Human-AI Interaction, Lecture 3

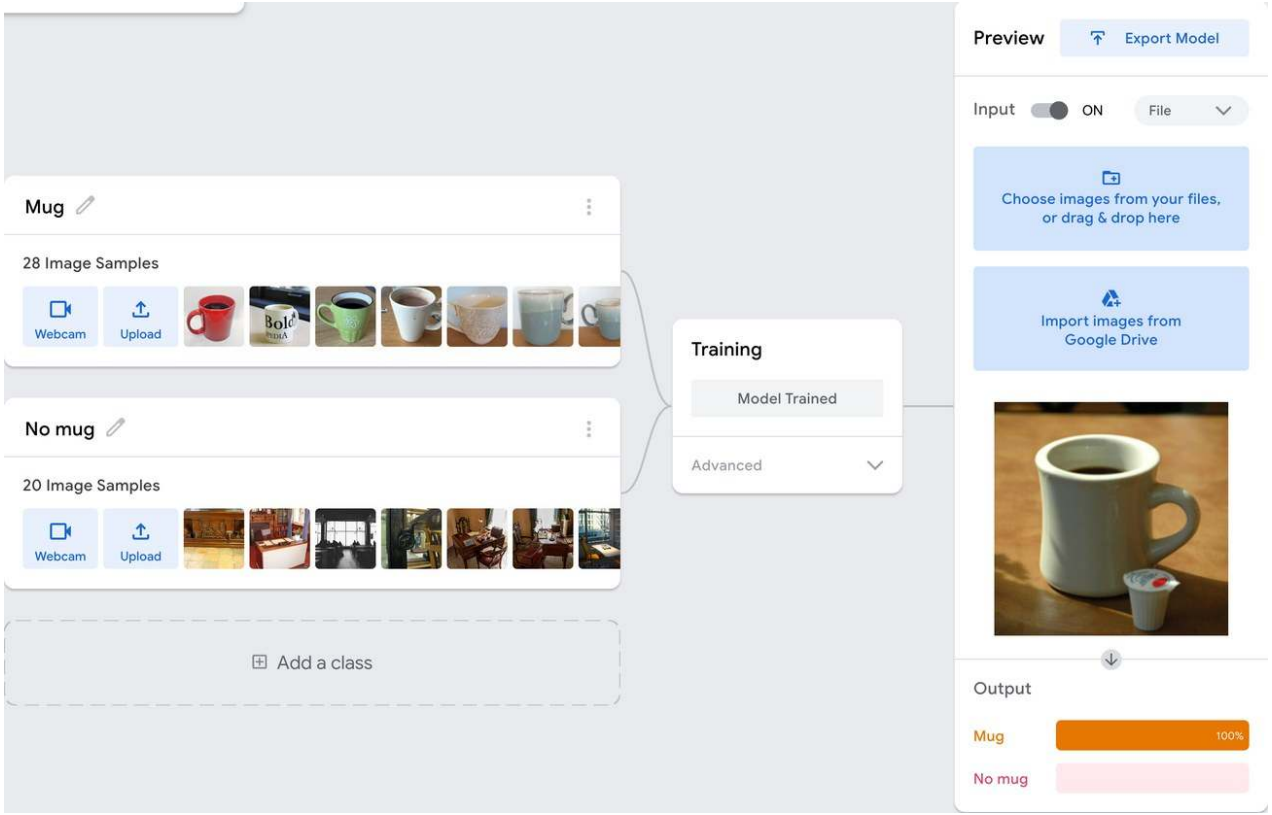
Predictive Models

Google's Teachable Machine



Google, “Now anyone can explore machine learning, no coding required,” The Keyword, October 3, 2017

Demo: training on mugs and desks



Teachable Machine. Photos from Wikimedia Commons.

Two new photos

A bowl





Teachable Machine. Photos from Wikimedia Commons.

Two new photos

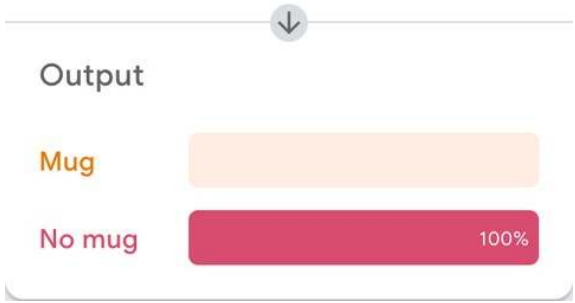
A bowl





A mug on a desk





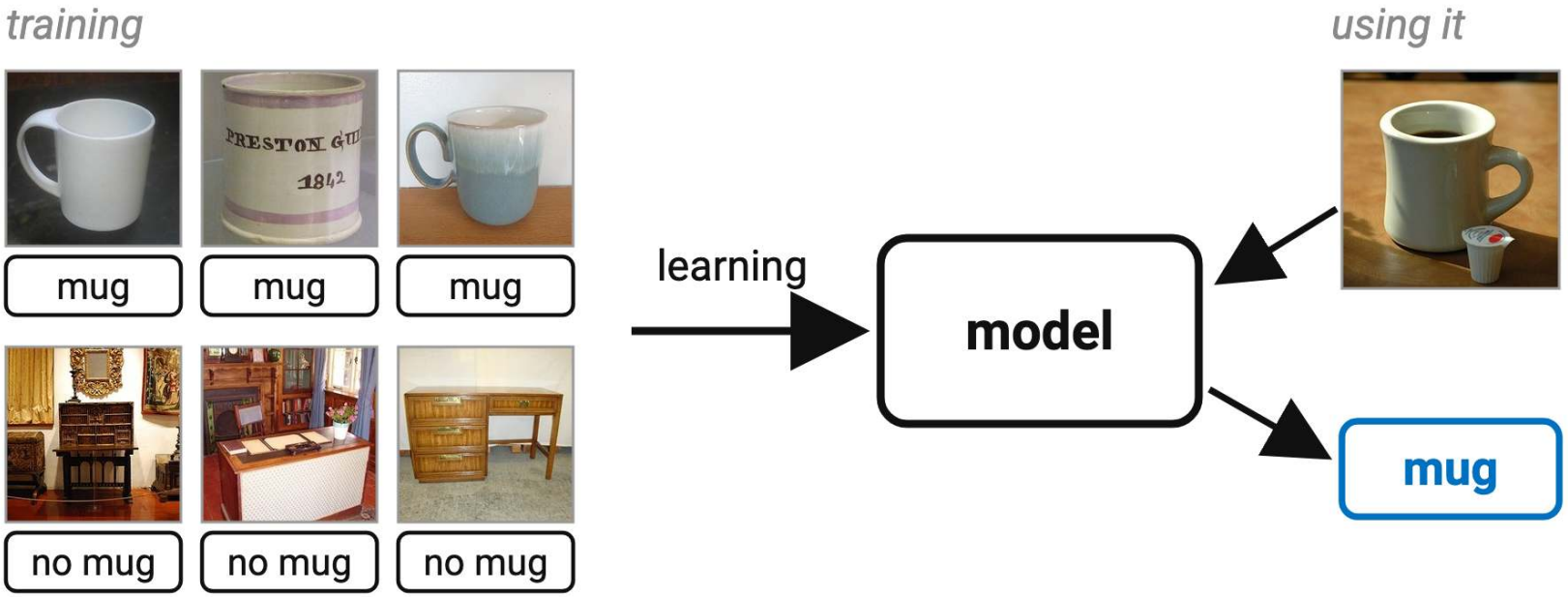
Teachable Machine. Photos from Wikimedia Commons.

What does a model learn from examples, and **where will it fail**?

Try writing the rule

```
def is_mug(photo):
    # photo: a grid of pixels,
    # each pixel is 3 numbers (red, green, blue)
    if ???:
        return "mug"
    return "no mug"
```

Learning from labeled examples



“Machine learning systems automatically learn programs from data.”

Domingos, CACM 2012, p. 78. Photos from Wikimedia Commons.

