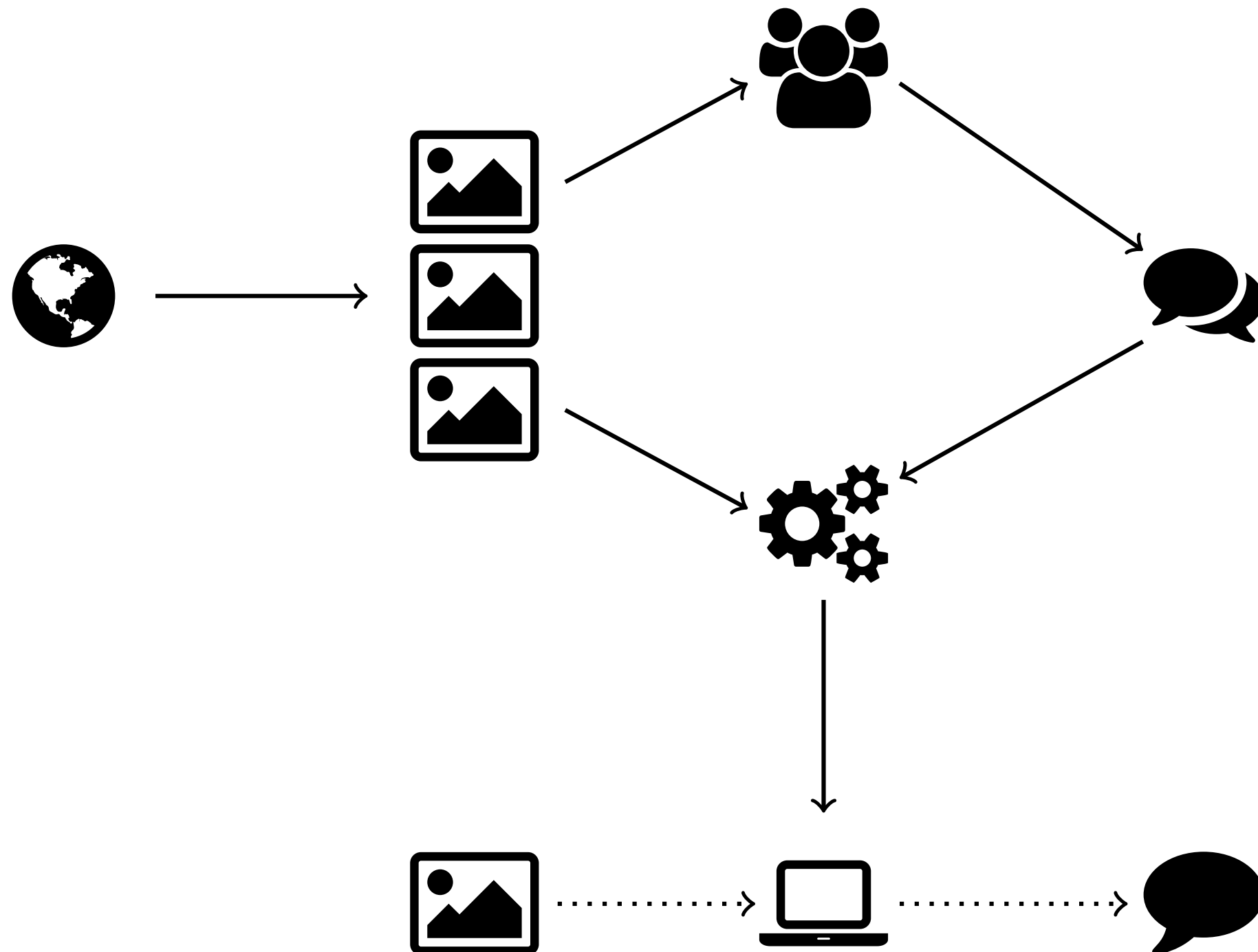


# **Stereotyping and Bias in the Flickr30k Dataset**

Emiel van Miltenburg  
Vrije Universiteit Amsterdam

# Background

Automatic image description



# Background

The Flickr30K dataset

- Standard in image description
- 30K images
- Collected from Flickr
- 5 descriptions per image
- Crowdsourced through Mechanical Turk
- Only workers from the United States



