Towards Transparency in AI: Methods and Challenges

Timnit Gebru
Google AI

Project will use AI to prevent or minimize electric grid failures

September 14, 2017 by Glennda Chui, SLAC National Accelerator Laboratory

A project led by the Department of Energy's SLAC National Accelerator Laboratory will combine artificial intelligence with massive amounts of data and industry experience from a dozen U.S. partners to identify places where the electric grid is vulnerable to disruption, reinforce those spots in advance and recover faster when failures do occur.

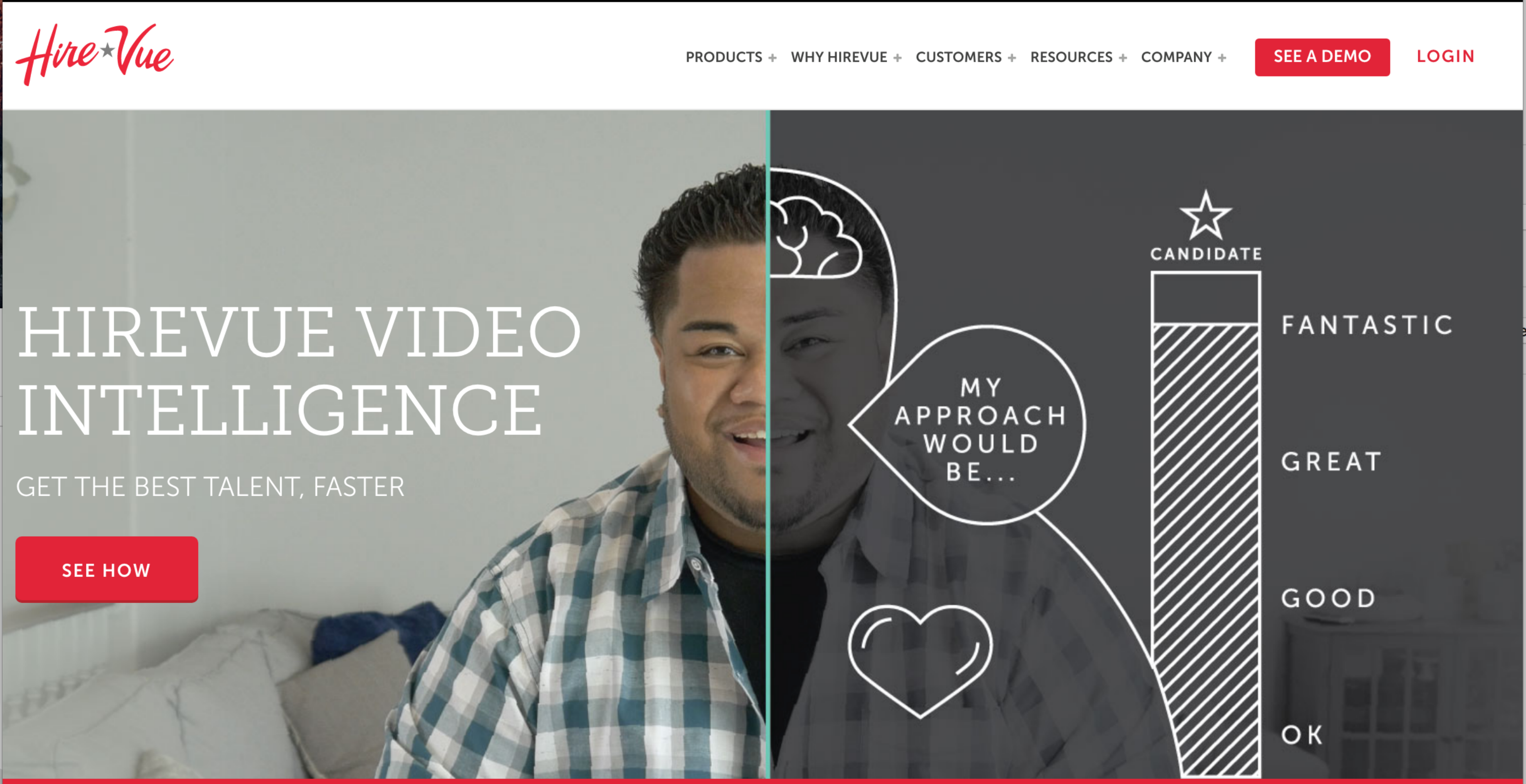
The eventual goal is an autonomous grid that seamlessly absorbs routine power fluctuations from clean energy sources like solar and wind and quickly responds to disruptive events—from major storms to eclipse-induced dips in solar power—with minimal intervention from humans.

"This project will be the first of its kind to use artificial intelligence and machine learning to improve the resilience of the grid," said Sila Kiliccote, director of SLAC's Grid Integration, Systems and Mobility lab, GISMo, and principal investigator for the project. "While the approach will be tested on a large scale in California, Vermont and the Midwest, we expect it to have national impact, and all the tools we develop will be made available either commercially or as open source code."

Called GRIP, for Grid Resilience and Intelligence Project, the project builds on other efforts to collect massive amounts of data and use it to fine-tune grid operations, including SLAC's

How AI Could Smarten Up Our Water System

Photo: Fernando Butcher


HireVue
PRODUCTS + WHY HIREVUE + CUSTOMERS + RESOURCES + COMPANY +
SEE A DEMO
LOGIN
HIREVUE VIDEO INTELLIGENCE
GET THE BEST TALENT, FASTER
SEE HOW
CANDIDATE
FANTASTIC
GREAT
GOOD
OK
MY APPROACH WOULD BE...

Extreme Vetting Initiative Objectives (cont.)

Performance Objectives of the Overarching Vetting Contract:

1. Centralizes screening and vetting processes to mitigate case backlog and provide law enforcement and field agents with timely, actionable information;

2. Allows ICE to develop richer case files that provide more value-added information to further investigations or support prosecutions in immigration or federal courts;

3. Allows ICE to perform regular, periodic and/or continuous review and vetting of nonimmigrants for changes in their risk profile after they enter the United States and;

4. Automates at no loss of data quality or veracity any manually-intensive vetting and screening processes that inhibit ICE from properly and thoroughly vetting individuals in a timely fashion.

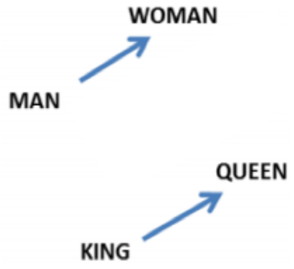
Israel Arrests Palestinian Because Facebook Translated 'Good Morning' to 'Attack Them'

No Arabic-speaking police officer read the post before arresting the man, who works at a construction site in a West Bank settlement

Gender analogies in "Word Embeddings"

Parallelograms capture semantics: [many papers]