Three terms

Training data: the examples

Label: the right answer

Feature: a measured number

Our training photos

Mug: 28 photos



No mug: 20 photos



Wikimedia Commons, categories Mugs, White mugs, Coffee mugs and Desks

Our training photos

Mug: 28 photos



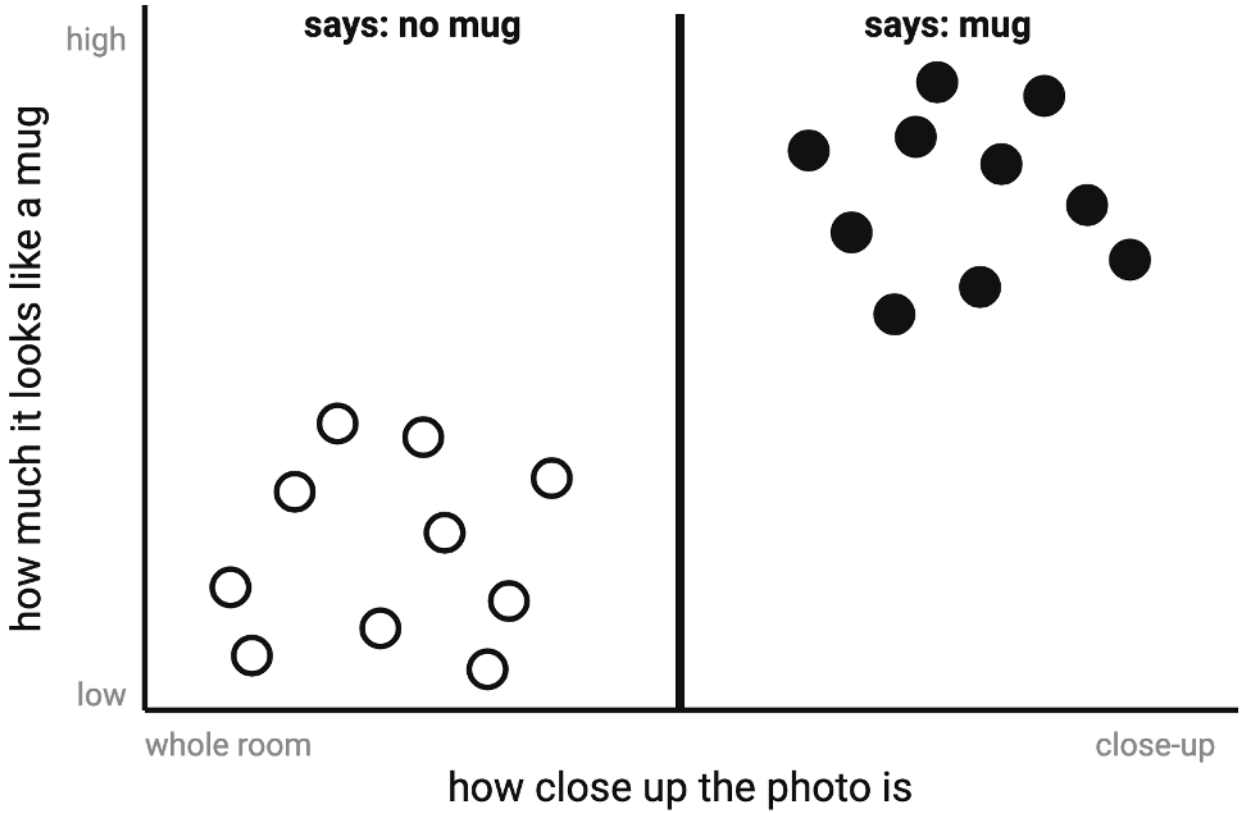
No mug: 20 photos



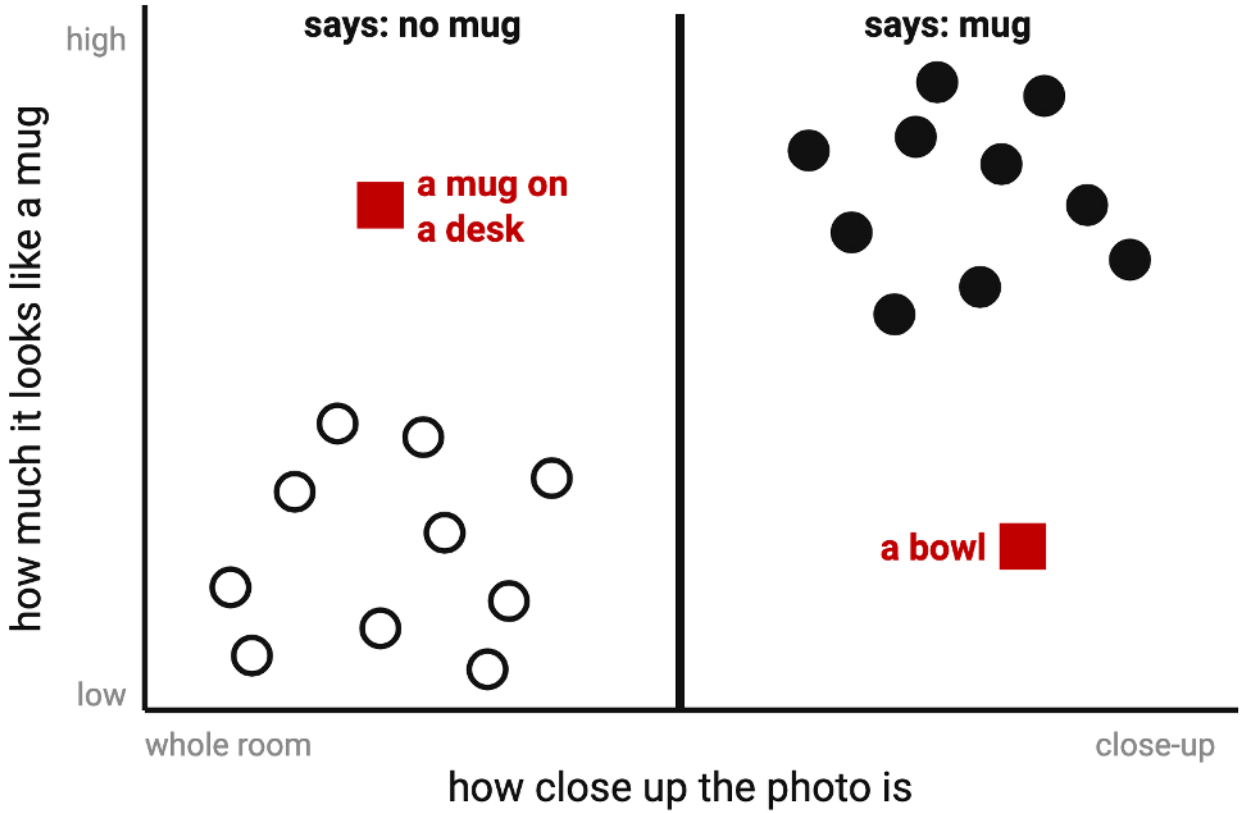
All 28 mug photos are **close-ups**, and all 20 desk photos show a whole room.

