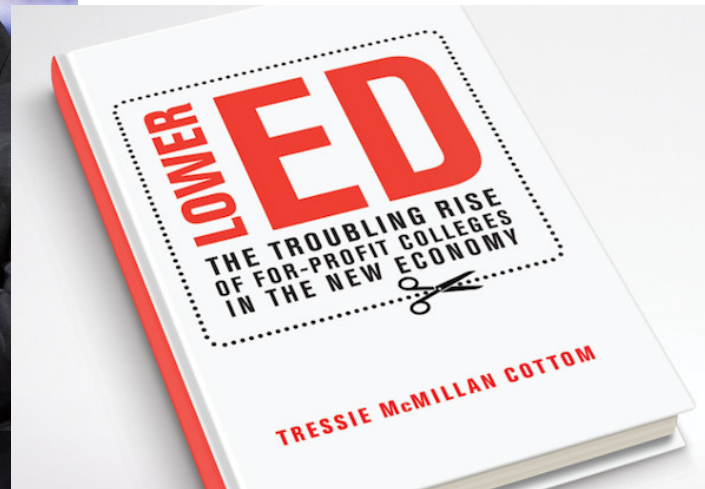


ML and Society

Feb 15, 2023

Passcode: **Tressie McMillan Cottom**

- A preeminent public scholar on race, digital media, and higher ed
If you want to see science writing done right, go read one of her books



Sit with your team!

Team 1	Afzal	Cole	Navid	Tim	
Team 2	Aishwarya	Herman	Mads	Melvin	
Team 3	Daphkar	Juliana	Ibtida	Monica	
Team 4	Joe	Ken	Vedant	Zach	
Team 5	Chaitanya	Evan	Hitesh	Sushanth	
Team 6	Hannah	Harinee	Gabriella	Suradhya	
Team 7	Alex	Connor	Gopi	Shane	Thanh
Team 8	Aditi	Connor	Jason	Mitali	
Team 9	Botsalano	Niharika	Vedang	Yunmei	
Team 10	Dhiraj	Frank	Kashyap	Michael	

Rage students in Green. ML+Soc students in black

Who does my machine
learning model serve?

How do I know?

**What can I do
about it?**

So who is right?



 **Ryan Saavedra** ✓
@RealSaavedra

Socialist Rep. Alexandria Ocasio-Cortez (D-NY) claims that algorithms, which are driven by math, are racist

Depends on what
you mean by
"algorithm"

9,455 3.2 22, 2019

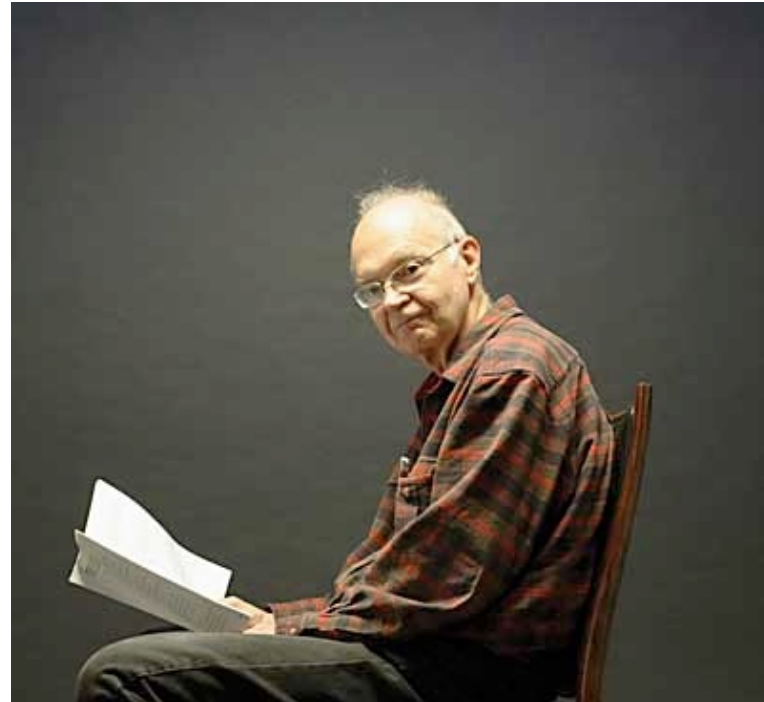
9,204 people are talking about this

"Algorithms can be biased"

