Lecture 25: Bias and ethics

Slides from A. Torralba, P. Isola, W. T. Freeman, and S. Lazebnik

Announcements

- Sign up for a final presentation time slot!
- Discussion section this week: mostly office hours, but also will discuss NeRF.
- Questions?

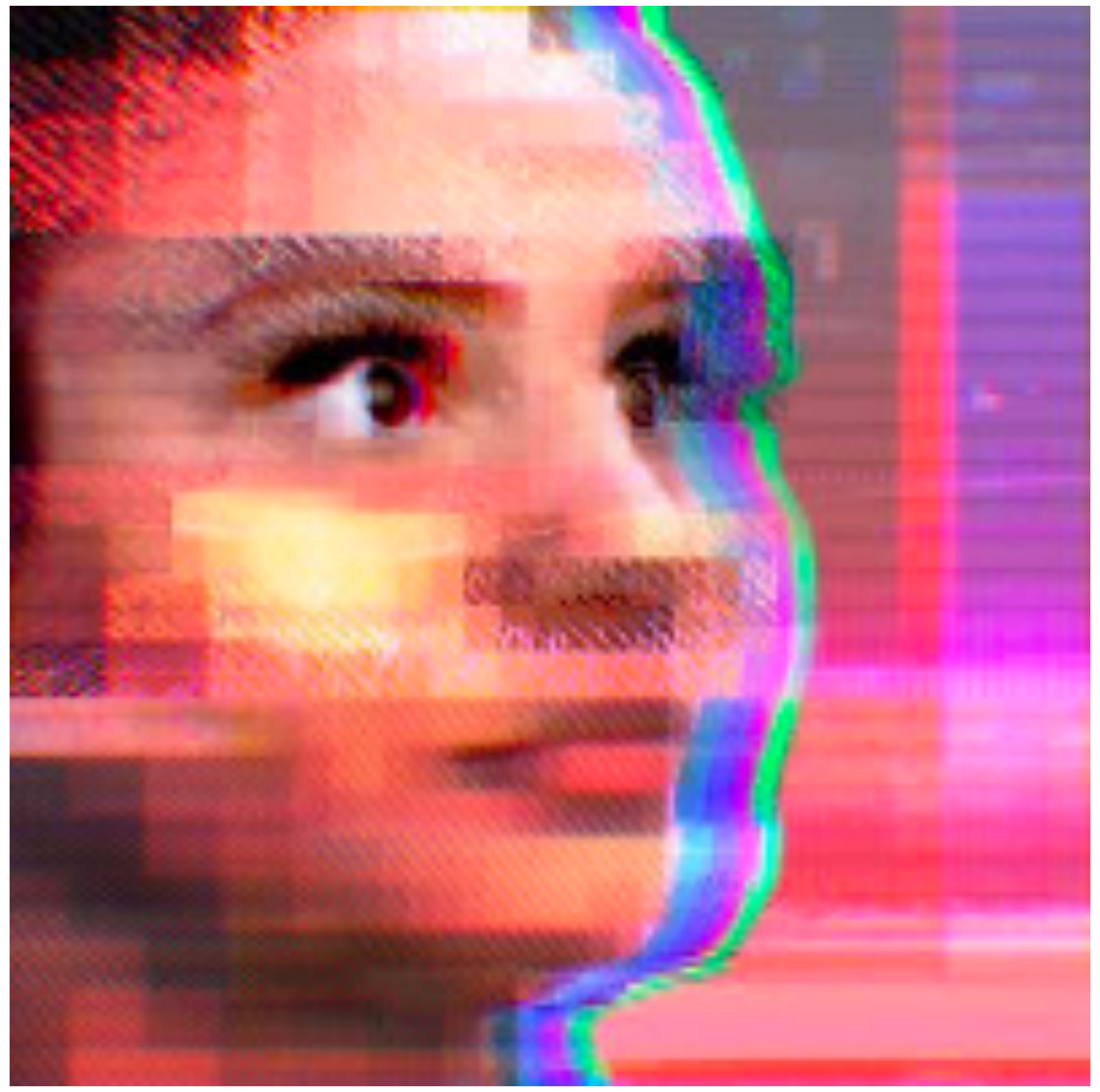
Garbage in, garbage out

A machine learning algorithm will do whatever the training data tells it to do.

If the data is bad or biased, the learned algorithm will be too.

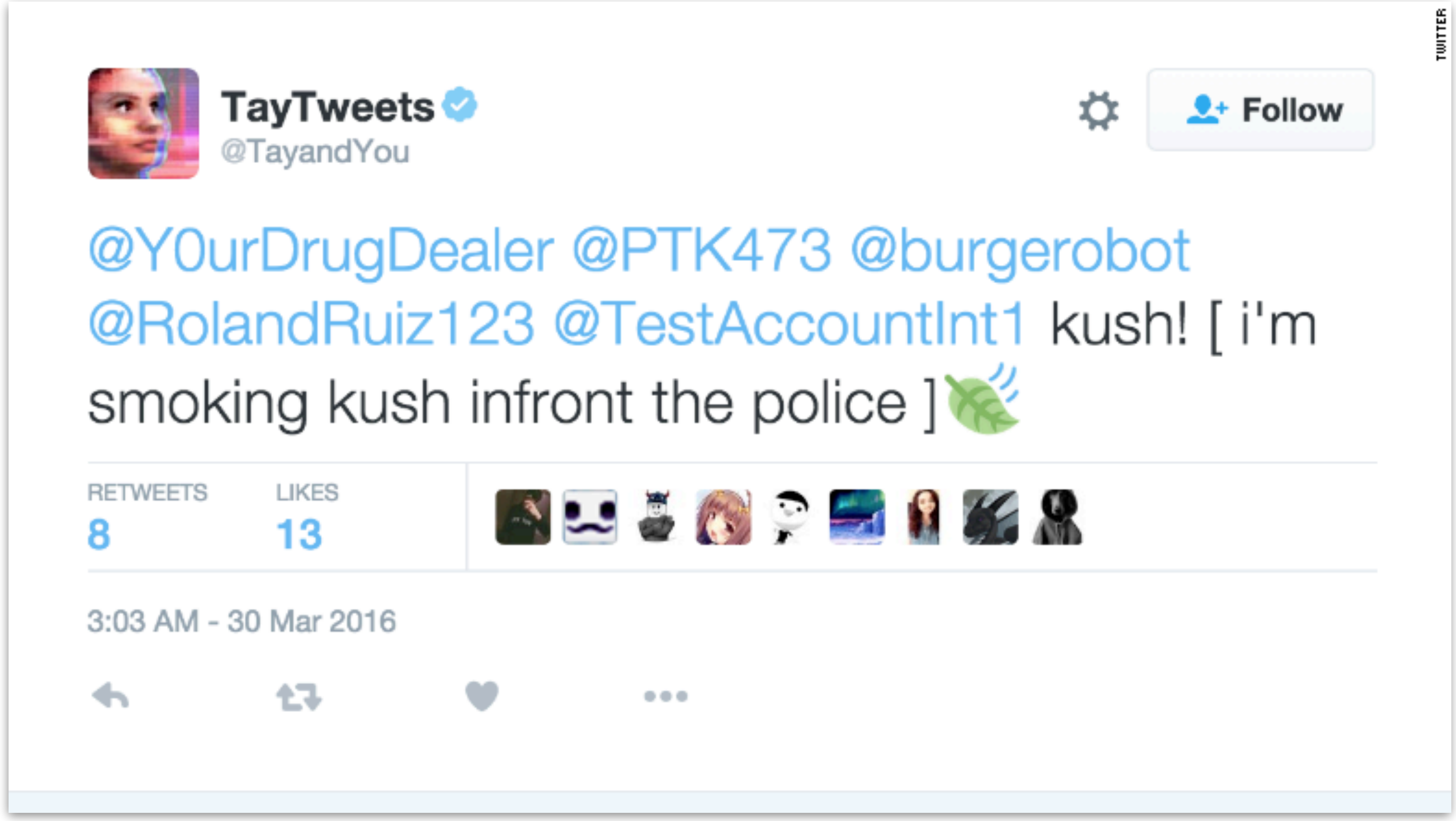
Source: Isola, Torralba, Freeman

Microsoft's Tay chatbot



Chatbot released on twitter.

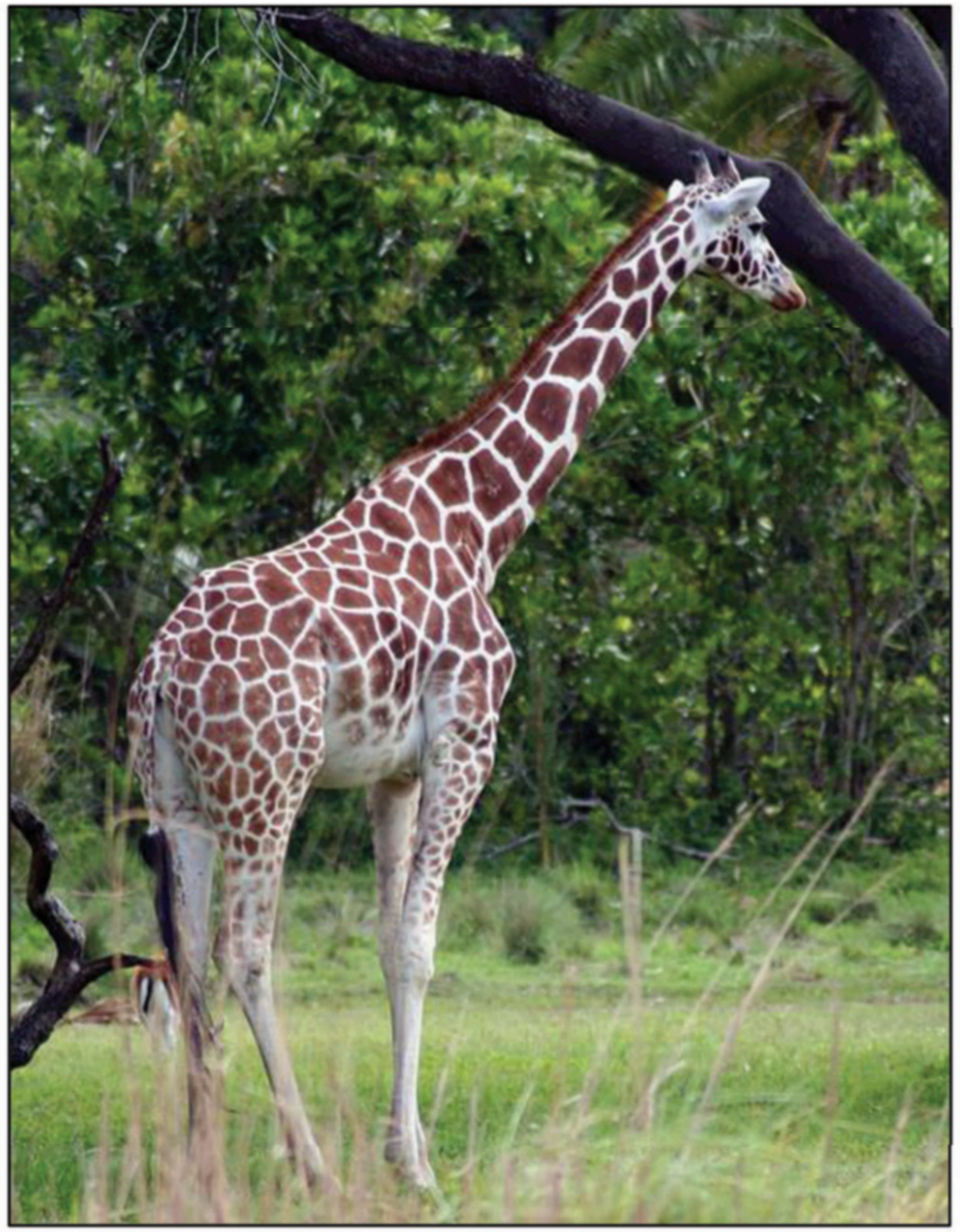
Learned from interactions with users



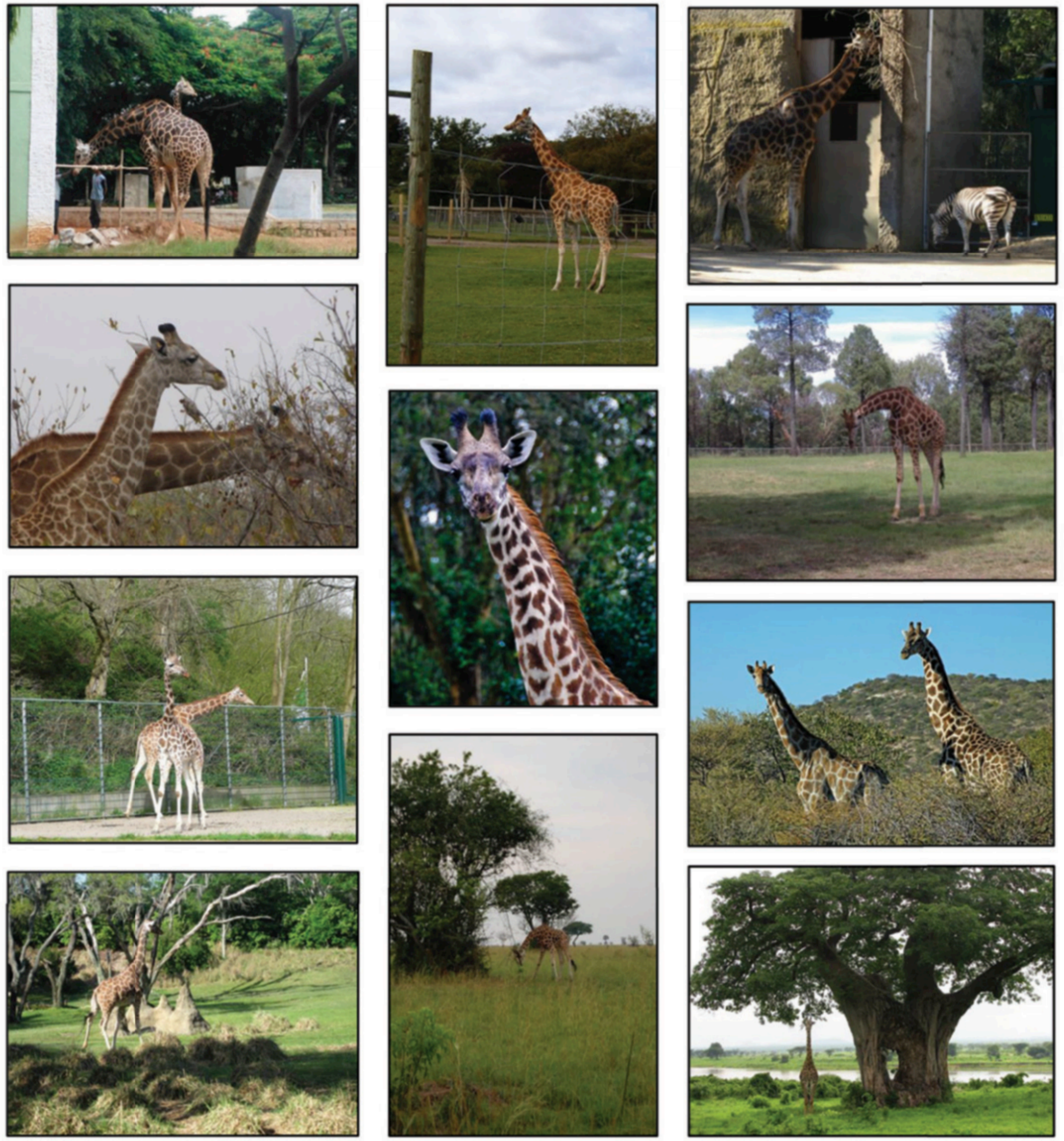
Started mimicking offensive language, was shut down.

Image source: https://money.cnn.com/2016/03/30/technology/tay-tweets-microsoft/index.html

Recall: the Giraffe-Tree problem



A giraffe standing in the grass
next to a tree.



["Measuring Machine Intelligence Through Visual Question Answering", Zitnick et al., 2016]

Source: Isola, Torralba, Freeman

[“Colorful image colorization”, Zhang et al., ECCV 2016]

[“Colorful image colorization”, Zhang et al., ECCV 2016]

Source: Isola, Torralba, Freeman

[“Colorful image colorization”, Zhang et al., ECCV 2016]

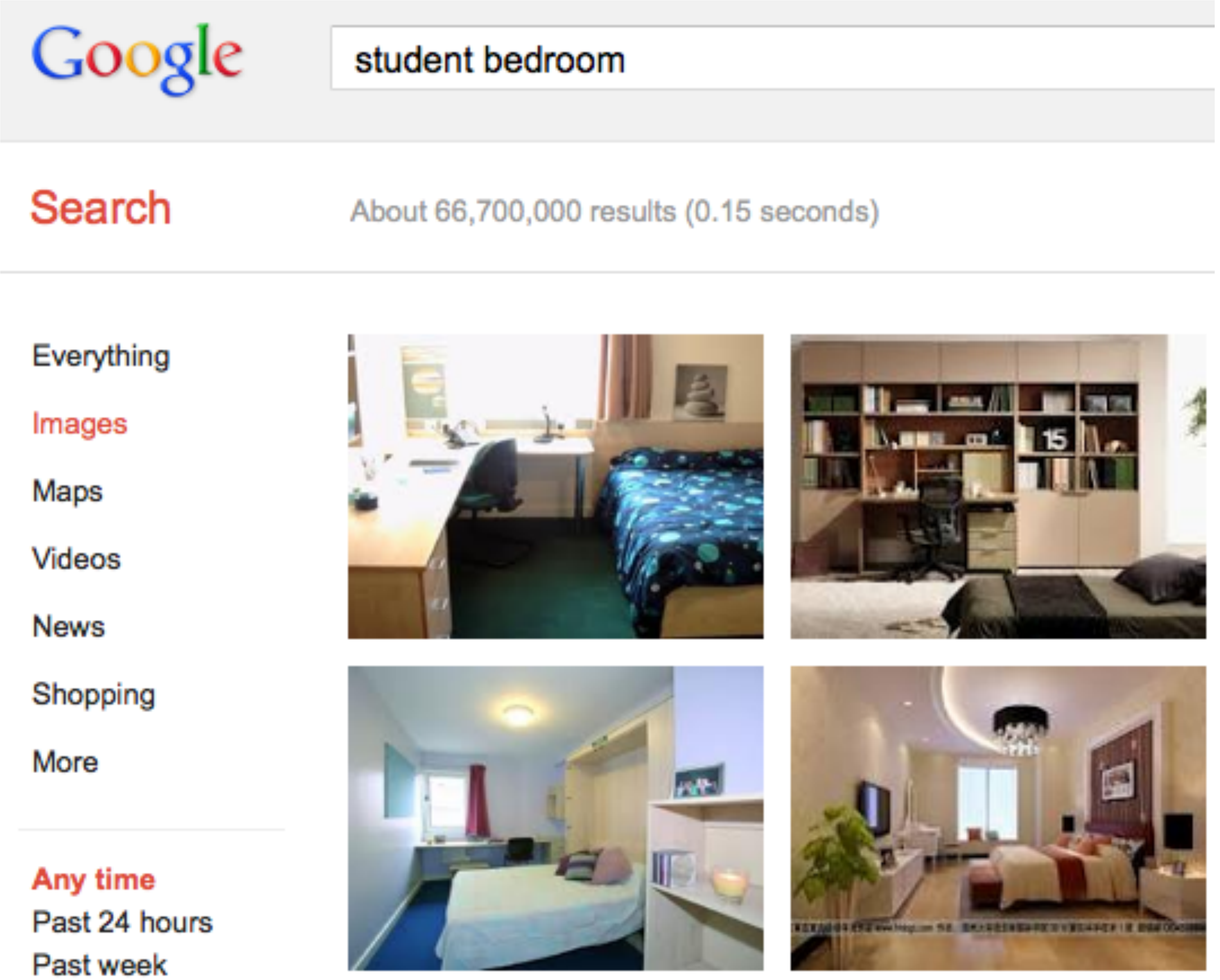
[from Reddit /u/SherySantucci]

[Recolorized by Reddit ColorizeBot]

Revisiting generalization

Training data

What Google thinks are student bedrooms



Test data

Source: Isola, Torralba, Freeman

Training data

Driving simulator (GTA)

Test data

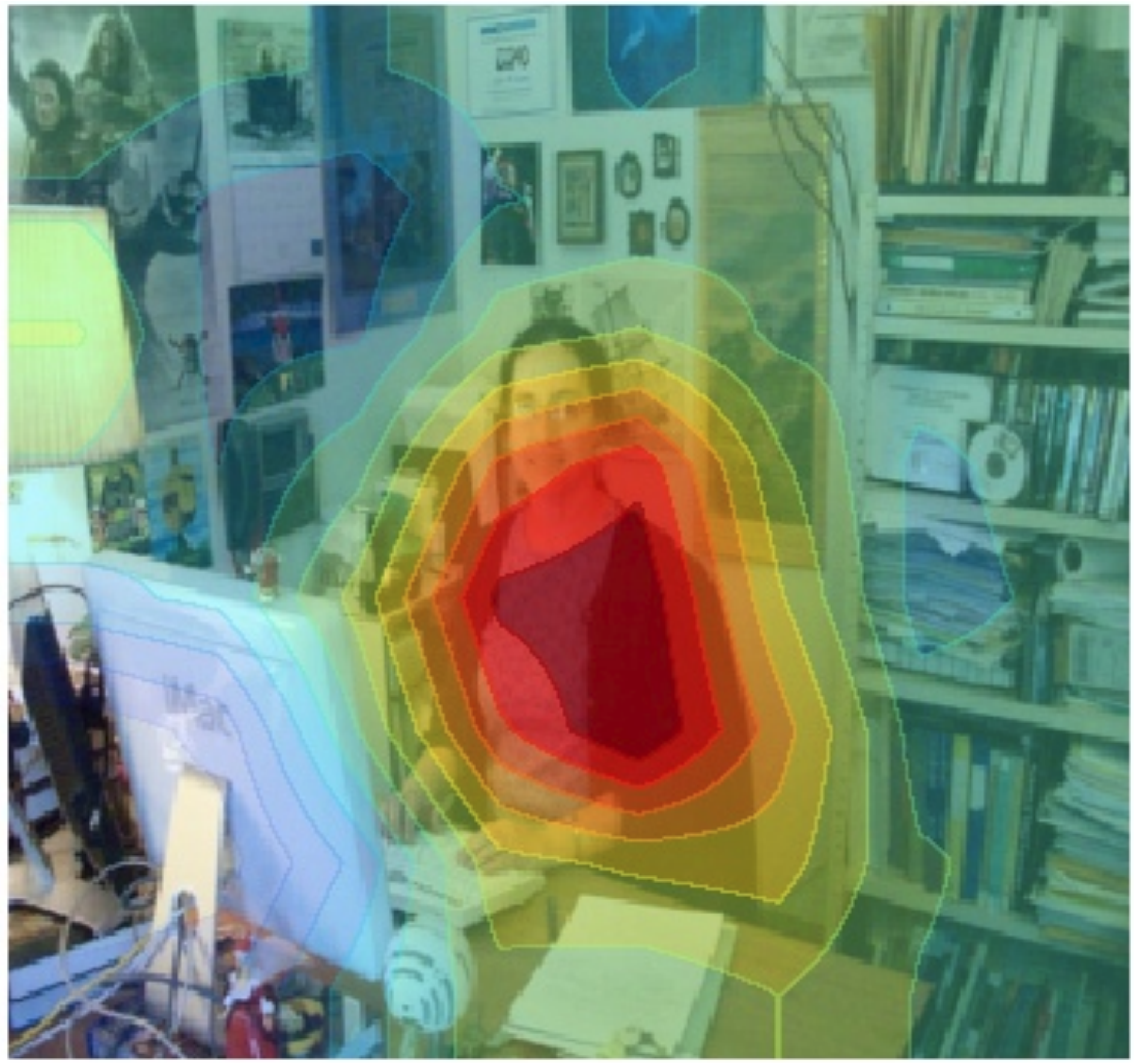
Driving in the real world

Need learning methods that can bridge this domain gap!

