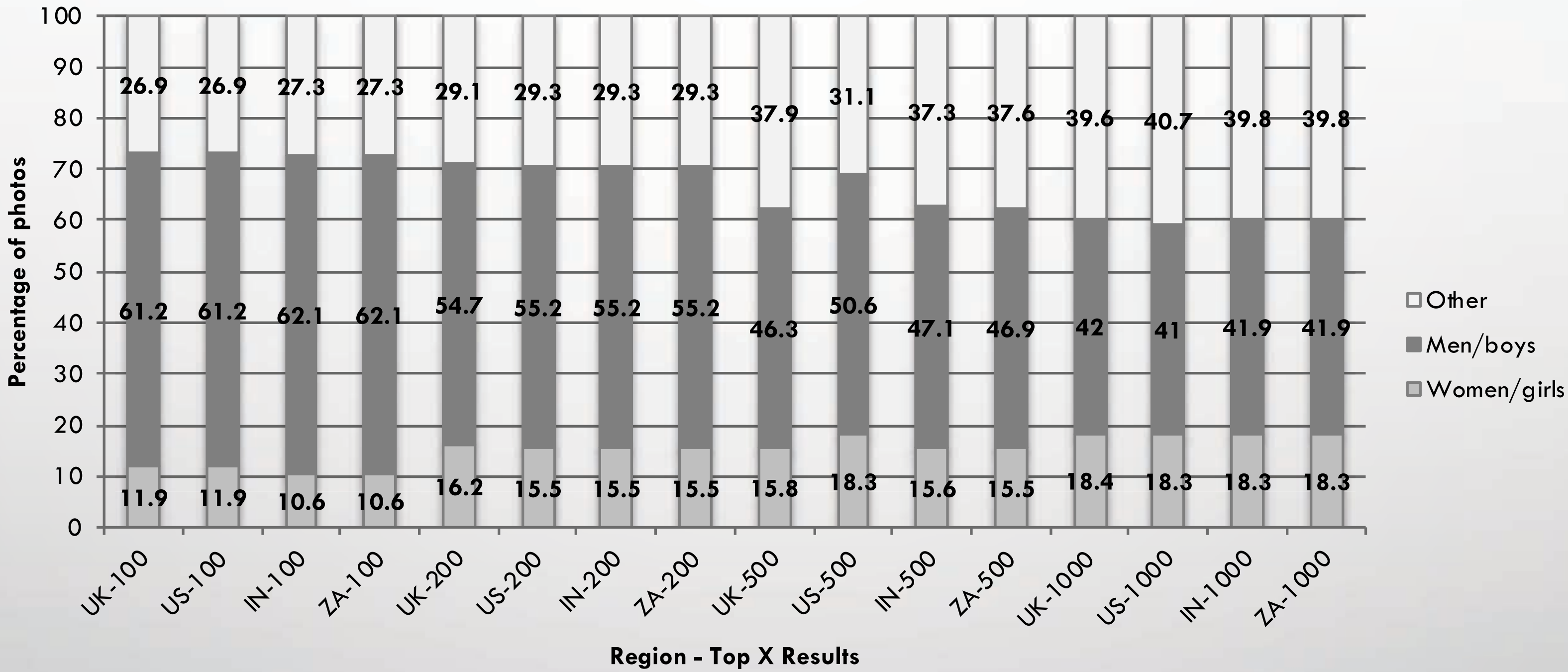
Person, man, famous, event, entertainment, talent, pop, fame, <u>portrait</u>, adult, one, serious, dark, guy, face, lid, human, young