Wikimedia Commons, categories Mugs, White mugs, Coffee mugs and Desks

Training the model



The learned rule: close-up means mug



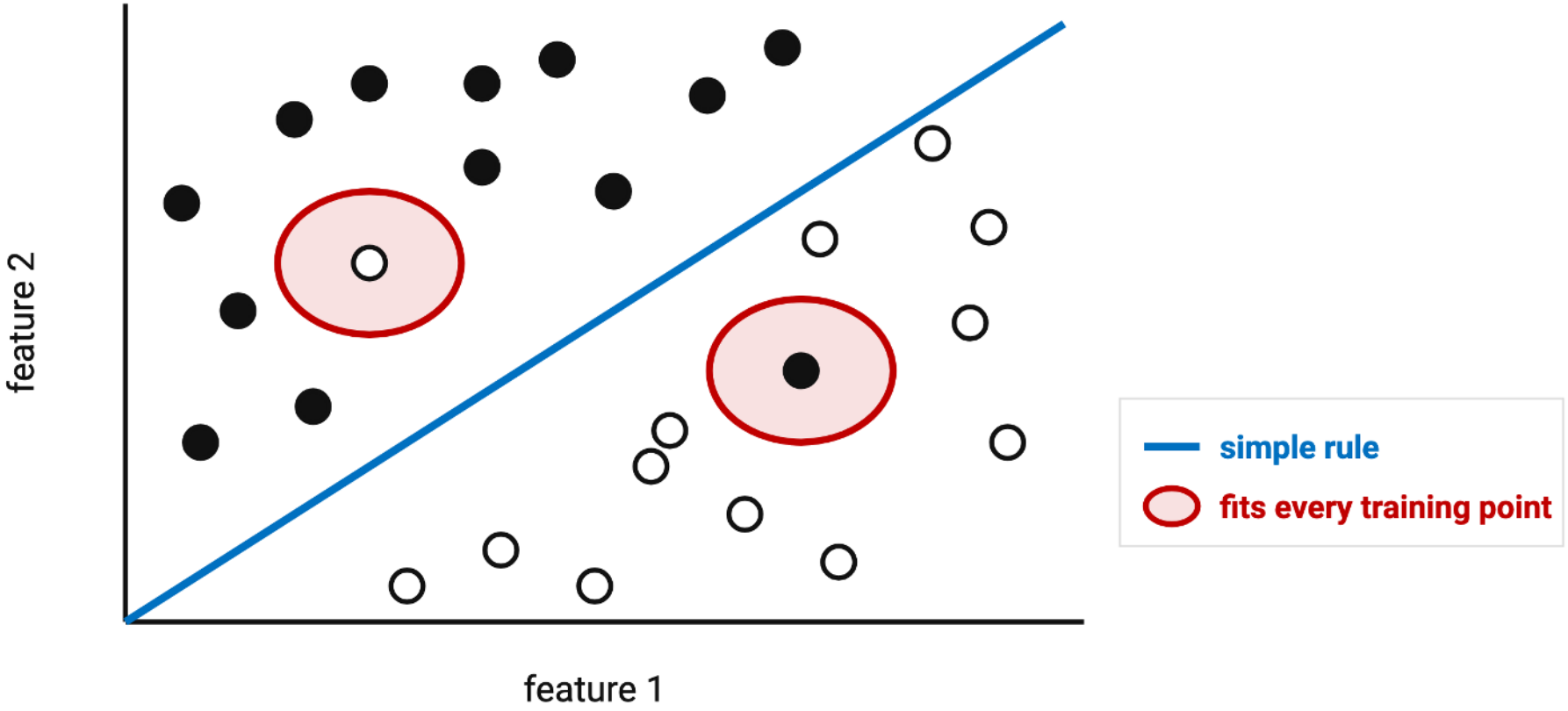
Generalization

“Doing well on the training set is easy (just **memorize** the examples).”

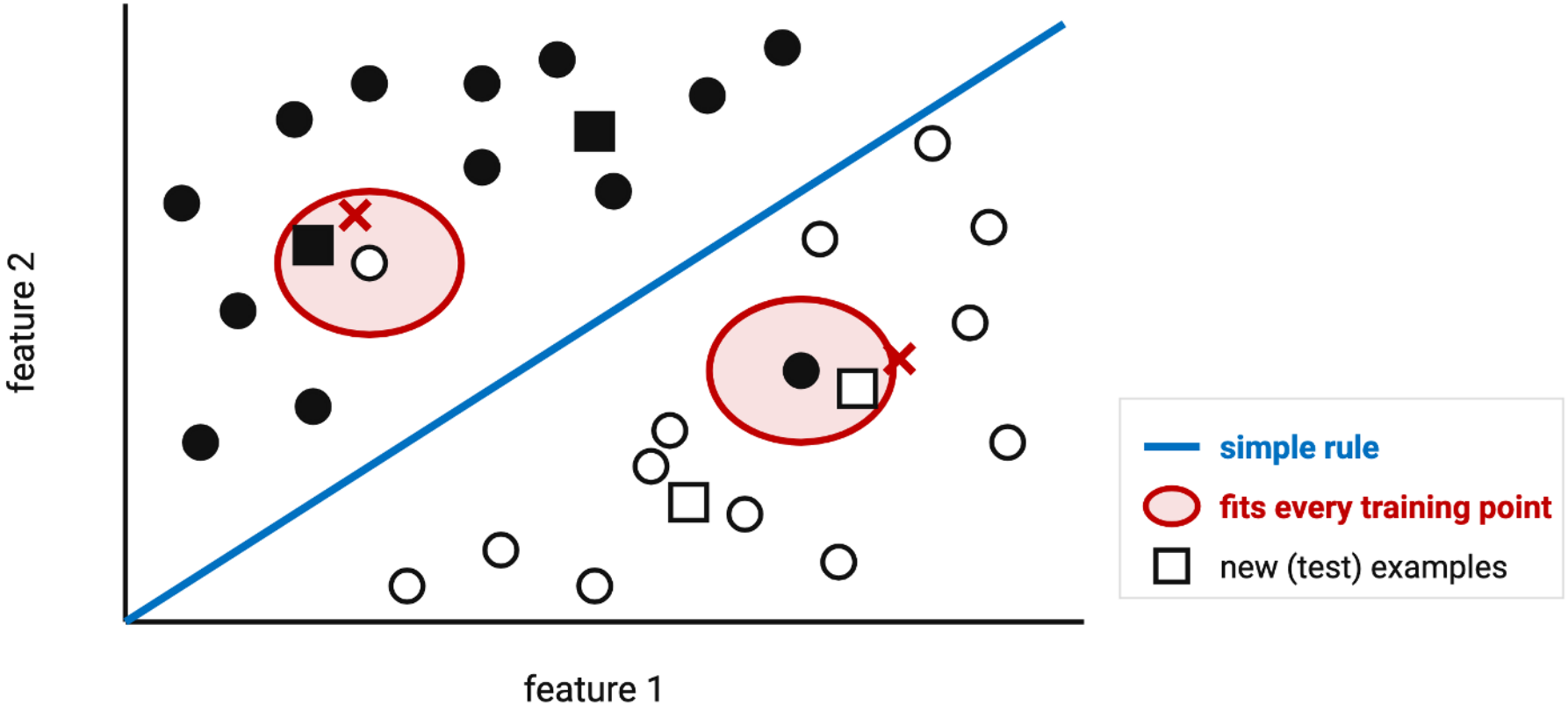
Domingos, CACM 2012, p. 80

Teachable Machine, Under the hood panel, Vocab

Two rules that fit the training examples



Overfitting



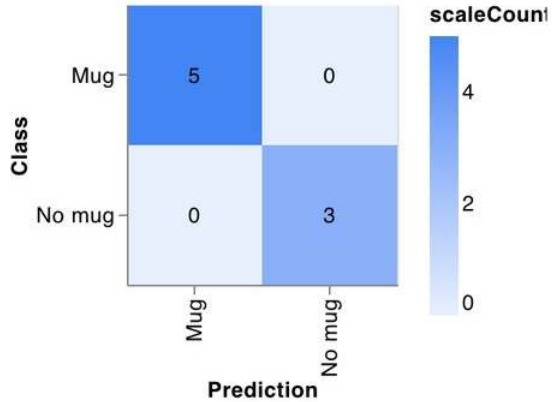
Teachable Machine: test results

The 8 test samples come from the same 48 photos.