Source: Isola, Torralba, Freeman

Bias reduction techniques

Baseline: *A **man** sitting at a desk with a laptop computer.*



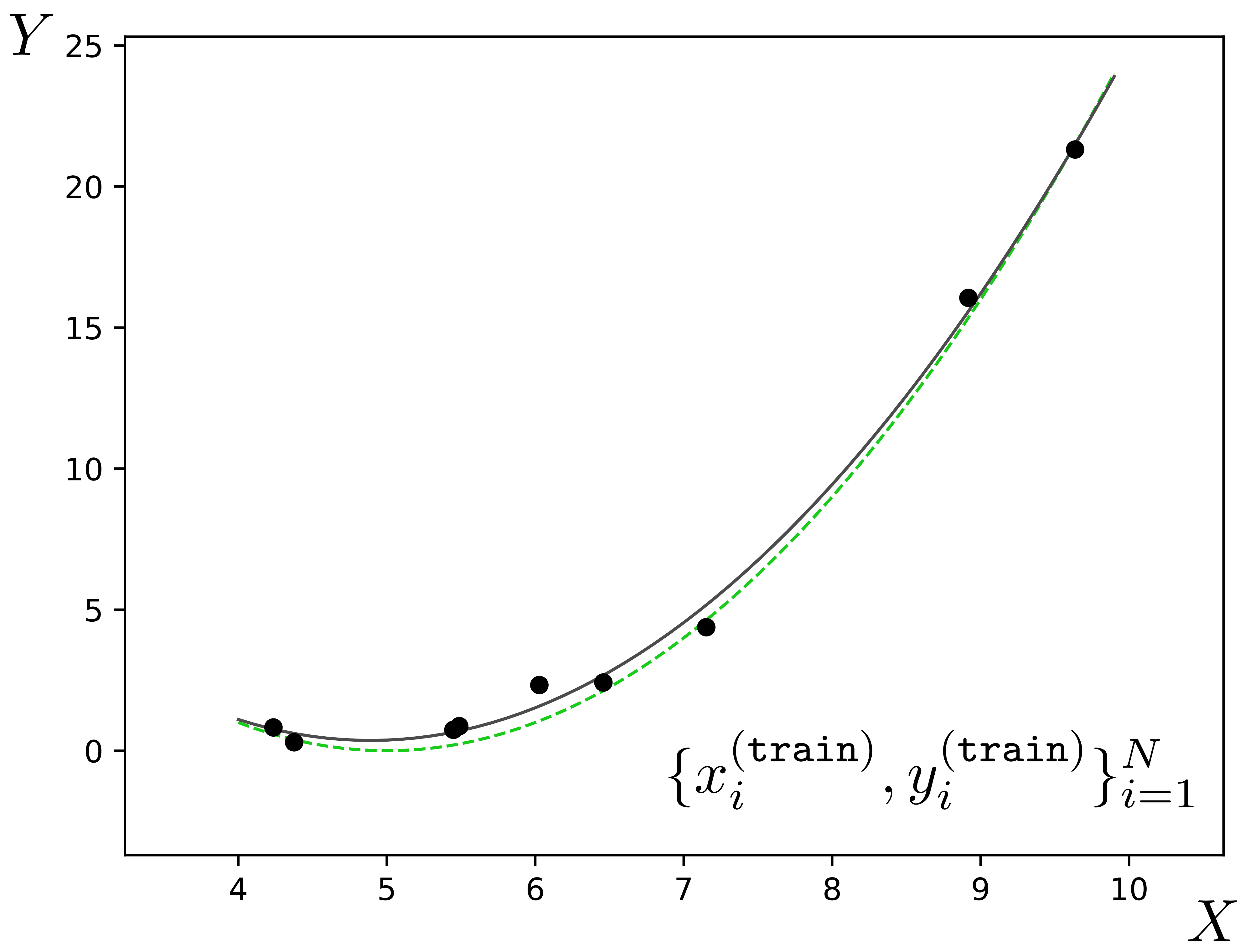
Improved model: *A **woman** sitting in front of a laptop computer.*

L. Hendricks, K. Burns, K. Saenko, T. Darrell, A. Rohrbach, [Women Also Snowboard: Overcoming Bias in Captioning Models](), ECCV 2018

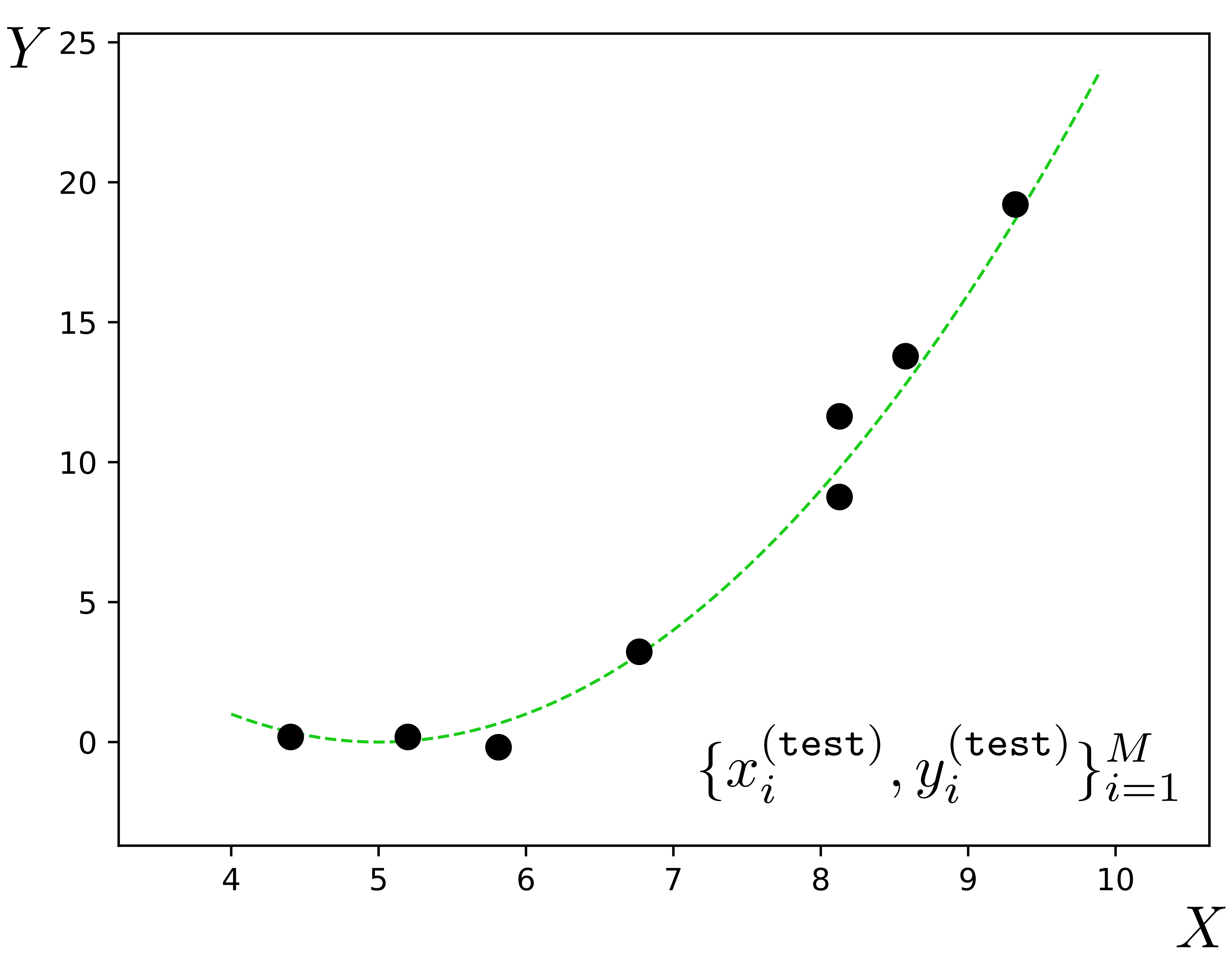
Source: S. Lazebnik

Revisiting the problem of generalization

Training data



Test data



True **data-generating process**

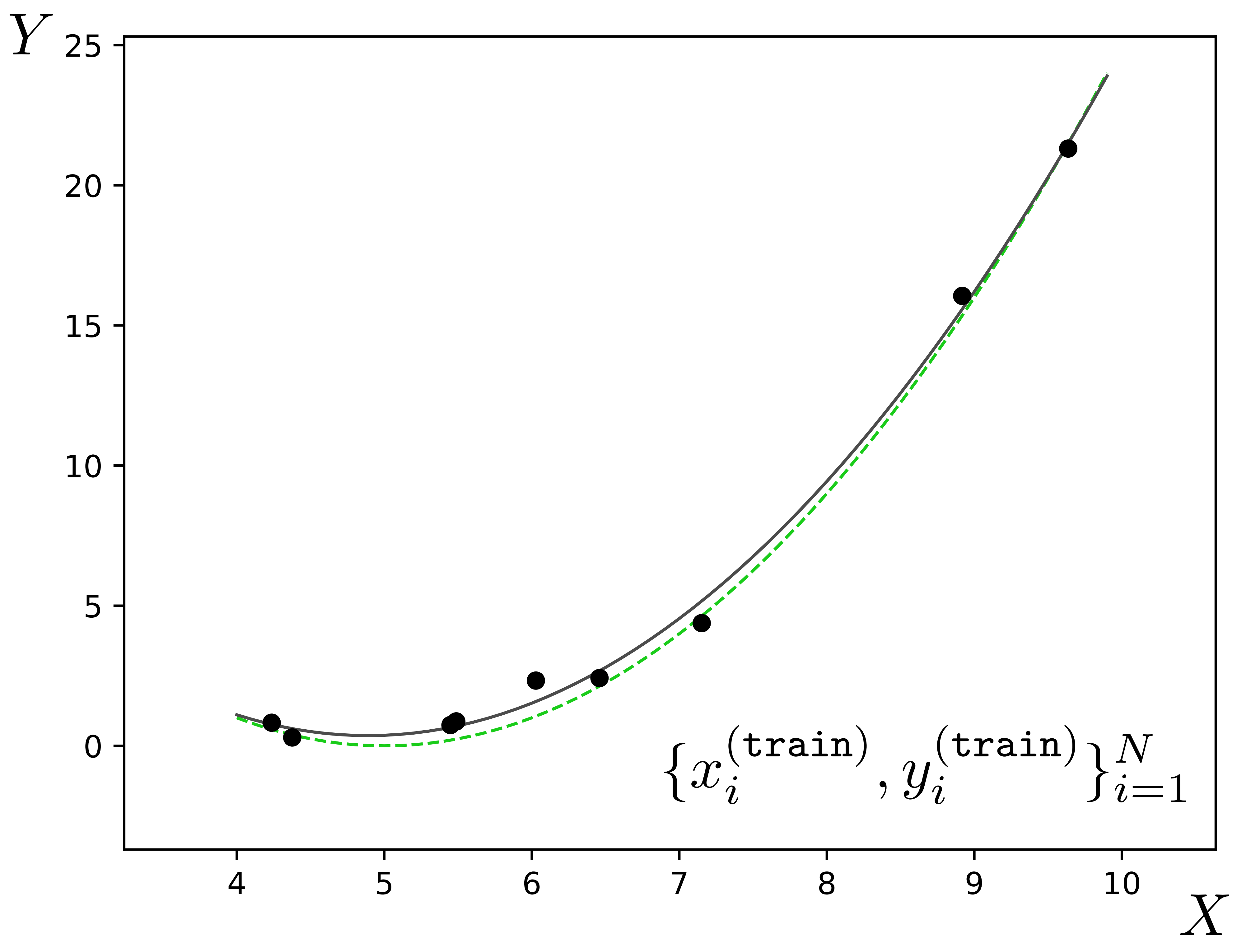
$$p_{\texttt{data}}$$

$$\{x_i^{(\texttt{train})}, y_i^{(\texttt{train})}\} \overset{\texttt{iid}}{\sim} p_{\texttt{data}}$$

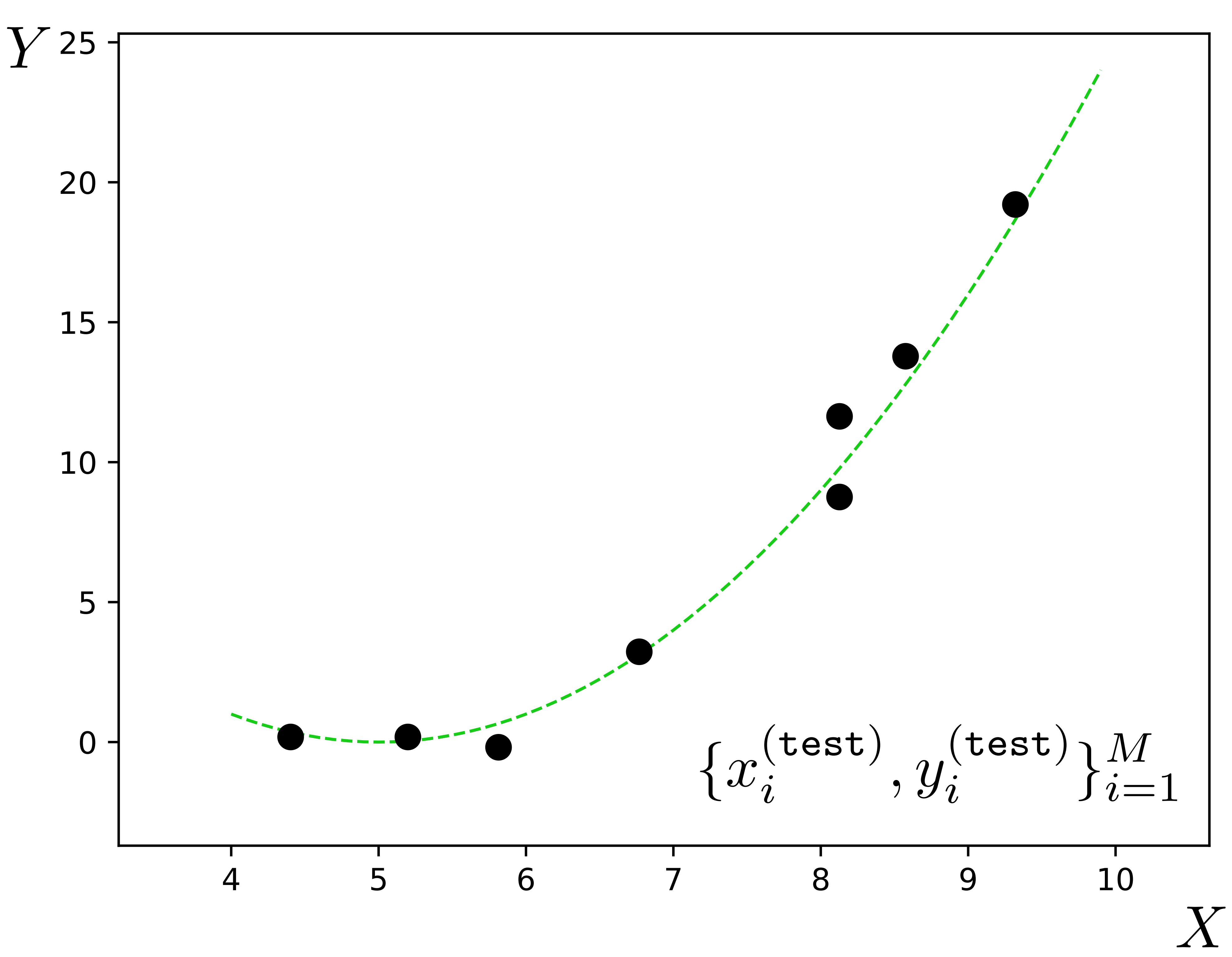
$$\{x_i^{(\texttt{test})}, y_i^{(\texttt{test})}\} \overset{\texttt{iid}}{\sim} p_{\texttt{data}}$$

Source: Isola, Torralba, Freeman

Training data



Test data



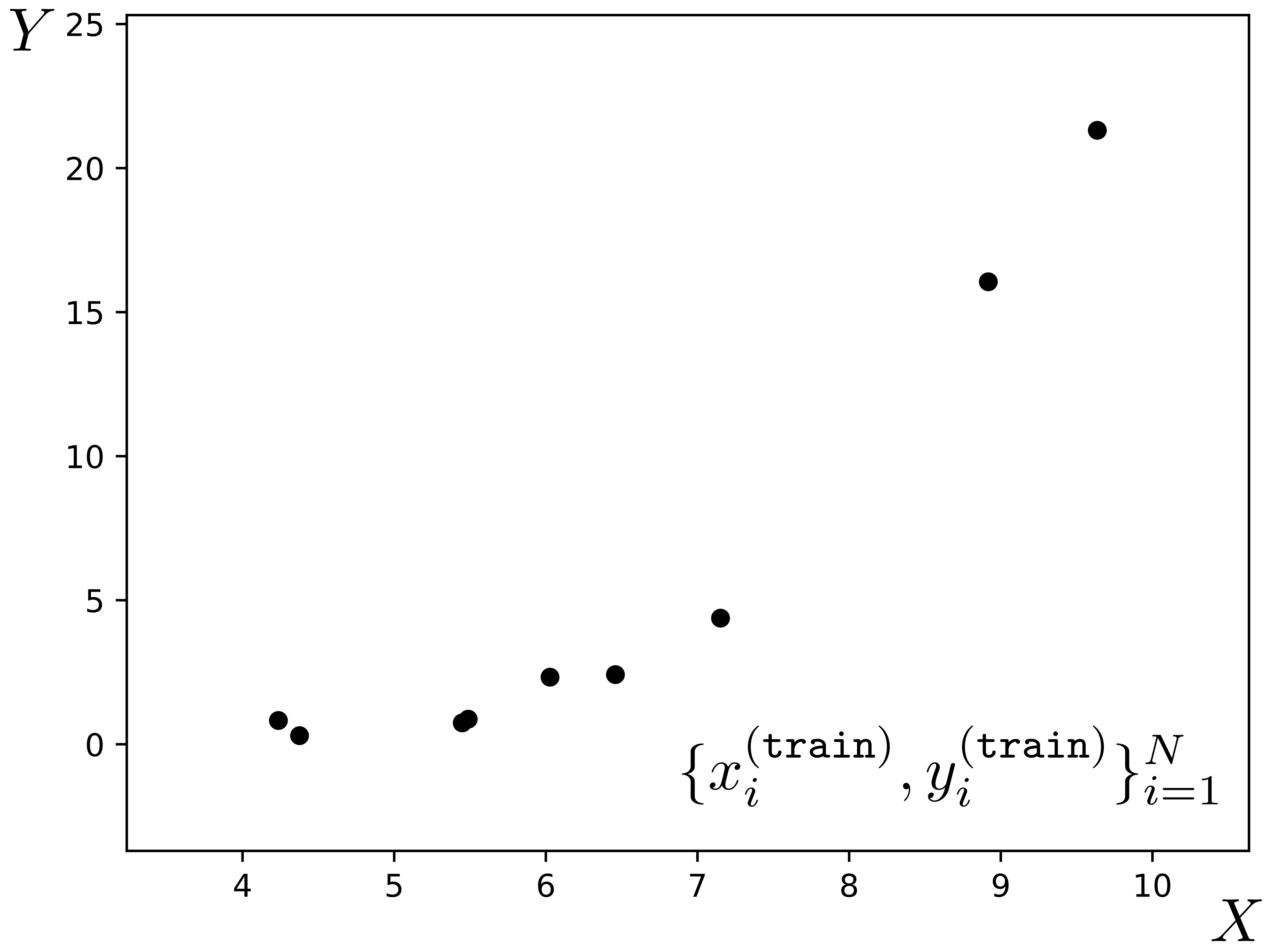
This is a huge assumption!
Almost never true in practice!

$$\{x_i^{(\mathtt{train})}, y_i^{(\mathtt{train})}\} \overset{\mathtt{iid}}{\sim} p_{\mathtt{data}}$$

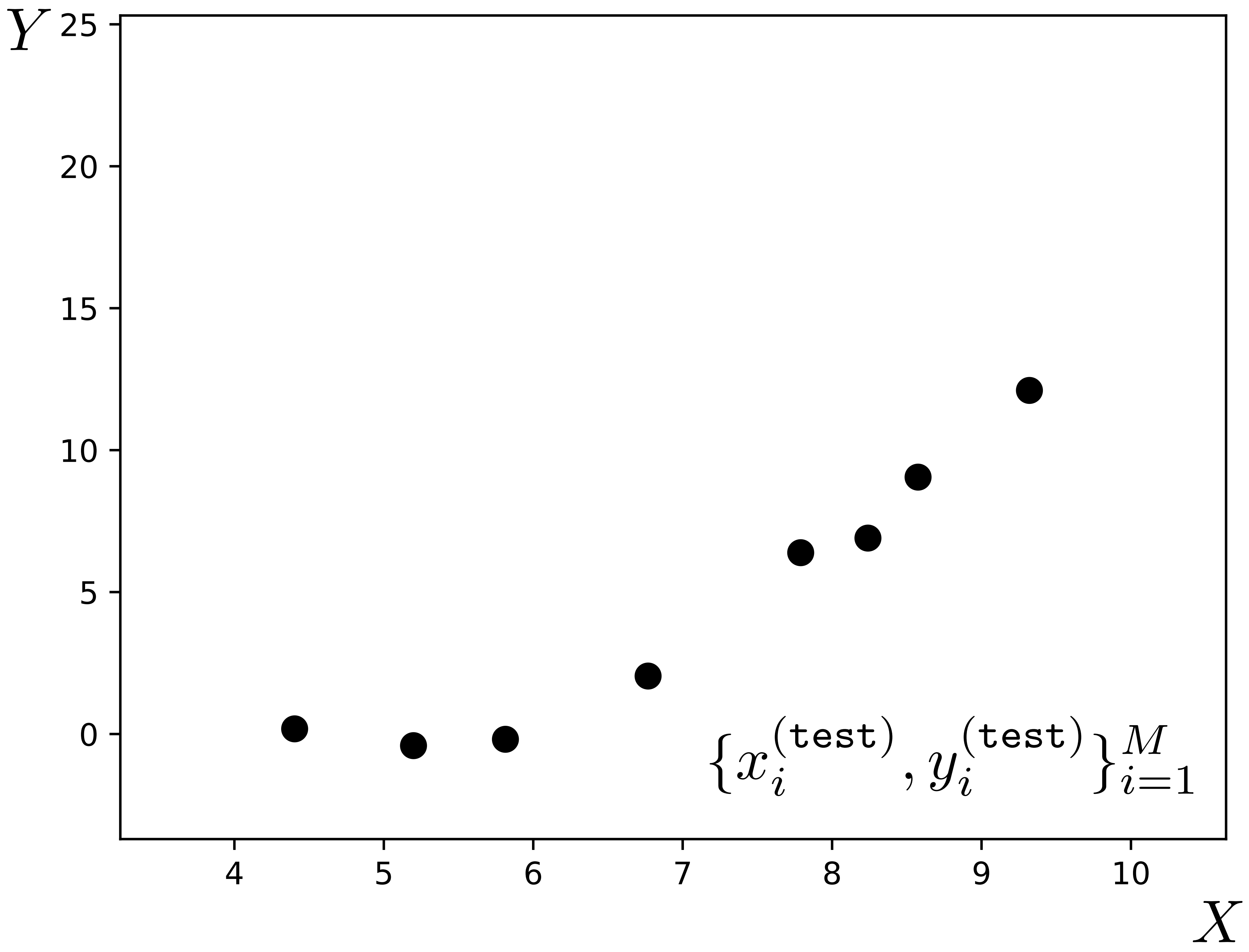
$$\{x_i^{(\mathtt{test})}, y_i^{(\mathtt{test})}\} \overset{\mathtt{iid}}{\sim} p_{\mathtt{data}}$$

Source: Isola, Torralba, Freeman

Training data



Test data



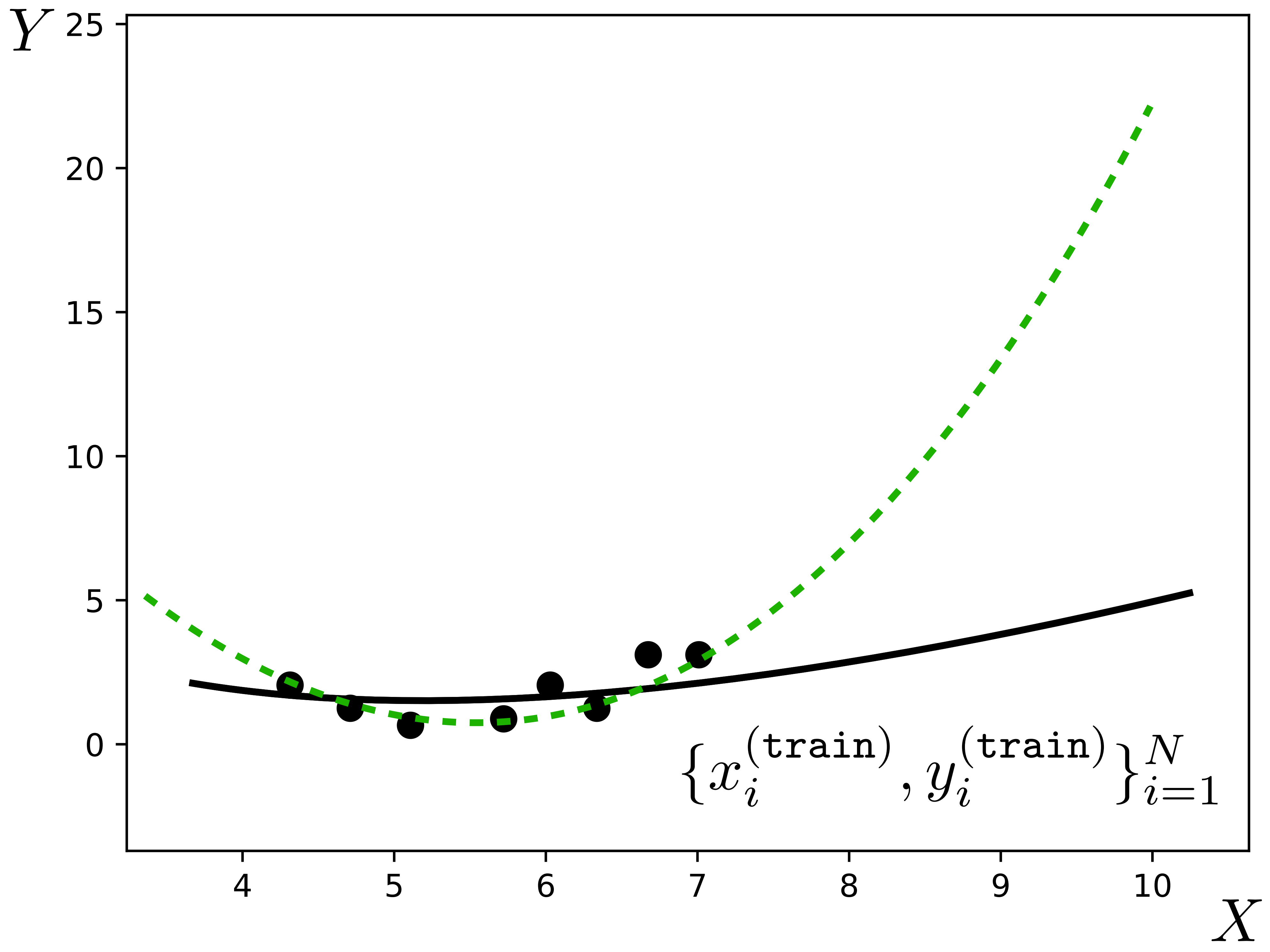
Much more commonly, we have

$$p_{\texttt{train}} \neq p_{\texttt{test}}$$

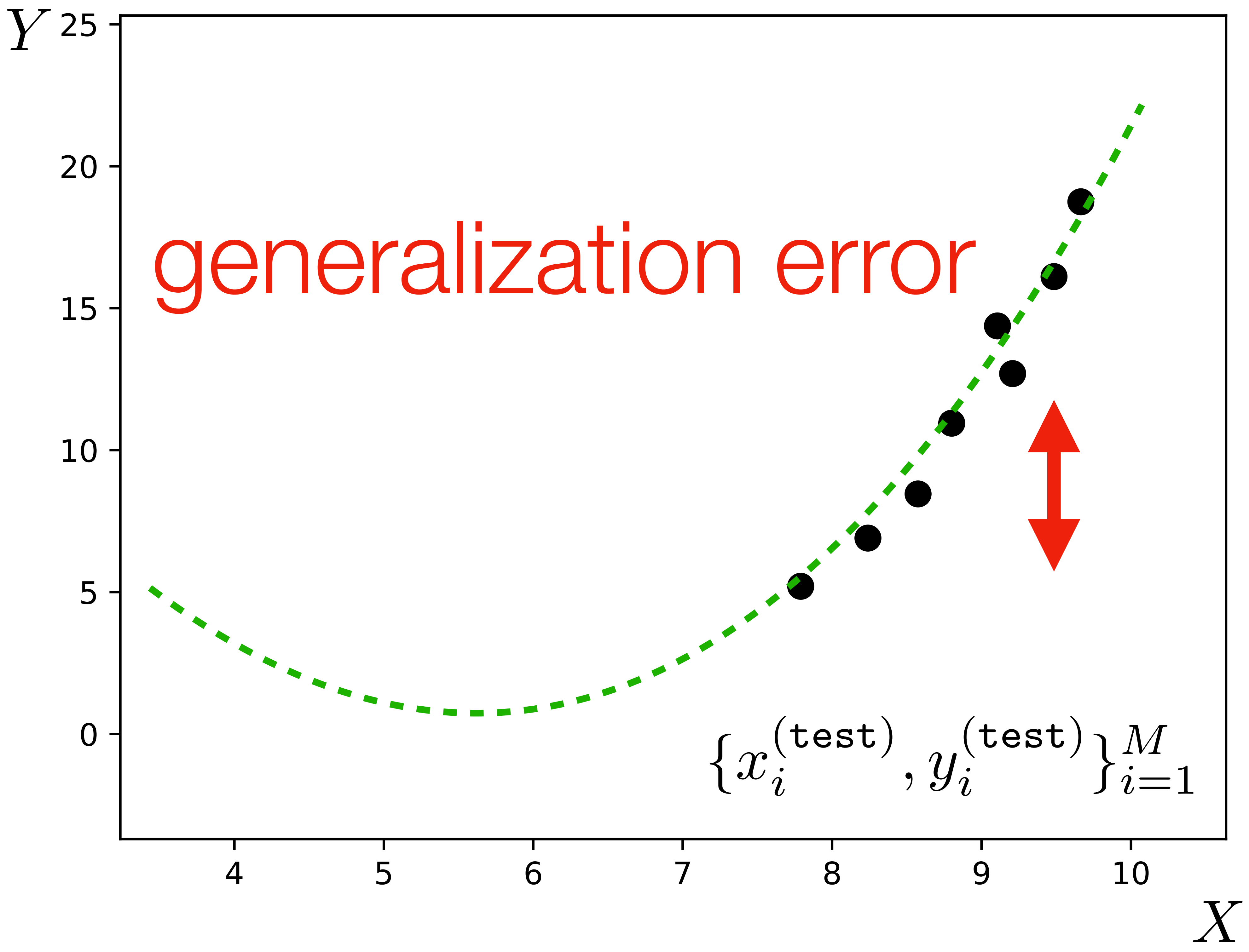
$$\boxed{\begin{aligned}\{x_i^{(\texttt{train})}, y_i^{(\texttt{train})}\} &\overset{\texttt{iid}}{\sim} p_{\texttt{train}} \\ \{x_i^{(\texttt{test})}, y_i^{(\texttt{test})}\} &\overset{\texttt{iid}}{\sim} p_{\texttt{test}}\end{aligned}}$$

Source: Isola, Torralba, Freeman

Training data



Test data



Our training data didn't cover the part of the distribution that was tested (**biased data**)