- Man:King :: Woman:*Queen*

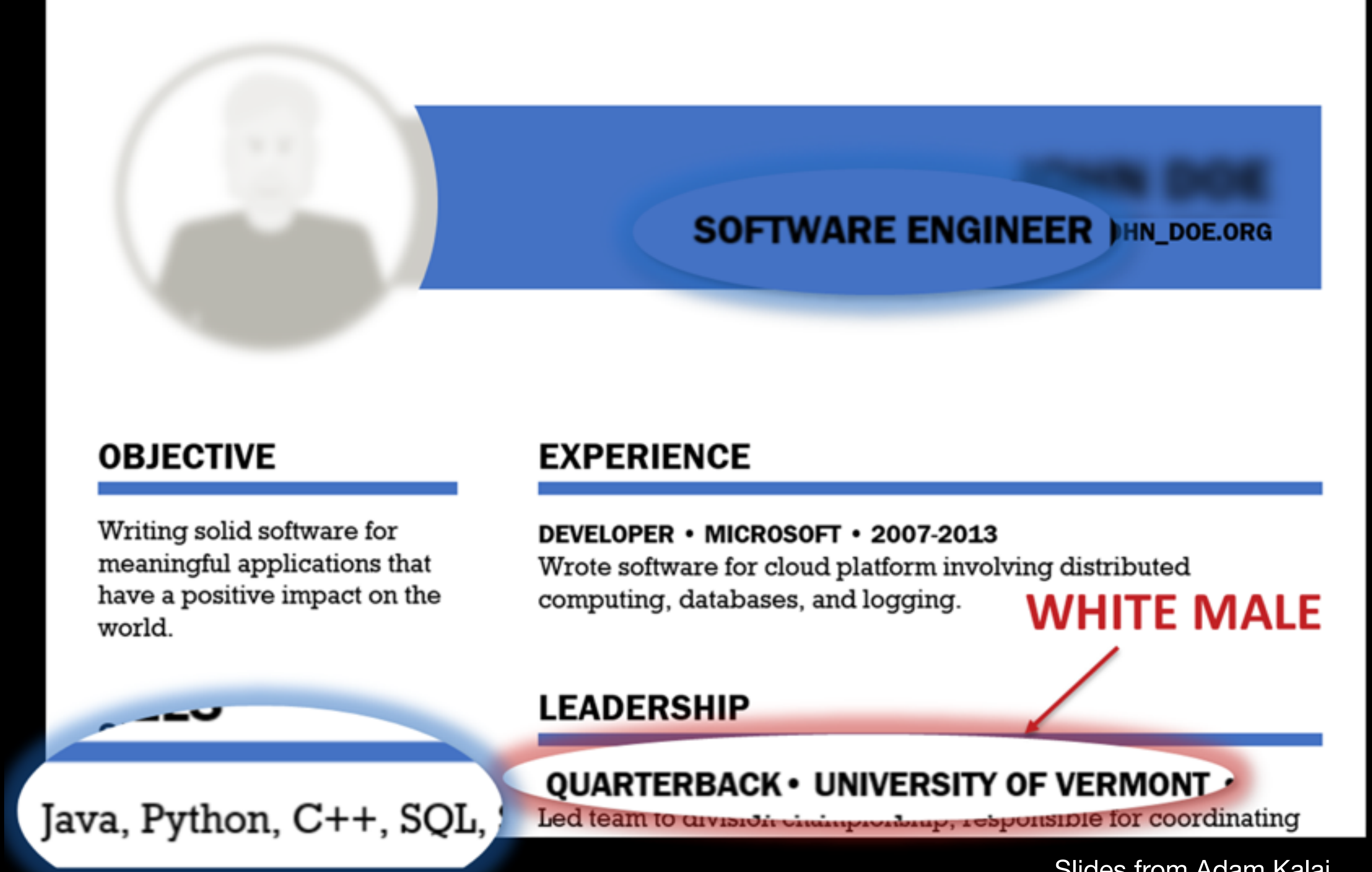


[Bolukbasi-Chang-Zou-Saligrama-**K**'16]:

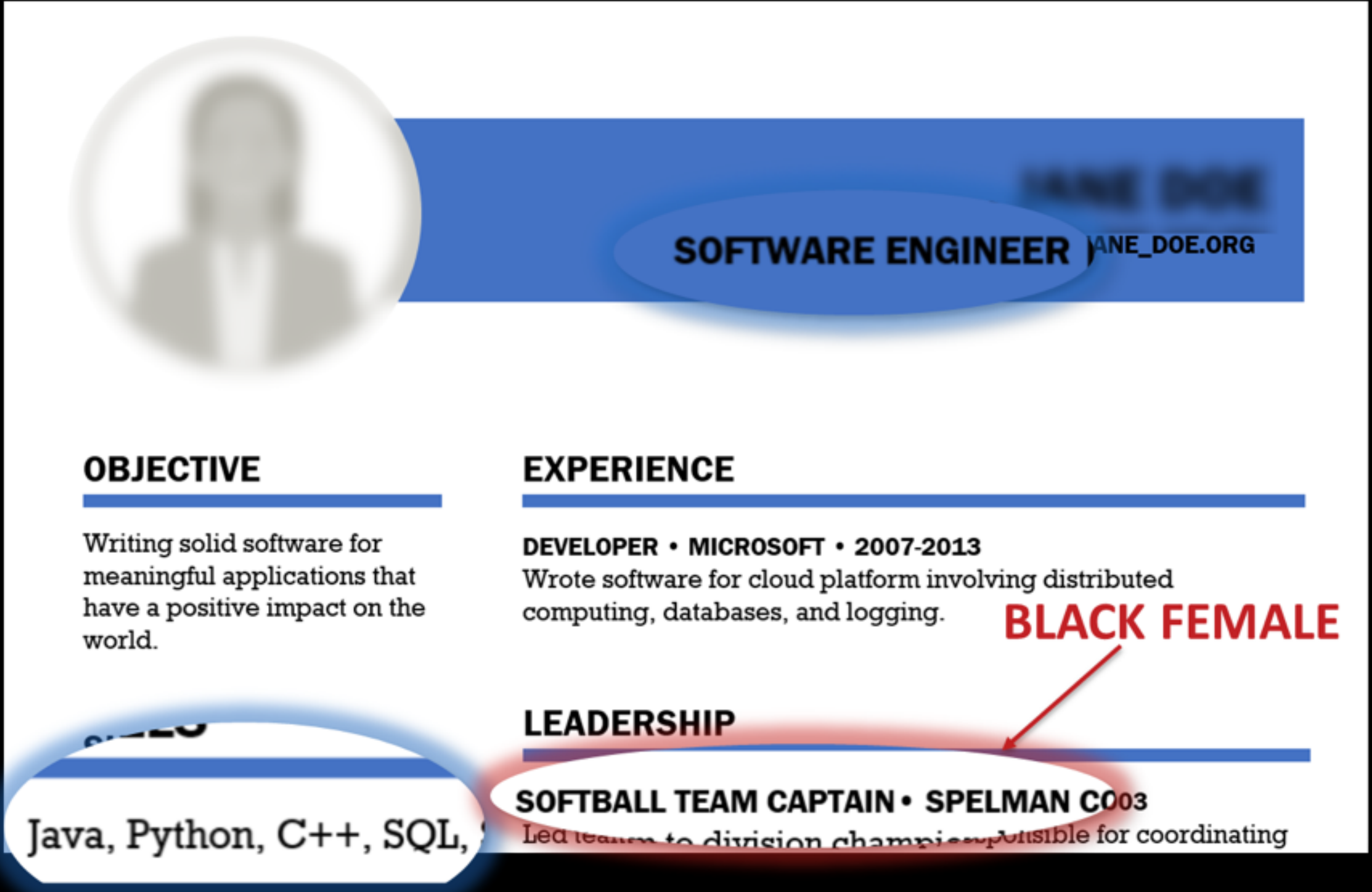
- He:*Brother* :: She:*Sister*
- He:*Doctor* :: She:*Nurse*
- He:*WTF* :: She:*OMG*
- He:*Realist* :: She:*Feminist*
- He:*Computer Programmer* :: She:*Homemaker*
- She:*Pregnancy* :: He:*Kidney stone*