Accuracy per class

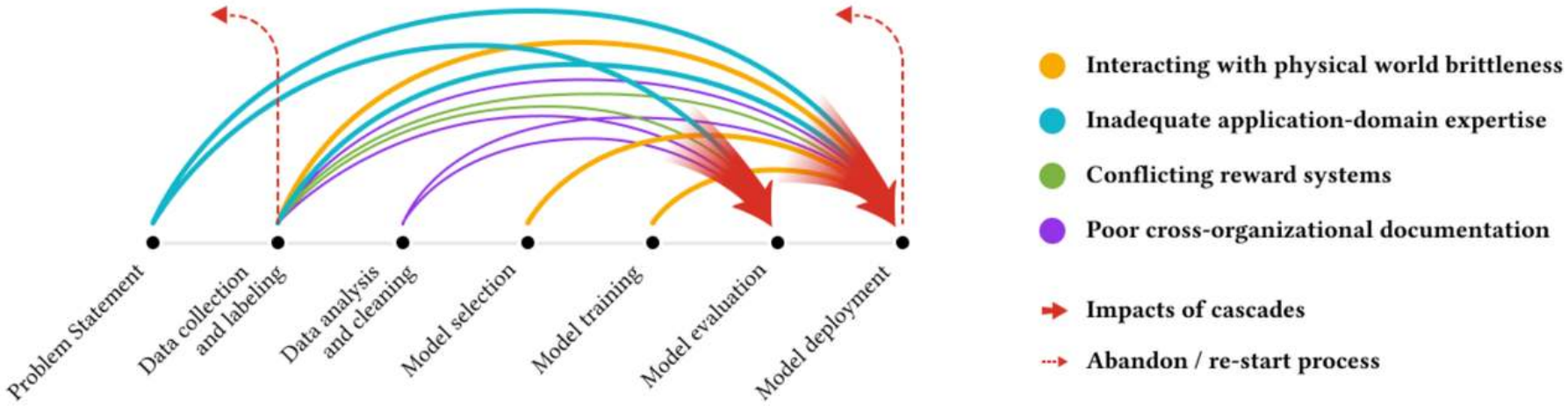
CLASS	ACCURACY	# SAMPLES
Mug	1.00	5
No mug	1.00	3

Confusion Matrix



Teachable Machine, Under the hood: accuracy per class and confusion matrix

Eye disease models trained on clean images failed when used on images with **small specks of dust**



Sambasivan et al., CHI 2021, p. 2 and Fig. 1 (data cascades: a problem in the data at one stage that causes harm at later stages)

Break the model: groups of 3, 15 minutes

Teach it: two classes, about 50 frames each

Break it: change one thing at a time, note what was missing

Fix it: add a few examples and retrain

teachablemachine.withgoogle.com, Image Project

Training runs in your browser, and the webcam frames stay on your computer (Carney et al., CHI 2020 Extended Abstracts). Ask before you record a classmate.

Writing down a failure



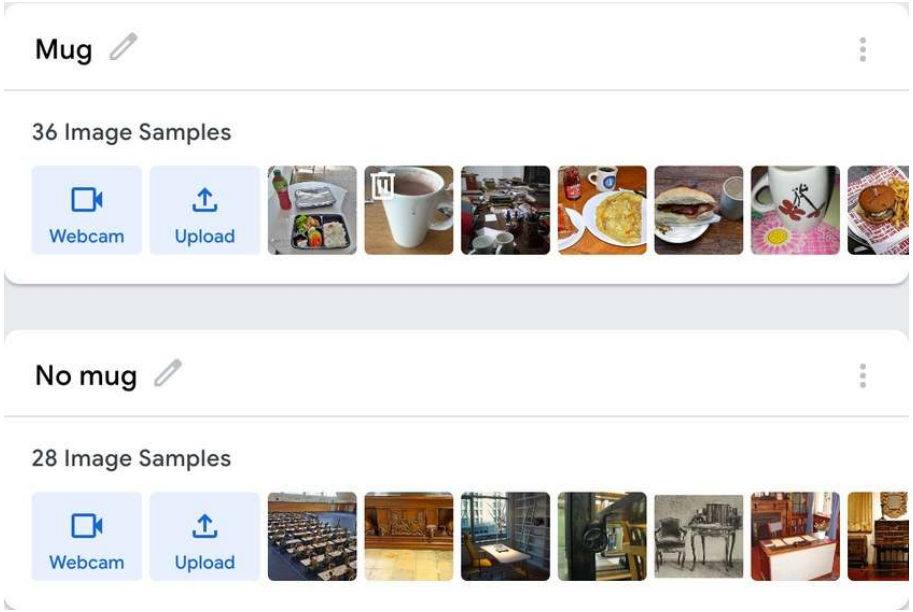
Changed: a glass instead of a mug

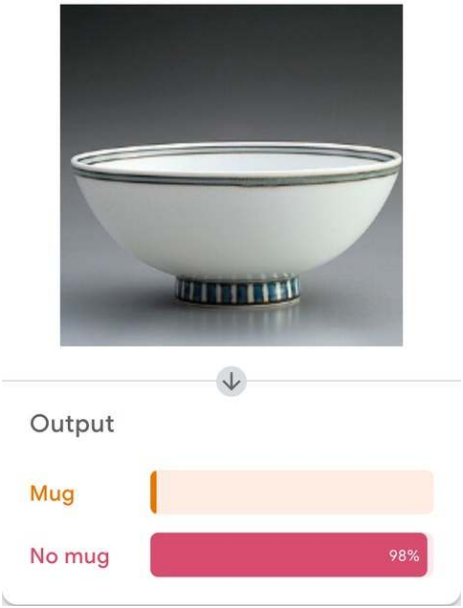
It said: Mug, 100%

Missing from training: close-ups that are not mugs

Teachable Machine. Photo from Wikimedia Commons.

Fixing a failure

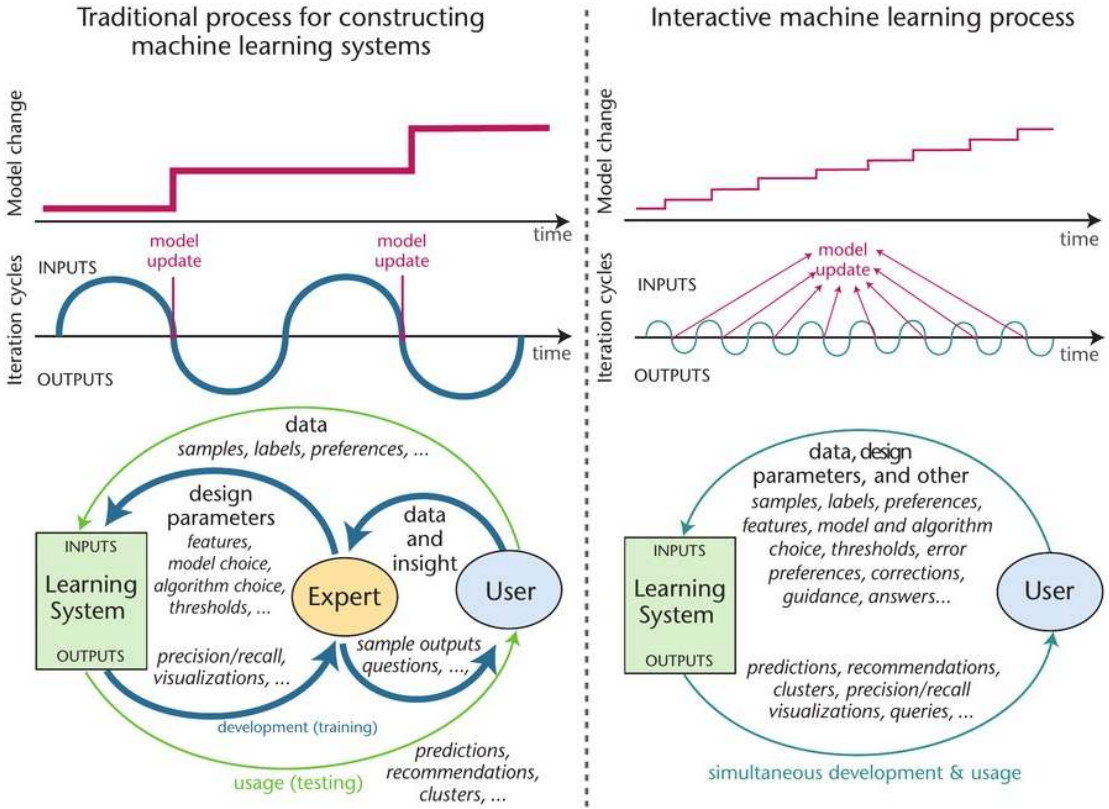




Fixed: both bowls and one glass. Still wrong: the vase, the other glass, and 3 of 5 mugs on desks.