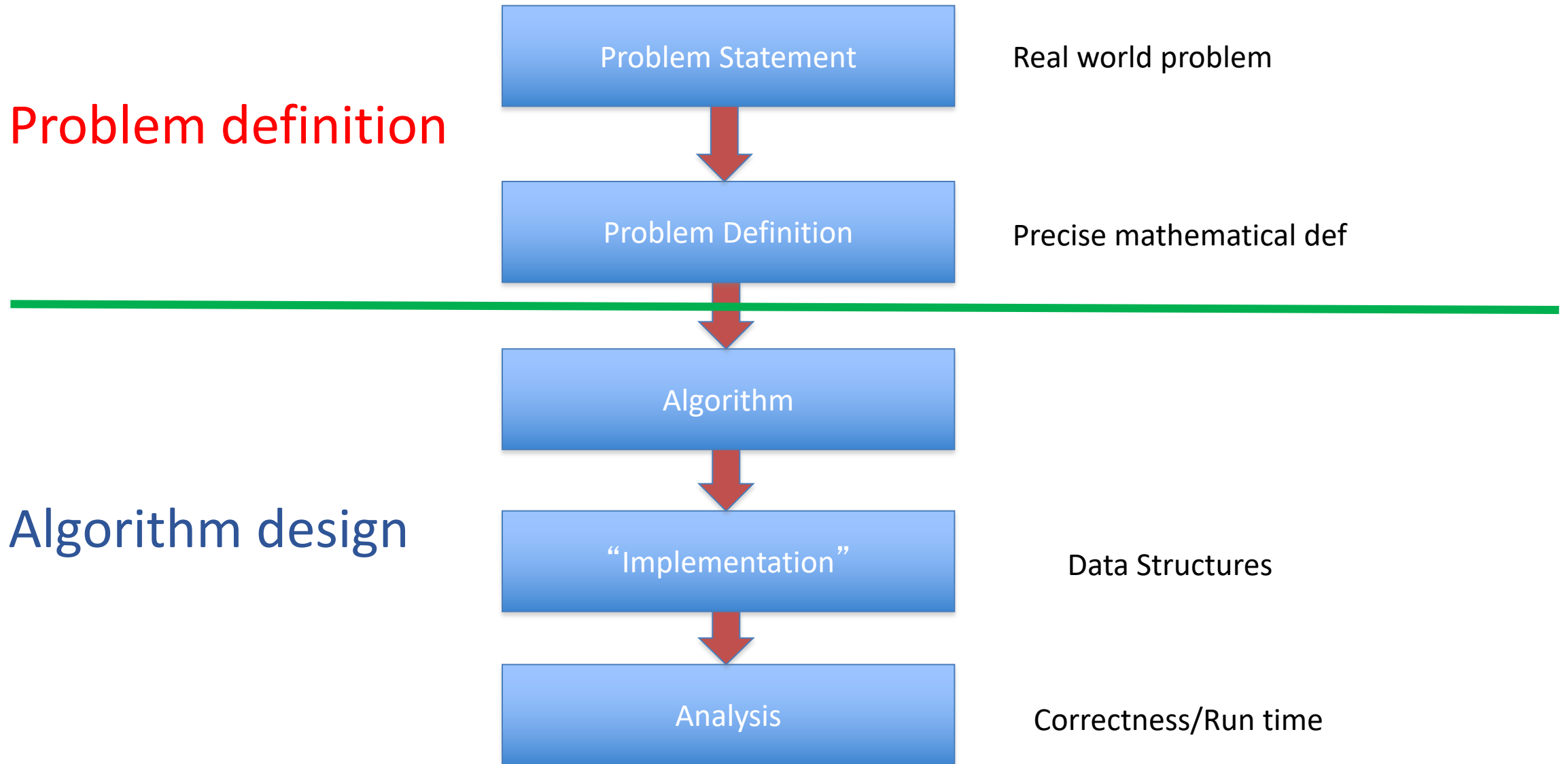
"Algorithms are based on math
and hence cannot be biased"

Knuth's Definition of an *algorithm*

An algorithm is a finite, definitive, effective procedure with some input and some output