Based on word2vec trained on Google News corpus

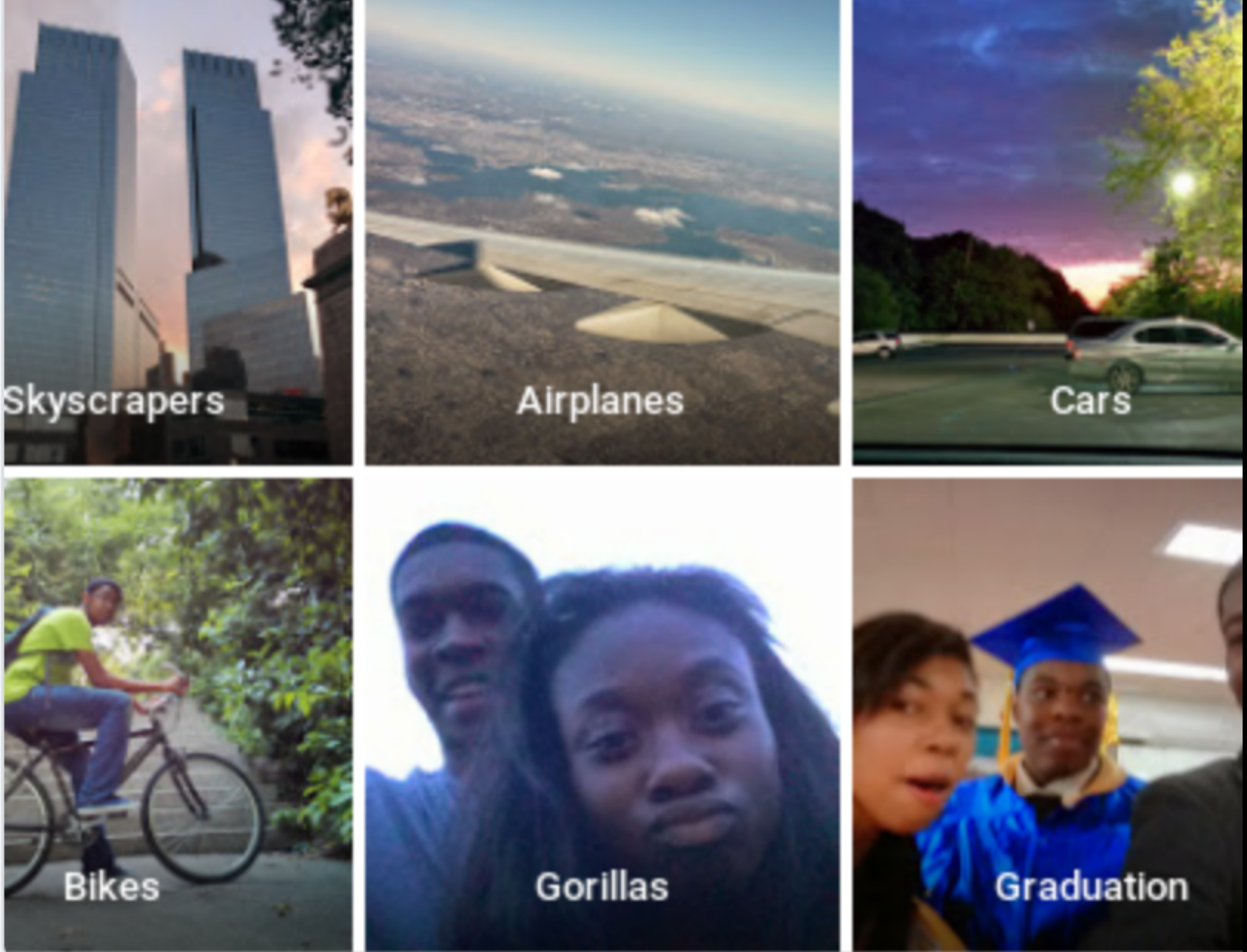
Slides from Adam Kalai



Slides from Adam Kalai



Slides from Adam Kalai


Skyscrapers
Airplanes
Cars
Bikes
Gorillas
Graduation

US ADULTS INDEXED

130 MILLION

One in two American adults is in a law enforcement face recognition network used in unregulated searches employing algorithms with unaudited accuracy.

The Perpetual Line Up

(Garvie , Bedoya, Frankle 2016)

ERROR RATE(1-PPV) BY FEMALE x SKIN TYPE

	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TYPE VI
	1.7%	**1.1%**	**3.3%**	**0%**	**23.2%**	**25.0%**
FACE++	**11.9%**	**9.7%**	**8.2%**	**13.9%**	**32.4%**	**46.5%**
IBM	**5.1%**	**7.4%**	**8.2%**	**8.3%**	**33.3%**	**46.8%**

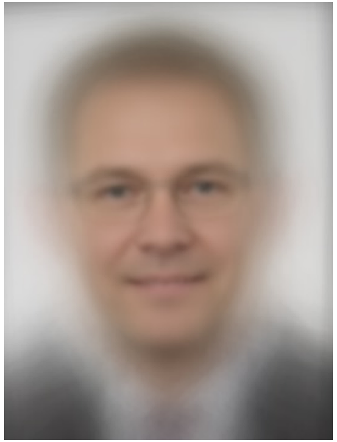
Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