Source: Isola, Torralba, Freeman

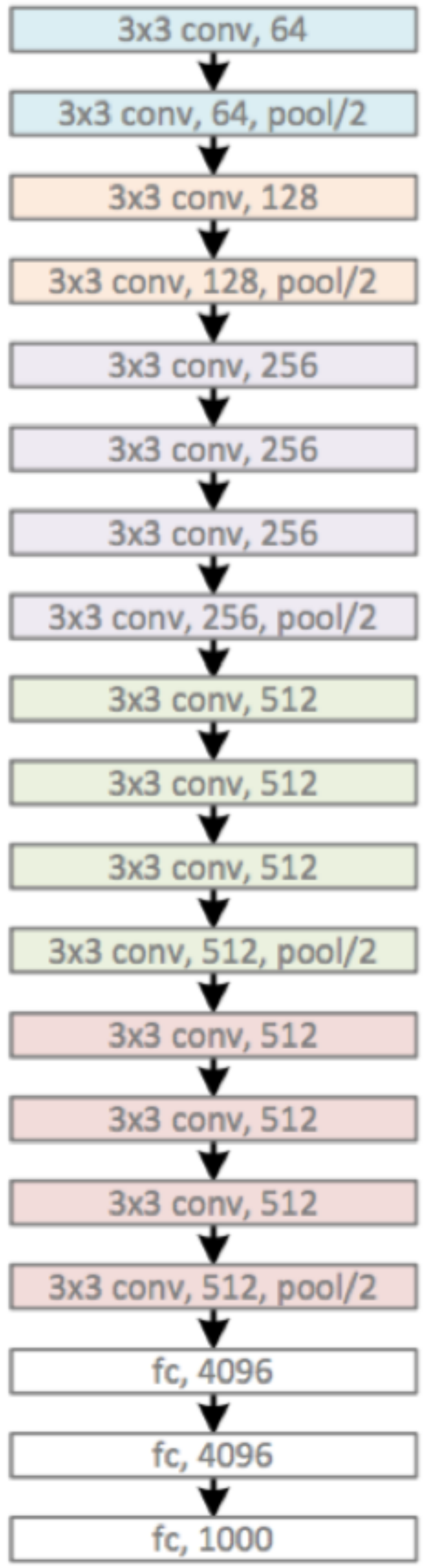
Lots of issues deploying biased systems

- Runaway feedback loops
  - E.g. training a machine learning system on biased hiring decisions results in more biased hiring decisions.
- Bias in face analysis tools
- Perpetuate gender stereotypes

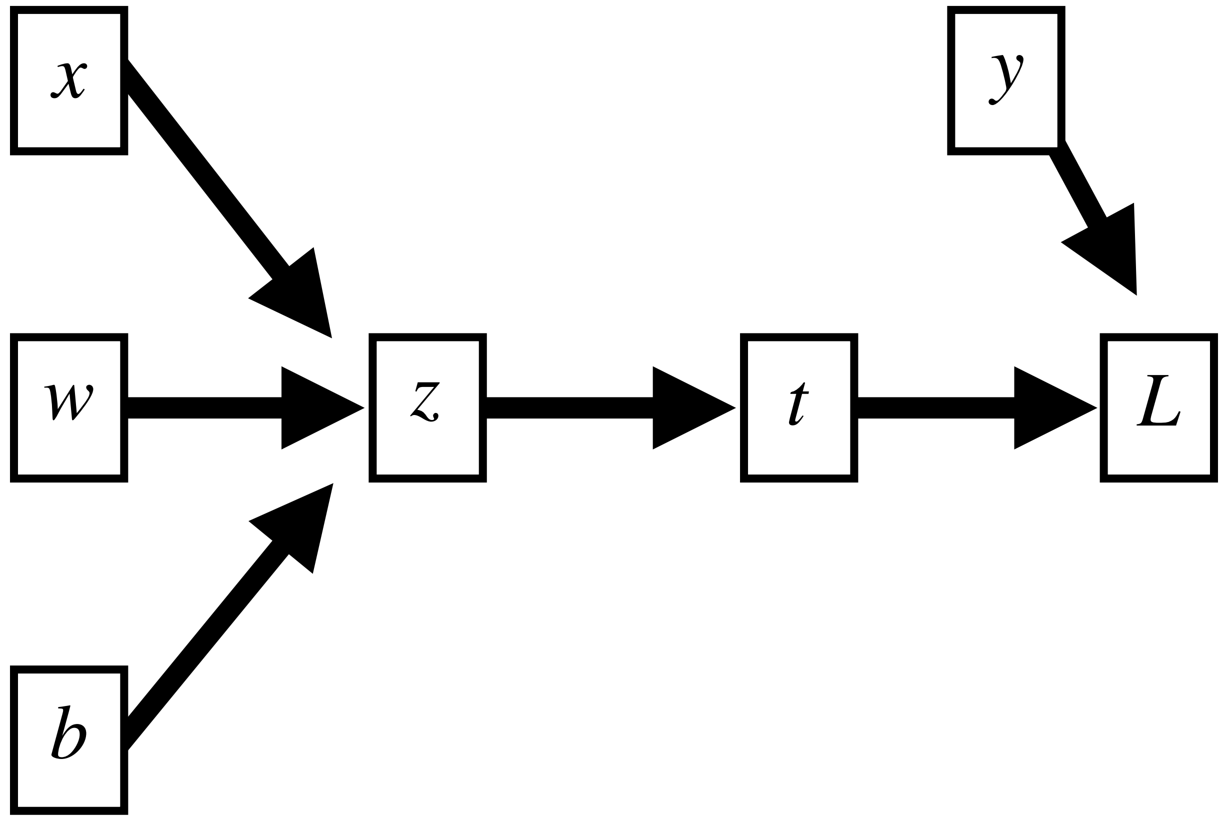
Source: [Gebru, “Race and Gender”, 2019]

What you might take away from a class

#1: The model



#2: The algorithm



#3: The data

IMAGENET

Source: Alexei Efros

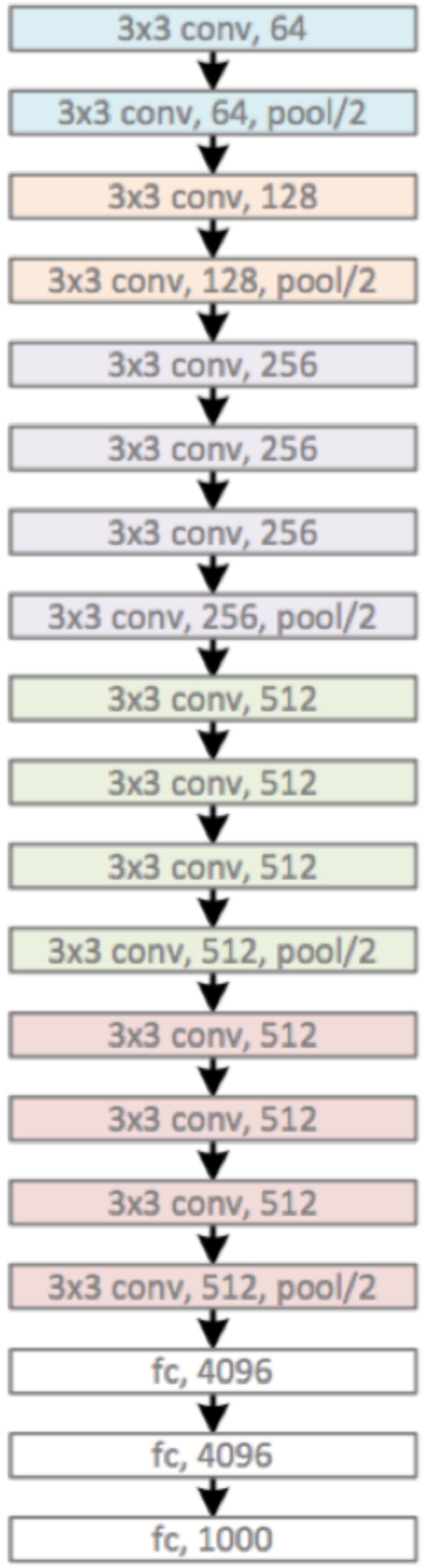
But in practice…

#1: The data

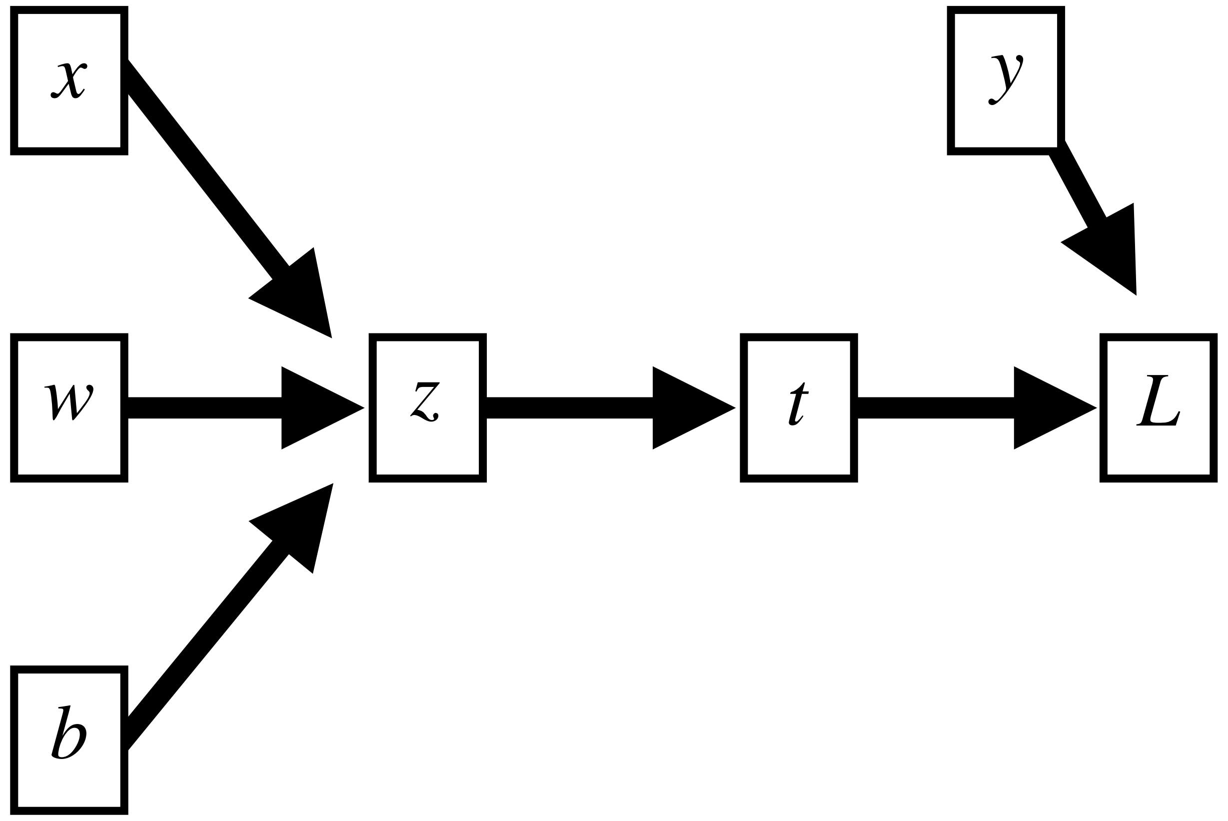
IMAGENET

…

#2: The model



#3: The algorithm



Source: Alexei Efros

How can we collect good data?

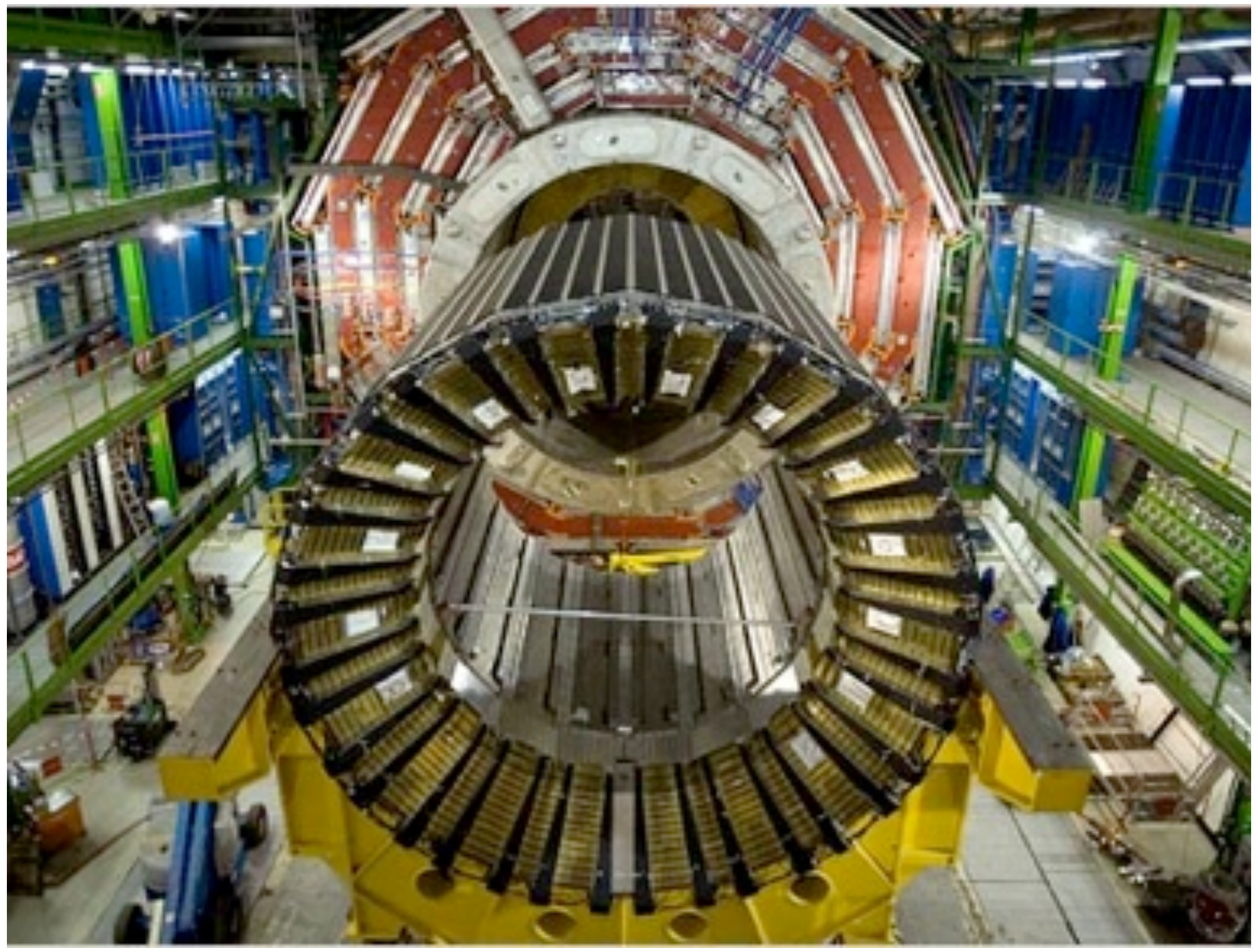
\+ Correctly labeled

\+ Unbiased (good coverage of all relevant kinds of data)

Source: Isola, Torralba, Freeman

Crowdsourcing

The value of data



The Large Hadron Collider

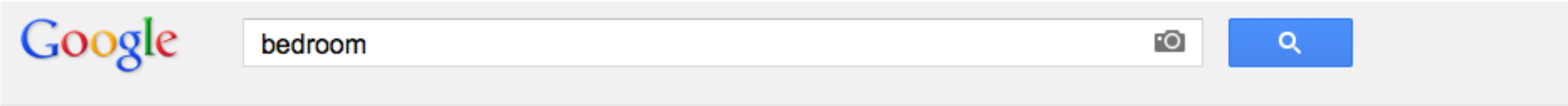
$ 10^{10}

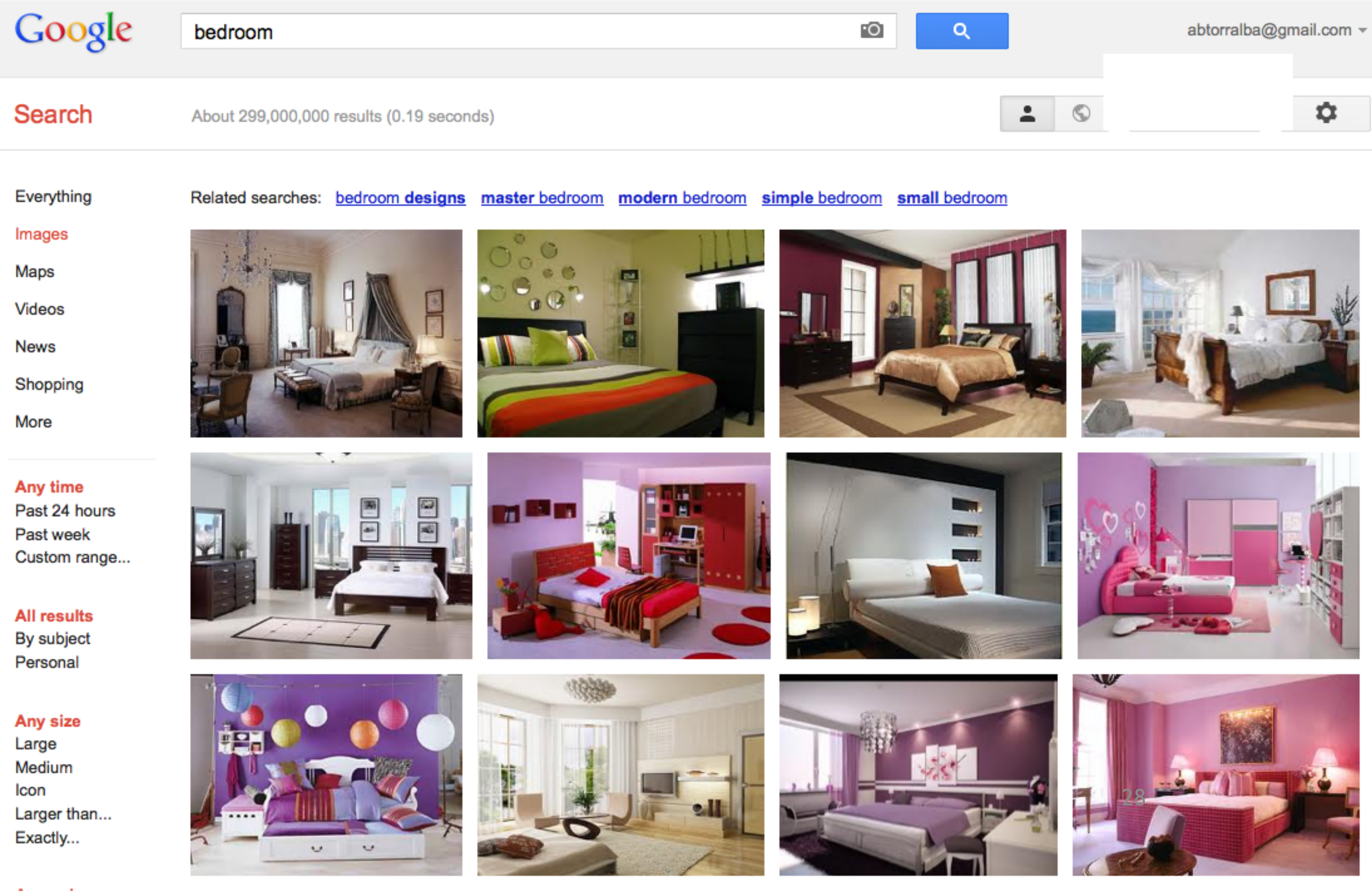
Amazon Mechanical Turk

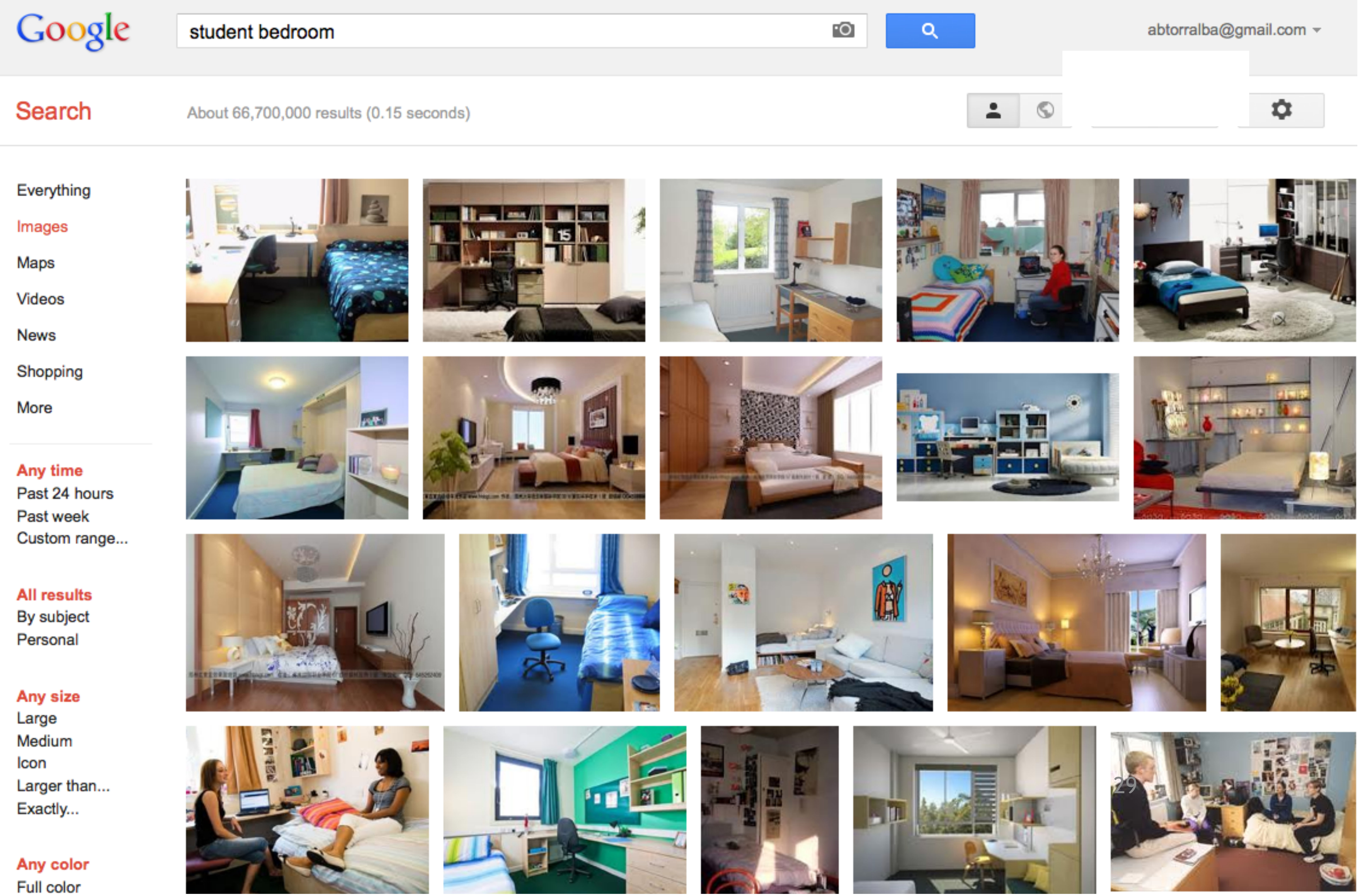
$ $10^{2} - 10^{4}$

Source: Isola, Torralba, Freeman

But can humans collect good data?


Google
bedroom


Google
bedroom
abtorralba@gmail.com
Search
About 299,000,000 results (0.19 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Any time
Past 24 hours
Past week
Custom range...
All results
By subject
Personal
Any size
Large
Medium
Icon
Larger than...
Exactly...
Related searches: bedroom designs master bedroom modern bedroom simple bedroom small bedroom


Google
student bedroom
abtorralba@gmail.com
Search
About 66,700,000 results (0.15 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Any time
Past 24 hours
Past week
Custom range...
All results
By subject
Personal
Any size
Large
Medium
Icon
Larger than...
Exactly...
Any color
Full color


www.bigstock.com · 7067629

Biases in data collection

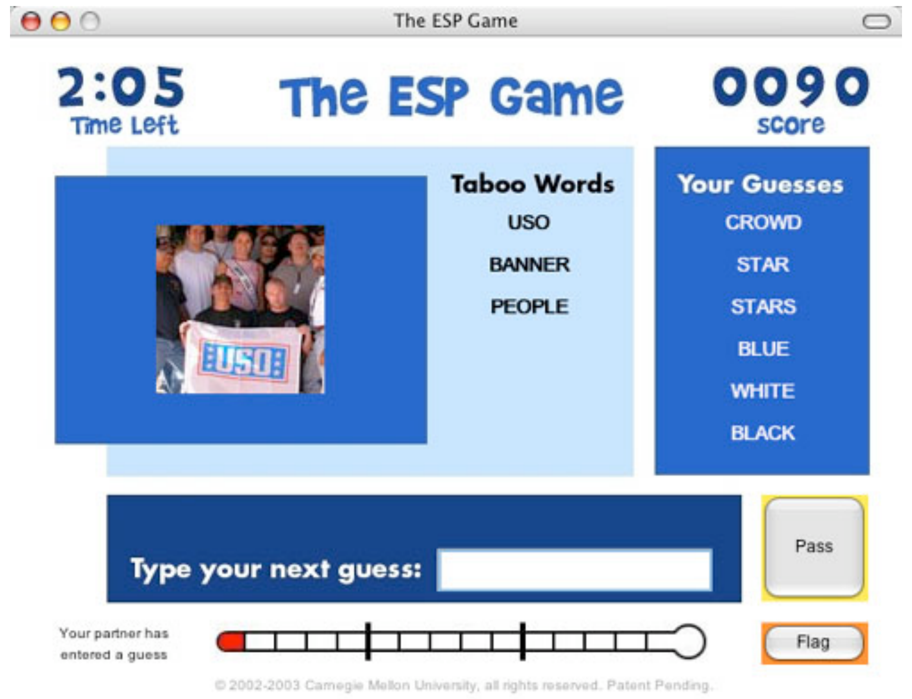
Source: Isola, Torralba, Freeman

Getting more humans in the annotation loop

Labeling to get a Ph.D.

Labeling for fun

Luis Von Ahn and Laura Dabbish 2004



Labeling for money

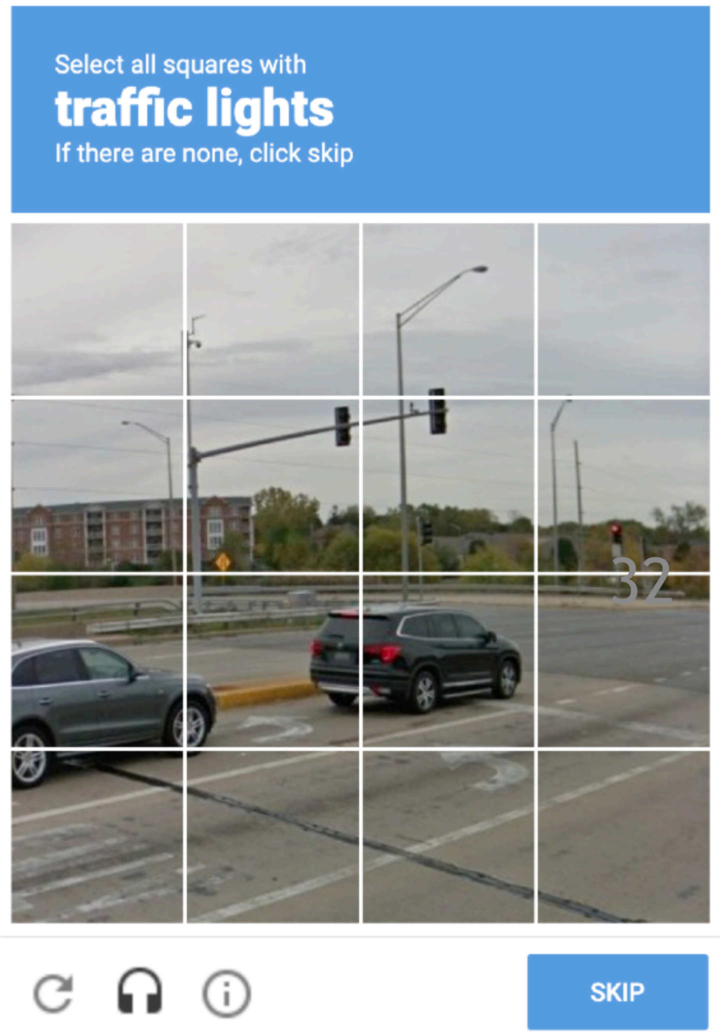
(Sorokin, Forsyth, 2008)

amazonmechanical turk

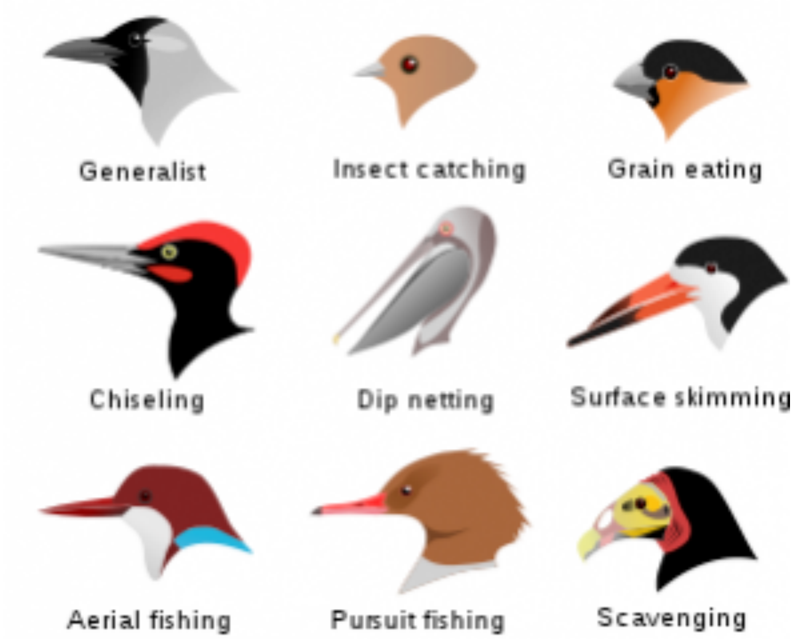
beta

Artificial Artificial Intelligence

Labeling to prove you’re human



Labeling because it gives you added value



Visipedia

(Belongie, Perona, et al)

Source: Isola, Torralba, Freeman

Beware of the human in your loop

- What do you know about them?
- Will they do the work you pay for?