<u>MAN</u>

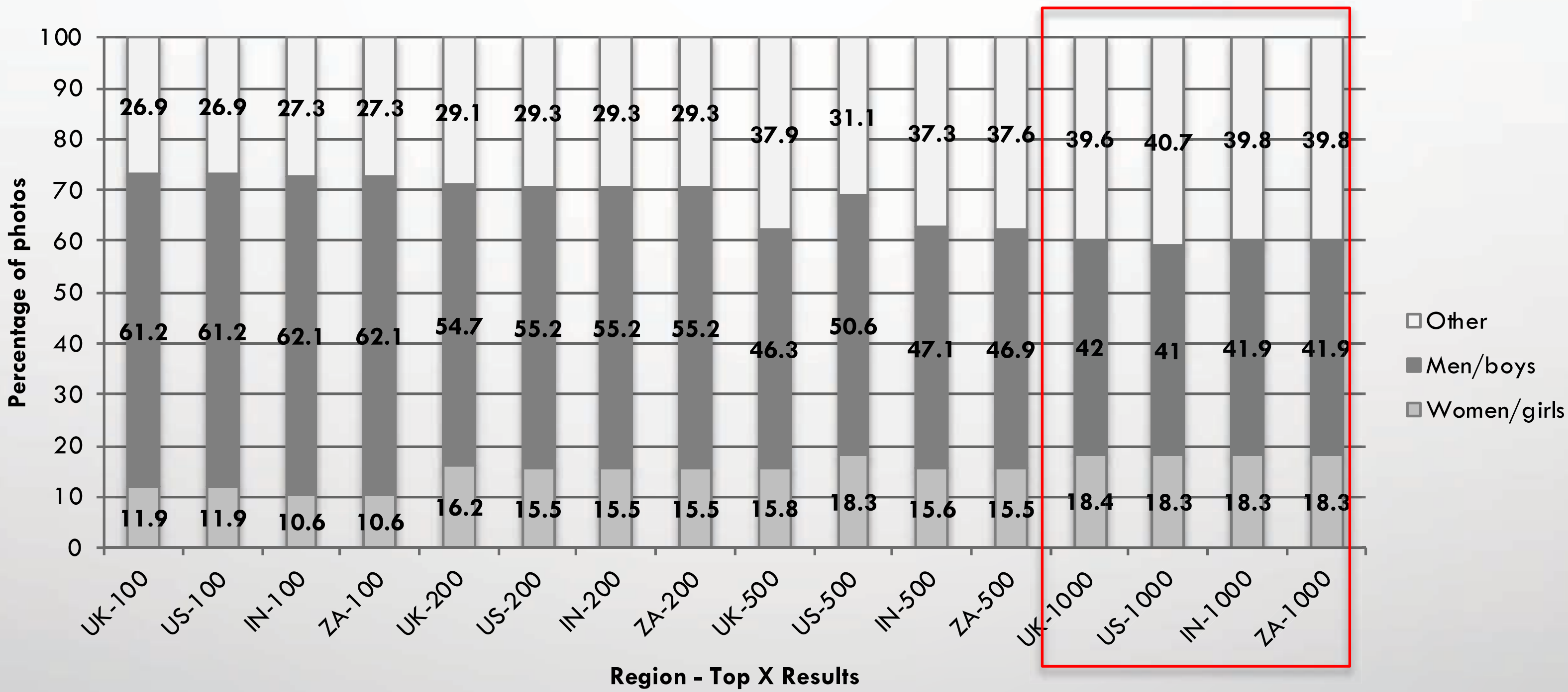
PERFORMANCE ON GENDER CLASSIFICATION

	