EXISTING BENCHMARKS

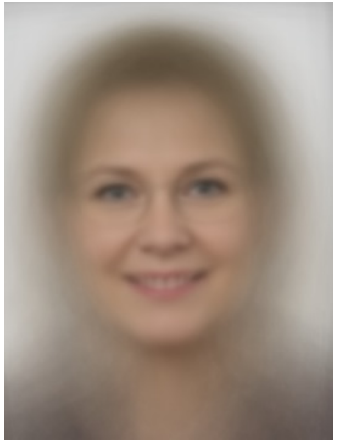
IJB-A

75.4% Male

79.6% Lighter



ADIENCE

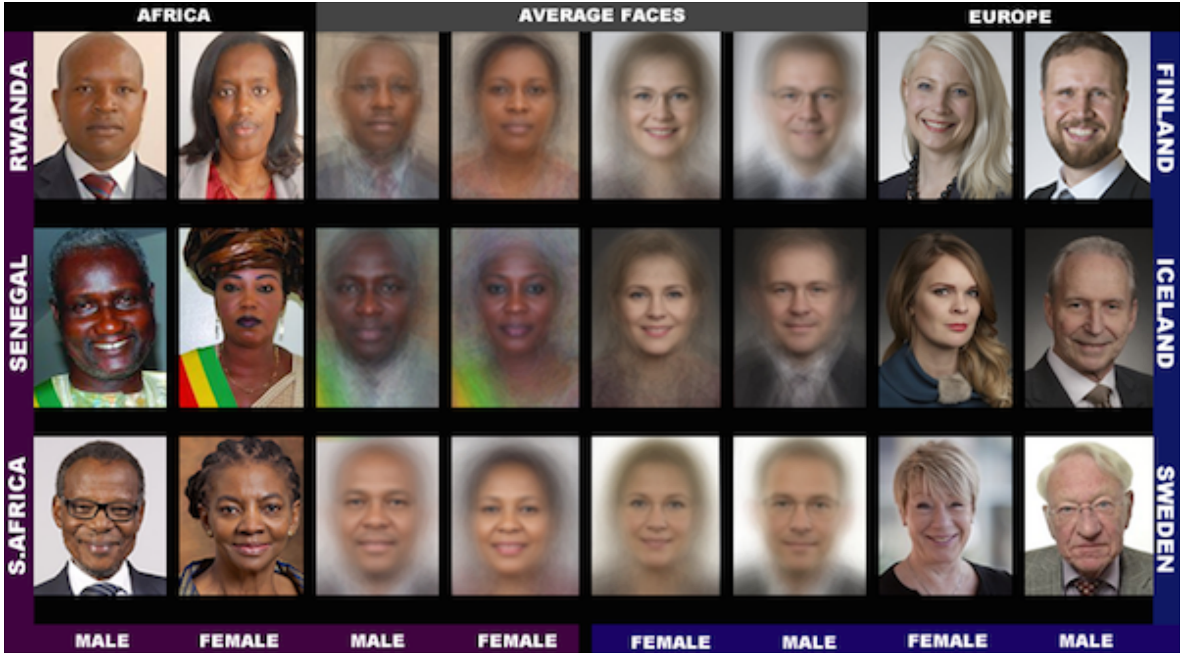


52% Female

86.2% Lighter

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

PILOT PARLIAMENTS BENCHMARK (PPB)



1270 Faces
6 Countries
54.4% Male
53.6% Lighter

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

DEMOGRAPHIC & PHENOTYPIC LABELS

- BINARY GENDER LABELS (M/F)
- FITZPATRICK SKIN TYPE LABELS

SKIN TYPE	one 	two 	three 	four 	five 	six 
Hair	red, blonde	blonde, red, light brown	chestnut, dark blonde	brown, medium brown, dark brown	dark brown	black
Eyes	blue, grey, green	blue, grey, green, hazel	brown, blue, grey, green, hazel	hazel, brown	brown	brown
Skin	very pale white, pale white	pale white	white, light brown	medium brown, dark brown	dark brown	black
Tanning Ability	burns very easily, never tans	burns easily, rarely tans	sometimes burns, gradually tans	hardly ever burn, tans very easily	Rarely burns, tans easily and quickly darkens	Never burns, tans very dark

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

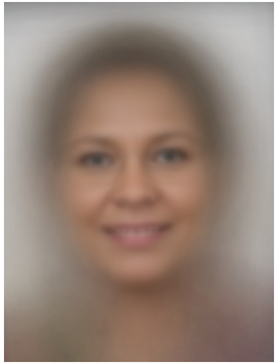
OVERALL ACCURACY(PPV) ON PPB

93.7%	90%	87.9%

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

ACCURACY BY GENDER

	FEMALE FACES	MALE FACES
	89.3%	97.4%
FACE++	78.7%	99.3%
IBM	79.7%	94.4%



FEMALE

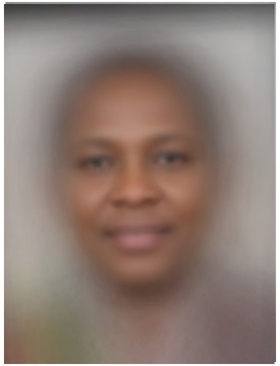
8-21%
ERROR
GAP

MALE

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

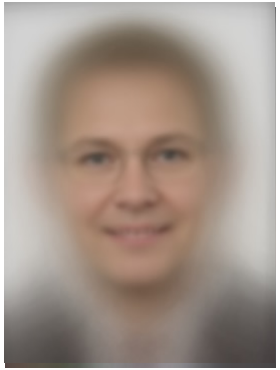
ACCURACY BY SKIN TYPE

	DARKER FACES	LIGHTER FACES
	87.1%	99.3%
FACE++	83.5%	95.3%
IBM	77.6%	96.8%



DARKER

12-19%
ERROR
GAP

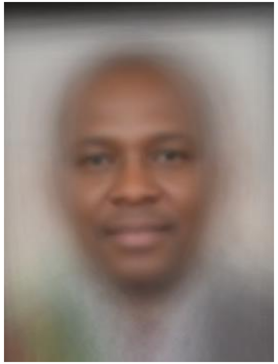


LIGHTER

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

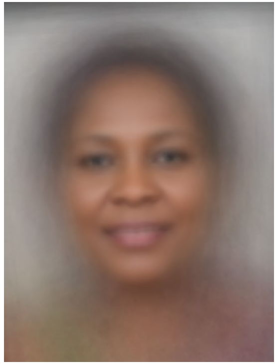
INTERSECTIONAL ACCURACY - MSFT

94%



DARKER
MALES

79.2%

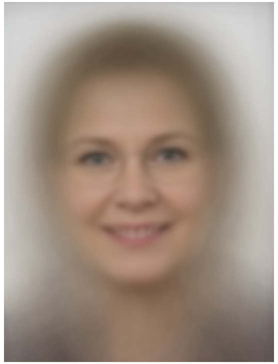


DARKER
FEMALES

100%

LIGHTER
MALES

98.3%



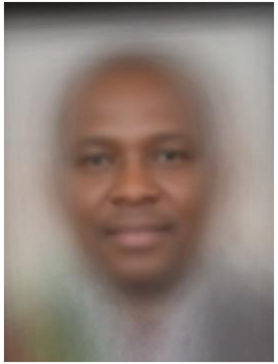
LIGHTER
FEMALES

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

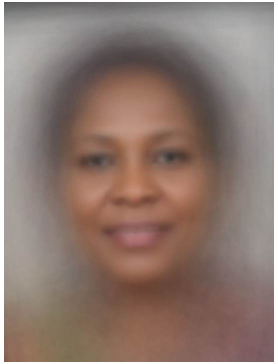
INTERSECTIONAL ACCURACY - FACE++

FACE++

99.3% | 65.5% | 99.2% | 98.3%

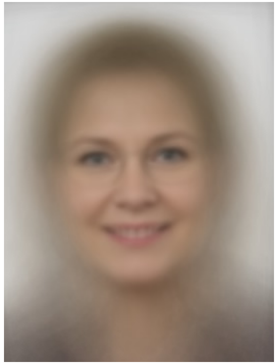


DARKER MALES



DARKER FEMALES

LIGHTER MALES



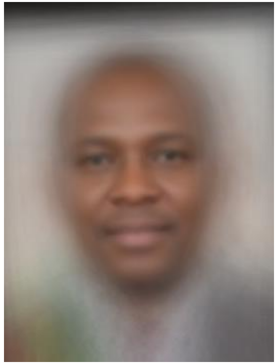
LIGHTER FEMALES

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

INTERSECTIONAL ACCURACY - IBM

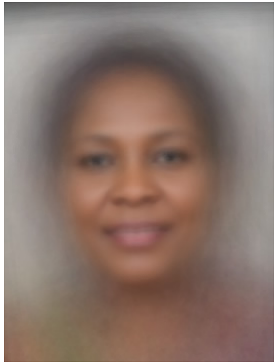
IBM

88%



DARKER MALES

65.3%

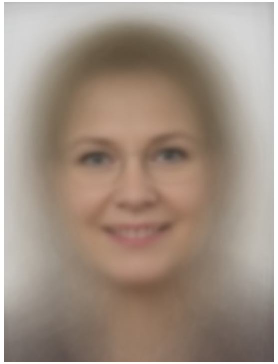


DARKER FEMALES

99.7%

LIGHTER MALES

92.9%



LIGHTER FEMALES

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

INTERSECTIONALITY MATTERS

558 F. 2d 480 - Emma Degraffenreid et Al. v. General Motors Assembly Division, St. Louis

558 F.2d 480

15 Fair Empl.Prac.Cas. 573, 14 Empl. Prac. Dec. P 7692

Emma DeGRAFFENREID et al., Appellants,

v.