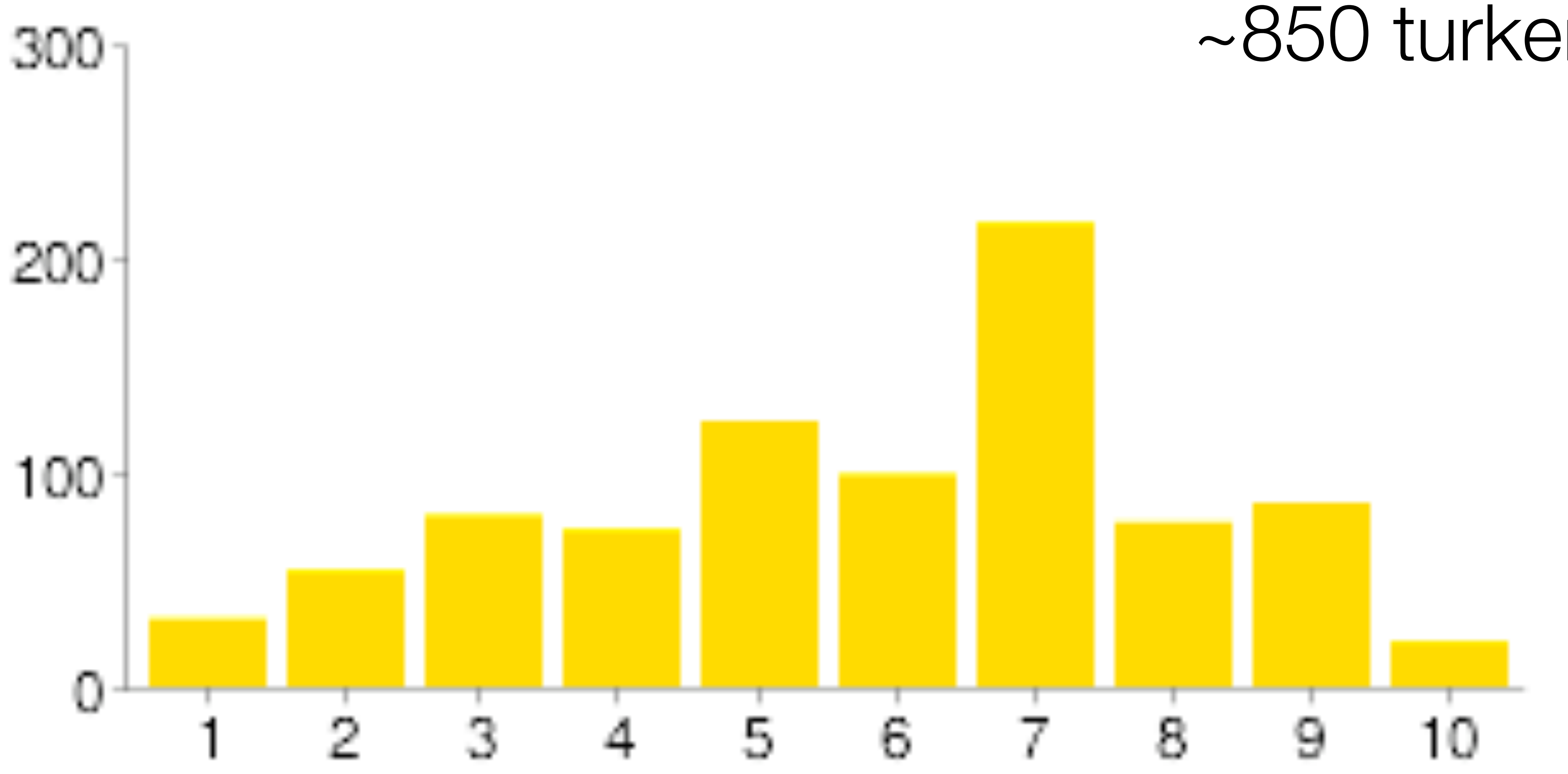
Let’s check a few simple experiments

Source: Isola, Torralba, Freeman

People have biases…

Turkers were offered 1 cent to pick a number from 1 to 10.

~850 turkers

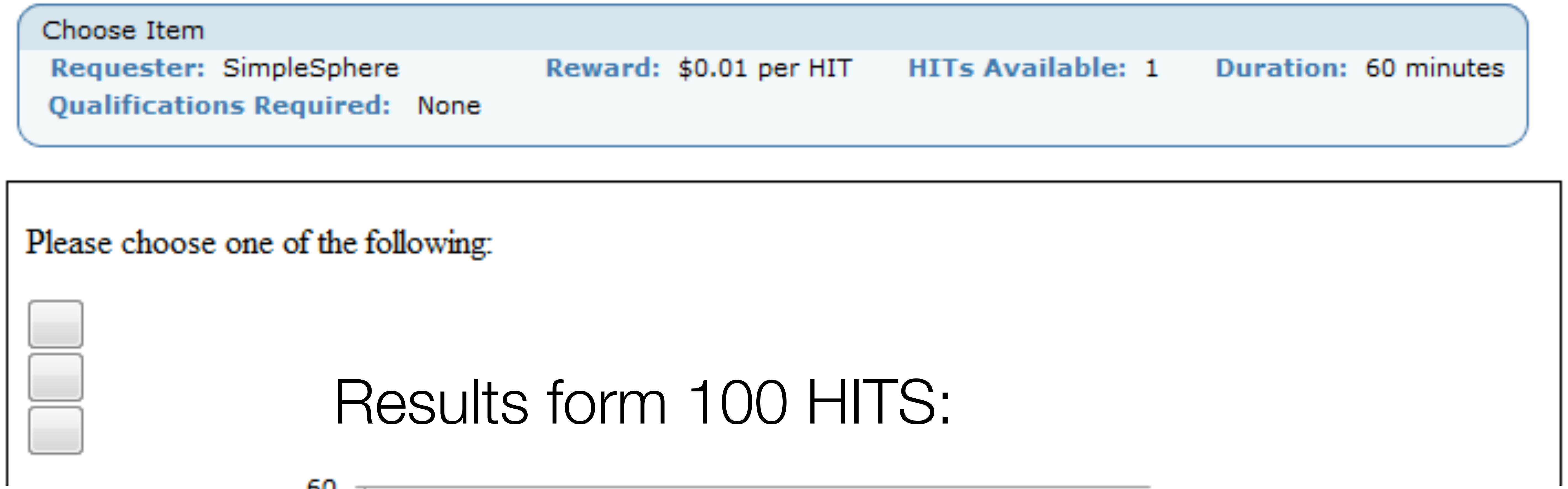


Experiment by Greg Little

From http://groups.csail.mit.edu/uid/deneme/

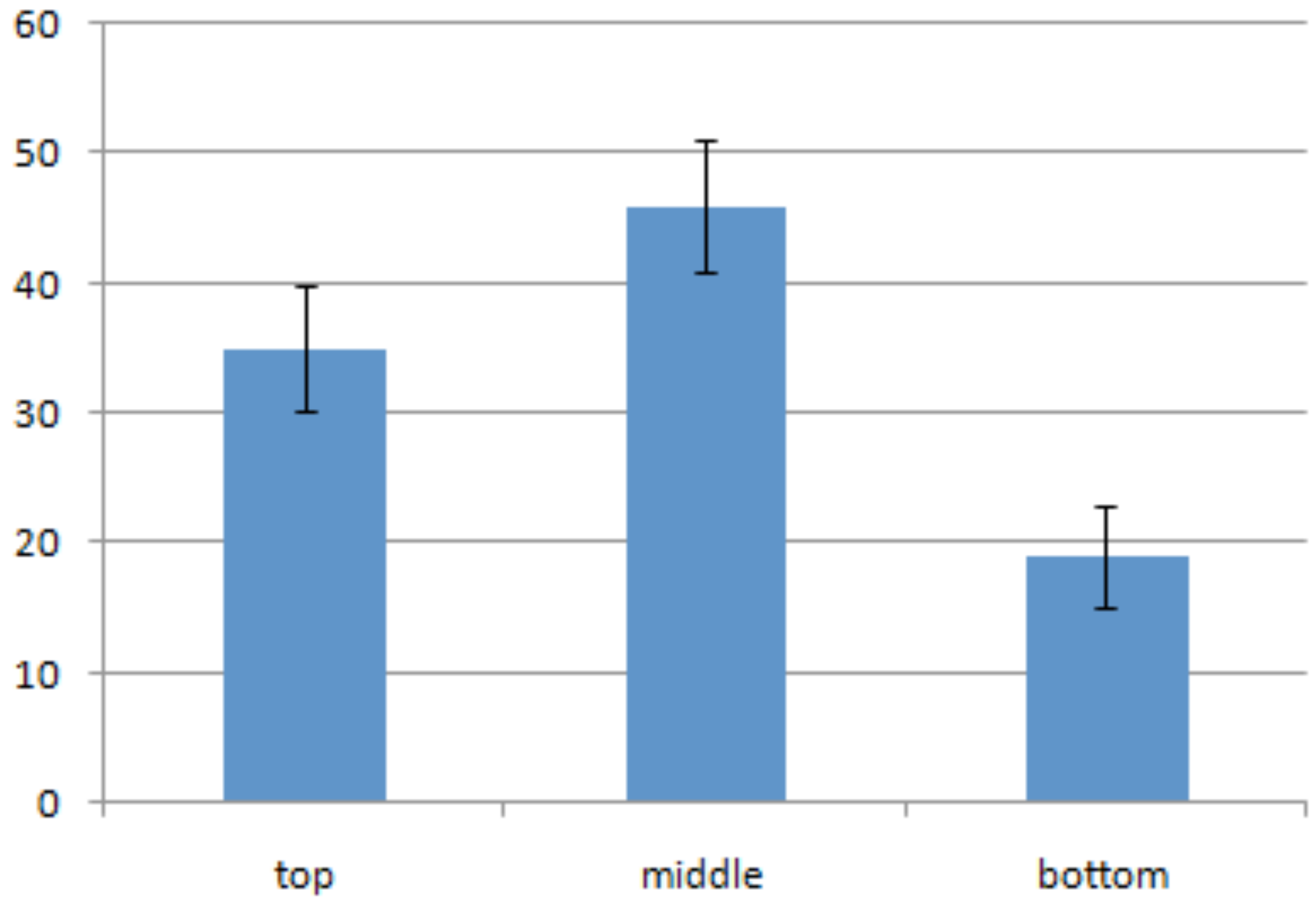
Source: Isola, Torralba, Freeman

Do humans have consistent biases?


Choose Item
Requester: SimpleSphere Reward: $0.01 per HIT HITs Available: 1 Duration: 60 minutes
Qualifications Required: None

Please choose one of the following:

Results form 100 HITS:



Experiment by Greg Little

From http://groups.csail.mit.edu/uid/deneme/

Source: Isola, Torralba, Freeman

Are humans reliable even in simple tasks?

Choose the given item.

Requester: SimpleSphere **Reward:** $0.01 per HIT **HITs Available:** 1 **Duration:** 60 minutes

Qualifications Required: None

Please click button B:

B

C

A

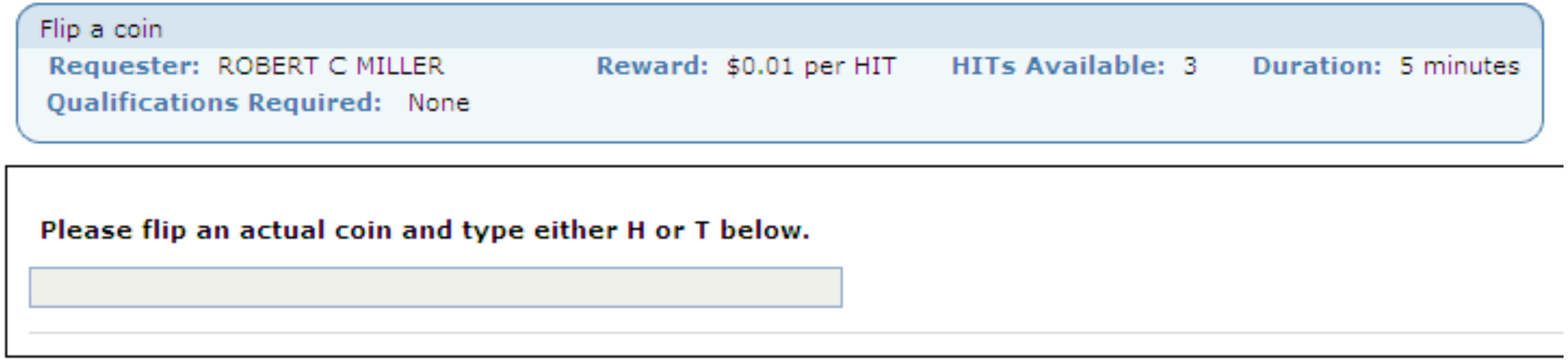
Results of 100 HITS:
A: 2
B: 96
C: 2

Experiment by Greg Little

From http://groups.csail.mit.edu/uid/deneme/

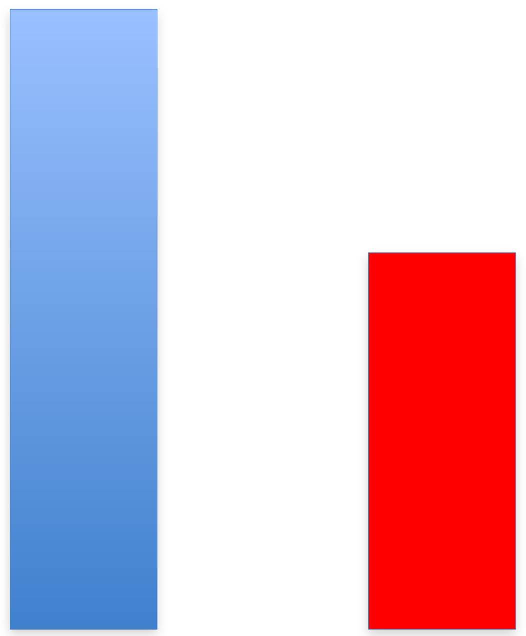
Source: Isola, Torralba, Freeman

Do humans do what you ask for?


Flip a coin
Requester: ROBERT C MILLER | Reward: $0.01 per HIT | HITs Available: 3 | Duration: 5 minutes
Qualifications Required: None

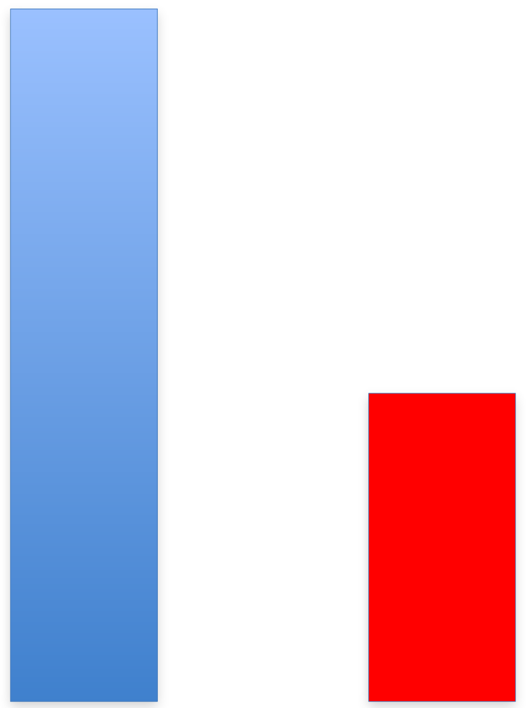
Please flip an actual coin and type either H or T below.

After 50 HITS:



31 heads, 19 tails

And 50 more:



34 heads, 16 tails

Experiment by Rob Miller

From http://groups.csail.mit.edu/uid/deneme/

Source: Isola, Torralba, Freeman

So we can sometimes collect good training data.

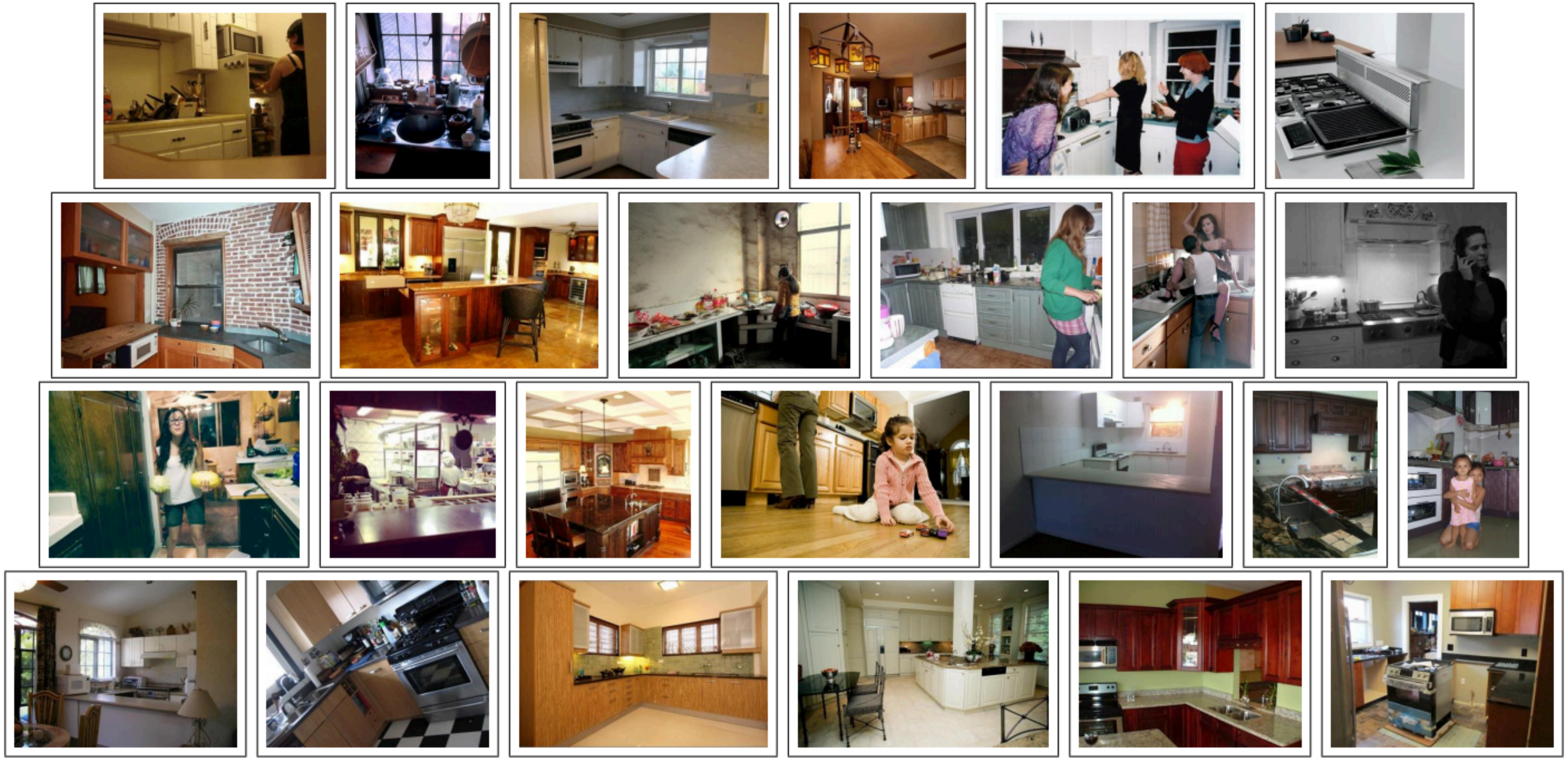
But suppose we messed up. Our test setting doesn't look like the training data!

How can we bridge the domain gap?

Source: Isola, Torralba, Freeman

Finding more representative images

Places365 Kitchen



[Fouhey et al., "From Lifestyle Vlogs to Everyday Actions", 2017]

Finding more representative images

VLOG Kitchen

[Fouhey et al., "From Lifestyle Vlogs to Everyday Actions", 2017]

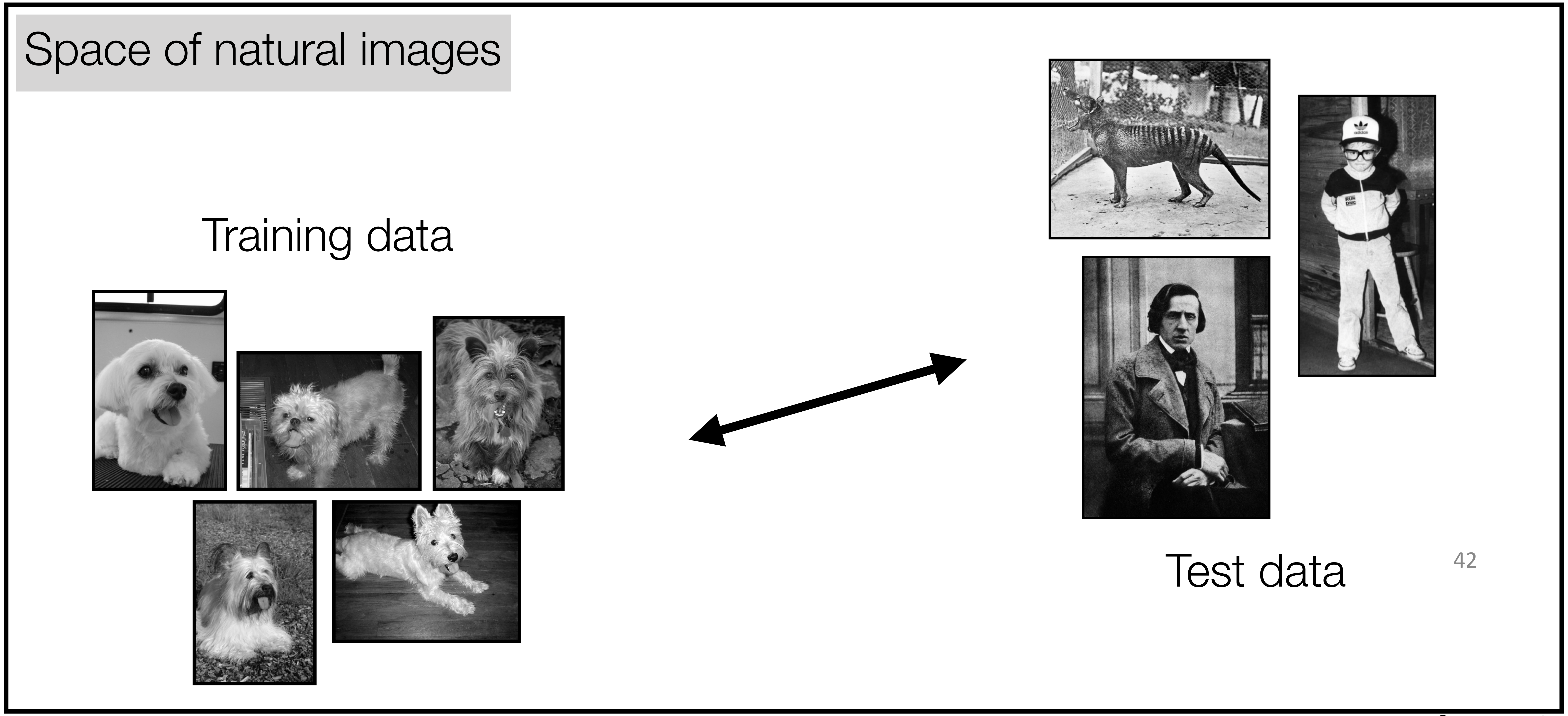
Name that dataset game

[Torralba and Efros, "An unbiased look at dataset bias," 2011]

testing domain
(where we actual use our model)

training domain

Domain gap between $p_{\texttt{train}}$ and $p_{\texttt{test}}$ will cause us to fail to generalize.