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

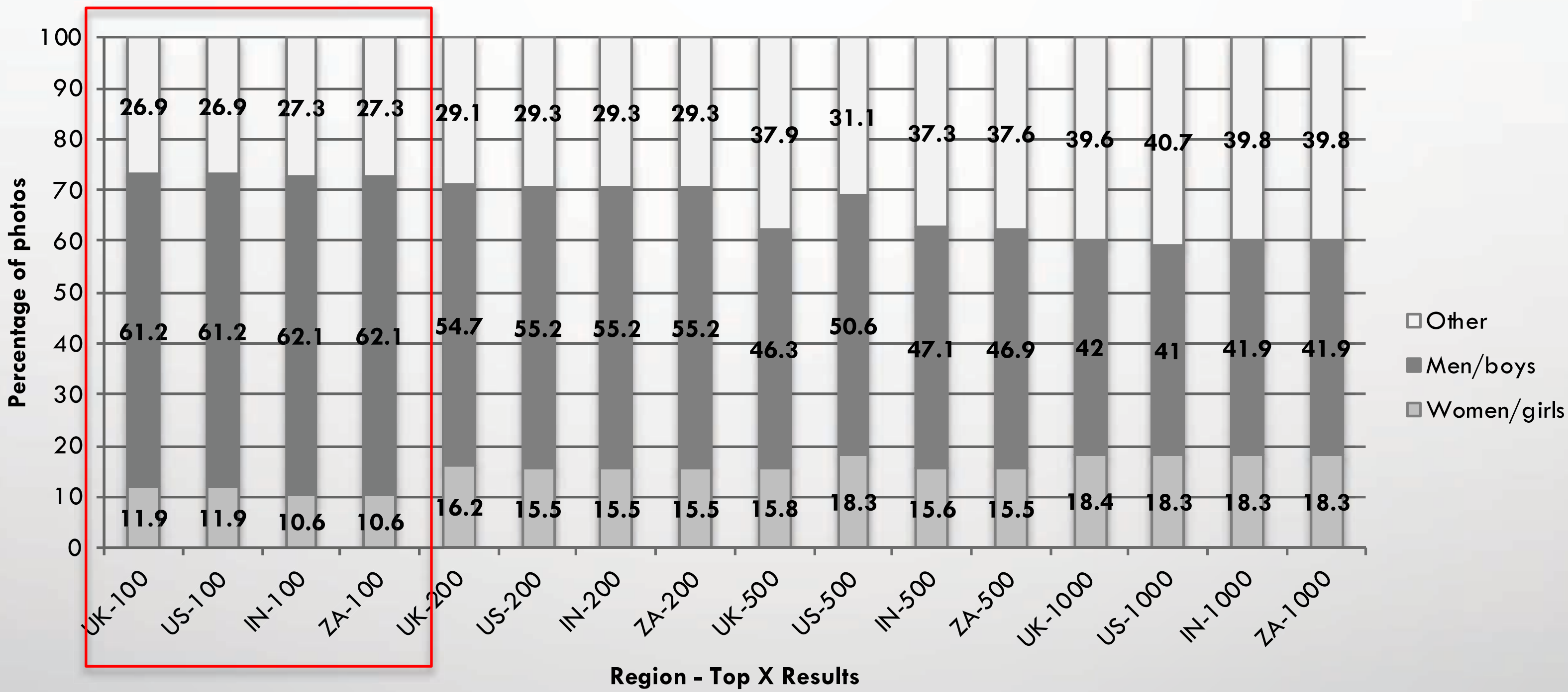
RQ1: WHO REPRESENTS A "PERSON"?