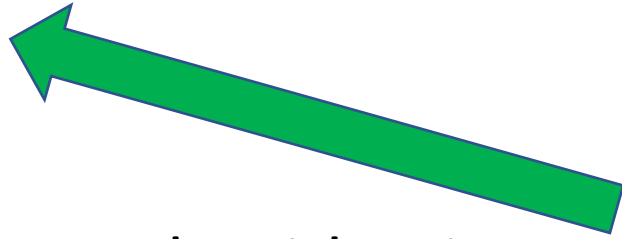


Main Steps in Traditional Algorithm Design

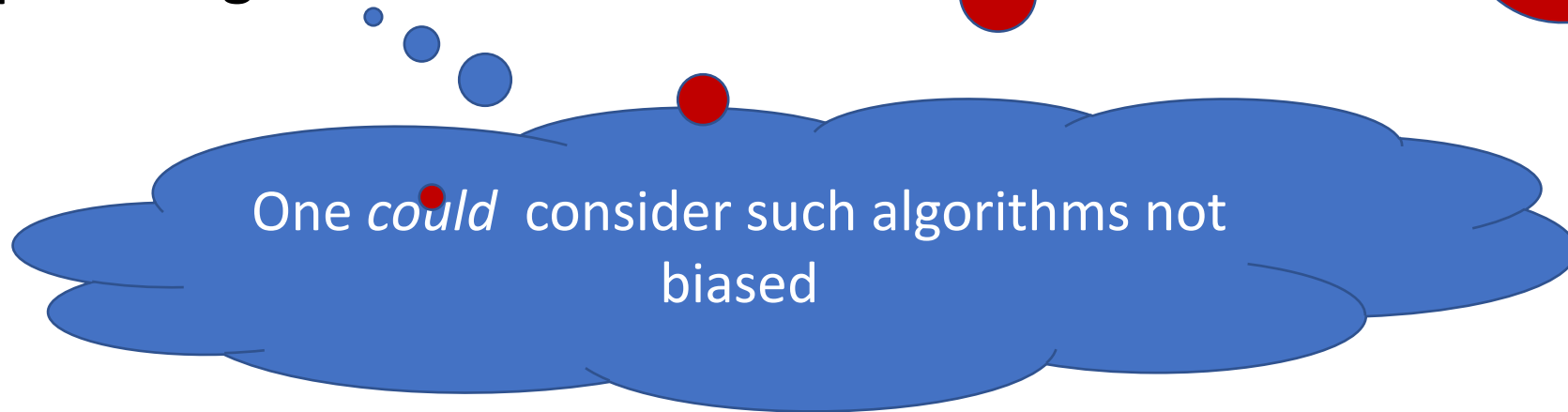
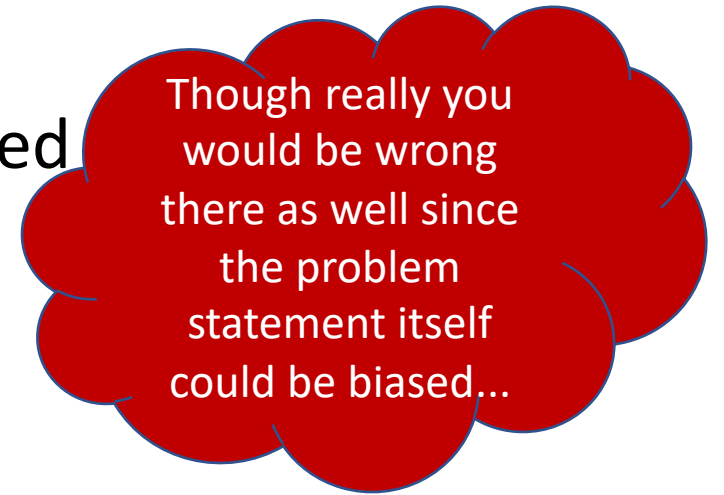