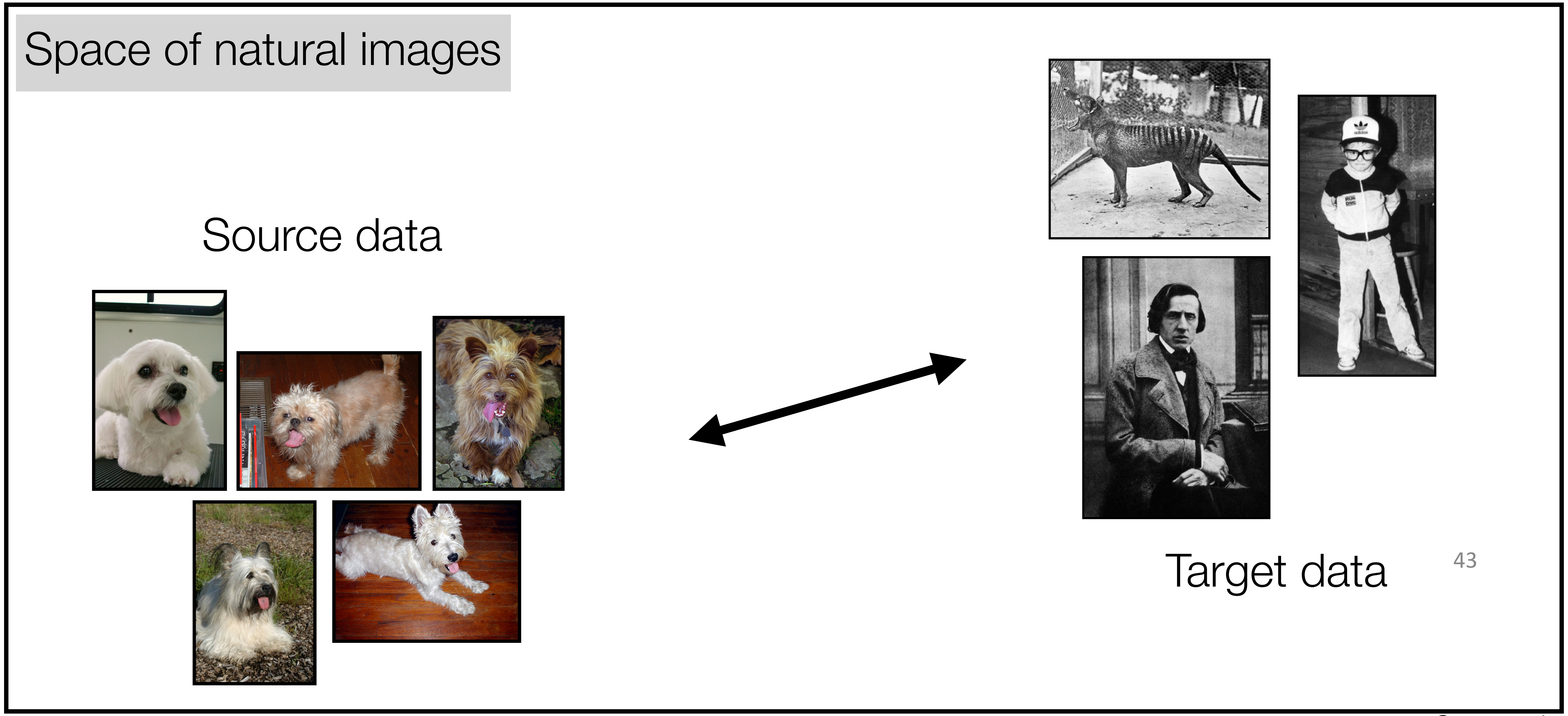


Source: Isola, Torralba, Freeman

source domain

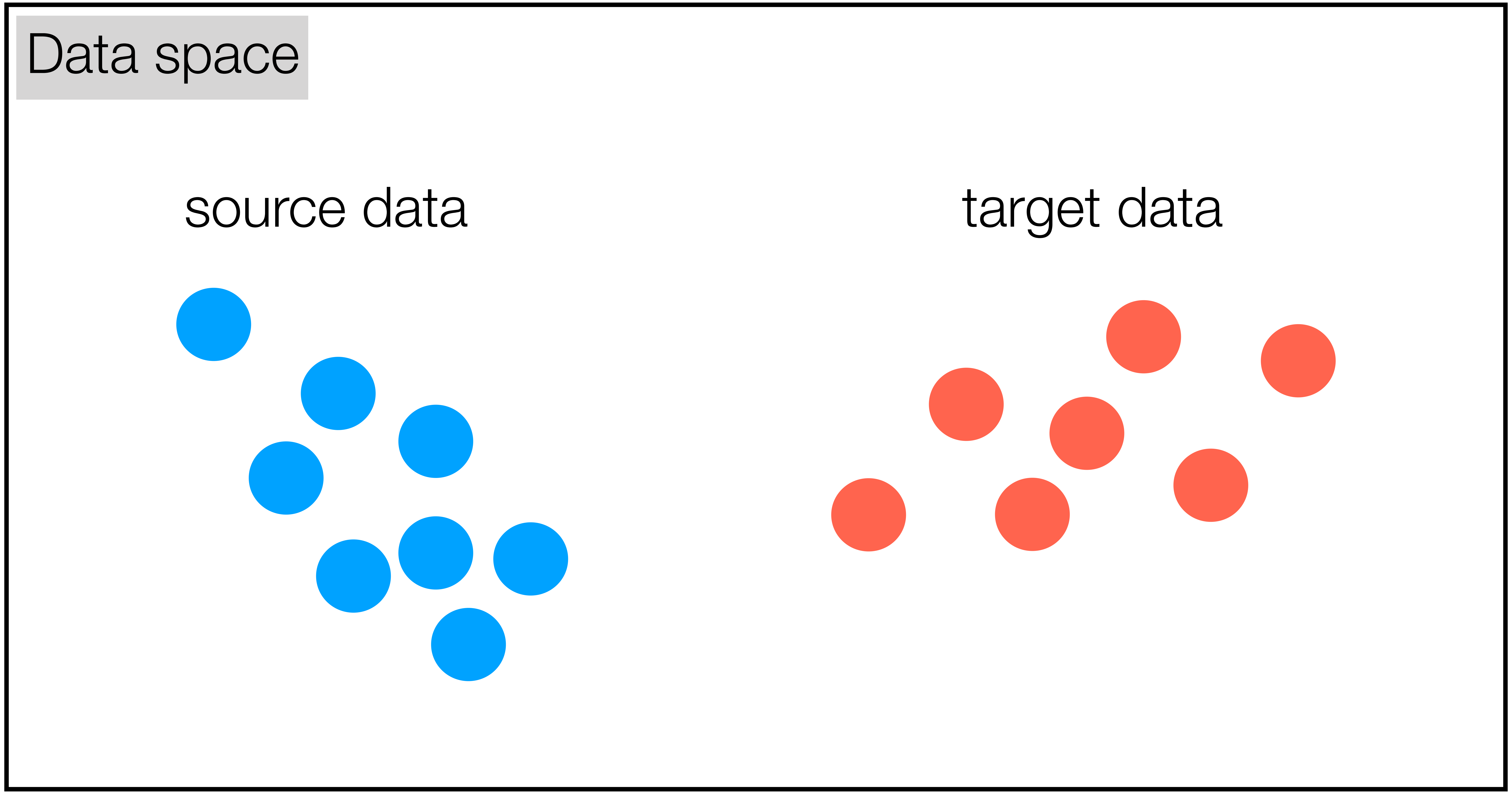
target domain
(where we actual use our model)

Domain gap between $p_{\texttt{source}}$ and $p_{\texttt{target}}$ will cause us to fail to generalize.



Source: Isola, Torralba, Freeman

Idea #1: transform the target domain to look like the source domain



(Or vice versa)

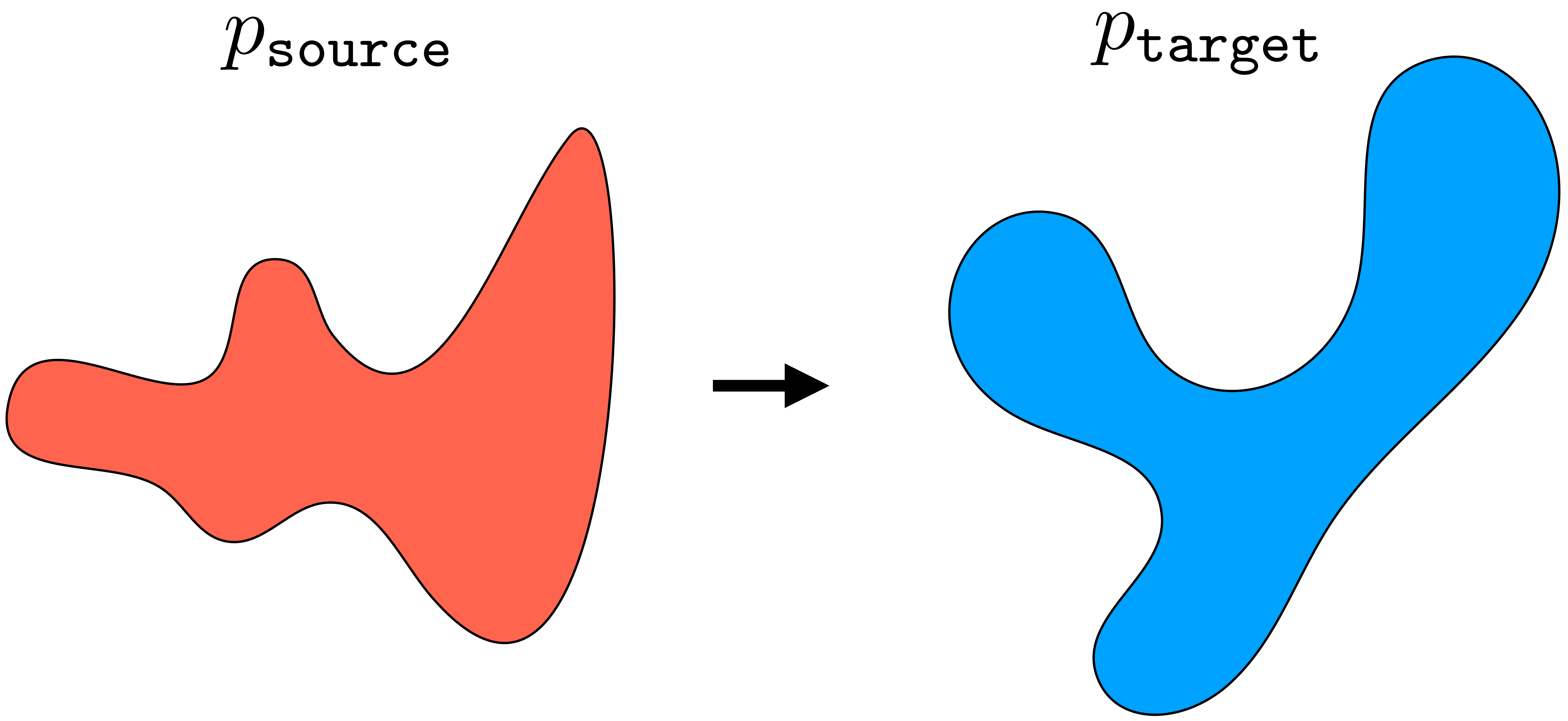
This is called **domain adaptation**

Source: Isola, Torralba, Freeman

Domain adaptation

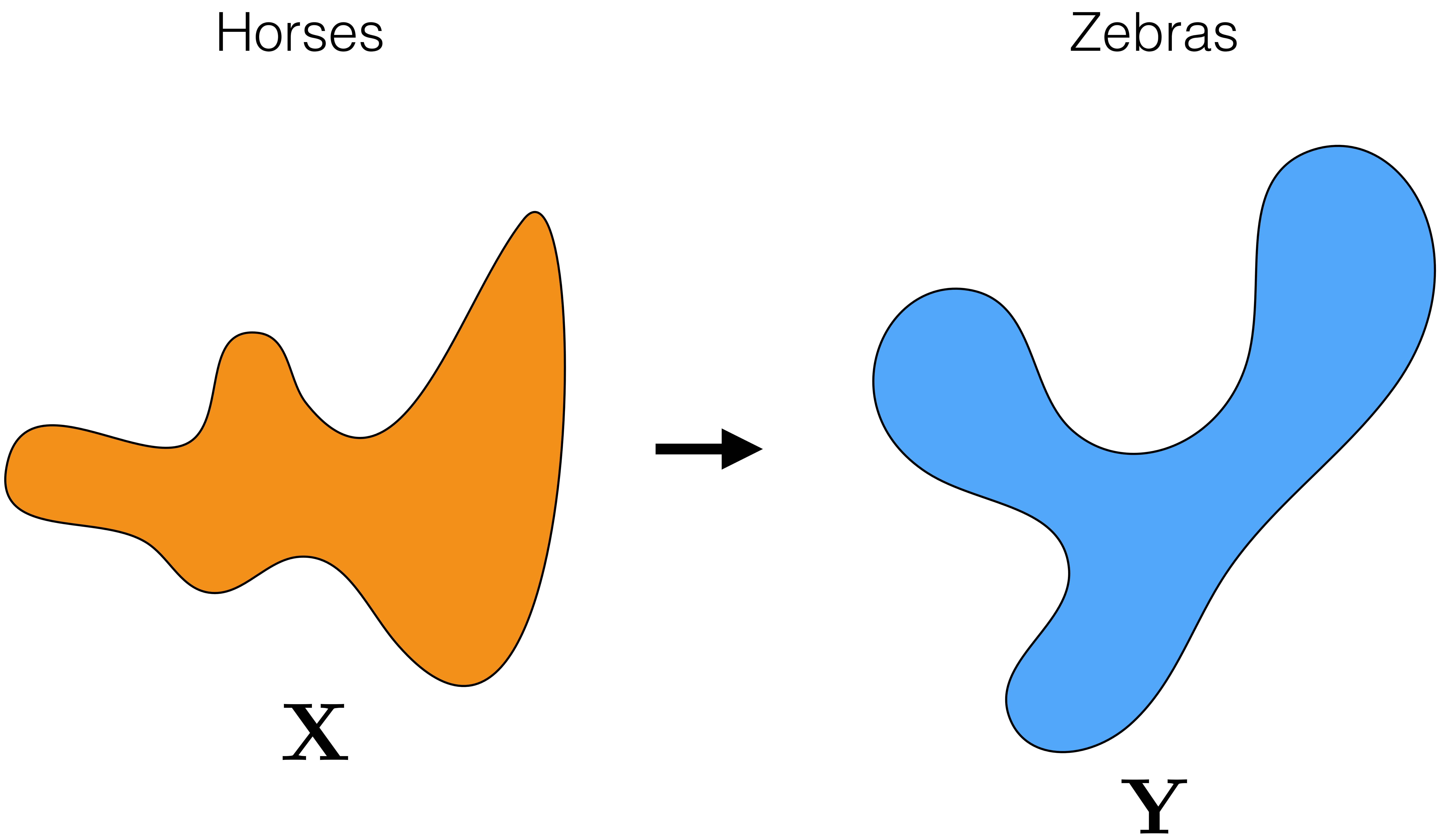
- We have source domain pairs $\{\mathbf{x}^{\mathbf{source}}, \mathbf{y}^{\mathbf{source}}\}$
- Learn a mapping F: $\mathbf{x}^{\mathbf{source}} \longrightarrow \mathbf{y}^{\mathbf{source}}$
- We want to apply F to target domain data $\mathbf{x}^{\mathbf{target}}$
- Find transformation T: $\mathbf{x}^{\mathbf{target}} \longrightarrow \mathbf{x}^{\mathbf{source}}$
- Now apply $F(T(\mathbf{x}^{\mathbf{target}}))$ to predict $\mathbf{y}^{\mathbf{target}}$

Source: Isola, Torralba, Freeman



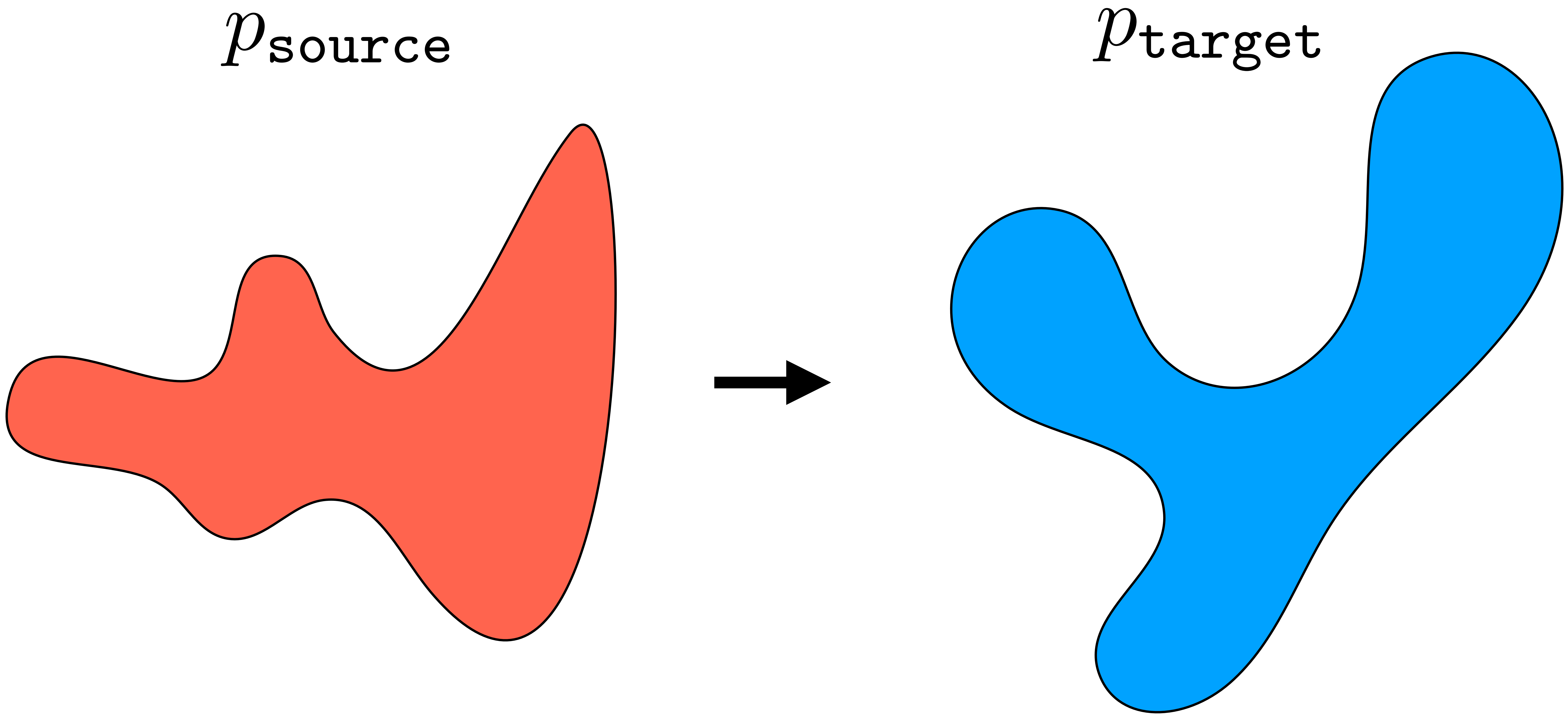
Source: Isola, Torralba, Freeman

CycleGAN



Source: Isola, Torralba, Freeman

Domain adaptation

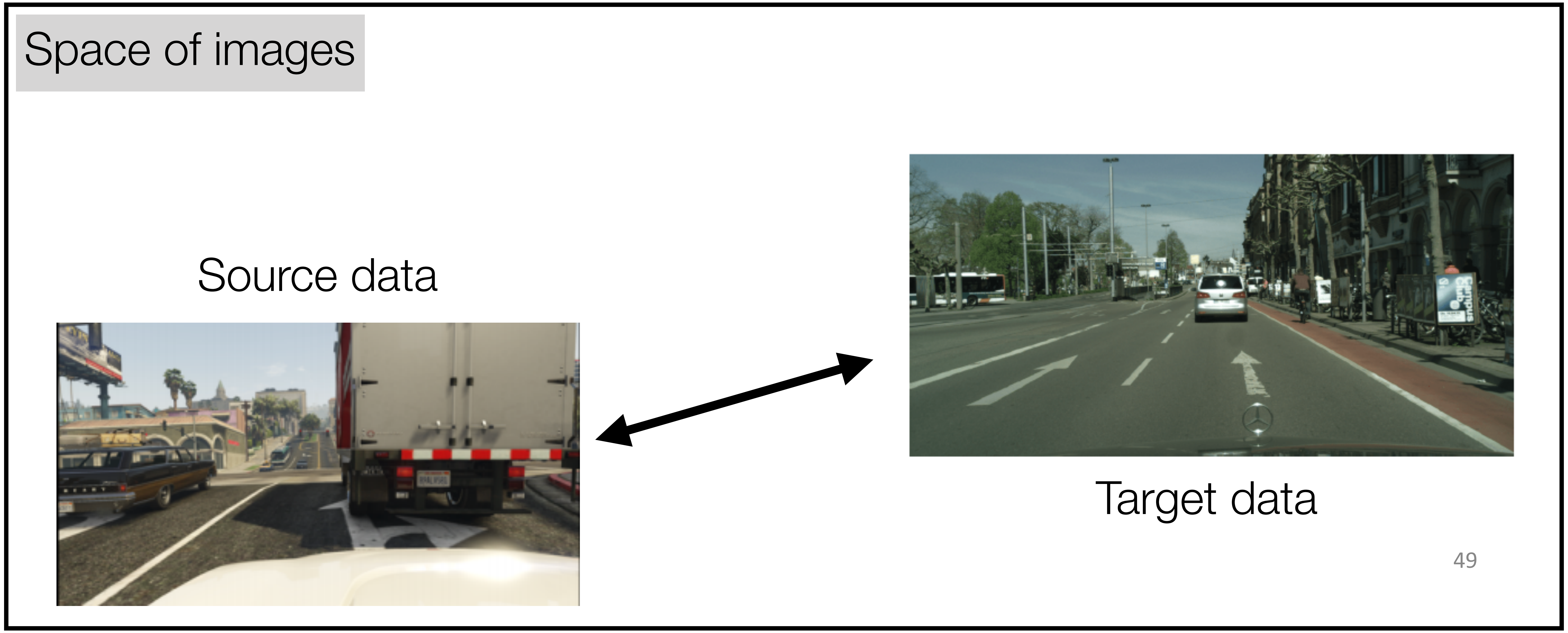


Source: Isola, Torralba, Freeman

source domain

target domain
(where we actual use our model)

Domain gap between $p_{\texttt{source}}$ and $p_{\texttt{target}}$ will cause us to fail to generalize.



Source: Isola, Torralba, Freeman

Cycle-Consistent Adversarial Domain Adaptation

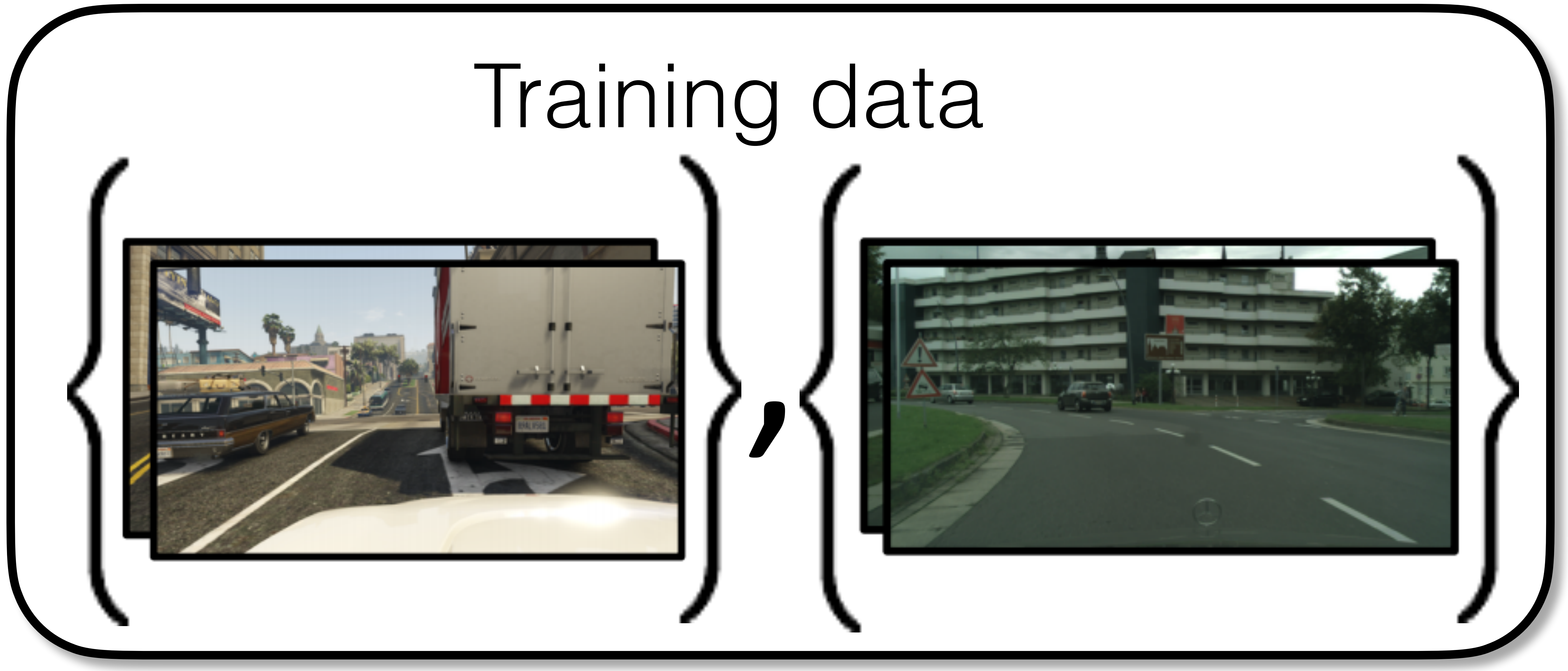
Source domain

Target domain

[Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Darrell, Efros, arXiv 2017]