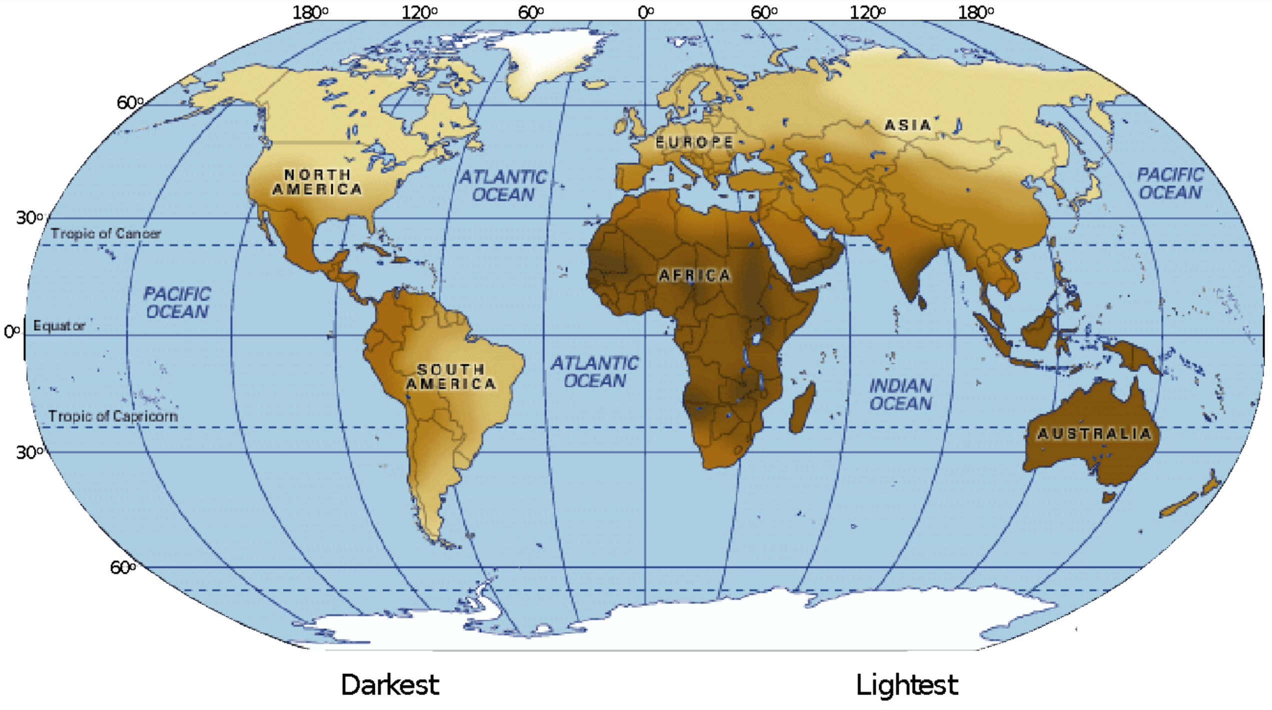
GENERAL MOTORS ASSEMBLY DIVISION, ST. LOUIS, et al., Appellees.

No. 76-1599.

United States Court of Appeals, Eighth Circuit.

Submitted March 18, 1977.

Decided July 15, 1977.


180°
120°
60°
0°
60°
120°
180°
60°
30°
0°
30°
60°
Tropic of Cancer
Equator
Tropic of Capricorn
NORTH AMERICA
SOUTH AMERICA
EUROPE
ASIA
AFRICA
AUSTRALIA
ATLANTIC OCEAN
ATLANTIC OCEAN
PACIFIC OCEAN
PACIFIC OCEAN
INDIAN OCEAN
Darkest
Lightest

We can’t ignore social & structural problems

I recommend reading The Moral Character of Cryptographic Work by Philip Rogaway

US ADULTS INDEXED

130 MILLION

One in two American adults is in a law enforcement face recognition network used in unregulated searches employing algorithms with unaudited accuracy.

The Perpetual Line Up

(Garvie , Bedoya, Frankle 2016)

Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.

There are no laws that restrict who can use our APIs & Datasets for what.

It is even possible that some algorithms are breaking existing laws (e.g. EEOC)