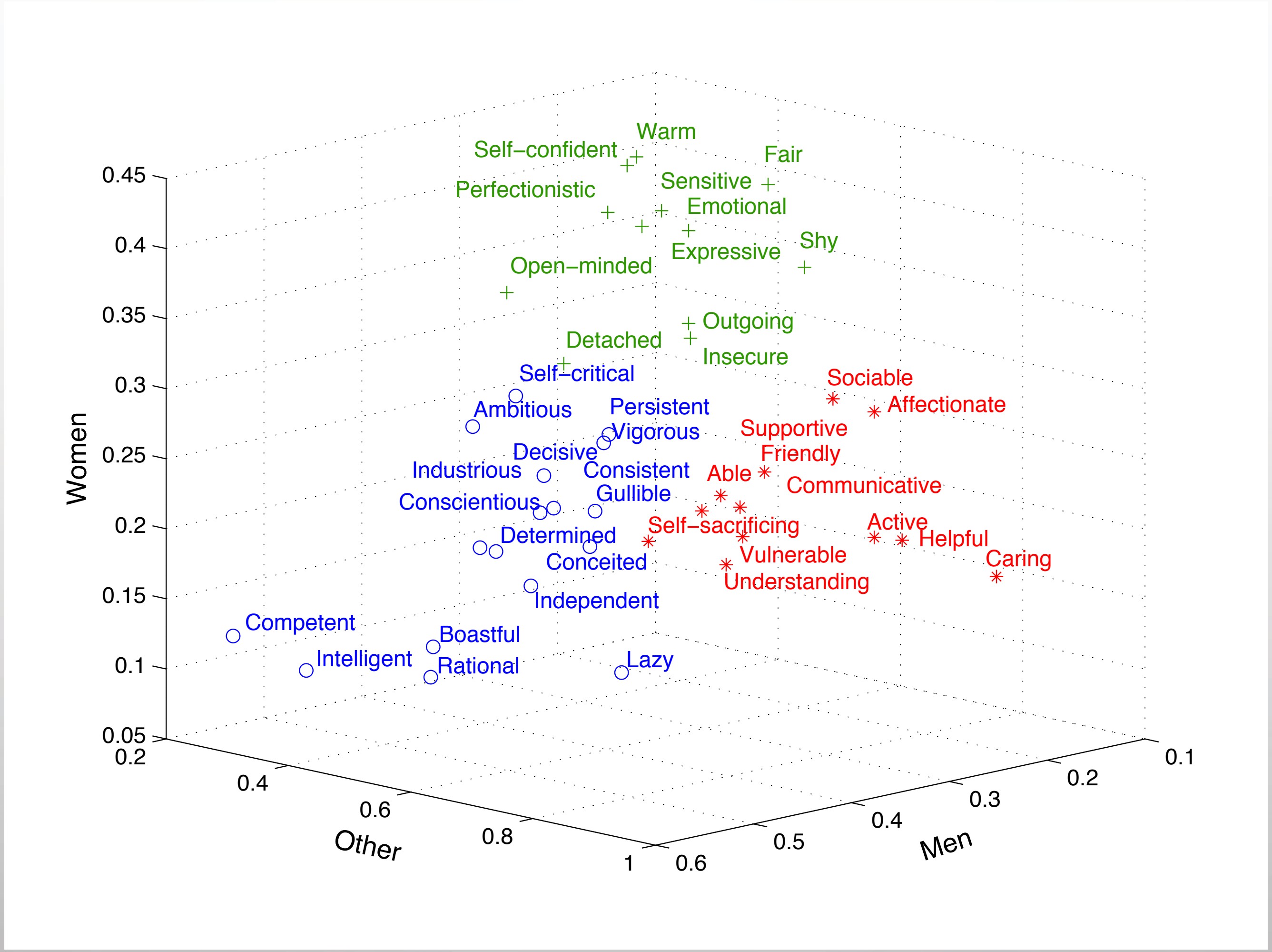
RQ1: WHO REPRESENTS A "PERSON"?



RQ1: WHO REPRESENTS A "PERSON"?



RQ2: WHICH TRAITS ARE GENDERED?