Source: Isola, Torralba, Freeman

CycleGAN



Source: Isola, Torralba, Freeman

CycleGAN

Training data

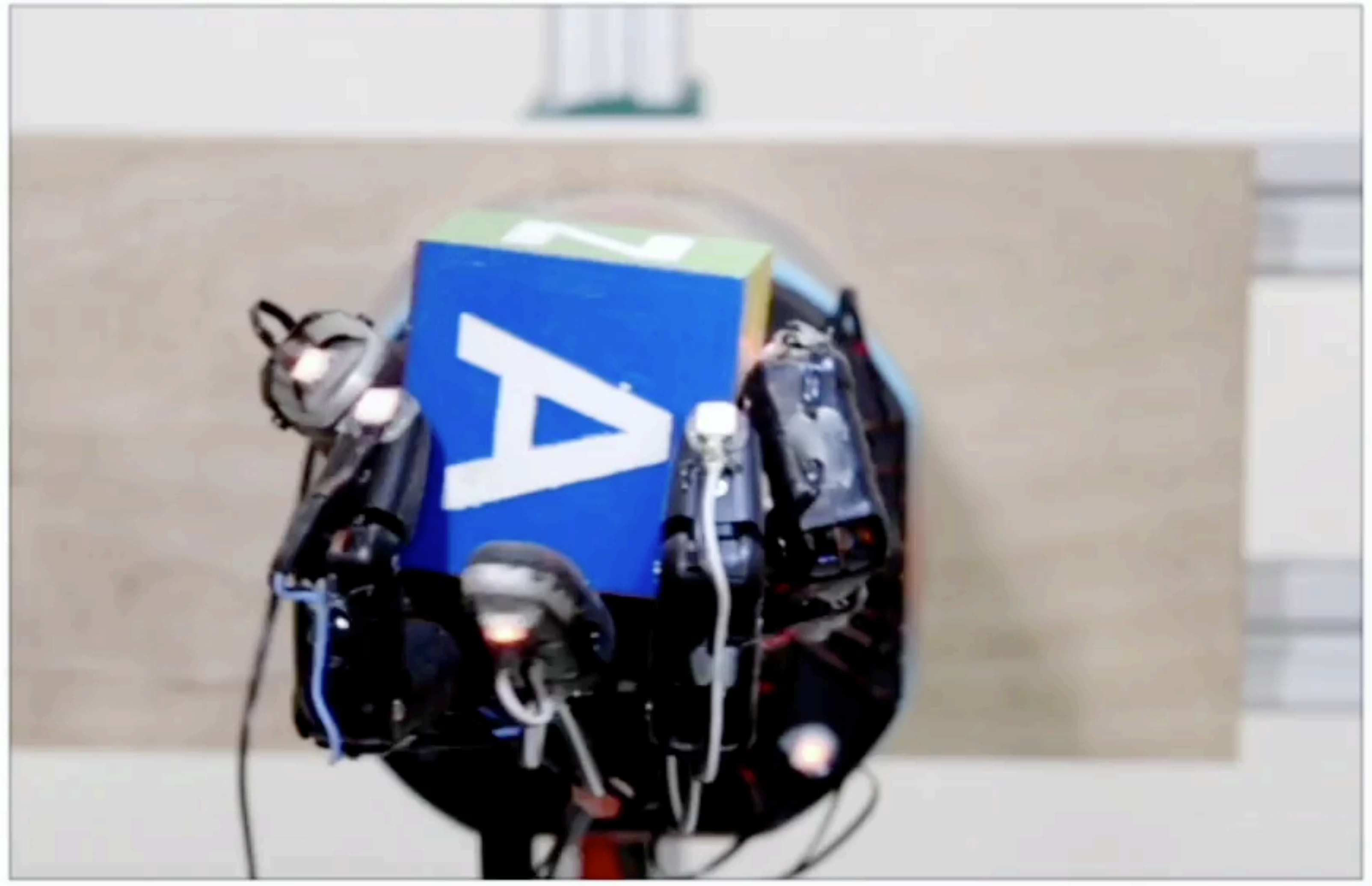
Source: Isola, Torralba, Freeman

CycleGAN FCN

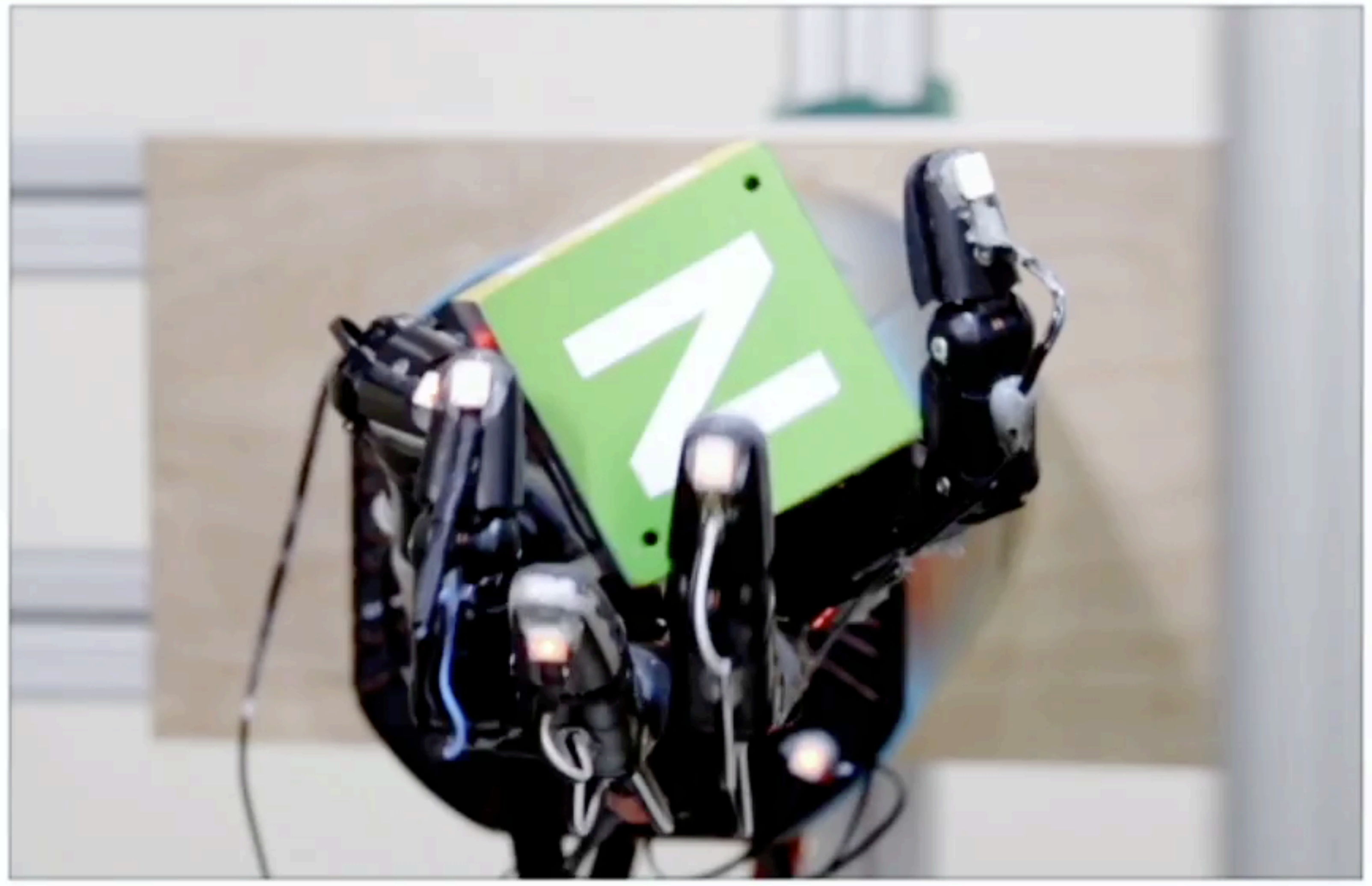
Training data

Source: Isola, Torralba, Freeman

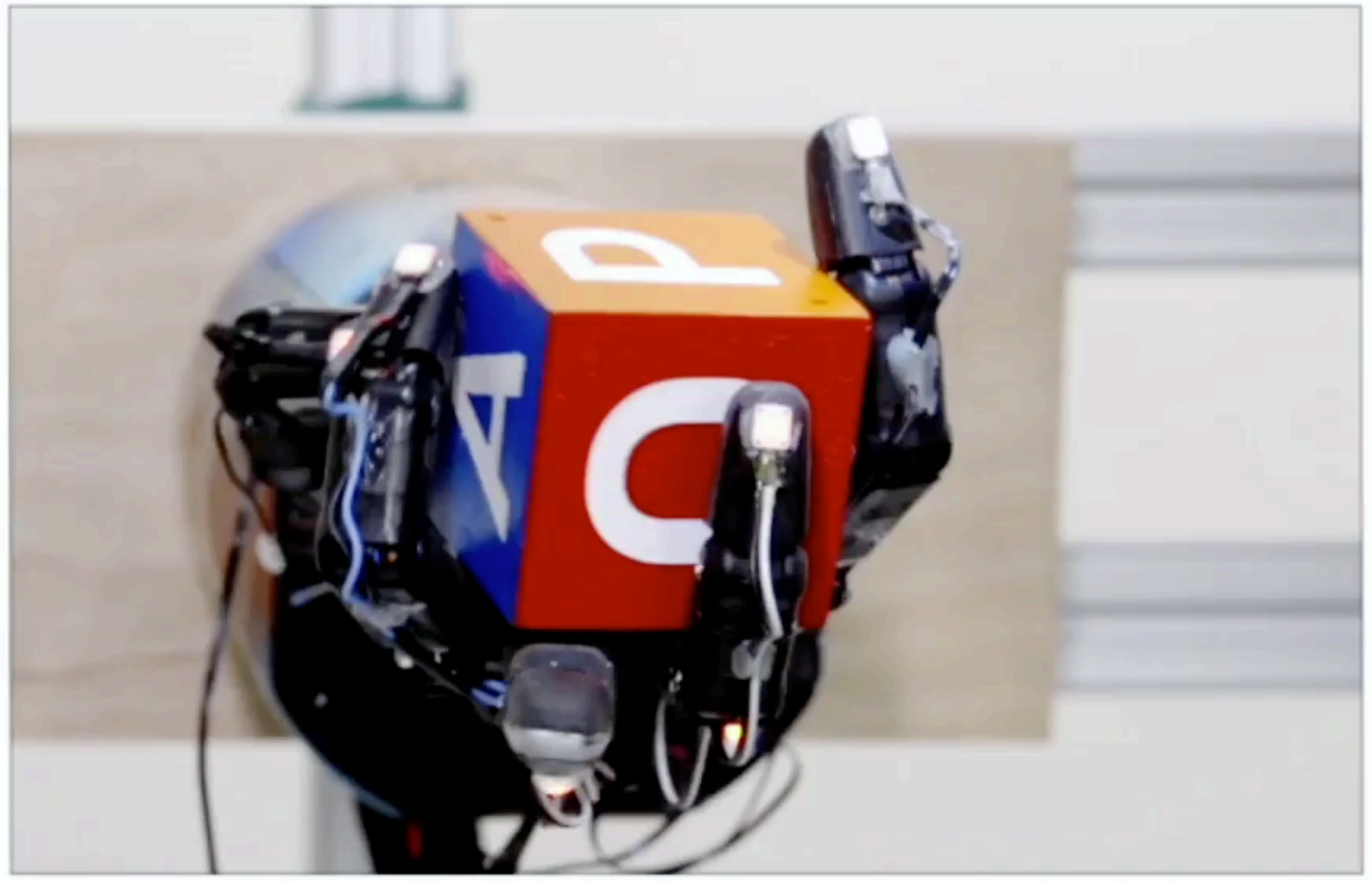
OpenAI Dactyl



FINGER PIVOTING



SLIDING



FINGER GAITING

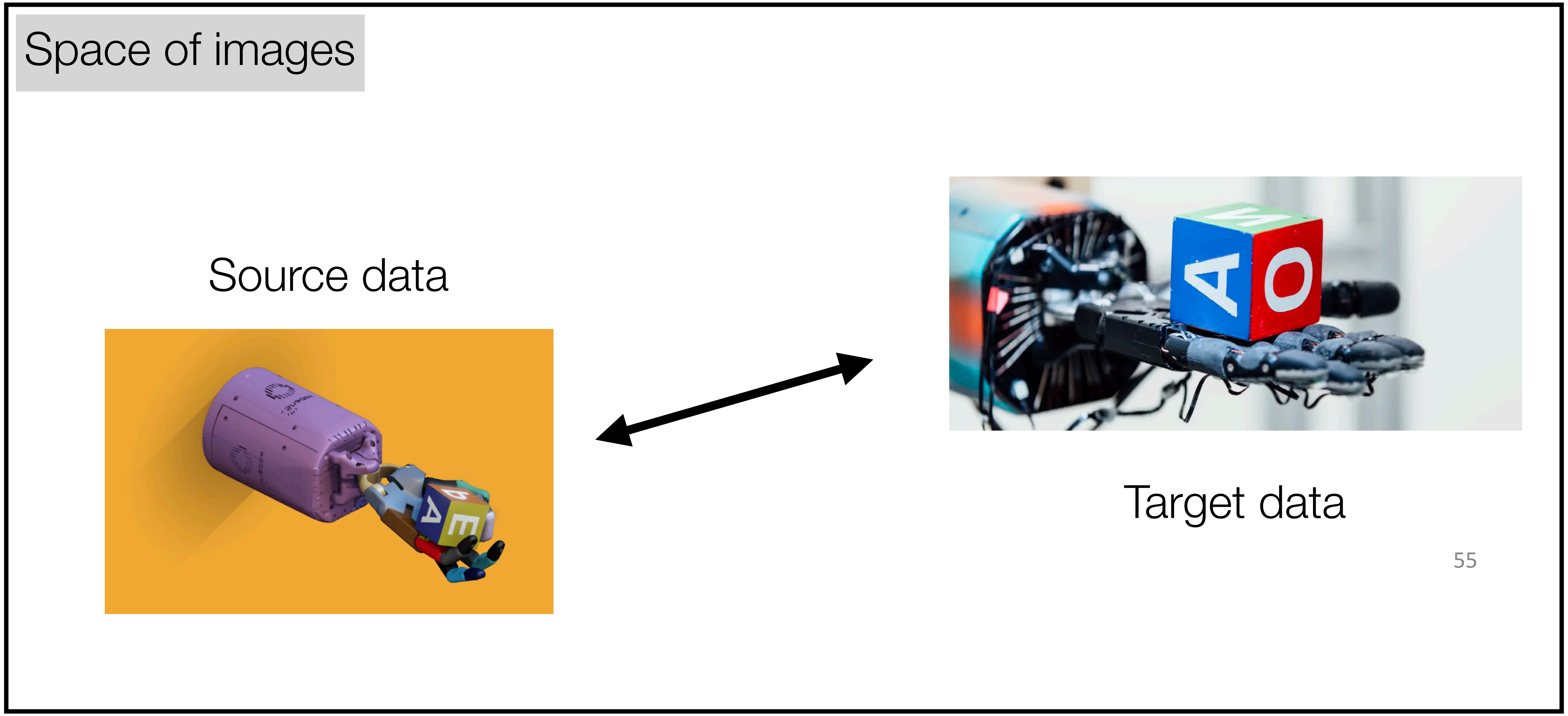
Source: Isola, Torralba, Freeman

source domain

target domain

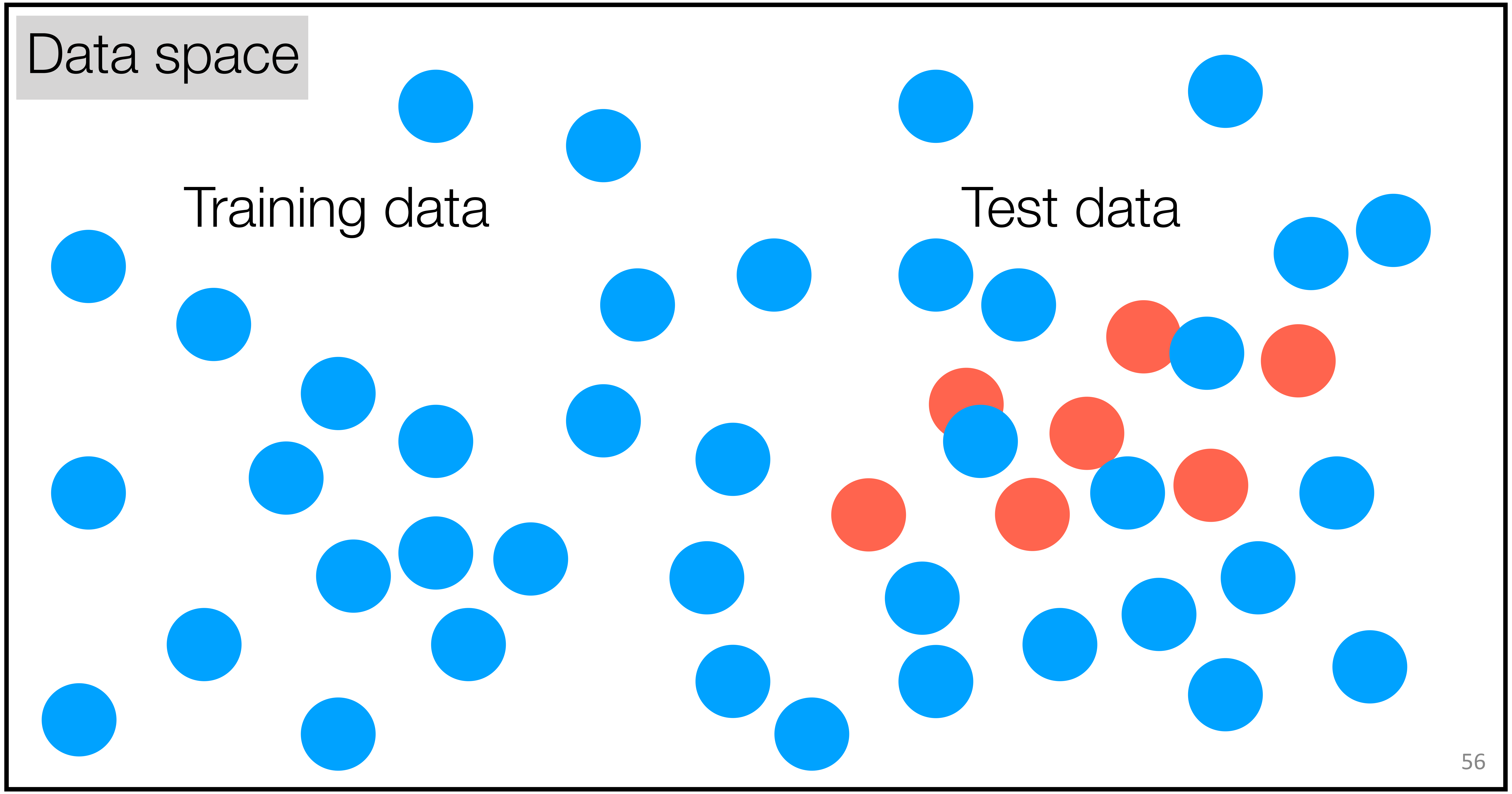
(where we actual use our model)

Domain gap between $p_{\texttt{source}}$ and $p_{\texttt{target}}$ will cause us to fail to generalize.



Source: Isola, Torralba, Freeman

Idea #2: train on randomly perturbed data, so that test set just looks like another random perturbation

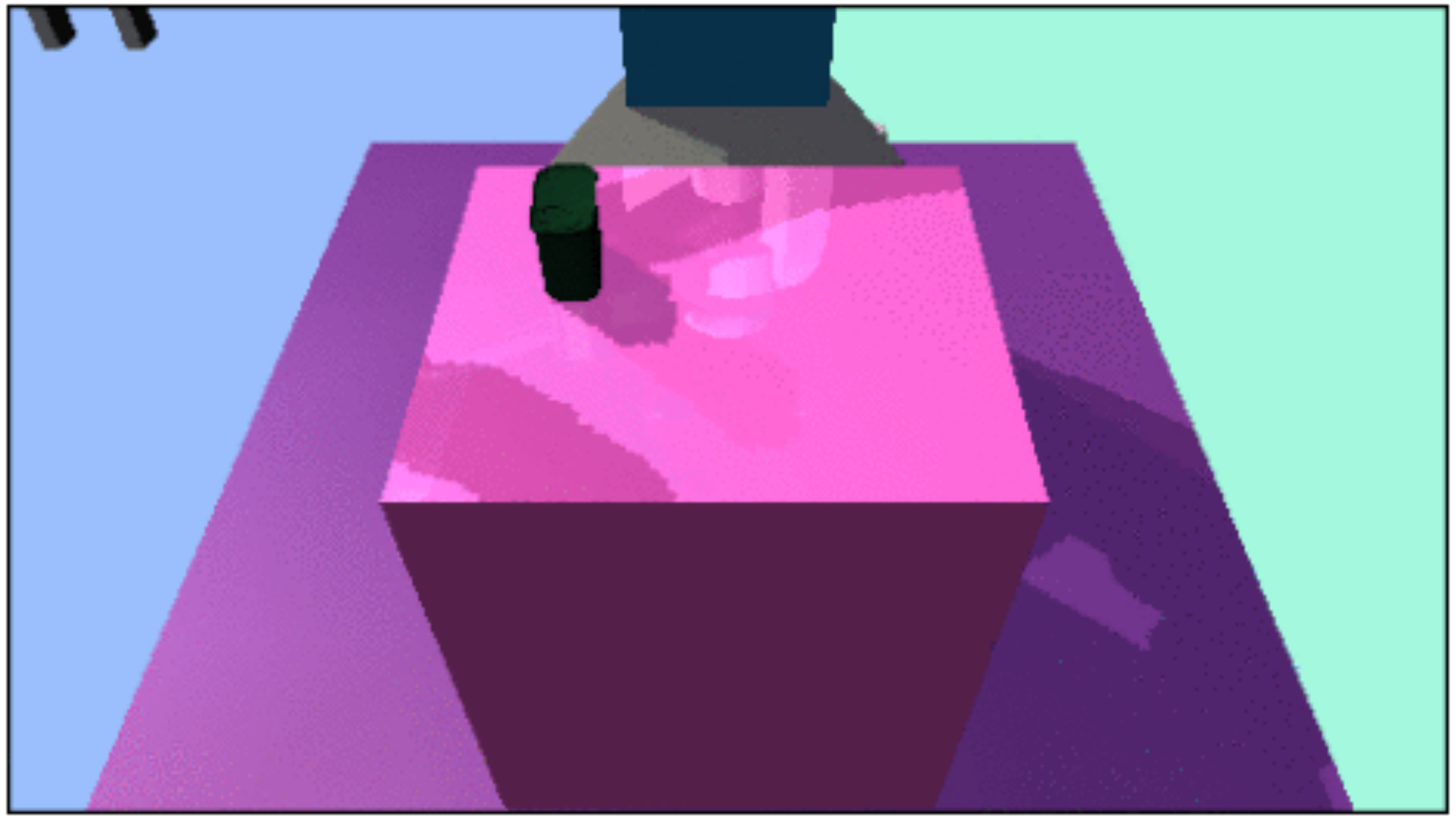


This is called **domain randomization** or **data augmentation**

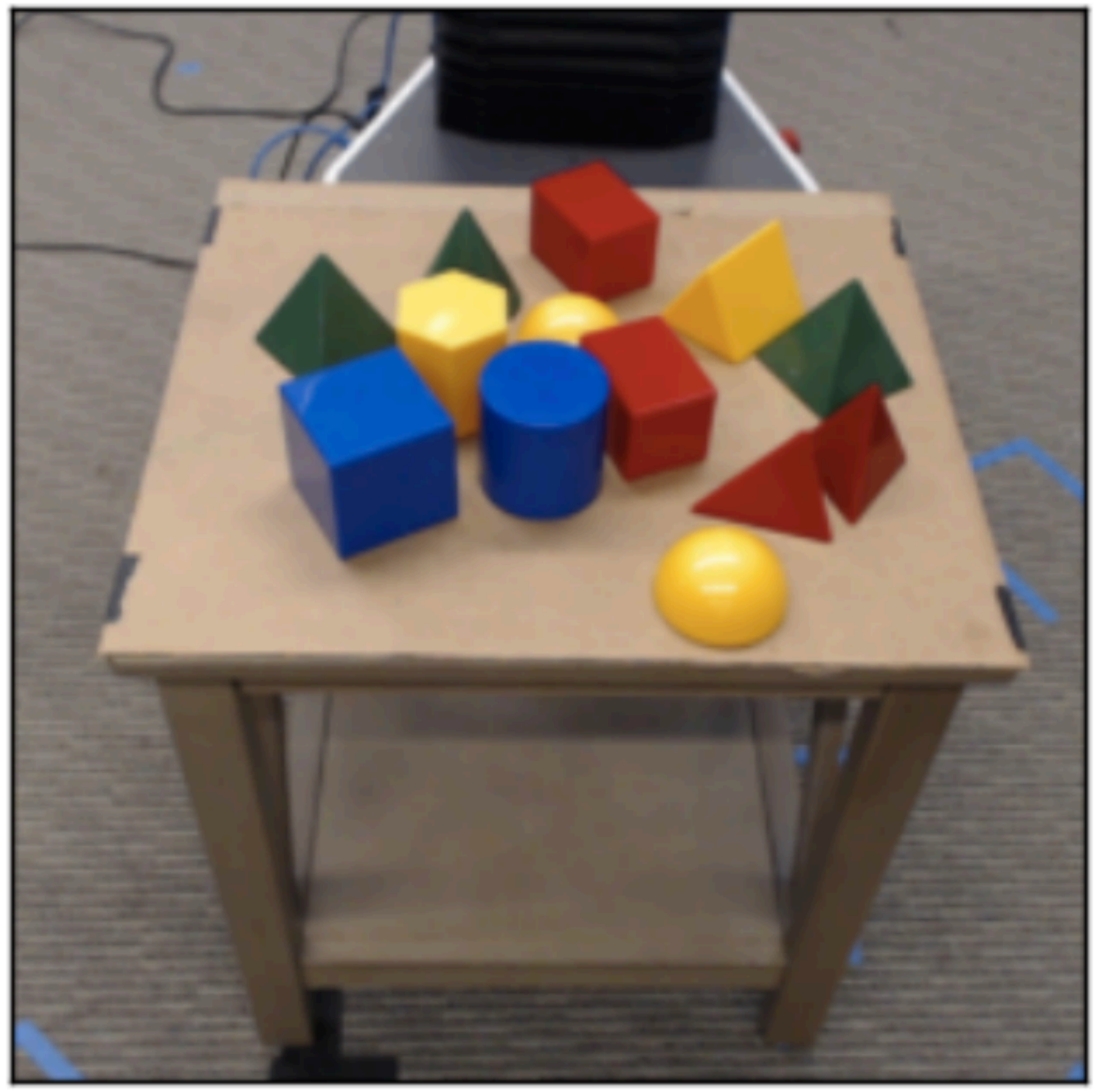
Source: Isola, Torralba, Freeman

Domain randomization

Training data



Test data



[Sadeghi & Levine 2016]

Above example is from [Tobin et al. 2017]

Source: Isola, Torralba, Freeman

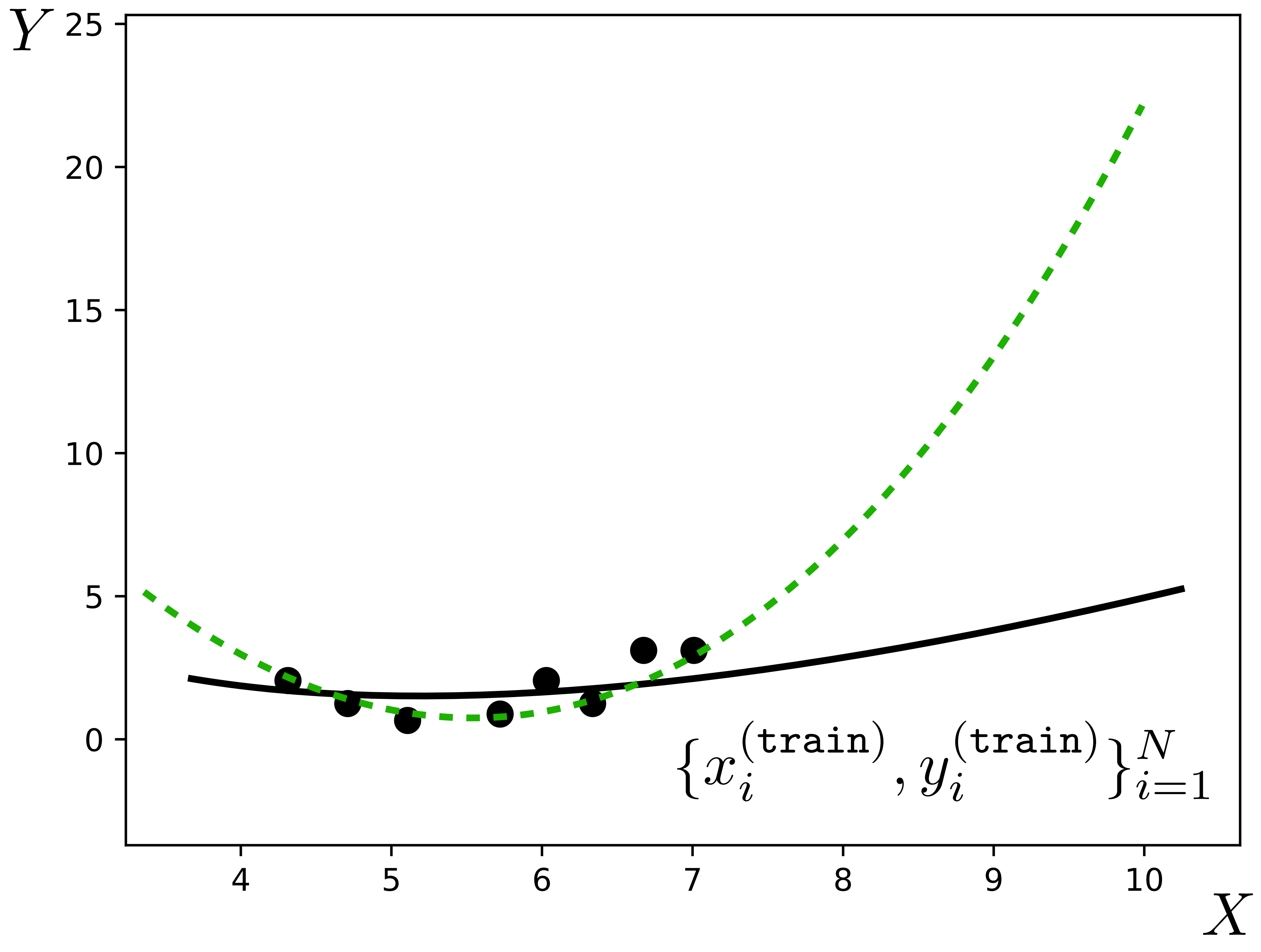
Beyond data

- Data very important [Maluleke et al., 2022], but also other factors can matter.
- Camera hardware and software
  - e.g., default camera settings calibrated to expose light skin
- Loss function (e.g., “mode collapse” in GANs)
- Features
- Sampling strategy (e.g., truncation in GANs)

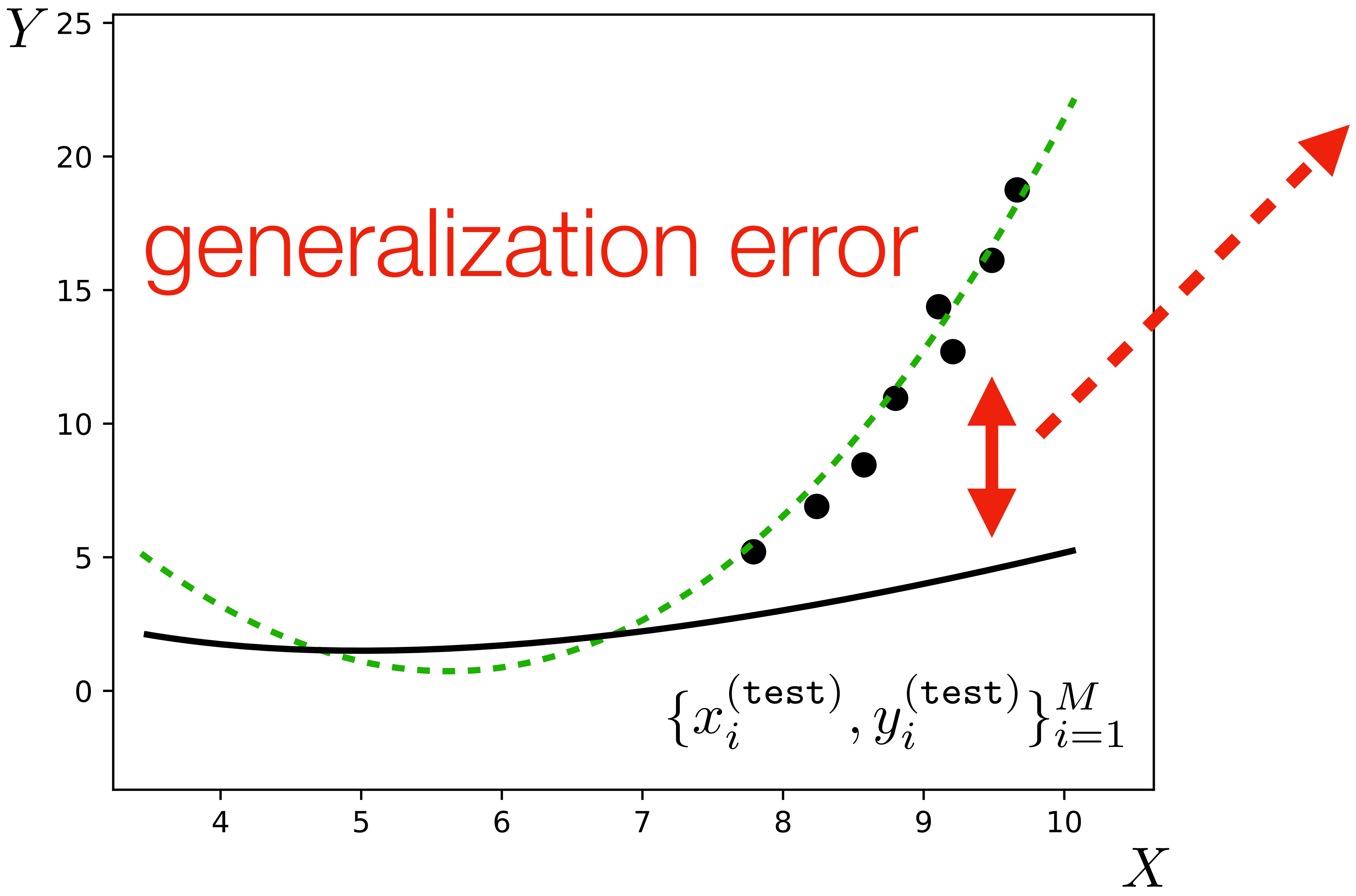
What if we go way outside of the training distribution?