Two main points about traditional algorithms

Problem is defined BEFORE algorithm is designed



Can prove algorithm is correct for ALL inputs

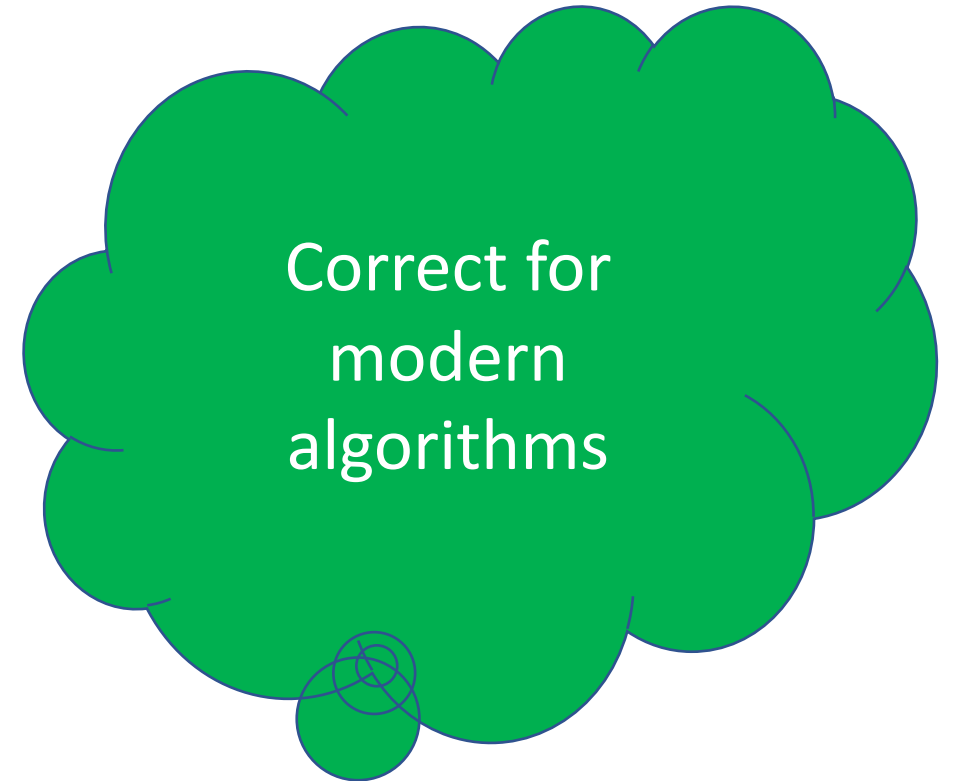


One *could* consider such algorithms not biased

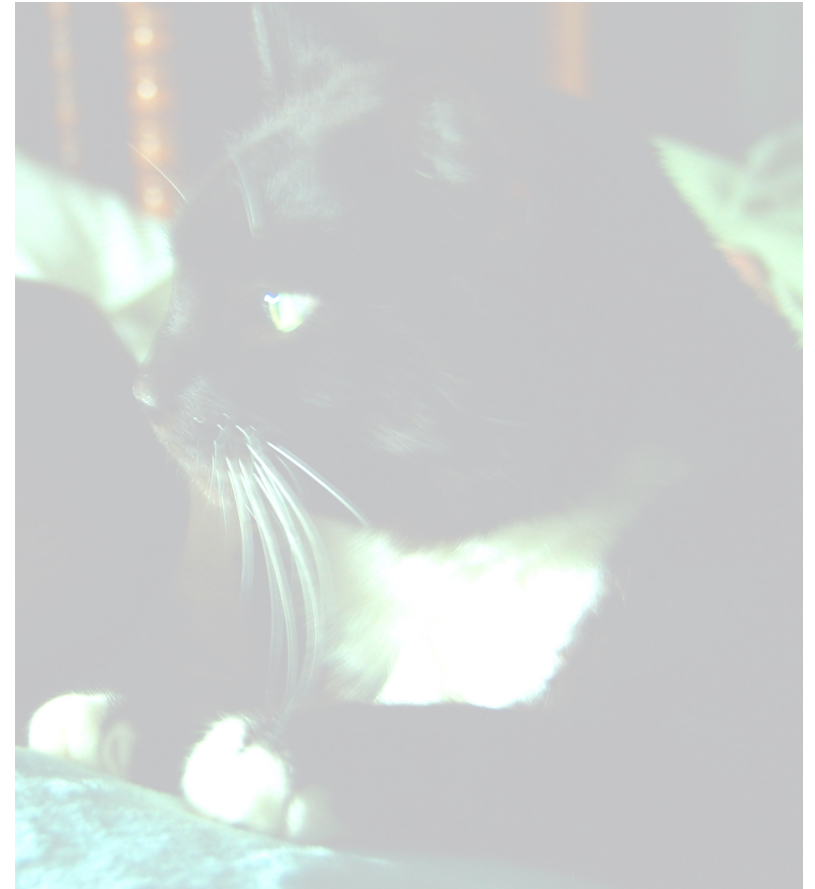
Today



“Algorithms can be biased”



Cat vs. Dogs



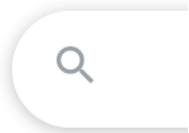
Warren and Billy



How do you “define” a dog vs cat image?




Google Images has “solved” this problem






Search by image ×

Search Google with an image instead of text. Try dragging an image here.

Paste image URL  **Upload an image**

Search by image

My result for Warren (Spring 20)

Google  warren.JPG.JPG  

[All](#) [Images](#) [Maps](#) [Shopping](#) [More](#) [Settings](#) [Tools](#)

About 2 results (0.47 seconds)




Image size:
1973 × 1895

No other sizes of this image found.

Possible related search: [pit bull](#)


[Pit bull - Wikipedia](#)
https://en.wikipedia.org/wiki/Pit_bull

Pit bull is the common name for a type of dog descended from bulldogs and terriers. The **pit bull**-type is particularly ambiguous, as it encompasses a range of ...

[American Pit Bull Terrier Dog Breed Information, Pictures ...](#)
<https://dogtime.com/dog-breeds/american-pit-bull-terrier>

The American **Pit Bull** Terrier is a companion and family dog breed. Originally bred to "bait" bulls, the breed evolved into all-around farm dogs, and later moved ...

[Visually similar images](#)



Pit bull

Dog

Pit bull is the common name for a type of dog descended from bulldogs and terriers. The pit bull-type is particularly ambiguous, as it encompasses a range of pedigree breeds, informal types and appearances that cannot be reliably identified. [Wikipedia](#)

Lifespan: [American Pit Bull Terrier](#): 8 – 15 years, [American Staffordshire Terrier](#): 12 – 16 years

Height: [American Pit Bull Terrier](#): 18 – 21 in., [MORE](#)

Mass: [American Pit Bull Terrier](#): 35 – 65 lbs, [Staffordshire Bull Terrier](#): 29 – 37 lbs

Life span: The average life span of the American Pit Bull Terrier ranges from 10 to 12 years. [petwave.com](#)

My result for Warren (Spring 22)

JPG x martingale x [camera icon] [magnifying glass icon]

All Images Maps Shopping More Tools

About 3 results (0.18 seconds)




Image size:
1973 x 1895

No other sizes of this image found.

Possible related search: [martingale](#)

[https://en.wikipedia.org/wiki/Martingale_\(probability_theory\)](https://en.wikipedia.org/wiki/Martingale_(probability_theory))

[Martingale \(probability theory\) - Wikipedia](#)


In probability theory, a **martingale** is a sequence of random variables (i.e., a stochastic process) for which, at a particular time, the conditional ...