Teachable Machine. Photos from Wikimedia Commons.

Interactive machine learning

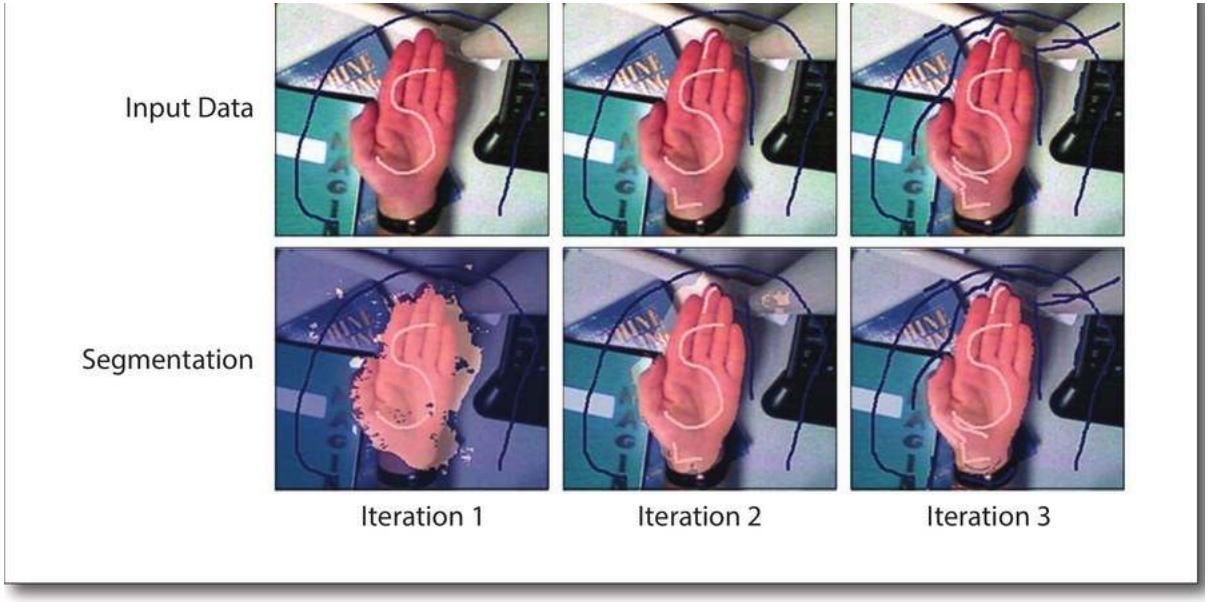


Left: experts, long cycles, big changes

Right: the person sees each small change

Amershi et al., AI Magazine 2014, Fig. 1

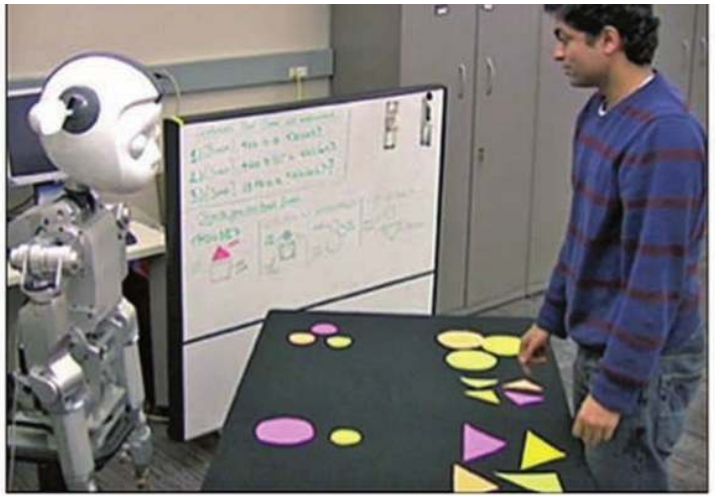
Rapid, focused and incremental updates

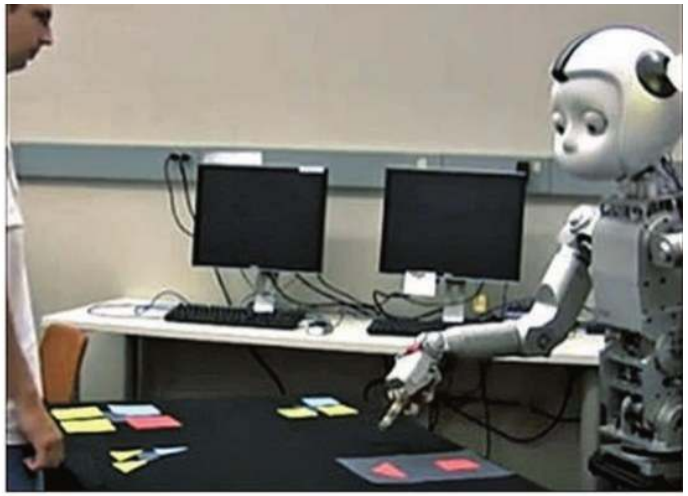


Crayons (Fails & Olsen, 2003)

Amershi et al., AI Magazine 2014, p. 106 and Fig. 2

People, not oracles





+ Wekinator students learned to spot noisy examples
- a robot that kept asking felt annoying
- people suggested features, not only labels

Cakmak et al. 2010 (left: the person picks the examples; right: the robot asks), Fiebrink et al. 2011 and Stumpf et al. 2007, in Amershi et al., AI Magazine 2014, Fig. 3

Discussion: Domingos and Amershi

1. What did our mug model pick up, and how could you have caught it earlier?

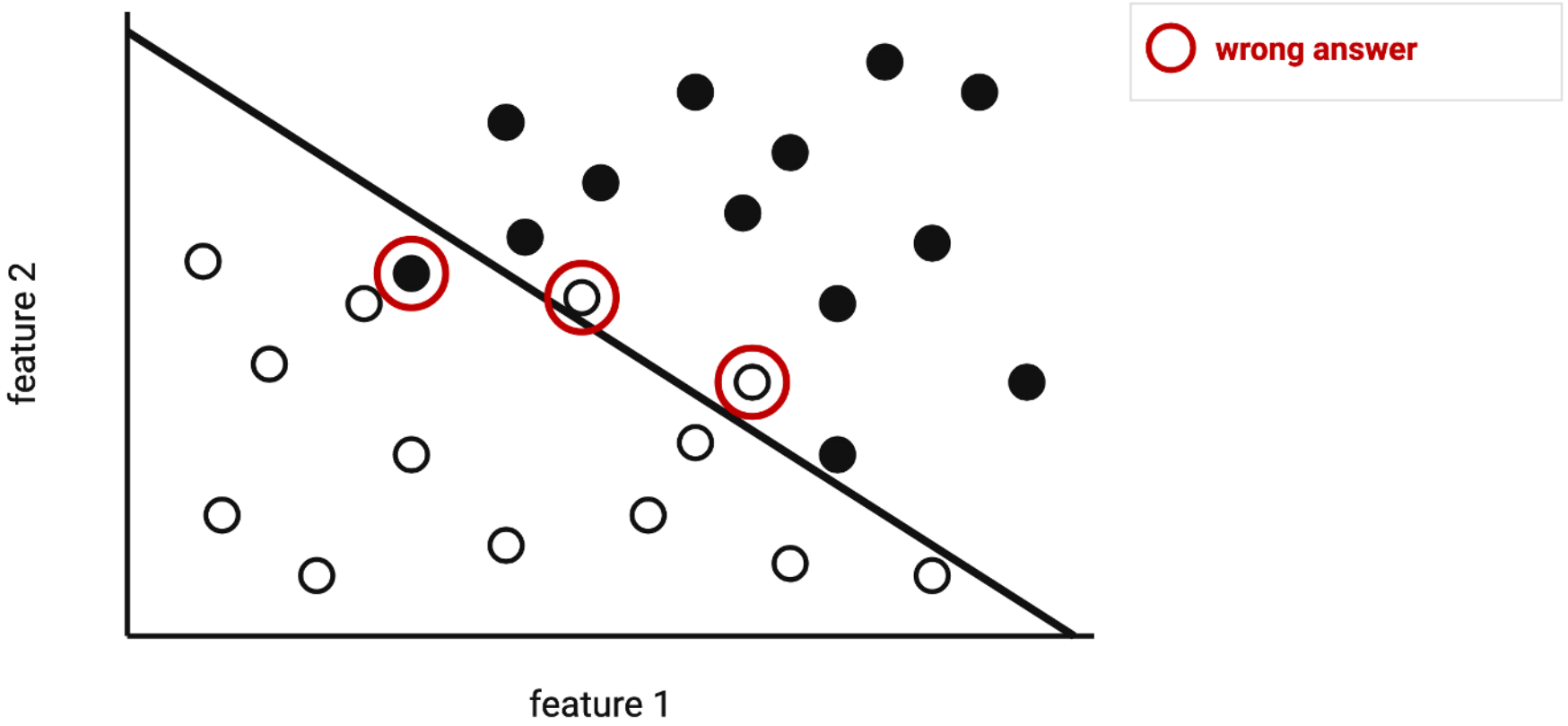
2. What would teaching a model feel like for the people you watch in milestone 1?