GENDERING OF TRAITS ACROSS ALL FOUR REGIONS

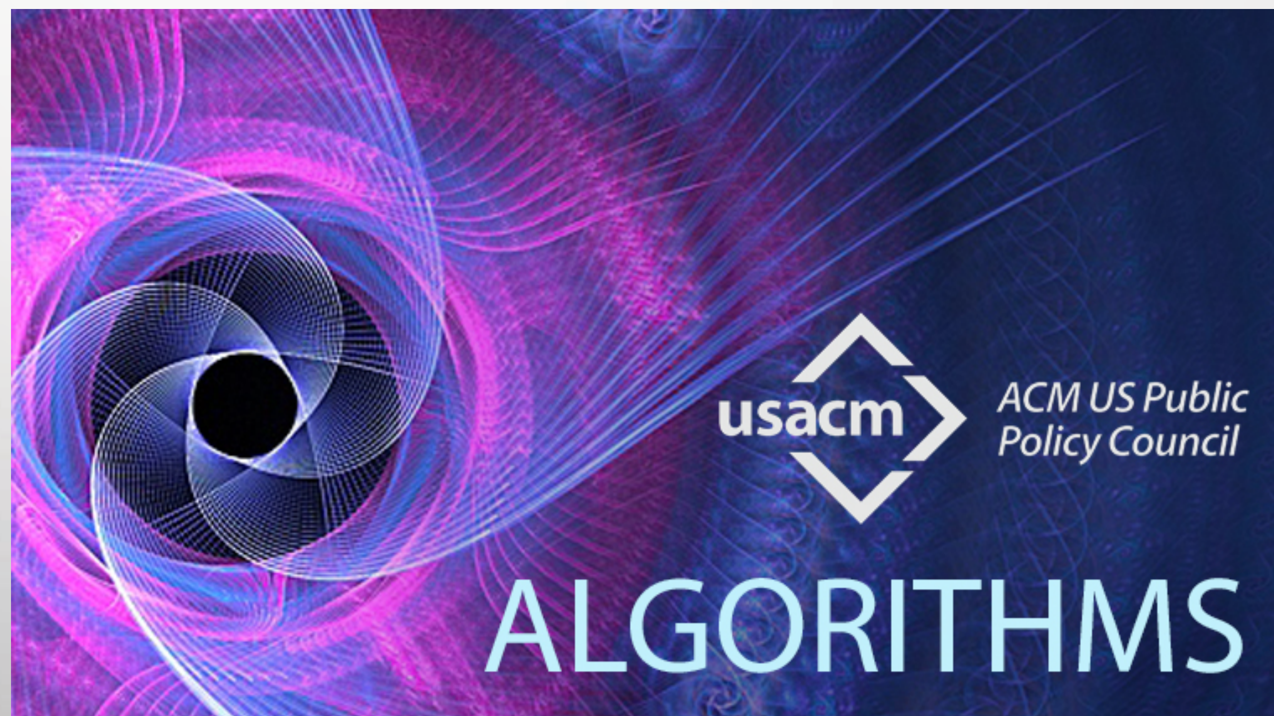
Men/boys: ambitious, boastful, competent, conceited, conscientious, consistent, decisive, determined, gullible, independent, industrious, intelligent, lazy, persistent, rational, self-critical, vigorous
Women/girls: detached, emotional, expressive, fair, insecure, open-minded, outgoing, perfectionistic,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

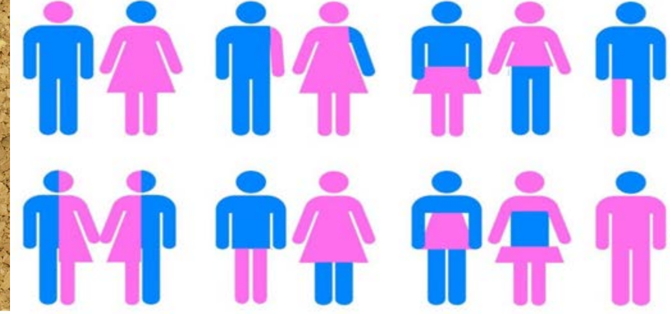
usacm ACM US Public Policy Council

acm Europe Council

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.






PART II: WHAT MIGHT BE THE UNDERLYING CAUSE OF SOCIAL BIAS IN IMAGE SEARCH?

OTTERBACHER, J. (2018, JUNE). SOCIAL CUES, SOCIAL BIASES: STEREOTYPES IN ANNOTATIONS ON PEOPLE IMAGES.
IN *PROCEEDINGS OF THE SIXTH AAAI CONFERENCE ON HUMAN COMPUTATION AND CROWDSOURCING* (HCOMP ‘18) (PP. 136-144). PALO ALTO: AAAI PRESS.