[https://en.wikipedia.org/wiki/Martingale_\(betting_system\)](https://en.wikipedia.org/wiki/Martingale_(betting_system))


[Martingale \(betting system\) - Wikipedia](#)

A **martingale** is a class of betting strategies that originated from and were popular in 18th-century France. The simplest of these strategies was designed ...

🖼️ Visually similar images



Martingale Collar



A martingale is a type of dog collar that provides more control over the animal without the choking effect of a slip collar. Martingale dog collars are also known as greyhound, whippet or humane choke collars.

[Wikipedia](#)

Feedback

My result for Warren (Spring 23)

Google

Upload



Find image source



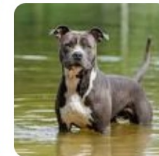
Search

Text

Translate



Pit bull



American Staffordshire...



Staffordshire Bull Terrier



American Pit Bull Terrier

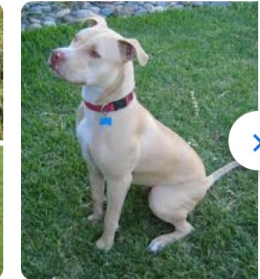
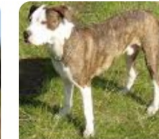


American Bully

Pit bull

Dog

Search



Visual matches




My result for Billy (Spring 20+22)

JPG x domestic short-haired cat

Q All **Images** Maps Shopping More Tools

About 3 results (0.64 seconds)

 Image size: 1763 x 1991
No other sizes of this image found.

Possible related search: [domestic short-haired cat](#)

[https://en.wikipedia.org/wiki/Domestic_short-haired...](https://en.wikipedia.org/wiki/Domestic_short-haired_cat)


Domestic short-haired cat - Wikipedia


Domestic short-haired cats are the most common **cat** in the United States, accounting for around 90–95% of their number. ... Other generic terms include house **cat** and ...

[https://www.hillspet.com/Cat_Care/What's_New?](https://www.hillspet.com/Cat_Care/What's_New/)


Domestic Shorthair Cat Breed: Personality & Info | Hill's Pet

Animal Planet affectionately refers to **Domestic** shorthair cats as the mutts of the **cat** world because they're a mix of various breeds, resulting in a vast range ...

 Visually similar images







Domestic short-haired cat



A domestic short-haired cat is a cat of mixed ancestry—thus not belonging to any particular recognised cat breed—possessing a coat of short fur. In Britain they are sometimes colloquially called moggies. [Wikipedia](#)

Shorthair cat breeds

[View 45+ more](#)

 British Shorthair	 American Shorthair	 Persian cat	 Maine Coon
--	---	--	---

[Feedback](#)

My result for Billy (Spring 23)

Google

Upload



Find image source



Search

Text

Translate



Bicolor cat



Black cat



Domestic short-haired...



Polydactyl cat



Manx Cat

Bicolor cat
Cat

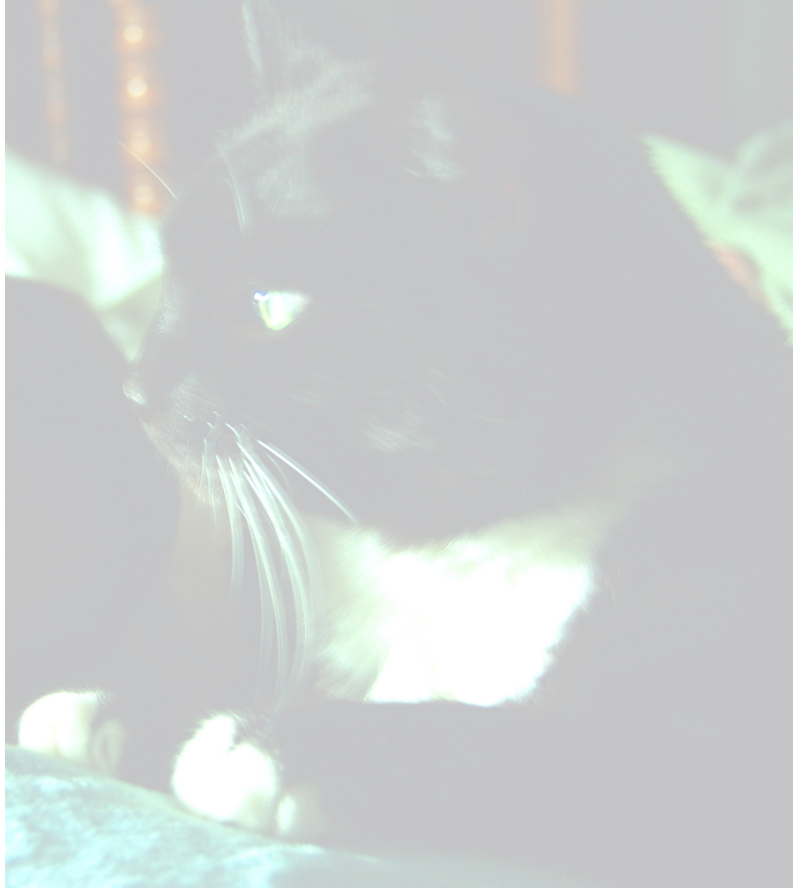
Search



Visual matches



So cats vs dogs problem solved?