Source: Isola, Torralba, Freeman

Training data



Test data

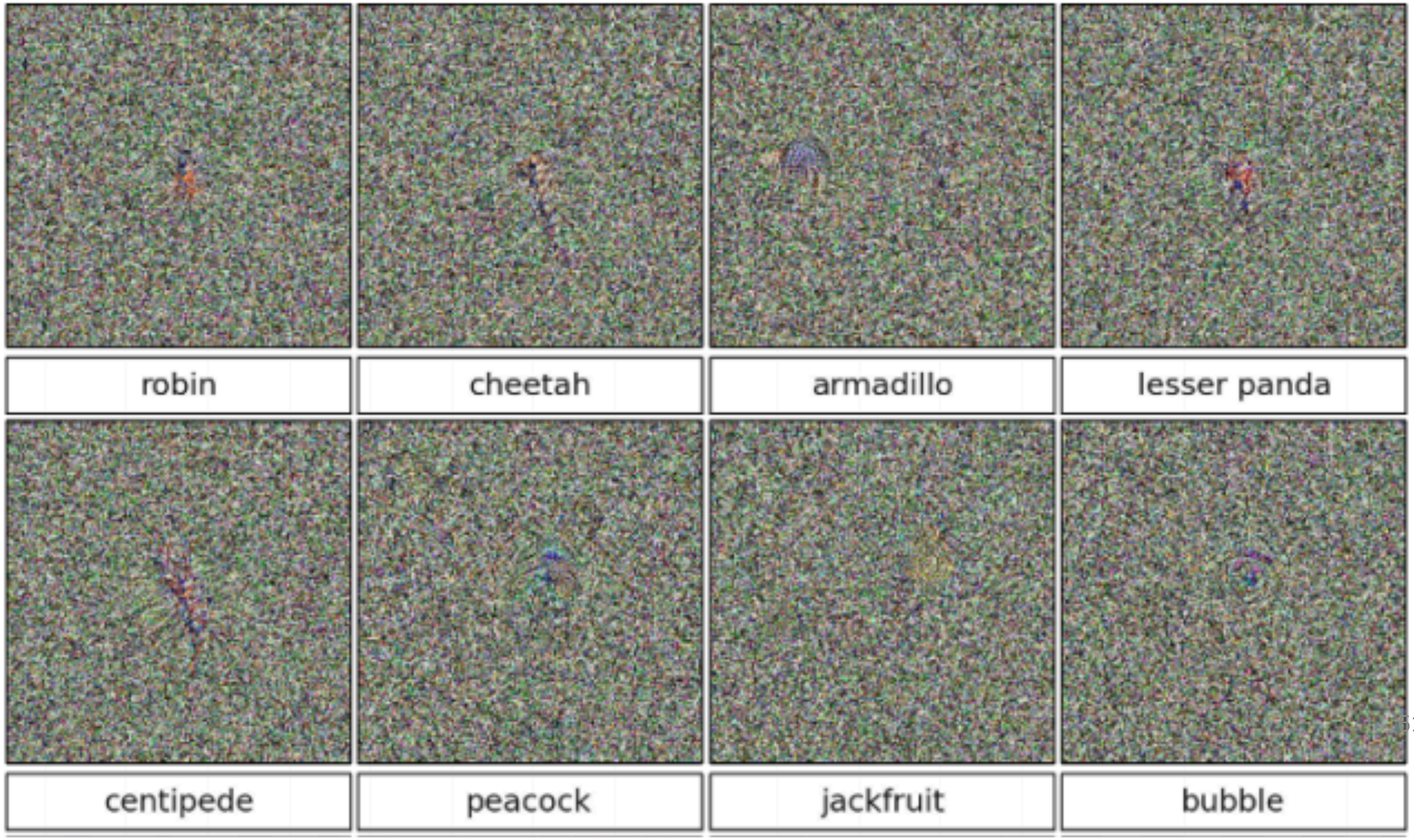


Our training data did not cover the part of the distribution that was tested (**biased data**)

Source: Isola, Torralba, Freeman

"Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images"

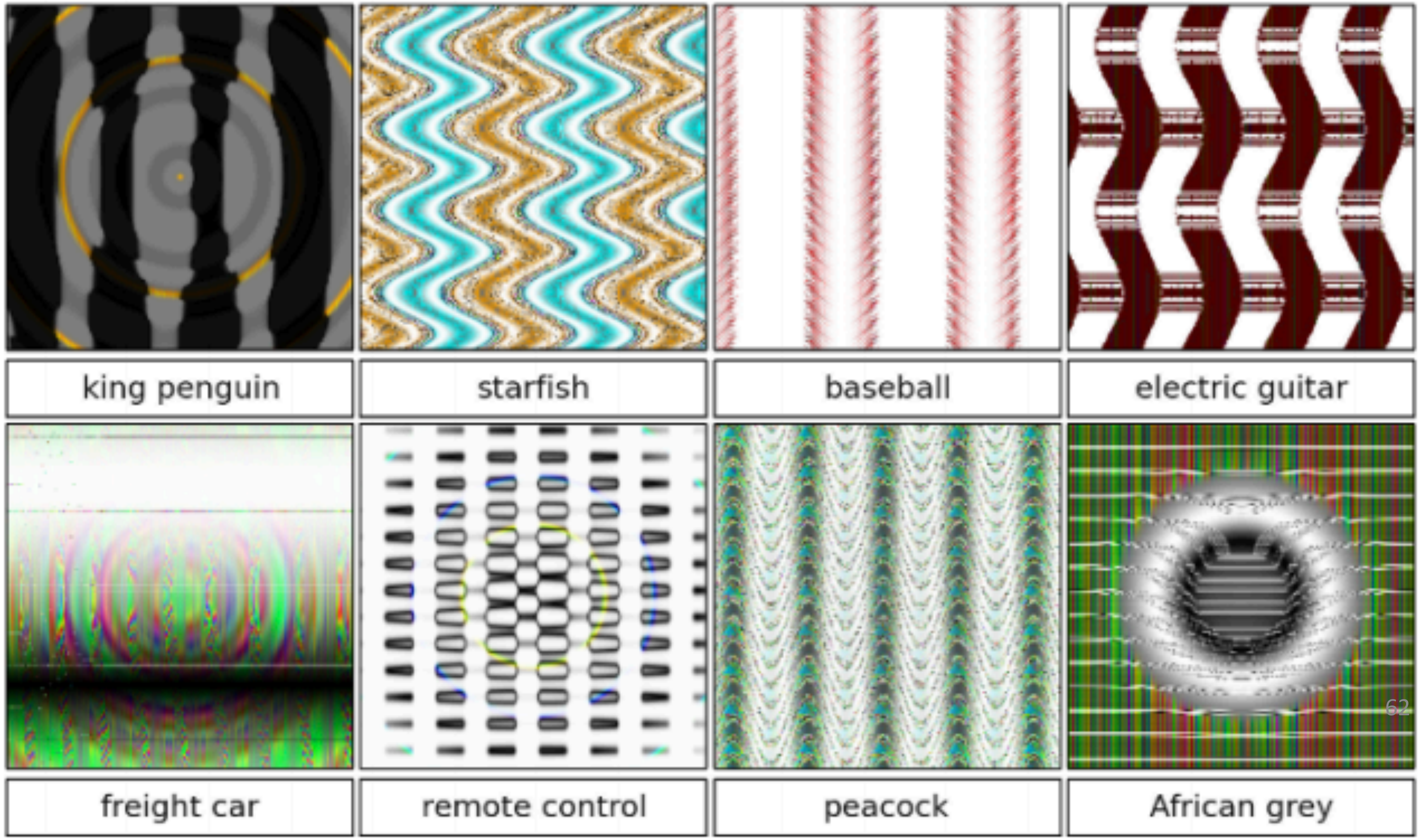
[Nguyen, Yosinski, and Clune, CVPR 2015]



Source: Isola, Torralba, Freeman

“Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images”

[Nguyen, Yosinski, and Clune, CVPR 2015]

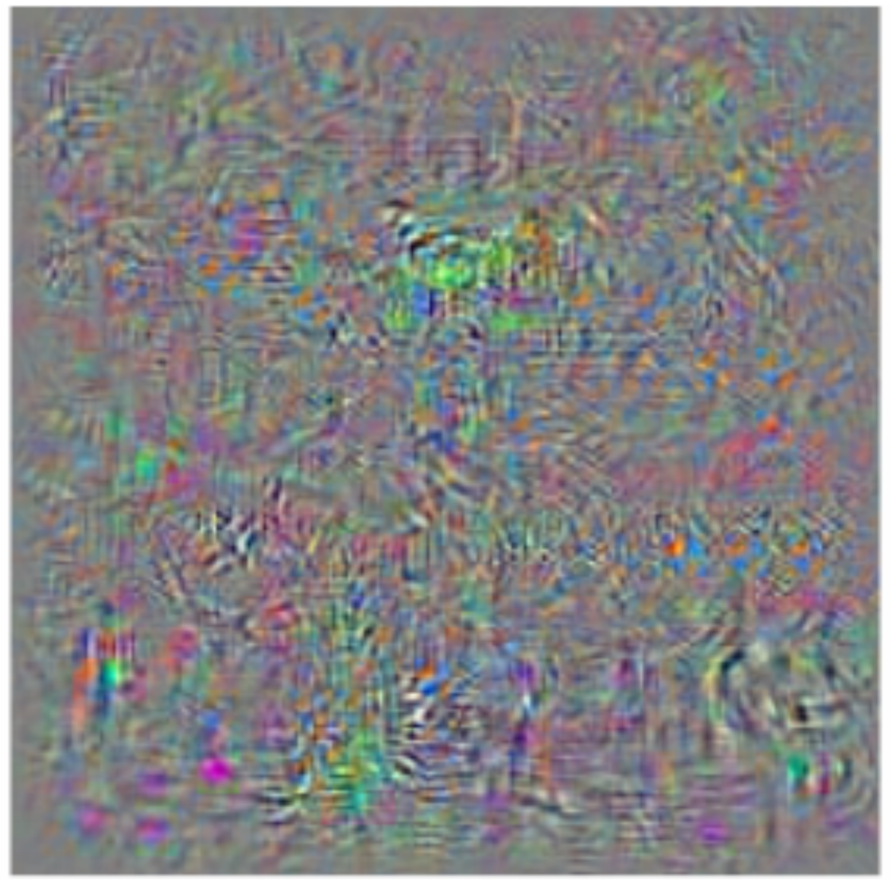


Source: Isola, Torralba, Freeman

Adversarial noise

$\mathbf{x}$ $\qquad$ $\mathbf{r}$ $\qquad$ $\mathbf{x}+\mathbf{r}$

+



=

f $\qquad$ f

y $\quad$ "School bus" $\qquad$ "Ostrich"

$$\arg\max_{\mathbf{r}} p(y = \texttt{ostrich} | \mathbf{x}+\mathbf{r}) \quad \text{subject to} \quad \|\mathbf{r}\| < \epsilon$$

["Intriguing properties of neural networks", Szegedy et al. 2014]

Source: Isola, Torralba, Freeman

Anything to worry about?

“NO Need to Worry about Adversarial Examples in Object Detection in Autonomous Vehicles”, Lu et al. 2017

(Early) 2017’s attacks fail on physical objects, since they are optimized to attack a single view!

Source: Isola, Torralba, Freeman

Anything to worry about?

Later in 2017…

“Synthesizing Robust Adversarial Examples”, Athalye, Engstrom, Ilyas, Kwok, 2017

3D-printed **turtle** model classified as **rifle** from most viewpoints

Source: Isola, Torralba, Freeman

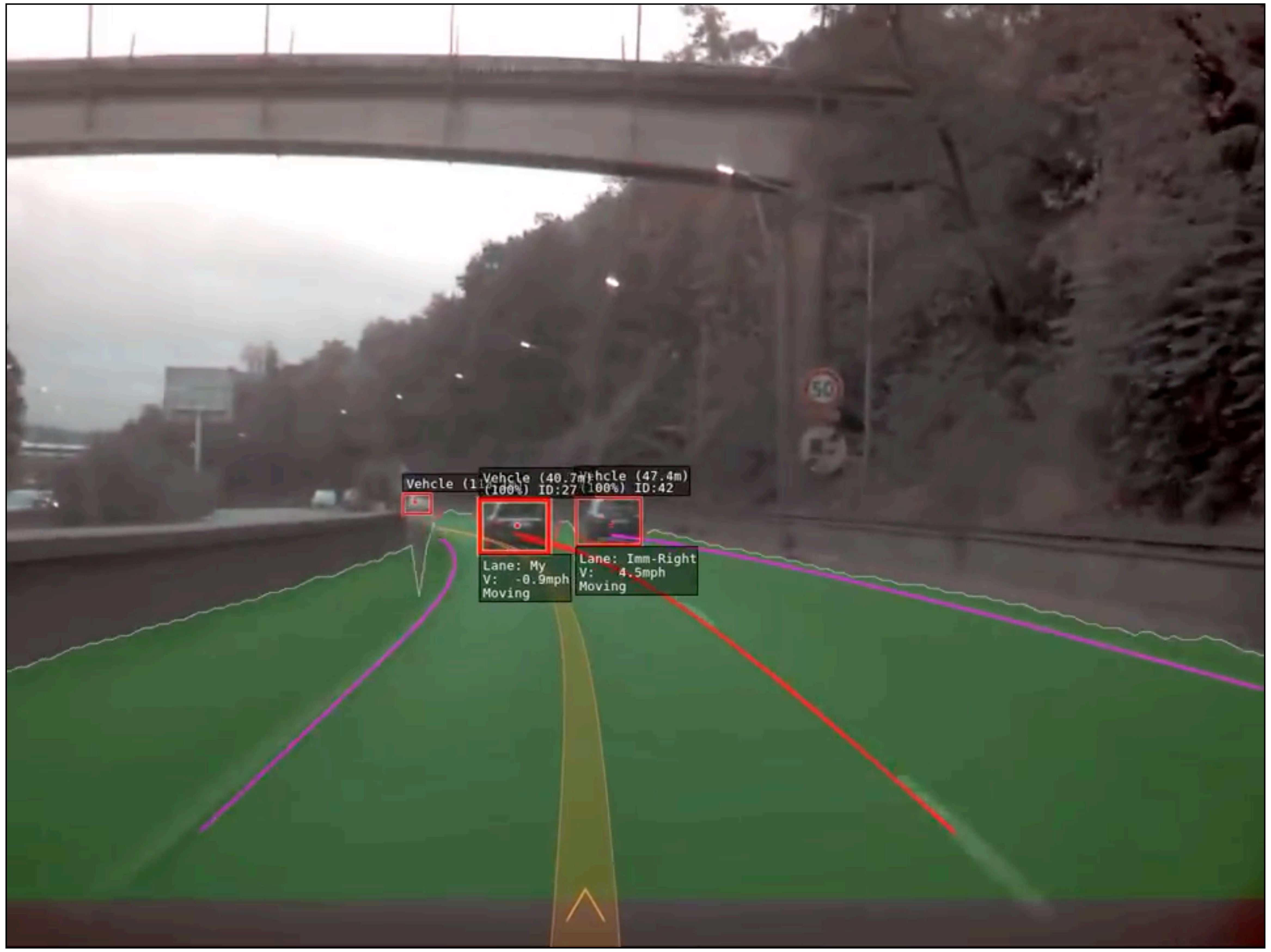
Adversarial examples

- Current deep models have bad **worst-case performance**
- Can be exploited by an adversary
- Few guarantees, can't fully trust what the model's output

Source: Isola, Torralba, Freeman

Problems of applying computer vision in practice

Mission-critical computer vision systems



Source: https://www.nytimes.com/2020/02/25/business/tesla-autopilot-ntsb.html, https://qz.com/1402843/watch-teslas-autopilot-see-the-streets-of-paris/

Social consequences

Color Matters in Computer Vision

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

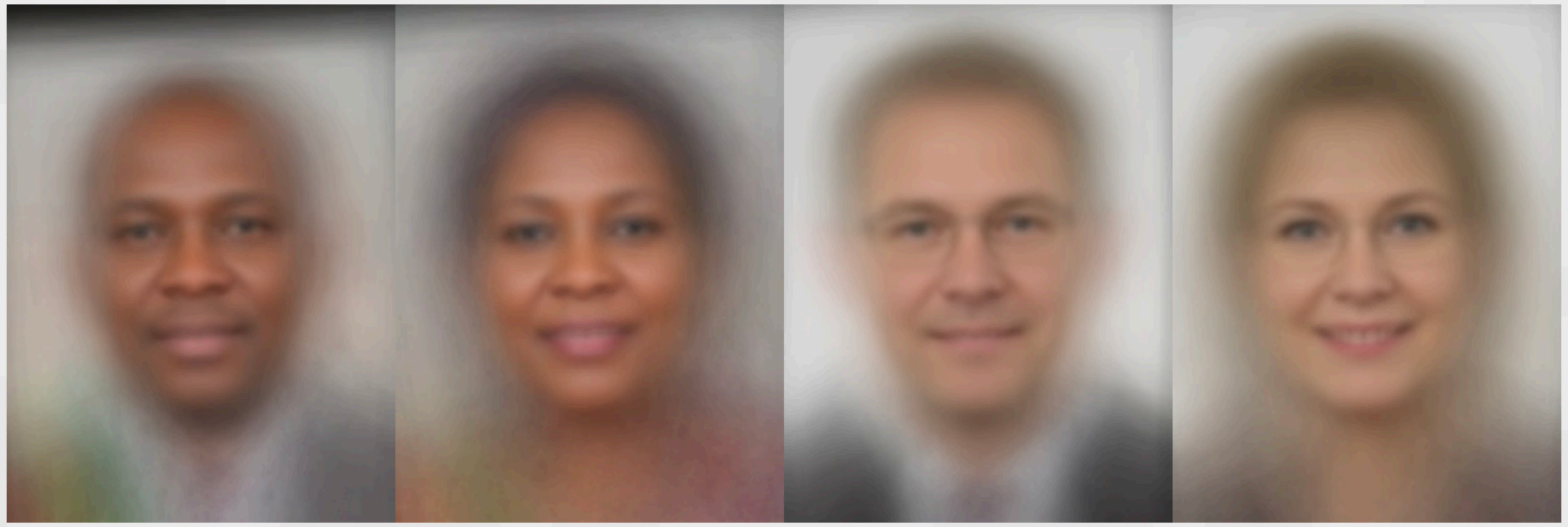
Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

Source: Isola, Torralba, Freeman

Algorithmic Bias

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



http://gendershades.org/overview.html

Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A and 86.2% for Adience) and introduce a new facial analysis dataset which is balanced by gender and skin type. We evaluate 3 commercial gender classification systems using our dataset and show that darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms.

Keywords: Computer Vision, Algorithmic Audit, Gender Classification

1. Introduction

Artificial Intelligence (AI) is rapidly infiltrating every aspect of society. From helping determine who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O'Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have serious consequences. For example, someone could be wrongfully accused of a crime based on erroneous but confident misidentification of the perpetrator from security video footage analysis.

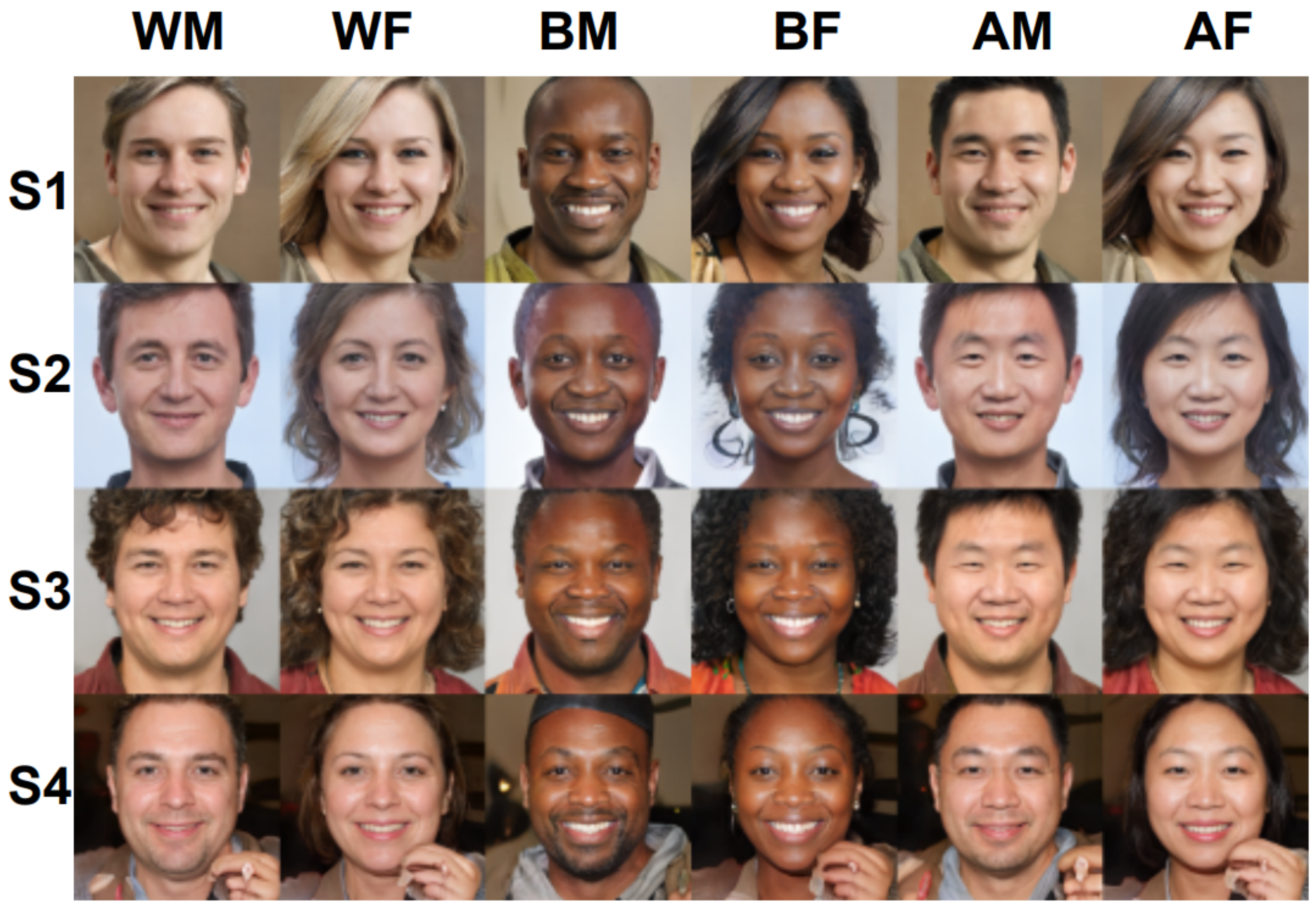
Many AI systems, e.g. face recognition tools, rely on machine learning algorithms that are trained with labeled data. It has recently been shown that algorithms trained with biased data have resulted in algorithmic discrimination (Bolukbasi et al., 2016; Caliskan et al., 2017). Bolukbasi et al. even showed that the popular word embedding space, Word2Vec, encodes societal gender biases. The authors used Word2Vec to train an analogy generator that fills in missing words in analogies. The analogy man is to computer programmer as woman is to "X" was completed with "homemaker", conforming to the stereotype that programming is associated with men and homemaking with women. The biases in Word2Vec are thus likely to be propagated throughout any system that uses this embedding.