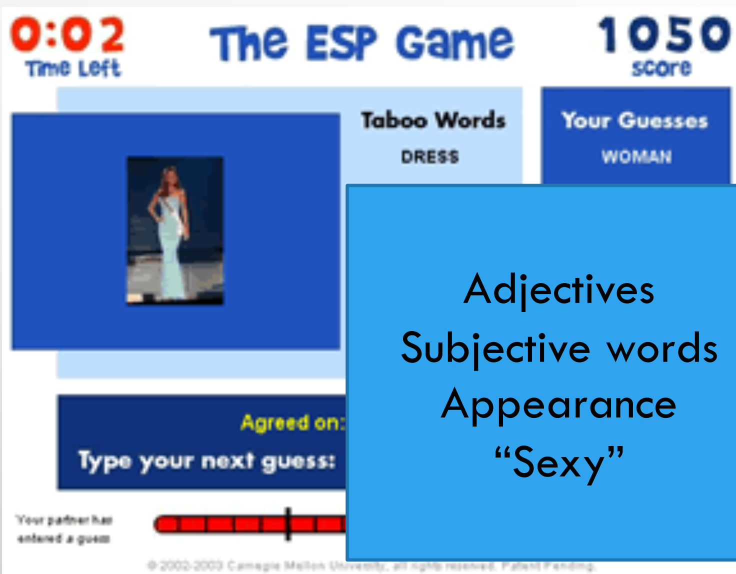
BIAS IN IMAGE METADATA?

LINGUISTIC BIAS IN IMAGE METADATA

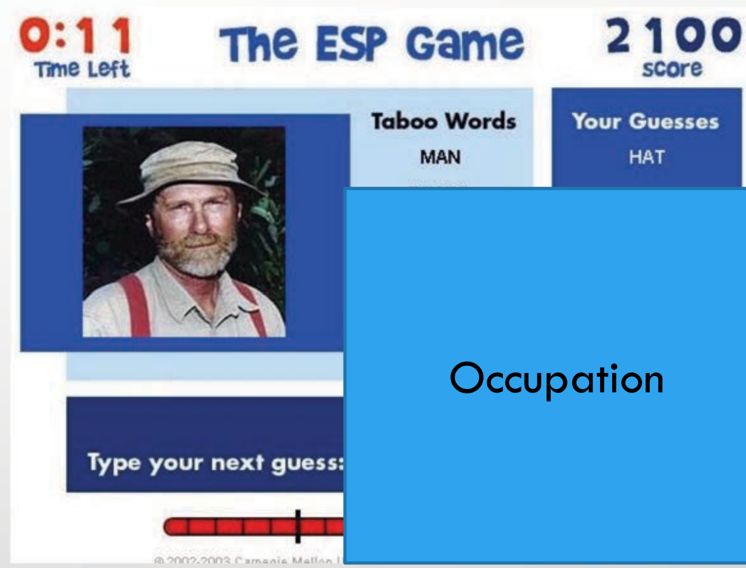
A SYSTEMATIC ASYMMETRY IN THE WAY ONE USES LANGUAGE, AS A FUNCTION OF THE SOCIAL GROUP OF THE PERSON(S) BEING DESCRIBED. [BEUKEBOOM, 2013]

- TWO LINGUISTIC PATTERNS THAT REVEAL EXPECTATIONS ABOUT OTHERS:
- -USE OF ABSTRACT VS. CONCRETE WORDS
- -USE OF SUBJECTIVE WORDS

LINGUISTIC BIAS IN IMAGE METADATA



Adjectives
Subjective words
Appearance
"Sexy"



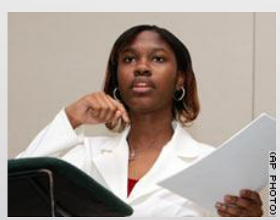
Occupation

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
- <u>CAVEAT</u>:
  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