My result for modified Warren (Spring 20)



JPG x companion dog  



All Images Maps Shopping More Settings Tools

About 2 results (2.32 seconds)



Image size:
1973 x 1895

No other sizes of this image found.

Possible related search: [companion dog](#)

Companion dog - Wikipedia

https://en.wikipedia.org/wiki/Companion_dog

A **companion dog** is a dog that does not work, providing only companionship as a pet, rather than usefulness by doing specific tasks. Many of the toy dog breeds ...

Best Companion Dog Breeds | Purina

https://www.purina.com/Dogs/Dog_Breeds/Collections

Whether you want a friendly face to come home to or the best **companion dog** breed for an elderly parent, get the complete list here.

Visually similar images

Companion dog



A companion dog is a dog that does not work, providing only companionship as a pet, rather than usefulness by doing specific tasks. Many of the toy dog breeds are used only for the pleasure of their company, not as workers. [Wikipedia](#)

Companion dog breeds

[View 4+ more](#)



Pekingese



Yorkshire Terrier



German Spaniel



Valley Bulldog




Borador

My result for modified Warren (Spring 22)

JPG x martingale x

All Images Maps Shopping More Tools


About 2 results (0.36 seconds)

 Image size: 1973 x 1895
No other sizes of this image found.

Possible related search: [martingale](#)


[https://en.wikipedia.org/wiki/Martingale_\(probability_theory\)](https://en.wikipedia.org/wiki/Martingale_(probability_theory))
Martingale (probability theory) - Wikipedia
In probability theory, a **martingale** is a sequence of random variables (i.e., a stochastic process) for which, at a particular time, the conditional ...


[https://en.wikipedia.org/wiki/Martingale_\(betting_system\)](https://en.wikipedia.org/wiki/Martingale_(betting_system))
Martingale (betting system) - Wikipedia
A **martingale** is a class of betting strategies that originated from and were popular in 18th-century France. The simplest of these strategies was designed ...

Martingale Collar 

A martingale is a type of dog collar that provides more control over the animal without the choking effect of a slip collar. Martingale dog collars are also known as greyhound, whippet or humane choke collars.
[Wikipedia](#)

Feedback

 Visually similar images



My result for modified Warren (Spring 23)

Google

Upload



Find image source



Search

Text

Translate



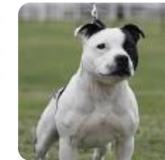
Pit bull



American Bulldog



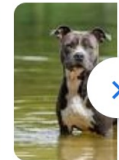
Exotic Bully



Staffordshire Bull Terrier



American Pit Bull Terrier



American Staffordshire Terrier

Pit bull

Dog

Search



Visual matches



My result for modified Billy (Spring 20)



Q All Images Maps Shopping More Settings Tools

About 2 results (1.40 seconds)



Image size:
1763 × 1991

No other sizes of this image found.

Possible related search: [fish](#)

Fish - Wikipedia

<https://en.wikipedia.org/wiki/Fish>

Fish are gill-bearing aquatic craniate animals that lack limbs with digits. They form a sister group to the tunicates, together forming the olfactores. Included in this ...

Pet Fish for Sale: Tropical and Freshwater Fish | PetSmart

<https://www.petsmart.com/fish/live-fish>

130 Items - Create or augment the perfect underwater community with our selection of freshwater and tropical **fish** for sale.

Visually similar images



Fish

Animal



Fish are gill-bearing aquatic craniate animals that lack limbs with digits. They form a sister group to the tunicates, together forming the olfactores. Included in this definition are the living hagfish, lampreys, and cartilaginous and bony fish as well as various extinct related groups. [Wikipedia](#)

Lifespan: [Common carp](#): 20 years, [MORE](#)

Phylum: [Chordate](#)

Mass: [Common carp](#): 4.4 – 31 lbs, [Northern pike](#): 34 lbs, [MORE](#)

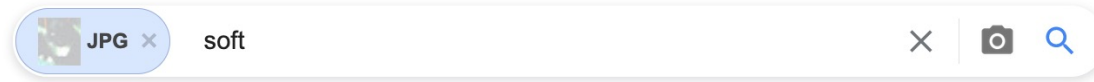
Encyclopedia of Life

Length: [Common carp](#): 16 – 31 in., [Siamese fighting fish](#): 2.8 in., [MORE](#)

Speed: [Ocean sunfish](#): 2 mph, [Great white shark](#): 35 mph

Clutch size: [Common carp](#): 300,000, [Siamese fighting fish](#): 10 – 40

My result for modified Billy (Spring 22)



Q All Images Maps Shopping More Tools

About 4 results (0.49 seconds)



Image size:
1763 × 1991

No other sizes of this image found.