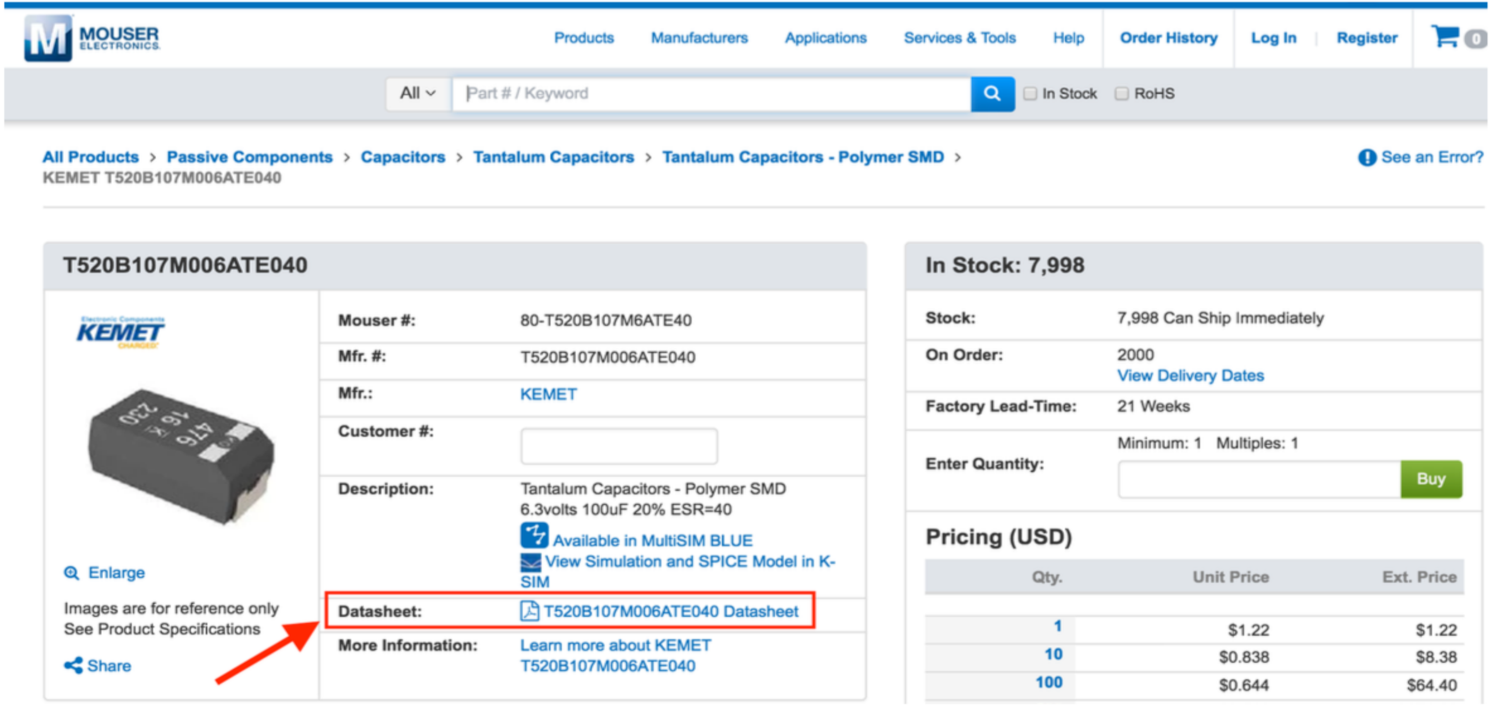
We need standards/Documentation

Other industries have been there

Electronics

Lots of standardization/testing/documentation

Electronics



Datasheets for Datasets

Electronics

XICON

Miniature Aluminum Electrolytic Capacitors — XRL Series

FEATURES

- Low impedance characteristics
- Case sizes are smaller than conventional general-purpose capacitors, with very high performance
- Can size larger than 9mm diameter has safety vents on rubber end seal
- RoHS Compliant

RoHS Compliant

CHARACTERISTICS

Item	Characteristics													
Operating Temperature Range	-40°C ~ +85°C													
Capacitance Tolerance	±20% at 120Hz, 20°C													
Leakage Current	≤100V	I = 0.01CWV or 3μA whichever is greater after 2 minutes of applied rated DC working voltage at 20°C Where: C = rated capacitance in μF; WV = rated DC working voltage												
	>100V	CWV ≤ 1000 μF: I= 0.03 CWV + 15uA; C= rated capacitance in uF CWV ≥ 1000 μF: I= 0.02 CWV + 25uA; WV= rated DC working voltage in V												
Dissipation Factor (Tan δ, at 20°C 120Hz)	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Tan δ		0.23	0.20	0.16	0.14	0.12	0.10	0.09	0.08	0.12	0.17	0.20	0.25
	For capacitors whose capacitance exceeds 1,000μF, the specification of tan δ is increased by 0.02 for every addition of 1,000μF													
Surge Voltage	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Surge voltage (SV)		8	13	20	32	44	63	79	125	200	300	400	500
Low Temperature Characteristics (Imp. ratio @ 120Hz)	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Z(-25°C)/Z(+20°C)	øD<16	6	4	3	3	2	2	2	2	3	8	12	16
		øD≥16	8	6	4	4	3	3	3	3	3	8	12	16
	Z(-40°C)/Z(+20°C)	øD<16	10	8	6	6	4	3	3	3	4	10	16	20
		øD≥16	18	16	12	10	8	8	6	6	4	10	16	20
Load Test	When returned to +20°C after 2,000 hours application of working voltage at +85°C, the capacitor will meet the following limits: Capacitance change is ≤ ±20% of initial value; tan δ is < 200% of specified value; leakage current is within specified value													
Shelf Life Test	When returned to +20°C after 1,000 hours at +85°C with no voltage applied, the capacitor will meet the following limits: Capacitance change is ≤ ±20% of initial value; tan δ is < 200% of specified value; leakage current is within specified value													

PART NUMBERING SYSTEM

1 4 0	–	X R L	1 6 V	1 0 0	–	R C
Prefix		Series	Voltage Actual Value	Capacitance (μF) Actual Value		Suffix RoHS Compliant

RIPPLE CURRENT AND FREQUENCY MULTIPLIERS

Capacitance (μF)	Frequency (Hz)				
	60 (50)	120	500	1K	≥10K
<100	0.70	1.0	1.30	1.40	1.50
100 ~ 1000	0.75	1.0	1.20	1.30	1.35
>1000	0.80	1.0	1.10	1.12	1.15

RIPPLE CURRENT AND TEMPERATURE MULTIPLIERS

Temperature (°C)	<50	70	85
Multiplier	1.78	1.4	1.0

XICON PASSIVE COMPONENTS • (800) 628-0544

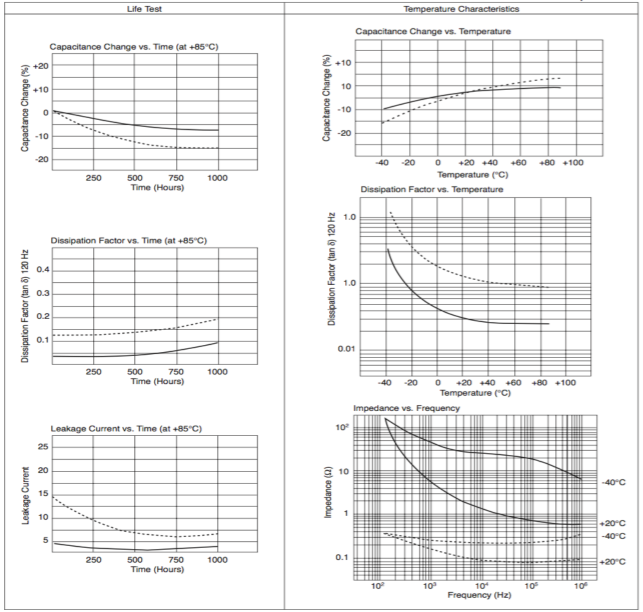
XC-600178 Date Revised: 1/8/07

Specifications are subject to change without notice. No liability or warranty implied by this information. Environmental compliance based on producer documentation.

XICON

Miniature Aluminum Electrolytic Capacitors — XRL Series

TYPICAL PERFORMANCE CHARACTERISTICS



XICON PASSIVE COMPONENTS • (800) 628-0544

XC-600178 Date Revised: 1/8/07

Specifications are subject to change without notice. No liability or warranty implied by this information. Environmental compliance based on producer documentation.

Datasheets for Datasets

Datasheets

We need datasheets for APIs, pretrained models and datasets

Datasheets

Need to have information regarding standard operating characteristics, recommended usage, how the dataset was gathered etc..

Datasheets

E.g. we do not expect Face API to accurately identify the gender of young children.

Datasheets for Datasets

Are there disclaimers in case someone uses this API for something it was not intended for?

E.g. in electronics, disclaimers for use of components in high stakes scenarios like nuclear power plants, life support…

Datasheets for Datasets

What are some of the characteristics of the data it was trained on?

Datasheets for Datasets

E.g. distribution of age, skin types, geography, gender

Datasheets for Datasets

Motivation for Dataset Creation

Why was the dataset created? (e.g., was there a specific task in mind? was there a specific gap that needed to be filled?)

Labeled Faces in the Wild was created to provide images that can be used to study face recognition in the unconstrained setting where image characteristics (such as pose, illumination, resolution, focus), subject demographic makeup (such as age, gender, race) or appearance (such as hairstyle, makeup, clothing) cannot be controlled. The dataset was created for the specific task of pair matching: given a pair of images each containing a face, determine whether or not the images are of the same person.[1]

What (other) tasks could the dataset be used for?

The LFW dataset can be used for the face identification problem. Some researchers have developed protocols to use the images in the LFW dataset for face identification.[2]

Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)?

Papers using this dataset and the specified evaluation protocol are listed in http://vis-www.cs.umass.edu/lfw/results.html

Who funded the creation of the dataset?