# Background

The Flickr30K dataset: example



*A bearded man, and a girl in a red dress are getting married.*

*The group of people are assembling for a wedding.*

*A man and woman dressed for a wedding function.*

*A woman holds a man 's arm at a formal event.*

*A wedding party walks out of a building.*

# Assumption

(Hodosh et al., 2013, p. 859)

“By asking people to describe the people, objects, scenes and activities that are shown in a picture without giving them any further information about the context in which the picture was taken, we were able to obtain **conceptual descriptions that focus only on the information that can be obtained from the image alone.**”

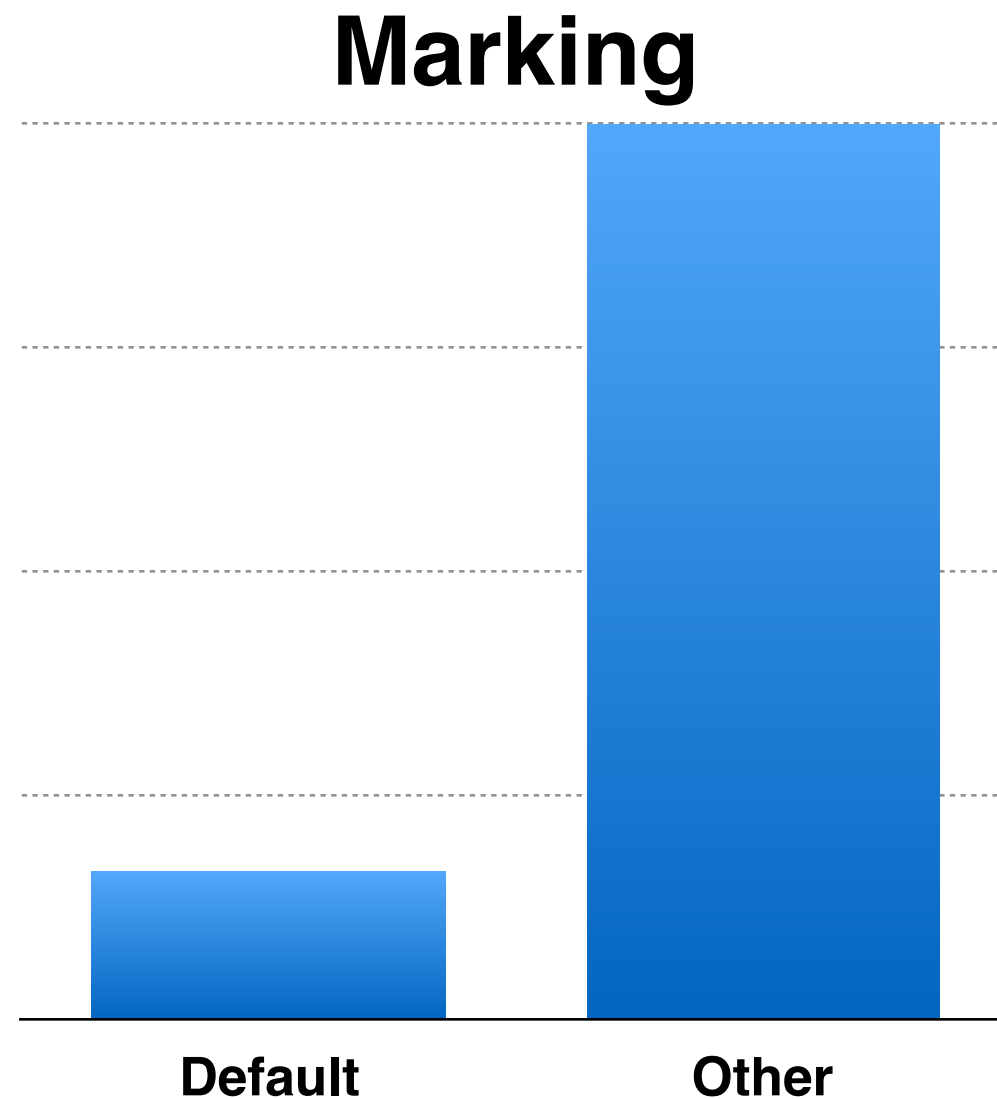
# Assumption

(Hodosh et al., 2013, p. 859)