3. A model scores 95% on its test set. What would you ask before giving it to people?

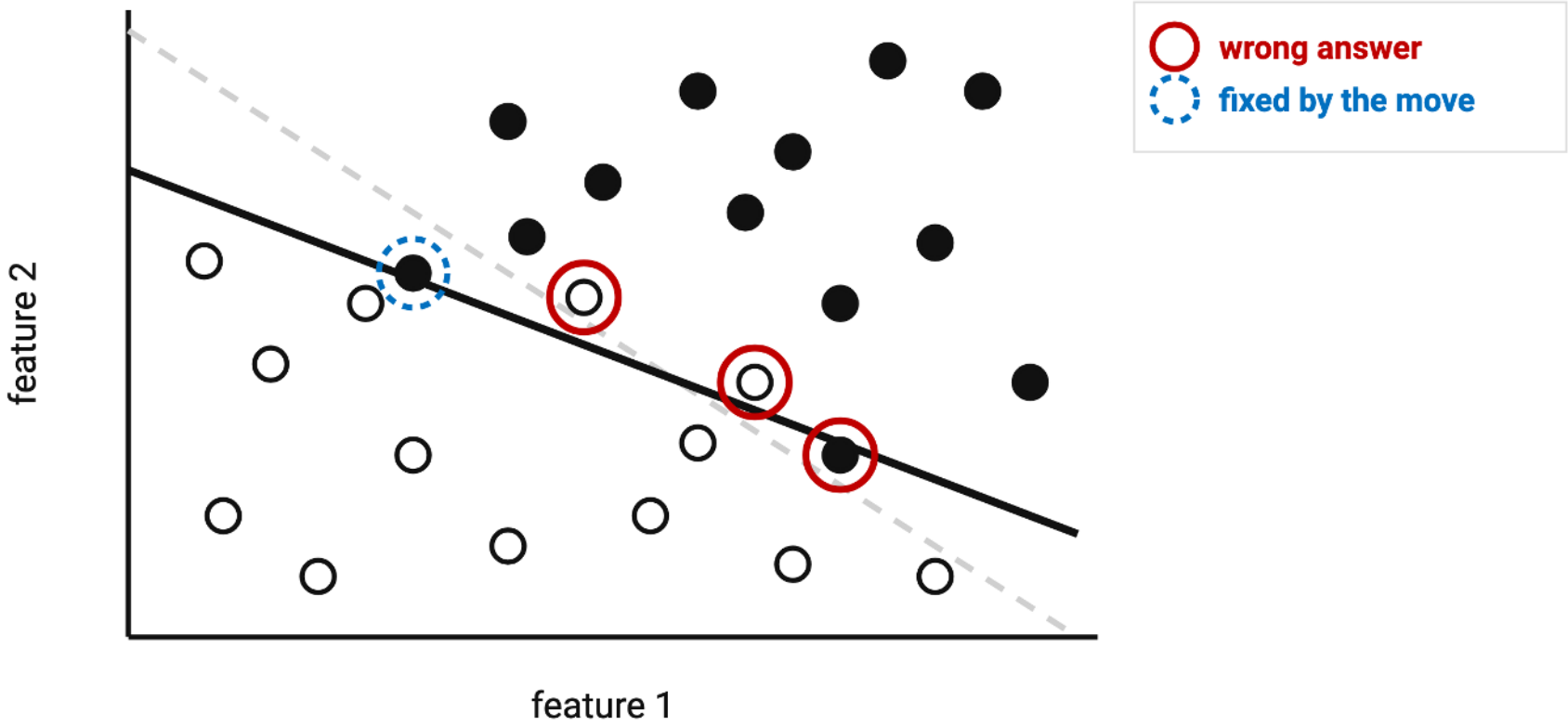


Domingos, CACM 2012, p. 87; Amershi et al., AI Magazine 2014, p. 109

Mistakes in a learned model

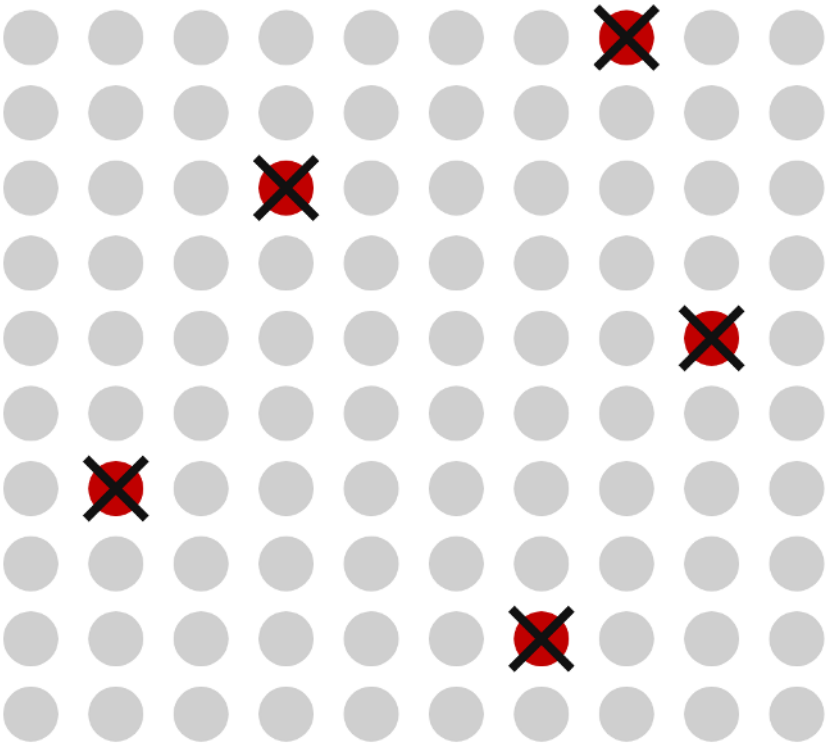


Side effects of a fix



See Kulesza et al. 2011, as described in Amershi et al. 2014