The building of the LFW database was supported by a United States National Science Foundation CAREER Award.

Dataset Composition

What are the instances? (that is, examples; e.g., documents, images, people, countries) Are there multiple types of instances? (e.g., movies, users, ratings; people, interactions between them; nodes, edges)

Each instance is a pair of images labeled with the name of the person in the image. Some images contain more than one face. The labeled face is the one containing the central pixel of the image—other faces should be ignored as "background".

Are relationships between instances made explicit in the data (e.g., social network links, user/movie ratings, etc.)?

There are no known relationships between instances except for the fact that they are all individuals who appeared in news sources on line, and some individuals appear in multiple pairs.

How many instances are there? (of each type, if appropriate)?

The dataset consists of 13,233 face images in total of 5749 unique individuals. 1680 of these subjects have two or more images and 4069 have single ones.

[1]All information in this datasheet is taken from one of five sources. Any errors that were introduced from these sources are our fault.
Original paper: http://www.cs.cornell.edu/people/pabo/movie-review-data/; LFW survey: http://vis-www.cs.umass.edu/lfw/lfw.pdf; Paper measuring LFW demographic characteristics : http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf; LFW website: http://vis-www.cs.umass.edu/lfw/.

[2]Unconstrained face recognition: Identifying a person of interest from a media collection: http://biometrics.cse.msu.edu/Publications/Face/BestRowdenetal_UnconstrainedFaceRecognition_TechReport_MSU-CSE-14-1.pdf

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)? Features/attributes? Is there a label/target associated with instances? If the instances related to people, are subpopulations identified (e.g., by age, gender, etc.) and what is their distribution?

Each instance contains a pair of images that are 250 by 250 pixels in JPEG 2.0 format. Each image is accompanied by a label indicating the name of the person in the image. While subpopulation data was not available at the initial release of the dataset, a subsequent paper[3] reports the distribution of images by age, race and gender. Table 2 lists these results.

Is everything included or does the data rely on external resources? (e.g., websites, tweets, datasets) If external resources, a) are there guarantees that they will exist, and remain constant, over time; b) is there an official archival version; c) are there access restrictions or fees?

Everything is included in the dataset.

Are there recommended data splits and evaluation measures? (e.g., training, development, testing; accuracy or AUC)

The dataset comes with specified train/test splits such that none of the people in the training split are in the test split and vice versa. The data is split into two views, View 1 and View 2. View 1 consists of a training subset (pairsDevTrain.txt) with 1100 pairs of matched and 1100 pairs of mismatched images, and a test subset (pairsDevTest.txt) with 500 pairs of matched and mismatched images. Practitioners can train an algorithm on the training set and test on the test set, repeating as often as necessary. Final performance results should be reported on View 2 which consists of 10 subsets of the dataset. View 2 should only be used to test the performance of the final model. We recommend reporting performance on View 2 by using leave-one-out cross validation, performing 10 experiments. That is, in each experiment, 9 subsets should be used as a training set and the 10^{th} subset should be used for testing. At a minimum, we recommend reporting the **estimated mean accuracy,** $\hat{\mu}$ and the **standard error of the mean:** S_E for View 2.
$\hat{\mu}$ is given by:

$$\hat{\mu} = \frac{\sum_{i=1}^{10} p_i}{10} \tag{1}$$

where p_i is the percentage of correct classifications on View 2 using subset i for testing. S_E is given as:

$$S_E = \frac{\hat{\sigma}}{\sqrt{10}} \tag{2}$$

Where $\hat{\sigma}$ is the estimate of the standard deviation, given by:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{10} (p_i - \hat{\mu})^2}{9}} \tag{3}$$

The multiple-view approach is used instead of a traditional train/validation/test split in order to maximize the amount of data available for training and testing.

[3]http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf

Training Paradigms: There are two training paradigms that can be used with our dataset. Practitioners should specify the training paradigm they used while reporting results.

- **Image-Restricted Training** This setting prevents the experimenter from using the name associated with each image during training and testing. That is, the only available information is whether or not a pair of images consist of the same person, not who that person is. This means that there would be no simple way of knowing if there are multiple pairs of images in the train/test set that belong to the same person. Such inferences, however, might be made by comparing image similarity/equivalence (rather than comparing names). Thus, to form training pairs of matched and mismatched images for the same person, one can use image equivalence to add images that consist of the same person.

  The files pairsDevTrain.txt and pairsDevTest.txt support image-restricted uses of train/test data. The file pairs.txt in View 2 supports the image-restricted use of training data.

- **Unrestricted Training** In this setting, one can use the names associated with images to form pairs of matched and mismatched images for the same person. The file people.txt in View 2 of the dataset contains subsets of of people along with images for each subset. To use this paradigm, matched and mismatched pairs of images should be formed from images in the same subset. In View 1, the files peopleDevTrain.txt and peopleDevTest.txt can be used to create arbitrary pairs of matched/mismatched images for each person. The unrestricted paradigm should only be used to create training data and not for performance reporting. The test data, which is detailed in the file pairs.txt, should be used to report performance. We recommend that experimenters first use the image-restricted paradigm and move to the unrestricted paradigm if they believe that their algorithm's performance would significantly improve with more training data. While reporting performance, it should be made clear which of these two training paradigms were used for particular test result.

What experiments were initially run on this dataset? Have a summary of those results.

The dataset was originally released without reported experimental results but many experiments have been run on it since then.

Any other comments?

Table 1 summarizes some dataset statistics and Figure 1 shows examples of images. Most images in the dataset are color, a few are black and white.

Property	Value
Database Release Year	2007
Number of Unique Subjects	5649
Number of total images	13,233
Number of individuals with 2 or more images	1680
Number of individuals with single images	4069
Image Size	250 by 250 pixels
Image format	JPEG
Average number of images per person	2.30

Table 1. A summary of dataset statistics extracted from the original paper: Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.

Demographic Characteristic	Value
Percentage of female subjects	22.5%
Percentage of male subjects	77.5%
Percentage of White subjects	83.5%
Percentage of Black subjects	8.47%
Percentage of Asian subjects	8.03%
Percentage of people between 0-20 years old	1.57%
Percentage of people between 21-40 years old	31.63%
Percentage of people between 41-60 years old	45.58%
Percentage of people over 61 years old	21.2%

Table 2. Demographic characteristics of the LFW dataset as measured by Han, Hu, and Anil K. Jain. *Age, gender and race estimation from unconstrained face images.* Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, MSU Tech. Rep.(MSU-CSE-14-5) (2014).

Data Collection Process

How was the data collected? (e.g., hardware apparatus/sensor, manual human curation, software program, software interface/API)

The raw images for this dataset were obtained from the Faces in the Wild database collected by Tamara Berg at Berkeley[4]. The images in this database were gathered from news articles on the web using software to crawl news articles.

Who was involved in the data collection process? (e.g., students, crowdworkers) and how were they compensated (e.g., how much were crowdworkers paid)?

Unknown

Over what time-frame was the data collected? Does the collection time-frame match the creation time-frame of the instances?