“By asking people to describe the people, objects, scenes and activities that are shown in a picture without giving them any further information about the context in which the picture was taken, we were able to obtain ~~conceptual descriptions that focus only on the information that can be obtained from the image alone~~”

# Stereotypes & Linguistic bias

What you say reflects the way you perceive the world



**Male** surgeon <-> **Female** surgeon

**Female** nurse <-> **Male** nurse



# Marking using negations



*“A little boy is eating pie without utensils”*



# Adjectives in the image data

<b>Asian</b>		Average 60%
2339632913	Asian child/baby	2
3208987435	Asian baby, Asian/oriental woman	3
7327356514	Asian girl/baby, Asian/oriental woman	4
<b>Black</b>		Average 40%
1319788022	African-American (AA)/black baby	3
149057633	African/AA child, black baby	3
3217909454	Dark-skinned baby	1
3614582606	AA baby	1
<b>White</b>		Average 20%
11034843	White baby boy	1
176230509	White baby boy	1
2058947638	White baby	1
3991342877	White baby	1
4592281294	White baby stroller	FP
661546153	White baby stroller	FP
442983801	Fair-skinned baby	1

n/5 descriptions mentioning race, if race is mentioned at all.

Category	Number
Asian	66
Black	36
White	504
Other	18
FP	73

Images with 'baby' in the description

# Unwarranted inferences

Inferences that stem from stereotypes & common knowledge

Annotators make assumptions about:

- **Activity**: what are they doing?
- **Ethnicity**: where are they from?
- **Event**: what is going on?
- **Goal**: why are they involved in the event?
- **Relation**: how do they relate to each other?
- **Status/occupation**: what is their position in society?

# In the image data (1)



- *A blond girl and a bald man with his arms crossed are standing inside looking at each other.*
- *A worker is being scolded by her boss in a stern lecture.*
- *A manager talks to an employee about job performance.*
- *A hot, blond girl getting criticized by her boss.*
- *Sonic employees talking about work.*