Accuracy with a rare class



safe: 95

phishing: 5

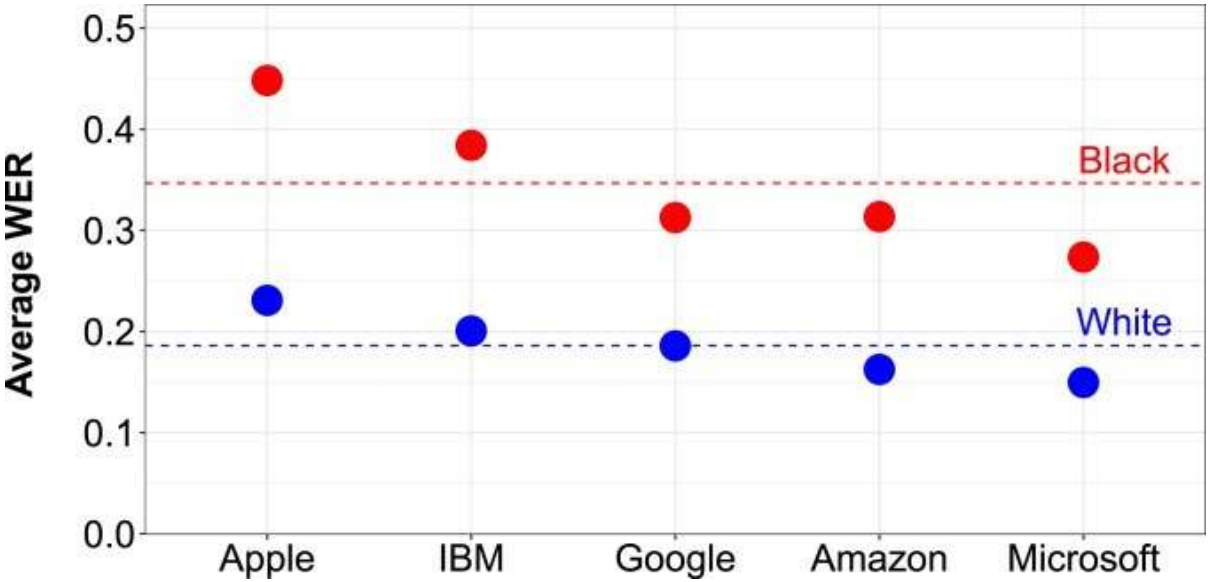
The model says
“safe” for all 100

right: 95

missed: all 5 phishing

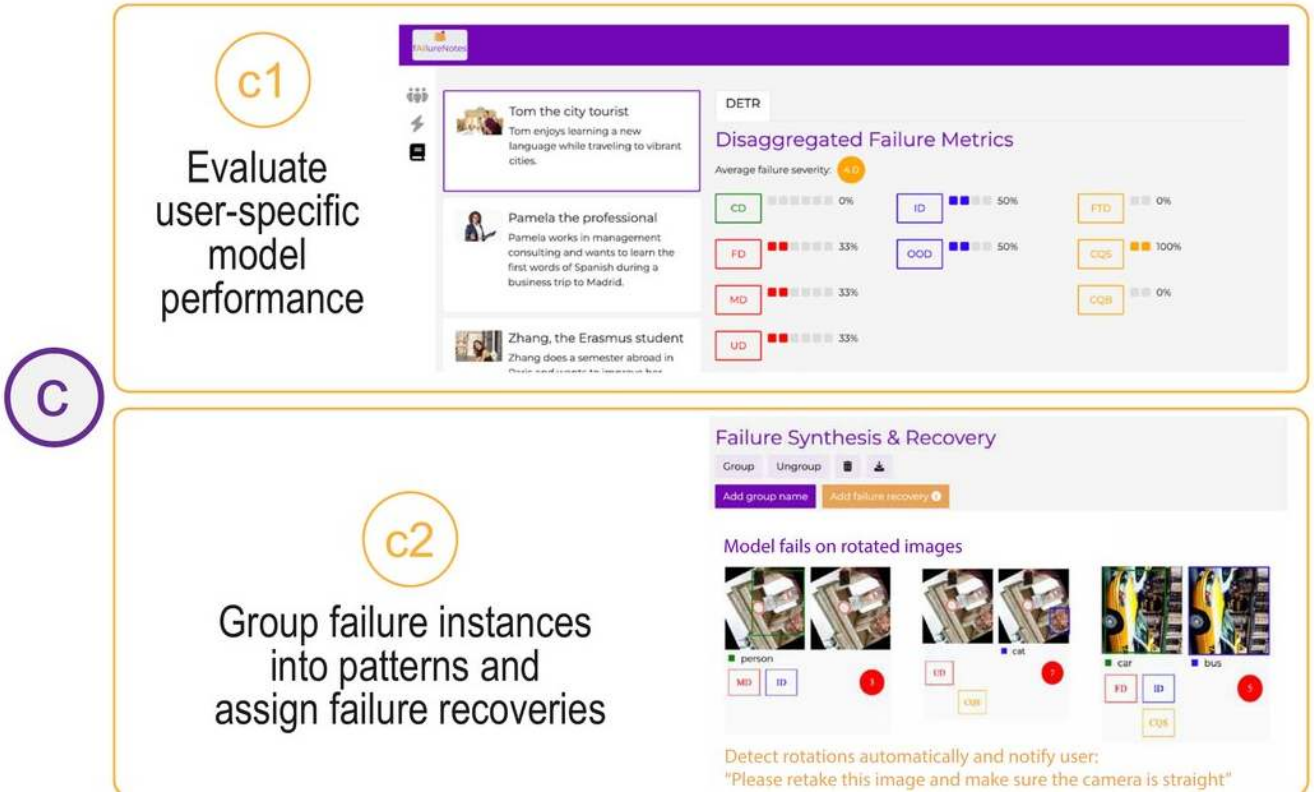
Speech recognition error rates

Average word error rate: **0.35** for Black speakers, **0.19** for white speakers



Koenecke et al., PNAS 2020, Fig. 1 (CC BY-NC-ND 4.0)