Possible related search: [soft](#)

<https://www.dictionary.com> > browse > soft ▾

[Soft Definition & Meaning | Dictionary.com](#)

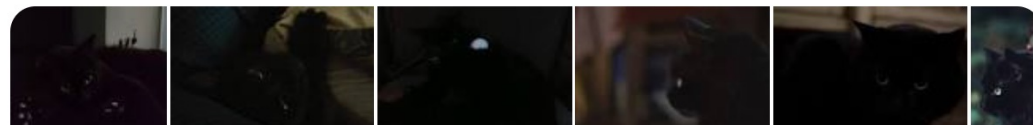
yielding readily to touch or pressure; easily penetrated, divided, or changed in shape; not hard or stiff: a **soft** pillow. · relatively deficient in hardness, as ...

<https://www.merriam-webster.com> > dictionary > soft ▾

[Soft Definition & Meaning - Merriam-Webster](#)

Definition of **soft** ; pleasing or agreeable to the senses : bringing ease, comfort, or quiet ; b · having a bland or mellow rather than a sharp or acid taste ; d ...

Visually similar images



My result for modified Billy (Spring 23)

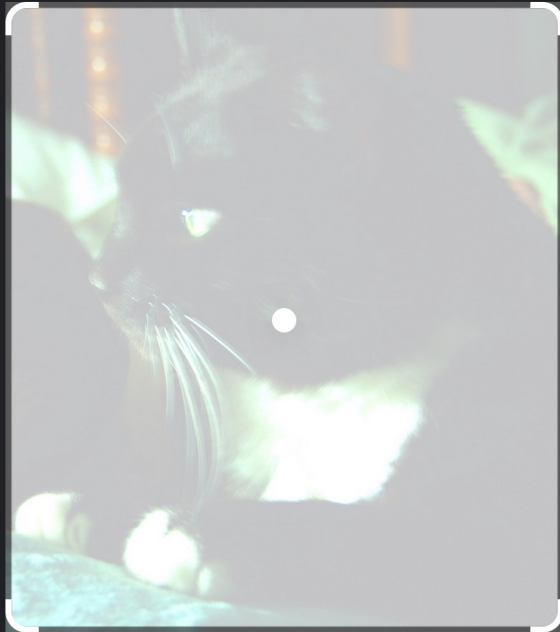
Google

Upload



A

Find image source



Search

Text

Translate



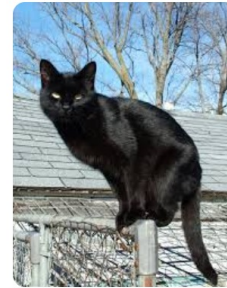
Black cat



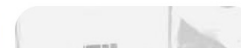
Bicolor cat

Black cat

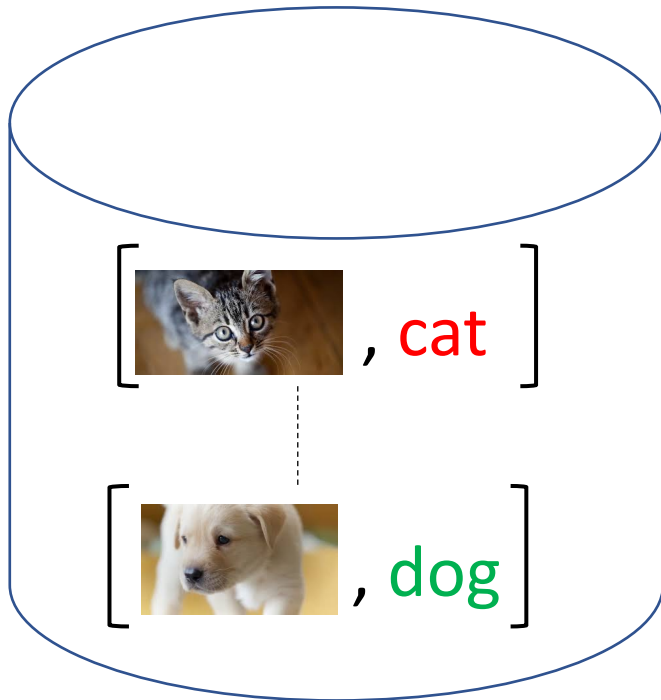
Search



Visual matches