# In the image data (2)

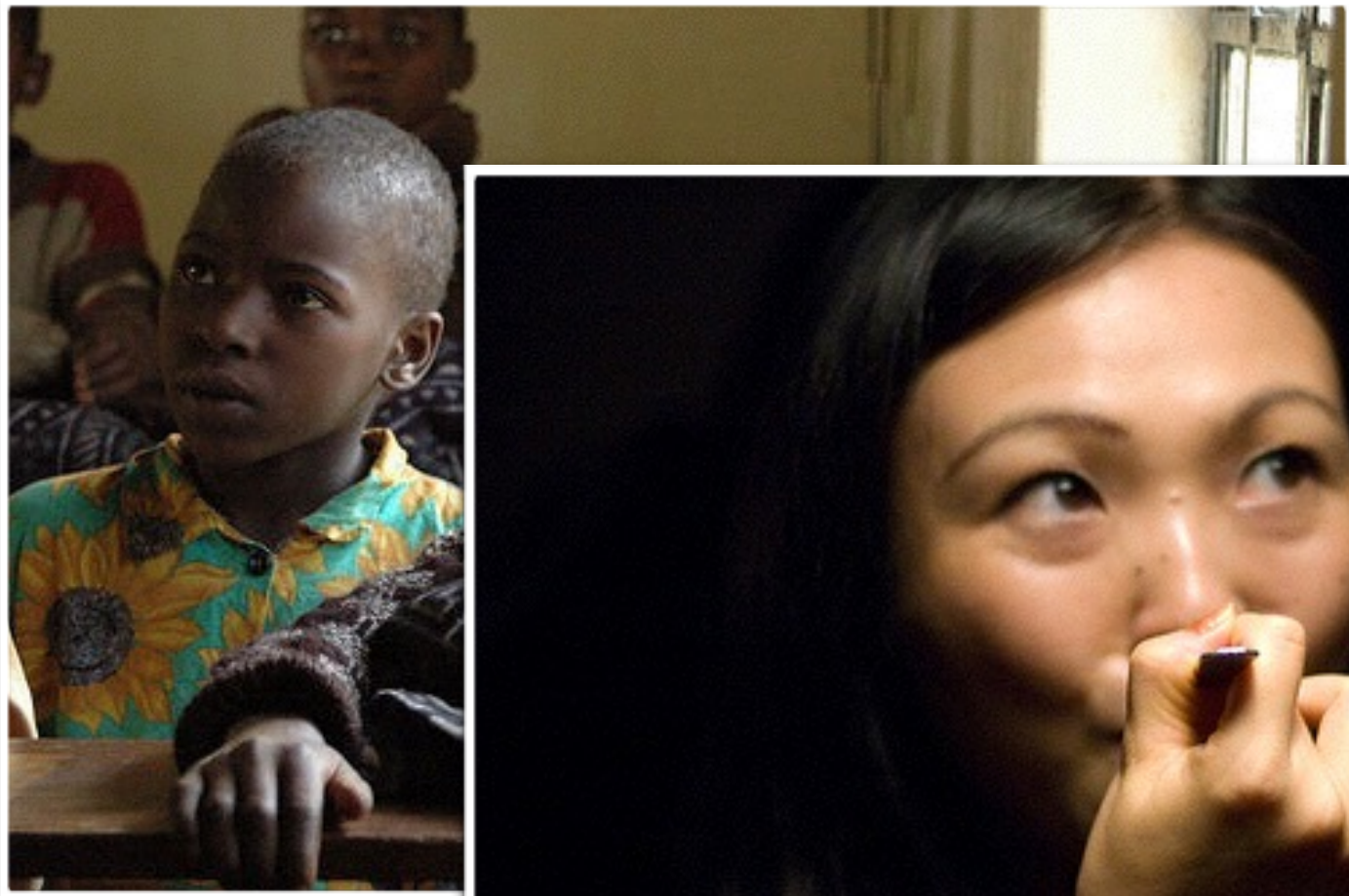
Ethnicity



Two young **African American** boys sitting at a desk in a classroom.

# In the image data (2)

Ethnicity



Two young  
boys sitting at a table in a



**A Chinese woman** is eating  
some kind of dessert.



# In the image data (2)

Ethnicity



Two young  
boys sit

**A blond woman** in an army green hat and dress sitting next to a **Japanese woman** with sunglasses on her head and a t-shirt.

some kind of desert.



# In the image data (3)



Bunches of teenagers **are watching a game** in a stadium.



# Conclusion

Image descriptions are **layered**

- ▶ basic description (maybe impossible?)
- ▶ interpretation + emphasis



**Problem:** if the data is biased, will machines be biased too?

**Open issue:** how to detect & deal with bias? —> Fair ML movement

# Guidelines

See: <http://www.jair.org/media/3994/live-3994-7275-jair.pdf>

For each question, indicate the sentence that best describes the image.

A good description...

- ...should provide an explicit description of prominent entities in the image.
- ...should not make unfounded assumptions about what is occurring in the image.
- ...should only talk about entities that appear in the image.

For example:



- The dog is wearing a red sombrero.  
**Very Good: This describes the two main objects concisely and accurately.**
- Dog wearing a red hat.  
**Good: Incomplete sentences like this are fine.**
- The white dog is wearing a pink collar.  
**Okay: This describes the dog, but it ignores the hat.**
- The red hat is adorned with gold sequins.  
**Bad: This is a detailed description of the hat, but ignores the dog.**
- The dog is trying to get away from the guy because he is angry that he has a hat on.  
**Bad: This is speculation. We do not know what the dog's intentions or feelings are.**
- The dog.  
**Very Bad: This does not give enough detail. It could describe any image of any dog.**

# Future work

- Cross-cultural study of stereotypes in descriptions
- Influence of crowdsourcing format on descriptions

Other examples



# Inferring goals



A man is hooking himself up to the tether line **in order to have some fun.**



The school is having a special event in **order to show the american culture on how other cultures are dealt with in parties.**

# Inferring relations



**A father** is spoon feeding **his baby girl**, as she smiles .



Little girl with backpack pushing **brother** in stroller.  
A child pushing their **younger sibling** in a stroller.



# Inferring status



**A group of college students** gathers to play texas hold em poker.



**Graphics Designer**, male (approximately 38 years old) drafting on design table in office.