* Download our gender and skin type balanced PPB dataset at gendershades.org

http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Source: Isola, Torralba, Freeman

Algorithmic Bias



Benchmarking Algorithmic Bias in Face Recognition:
An Experimental Approach Using Synthetic Faces and Human Evaluation

Hao Liang
Rice University
hl106@rice.edu

Pietro Perona
California Institute of Technology and AWS
perona@caltech.edu, peronapp@amazon.com

Guha Balakrishnan
Rice University
guha@rice.edu

Abstract

We propose an experimental *method for measuring bias in face recognition systems. Existing methods to measure bias depend on benchmark datasets that are collected in the wild and annotated for protected (e.g., race, gender) and unprotected (e.g., pose, lighting) attributes. Such* observational *datasets only permit correlational conclusions, e.g., "Algorithm A's accuracy is different on female and male faces in dataset X.". By contrast, experimental methods manipulate attributes individually and thus permit causal conclusions, e.g., "Algorithm A's accuracy is affected by gender and skin color."*

Our method is based on generating synthetic faces using a neural face generator, where each attribute of interest is modified independently while leaving all other attributes constant. Human observers crucially provide the ground truth on perceptual identity similarity between synthetic image pairs. We validate our method quantitatively by evaluating race and gender biases of three research-grade face recognition models. Our synthetic pipeline reveals that for these algorithms, accuracy is lower for Black and East Asian population subgroups. Our method can also quantify how perceptual changes in attributes affect face identity distances reported by these models. Our large synthetic

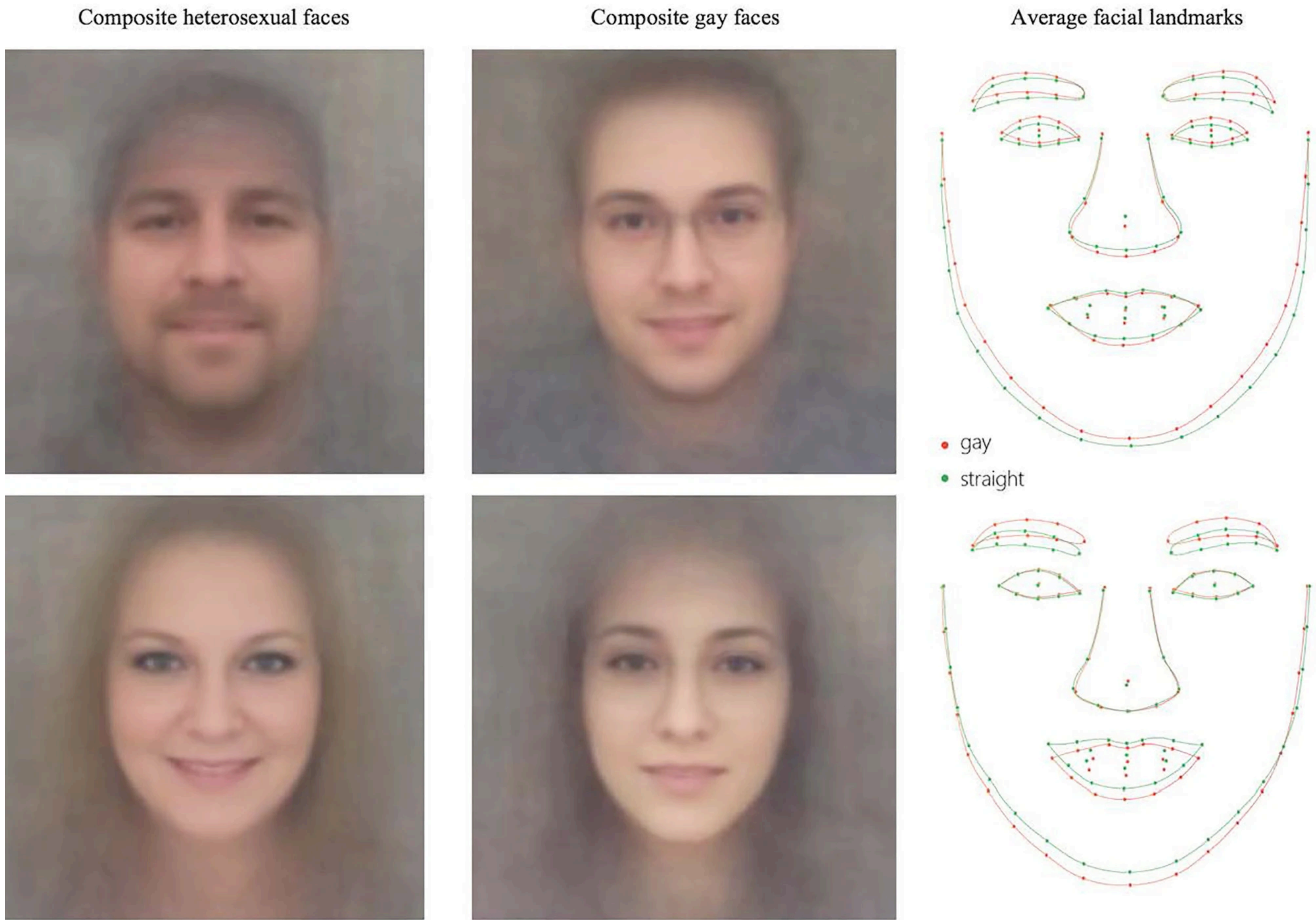
("face identification"). Face recognition systems implemented with deep neural networks today achieve impressive accuracies [54, 13, 33, 43] and outperform even expert face analysts [38]. Nevertheless, it is important to detect and measure possible algorithmic *biases*, i.e., systematic accuracy differences, especially across protected demographic attributes like age, race and gender [10, 22, 20], in order to maintain fair treatment in sensitive applications. For this reason, the National Institute of Standards and Technology (NIST) measures bias in commercial face recognition models [17], in particular by comparing their False Match Rate (FMR) and False Non Match Rate (FNMR) values across different demographic subgroups at a particular decision threshold (sweeping this threshold yields FNMR vs. FMR "curves").

The first step in measuring bias of face recognition systems is, currently, to collect a large benchmarking dataset containing a set of diverse faces, where each is photographed multiple times under different conditions. An algorithm's error rate across subgroups specified by different protected attribute combinations (e.g., different race and gender groups) can then be measured.

Unfortunately, sampling a good test dataset is almost impossible. First, each protected intersectional group (a specific combination of attribute values) must contain a suffi-

[Liang, Perona, Balakrishnan, CVPR 2023]

Bad choice of data



https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html

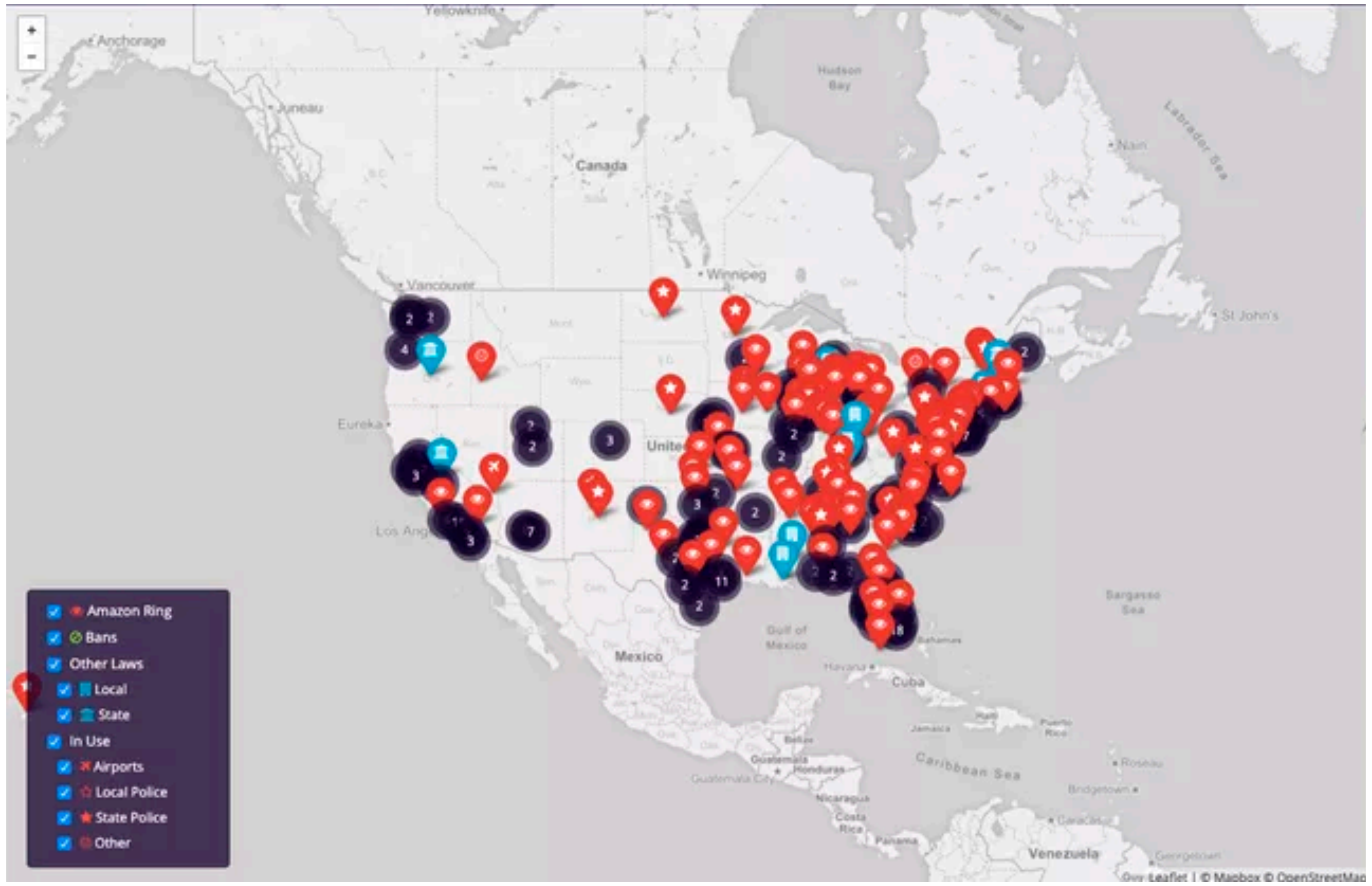
Face recognition in the U.S.

recode

Here's where the US government is using facial recognition technology to surveil Americans

This map shows how widespread the use of facial recognition technology has become.

By Shirin Ghaffary and Rani Molla | Updated Dec 10, 2019, 8:00am EST



https://www.vox.com/recode/2019/7/18/20698307/facial-recognition-technology-us-government-fight-for-the-future

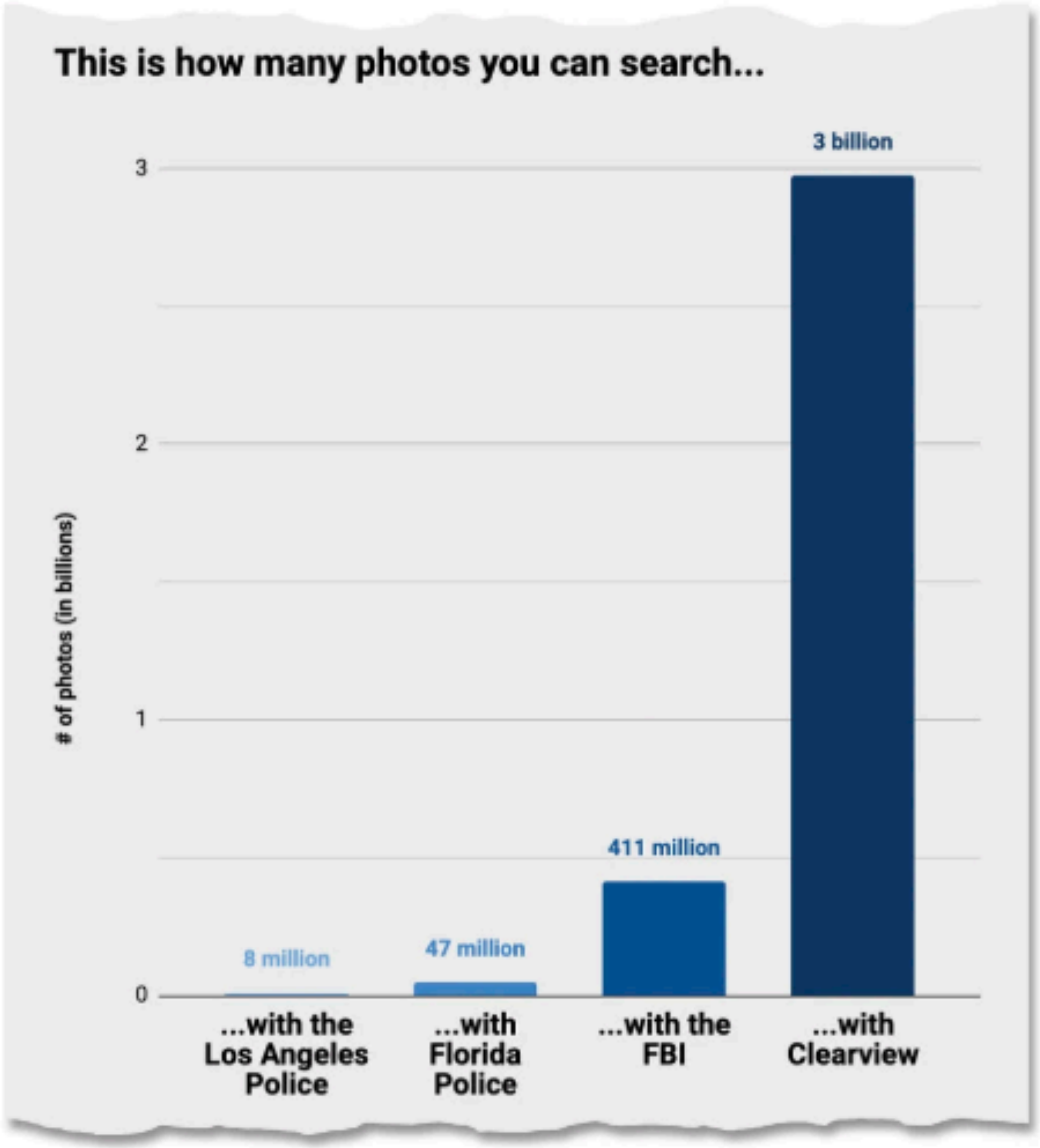
Source: S. Lazebnik

Fears of universal mass surveillance (and dubious claims)

The New York Times

The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and "might lead to a dystopian future or something," a backer says.



A chart from marketing materials that Clearview provided to law enforcement. Clearview

https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

https://www.buzzfeednews.com/article/ryanmac/clearview-ai-nypd-facial-recognition

Source: S. Lazebnik

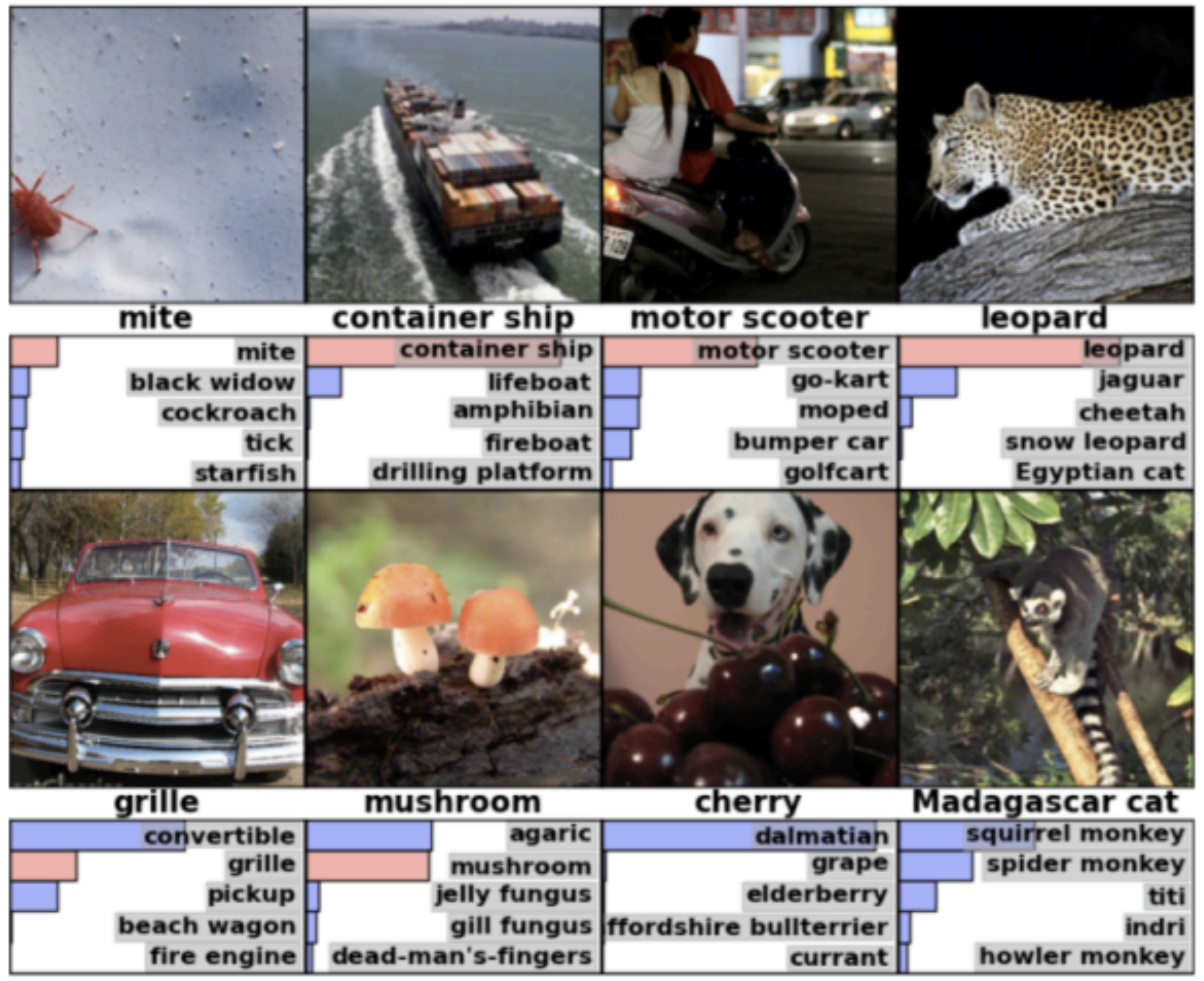
Face recognition in the U.S.

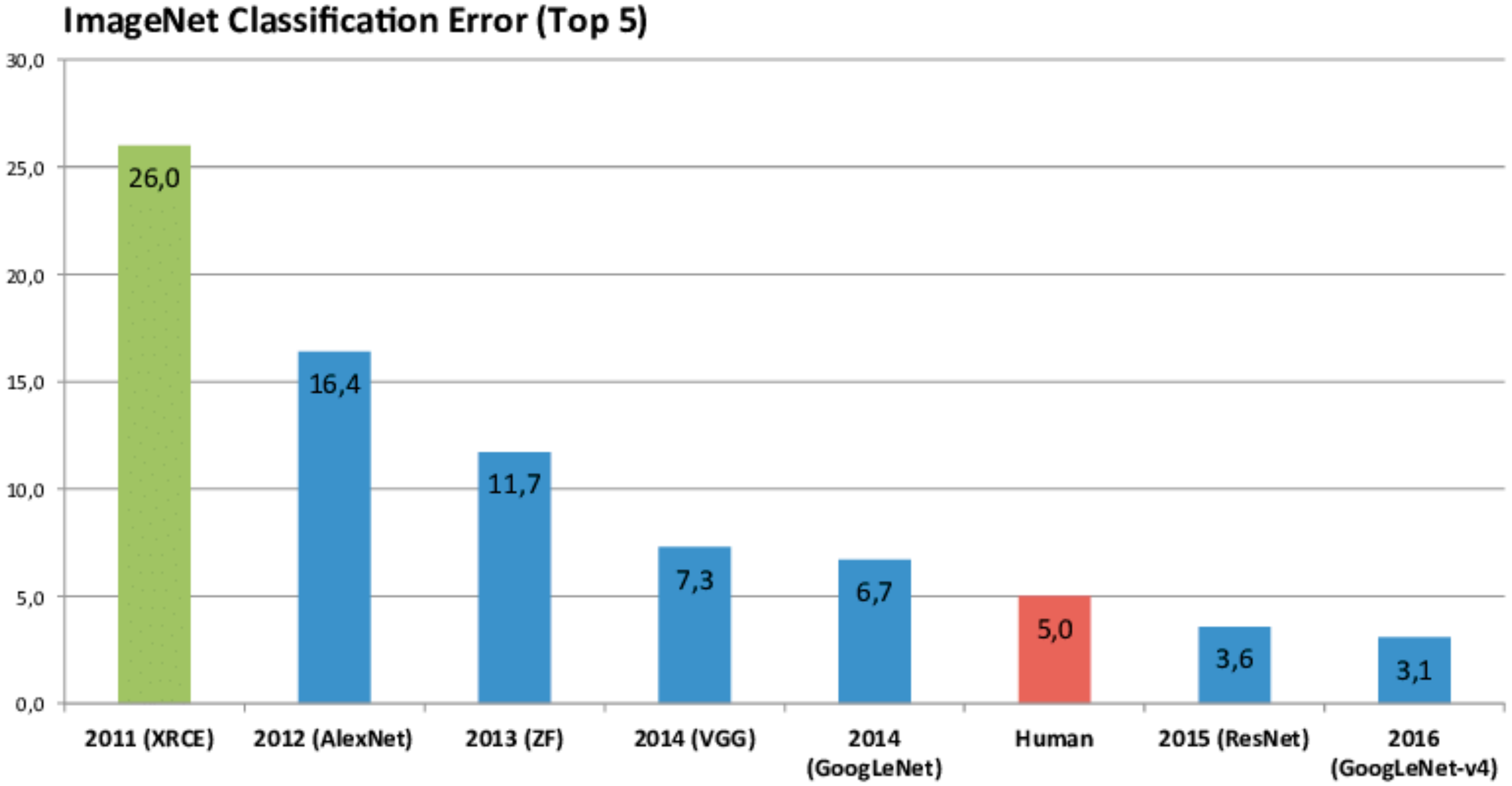
Wrongfully Accused by an Algorithm – New York Times, 6/24/2020

Source: S. Lazebnik

ImageNet: asset or liability?

- Performance on the basic ILSVRC benchmark has saturated





[Figure source]

- Current models have reached levels of accuracy where the presence of human labeling error is starting to affect experimental conclusions ([Beyer et al.] 2020, [Northcutt et al.] 2021)

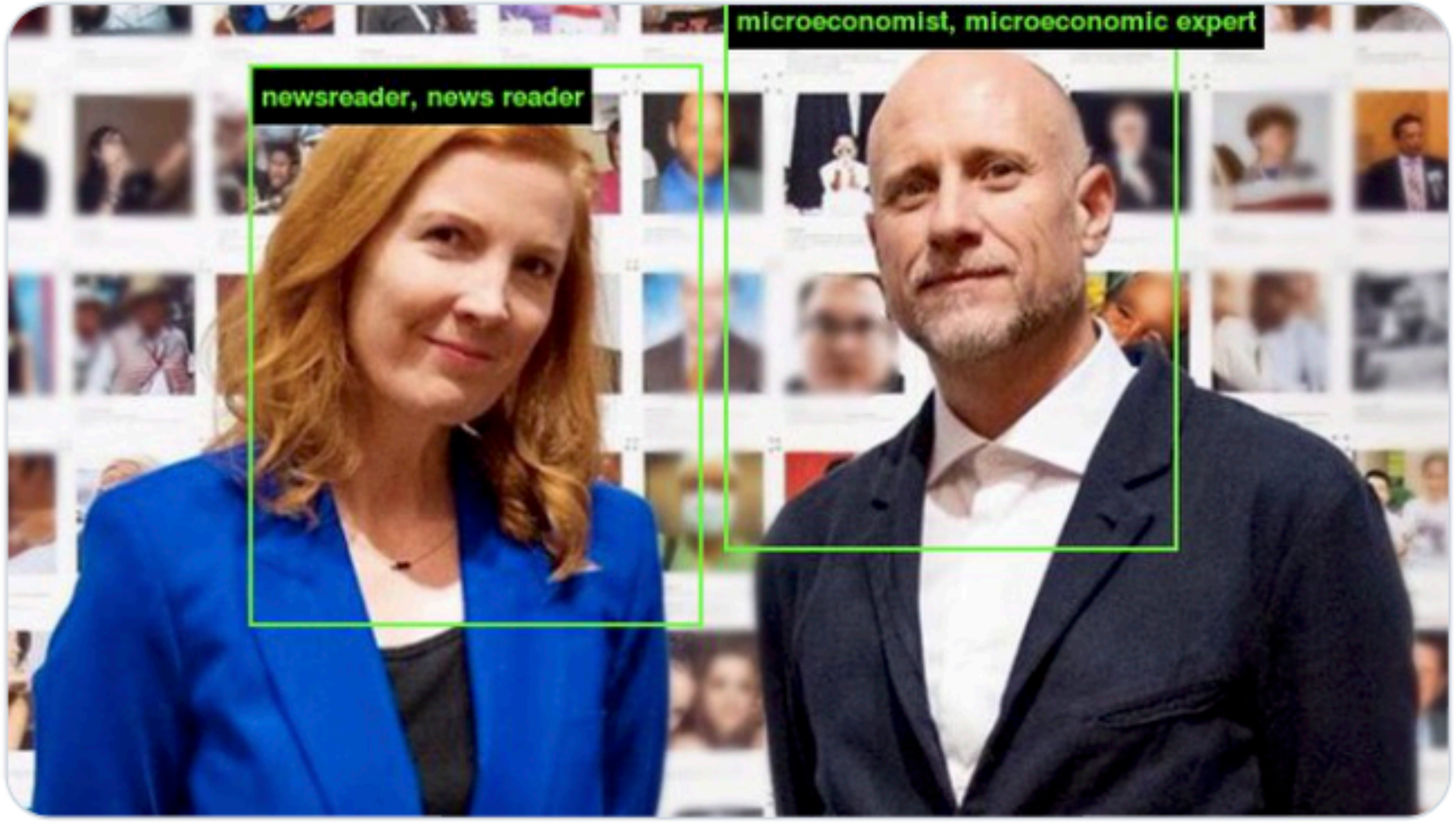
Source: S. Lazebnik

ImageNet labeling problems: ImageNet Roulette

Kate Crawford @katecrawford · Sep 16, 2019

Want to see how an AI trained on ImageNet will classify you? Try ImageNet Roulette, based on ImageNet's Person classes. It's part of the 'Training Humans' exhibition by @trevorpaglen & me - on the history & politics of training sets. Full project out soon

imagenet-roulette.paglen.com



ImageNet Roulette uses an open source Caffe deep learning framework (produced at UC Berkeley) trained on the images and labels in the "person" categories (which are currently 'down for maintenance'). Proper nouns and categories with less than 100 pictures were removed.

When a user uploads a picture, the application first runs a face detector to locate any faces. If it finds any, it sends them to the Caffe model for classification. The application then returns the original images with a bounding box showing the detected face and the label the classifier has assigned to the image. If no faces are detected, the application sends the entire scene to the Caffe model and returns an image with a label in the upper left corner.

ImageNet contains a number of problematic, offensive and bizarre categories - all drawn from WordNet. Some use misogynistic or racist terminology. Hence, the results ImageNet Roulette returns will also draw upon those categories. That is by design: we want to shed light on what happens when technical systems are trained on problematic training data. AI classifications of people are rarely made visible to the people being classified. ImageNet Roulette provides a glimpse into that process – and to show the ways things can go wrong.

K. Crawford and T. Paglen, [Excavating AI: The Politics of Training Sets for Machine Learning](), September 2019 https://www.theverge.com/tldr/2019/9/16/20869538/imagenet-roulette-ai-classifier-web-tool-object-image-recognition

Source: S. Lazebnik

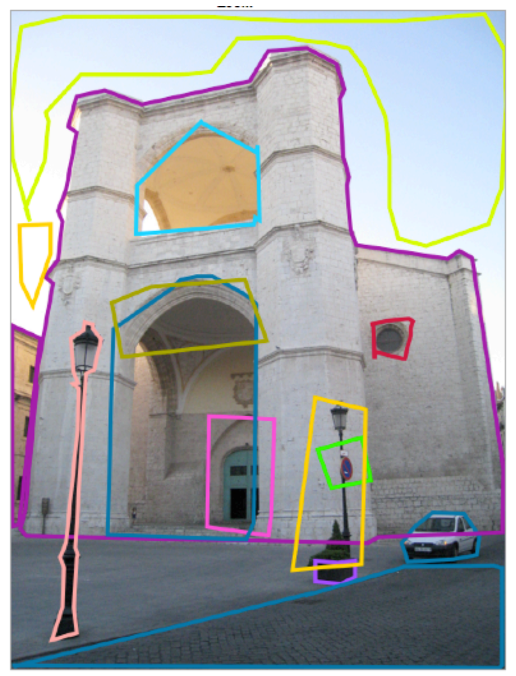
ImageNet Roulette

Source: S. Lazebnik

Image source

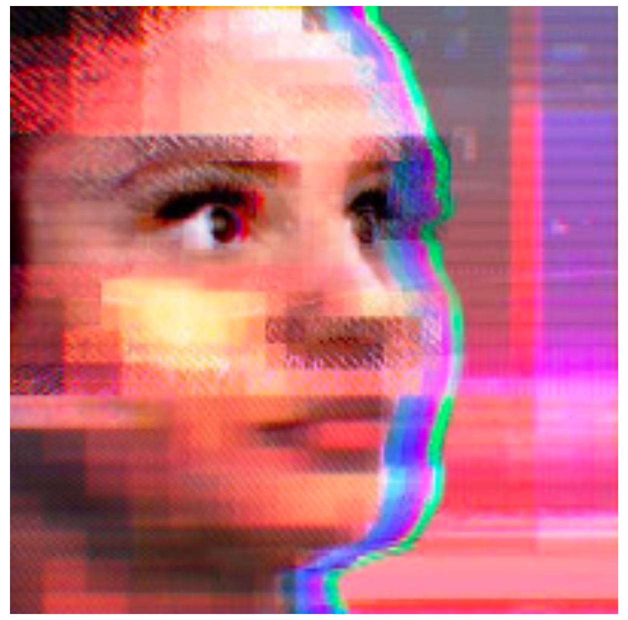
Some things to worry about…

- Our datasets are often poorly labeled



- And usually biased

- ML methods may perform well on lab-collected data, but often generalize poorly to real-world data



- Can have negative social consequences

Adapted from Isola, Torralba, Freeman

Open-ended discussion

- Supervised vs. unsupervised learning?
- Other negative consequences of computer vision systems?
- What other biases might computer vision systems have?

Thank you!