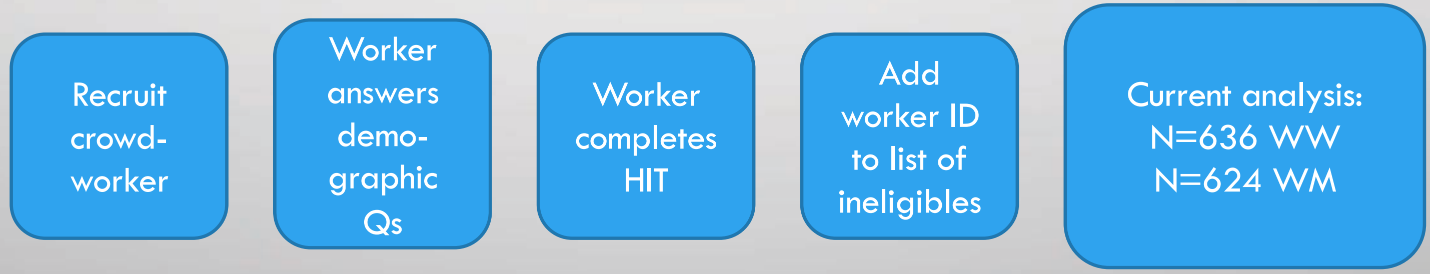
Describe the image as accurately as you can in your own words:

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

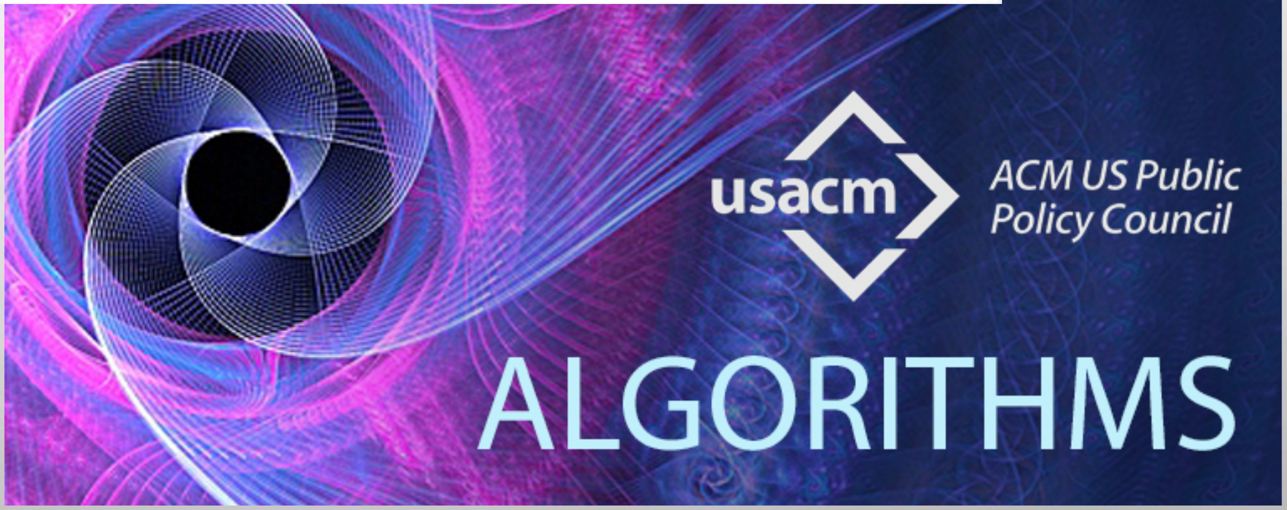
usacm ACM US Public Policy Council

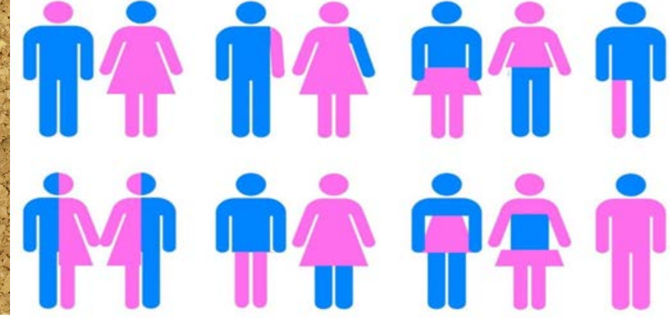
acm Europe Council

Principles for Algorithmic Transparency and Accountability

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.






THANK YOU

JAHNA.OTTERBACHER@OUC.AC.CY

OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

CY. center for algorithmic transparency

Research Centre on Interactive Media Smart Systems and Emerging Technologies