Unknown

Datasheets for Datasets

Datasheets for Datasets

Motivation for Dataset Creation

Why was the dataset created? (e.g., was there a specific task in mind? was there a specific gap that needed to be filled?)
Labeled Faces in the Wild was created to provide images that can be used to study face recognition in the unconstrained setting where image characteristics (such as pose, illumination, resolution, focus), subject demographic makeup (such as age, gender, race) or appearance (such as hairstyle, makeup, clothing) cannot be controlled. The dataset was created for the specific task of pair matching: given a pair of images each containing a face, determine whether or not the images are of the same person.[1]

What (other) tasks could the dataset be used for?
The LFW dataset can be used for the face identification problem. Some researchers have developed protocols to use the images in the LFW dataset for face identification.[2]

Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)?
Papers using this dataset and the specified evaluation protocol are listed in http://vis-www.cs.umass.edu/lfw/results.html

Who funded the creation of the dataset?
The building of the LFW database was supported by a United States National Science Foundation CAREER Award.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)? Features/attributes? Is there a label/target associated with instances? If the instances related to people, are subpopulations identified (e.g., by age, gender, etc.) and what is their distribution?
Each instance contains a pair of images that are 250 by 250 pixels in JPEG 2.0 format. Each image is accompanied by a label indicating the name of the person in the image. While subpopulation data was not available at the initial release of the dataset, a subsequent paper[3] reports the distribution of images by age, race and gender. Table 2 lists these results.

Is everything included or does the data rely on external resources? (e.g., websites, tweets, datasets) If external resources, a) are there guarantees that they will exist, and remain constant, over time; b) is there an official archival version; c) are there access restrictions or fees?
Everything is included in the dataset.

Are there recommended data splits and evaluation measures? (e.g., training, development, testing; accuracy or AUC)
The dataset comes with specified train/test splits such that none of the people in the training split are in the test split and vice versa. The data is split into two views, View 1 and View 2. View 1 consists of a training subset (pairsDevTrain.txt) with 1100 pairs of matched and 1100 pairs of mismatched images, and a test subset (pairsDevTest.txt) with 500 pairs of matched and mismatched images. Practitioners can train an algorithm on the training set and test on the test set, repeating as often as necessary. Final

Datasheets for Datasets

Datasheets for Datasets

Dataset Composition

What are the instances? (that is, examples; e.g., documents, images, people, countries) Are there multiple types of instances? (e.g., movies, users, ratings; people, interactions between them; nodes, edges)

Each instance is a pair of images labeled with the name of the person in the image. Some images contain more than one face. The labeled face is the one containing the central pixel of the image—other faces should be ignored as "background".

Are relationships between instances made explicit in the data (e.g., social network links, user/movie ratings, etc.)?

There are no known relationships between instances except for the fact that they are all individuals who appeared in news sources on line, and some individuals appear in multiple pairs.

How many instances are there? (of each type, if appropriate)?

The dataset consists of 13,233 face images in total of 5749 unique individuals. 1680 of these subjects have two or more images and 4069 have single ones.

Data Collection Process

How was the data collected? (e.g., hardware apparatus/sensor, manual human curation, software program, software interface/API)

The raw images for this dataset were obtained from the Faces in the Wild database collected by Tamara Berg at Berkeley[4]. The images in this database were gathered from news articles on the web using software to crawl news articles.

Who was involved in the data collection process? (e.g., students, crowdworkers) and how were they compensated (e.g., how much were crowdworkers paid)?

Unknown

Over what time-frame was the data collected? Does the collection time-frame match the creation time-frame of the instances?

Unknown

Datasheets for Datasets

Datasheets for Datasets

If the dataset is a sample, then what is the population? What was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? Is the sample representative of the larger set (e.g., geographic coverage)? If not, why not (e.g., to cover a more diverse range of instances)? How does this affect possible uses?

The original Faces in the Wild dataset is a sample of pictures of people appearing in the news on the web. Labeled Faces in the Wild is thus also a sample of images of people found on the news on line. While the intention of the dataset is to have a wide range of demographic (e.g. age, race, ethnicity) and image (e.g. pose, illumination, lighting) characteristics, there are many groups that have few instances (e.g. only 1.57% of the dataset consists of individuals under 20 years old).

Is there information missing from the dataset and why? (this does not include intentionally dropped instances; it might include, e.g., redacted text, withheld documents) Is this data missing because it was unavailable?

Unknown

Data Preprocessing

What preprocessing/cleaning was done? (e.g., discretization or bucket

Legal & Ethical Considerations

If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection? (e.g., datasets that collect writing, photos, interactions, transactions, etc.)

No. The data was crawled from public web sources, and the individuals appeared in news stories. But there was no explicit informing of these individuals that their images were being assembled into a dataset.

If it relates to people, were they told what the dataset would be used for and did they consent? If so, how? Were they provided with any mechanism to revoke their consent in the future or for certain uses?

No (see first question).

If it relates to people, could this dataset expose people to harm or legal action? (e.g., financial social or otherwise) What was done to mitigate or reduce the potential for harm?

There is minimal risk for harm: the data was already public.

If it relates to people, does it unfairly advantage or disadvantage a particular social group? In what ways? How was this mitigated?

Unknown

Datasheets for Datasets

Challenges

What are the questions that can be answered?

E.g. Companies may not want to disclose the exact size of their training data because of competition

Challenges

How can we incentivize the entire field to move in this direction?

Israel Arrests Palestinian Because Facebook Translated 'Good Morning' to 'Attack Them'

No Arabic-speaking police officer read the post before arresting the man, who works at a construction site in a West Bank settlement

NEWS | አማርኛ

ዜና | ኢትዮጵያ | ቪዲዮ | በጣም የተወደዱ

Translate this page?

Options Translate

በቴክኖሎጂ ኢ-ፍትሐዊነትን የምትታገለው ኢትዮጵያዊት

1 ኦክተውበር 2018

ያጋሩ

ትምኒት ገብሩ

ው። ትምኒት ገብሩ ትባላለች። በ 'አርቴፊሻል ኢንተለጀንስ'

ws-45650024#

እንዳያመልጥ

ያልታበሰው የላሊ

ኤልቲቪ ብሮድካስ
ማስጠንቀቂያ እ

በመኪና አደጋ የሞ
ቁጥር አራት ደረሰ

Born in 1983 as a European citizen. It's called "basketball". She is one of the leading innovators in the "artificial intelligence" field. He is a Research Scientist in Australia, now living in California.

It is one of the foundations for the creation of the BlackImIA, a state-of-the-art building that is dedicated to increasing the participation of black women. She is also known for her research of the authorship of inventions and the question of technology and human rights. Here is my interview with **Habib.**

Learn from other industries

Datasheets for Datasets

Automobile

- No stop signs, drivers licenses drunk driving laws, seatbelts etc.
- Lots of accidents
- Crash tests done on male dummies
- Studies show that accidents disproportionately affected women

Automobile

Court opinion and news paper editorials on whether it was evil

Clinical Trials

- Used to be illegal
- Illegal experimentation on vulnerable populations
- Women were not required to be part of clinical trials until recently
- Study shows that 8-10 drugs that were pulled from circulation between 1997-2001 disproportionately affected women

Key Lessons

It took many years for standards to be placed and we are still suffering consequences from bias in automobile design and clinical trials

Newsweek

Cancer Scientists Have Ignored African DNA in the Search for Cures

BY JESSICA WAPNER ON 7/18/18 AT 9:01 AM

Key Lessons

We should learn from other industries while thinking about standardization/documentation

Key Lessons

Work on fairness accountability and transparency does not just involve intellectual interesting mathematical formulations. If you do not do anything in practice to ease the burden on the *subjects* of your research, consider reading about the term *helicopter science*

Questions?