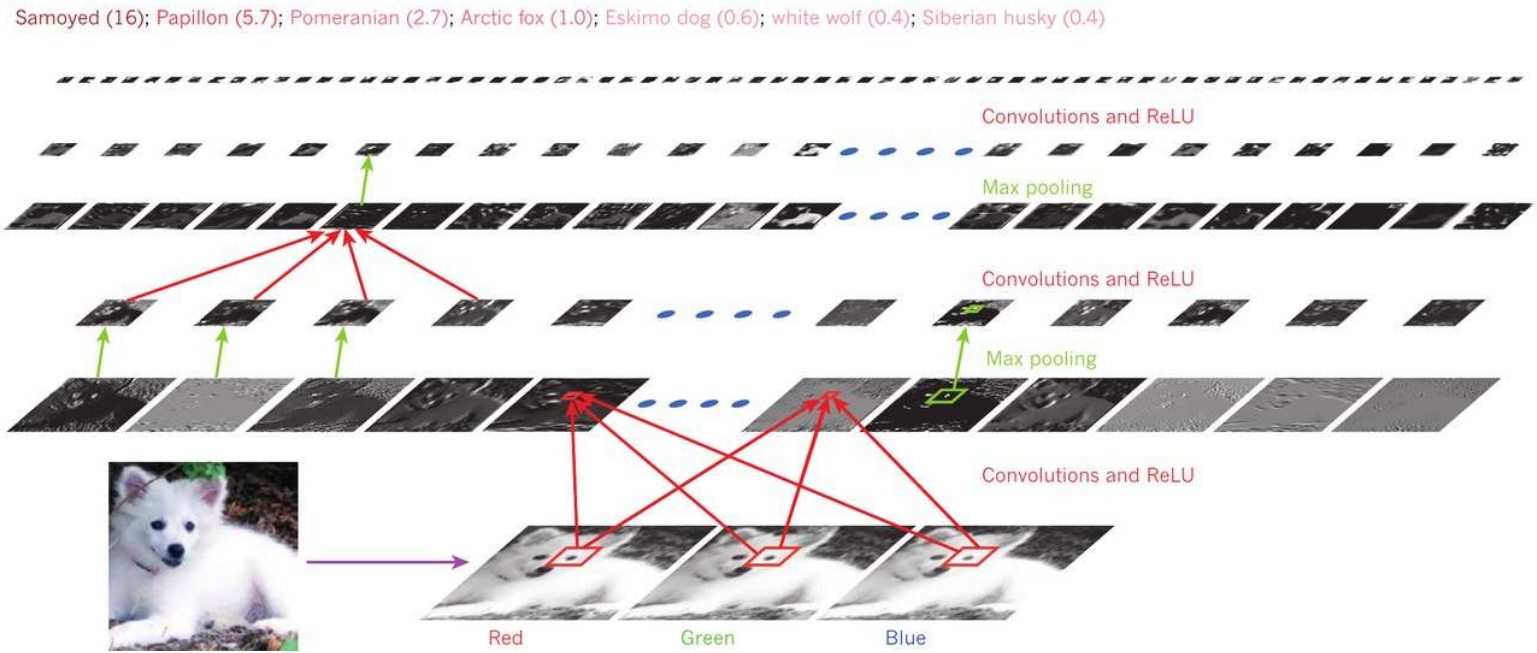
fAllureNotes



Sorts failures by user and by kind

Moore, Liao & Subramonyam, CHI 2023, Fig. 4

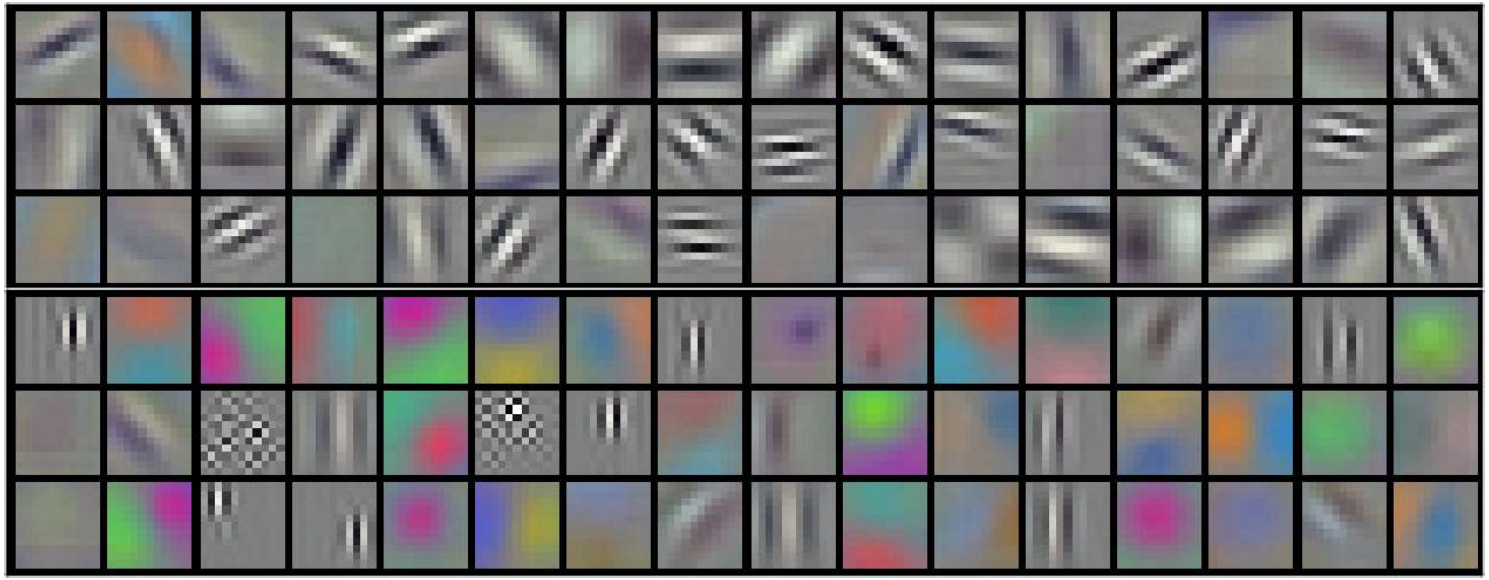
Deep learning: learned features



“these layers of features are not designed by human engineers: they are learned from data”

LeCun, Bengio & Hinton, Nature 2015 (optional reading), Fig. 2 and p. 436

First-layer filters



Mostly edges at different angles, and patches of color.

Krizhevsky, Sutskever & Hinton, NIPS 2012, Fig. 3

2012: the ImageNet challenge



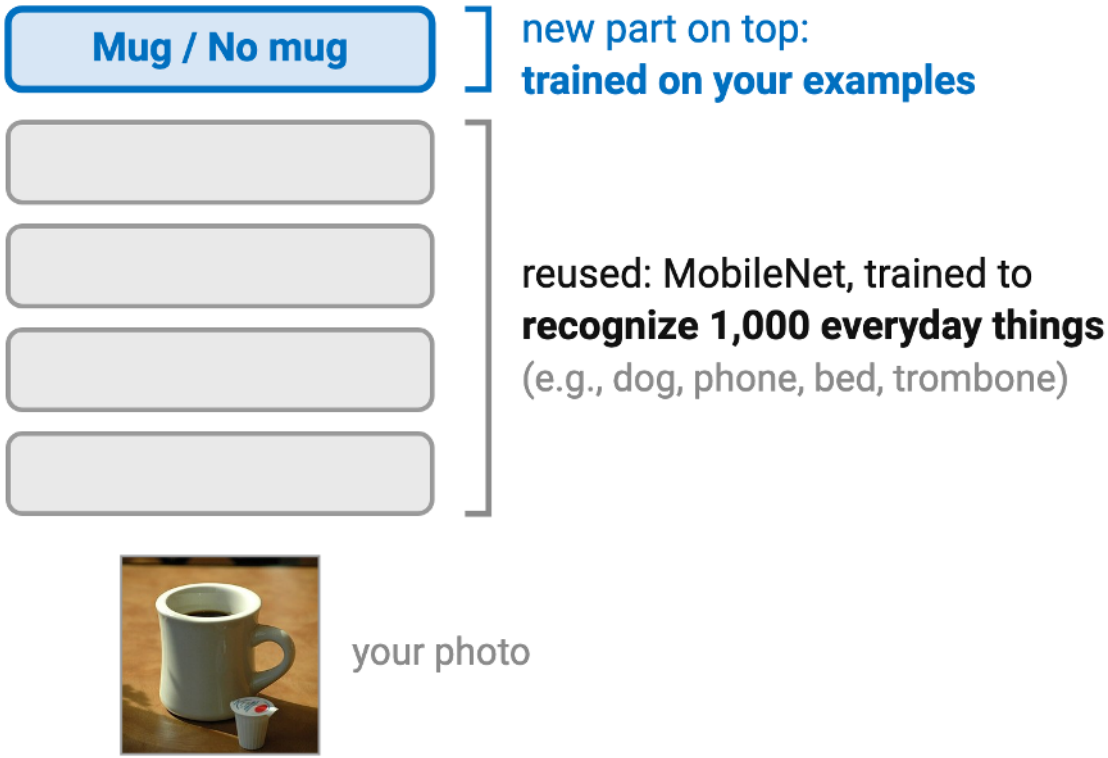
Top-5 error:

15.3% for this network

26.2% for the next best entry

Krizhevsky, Sutskever & Hinton, NIPS 2012, abstract and Fig. 4; LeCun, Bengio & Hinton, Nature 2015, p. 440

Transfer learning



Carney et al., CHI 2020 Extended Abstracts

Milestone 1: running a session

1. Ask for consent first
2. Ask them to think aloud while they work
3. Do not help or explain the tool

Milestone 1: interview questions

“What do you think the AI is doing?”

“When would you not use it?”

Milestone 1: Watch people use AI

Due Thursday of week 3

Takeaways

1. A model learns what separates its examples
2. Test on examples like the ones in real use
3. Check accuracy for each class and group
4. Deep networks learn their own features

Next: Lecture 4, Generative Models

Read and watch before class:

“Large language models, explained with a minimum of math and jargon”
Lee & Trott, Understanding AI, 2023
“Large Language Models explained briefly”
3Blue1Brown, video, 2024
“Diffusion Explainer,” and try its live tool
Lee et al., IEEE VIS 2024