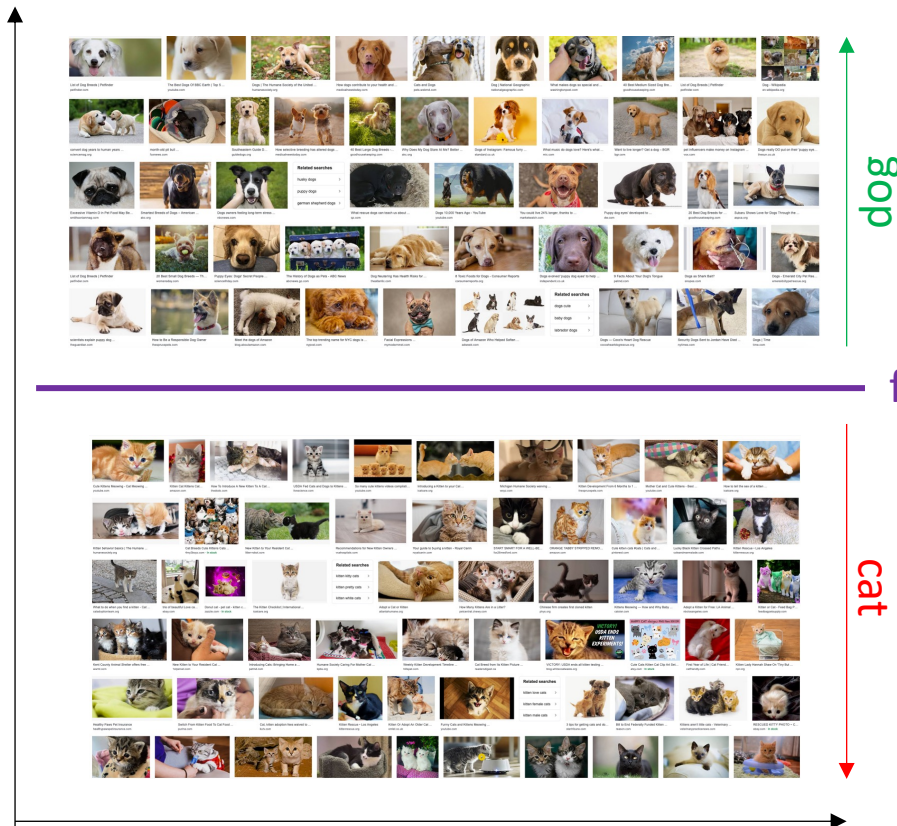
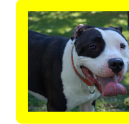
How does Google Images work?



Training



When a new image comes in



When an algorithm isn't...

Cc



Suresh Venkat [Follow](#)

Oct 2, 2015 · 5 min read



Go

The popular press is full of articles about “algorithms” and “algorithmic fairness” and “algorithms that discriminate, (or don’t)”. As a computer scientist (and one who studies algorithms to boot), I find all this attention to my field rather gratifying, and not a bit terrifying.

igs

What’s even more pleasing is that the popular explanation of an algorithm follows along the lines of the definition we’ve been using since, well, forever

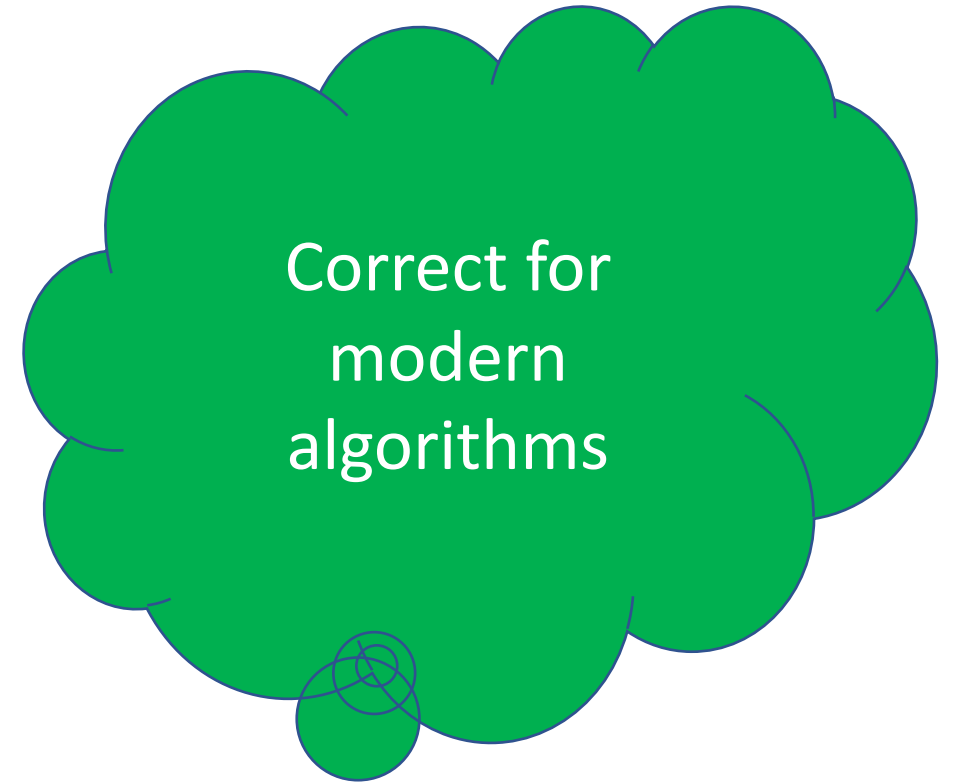
An algorithm is a set of steps (the instructions) each of which is simple and well defined, and that stops after a finite number of these steps.

If we wanted a less intimidating definition of an algorithm, we turn to the kitchen:

AOC is right!



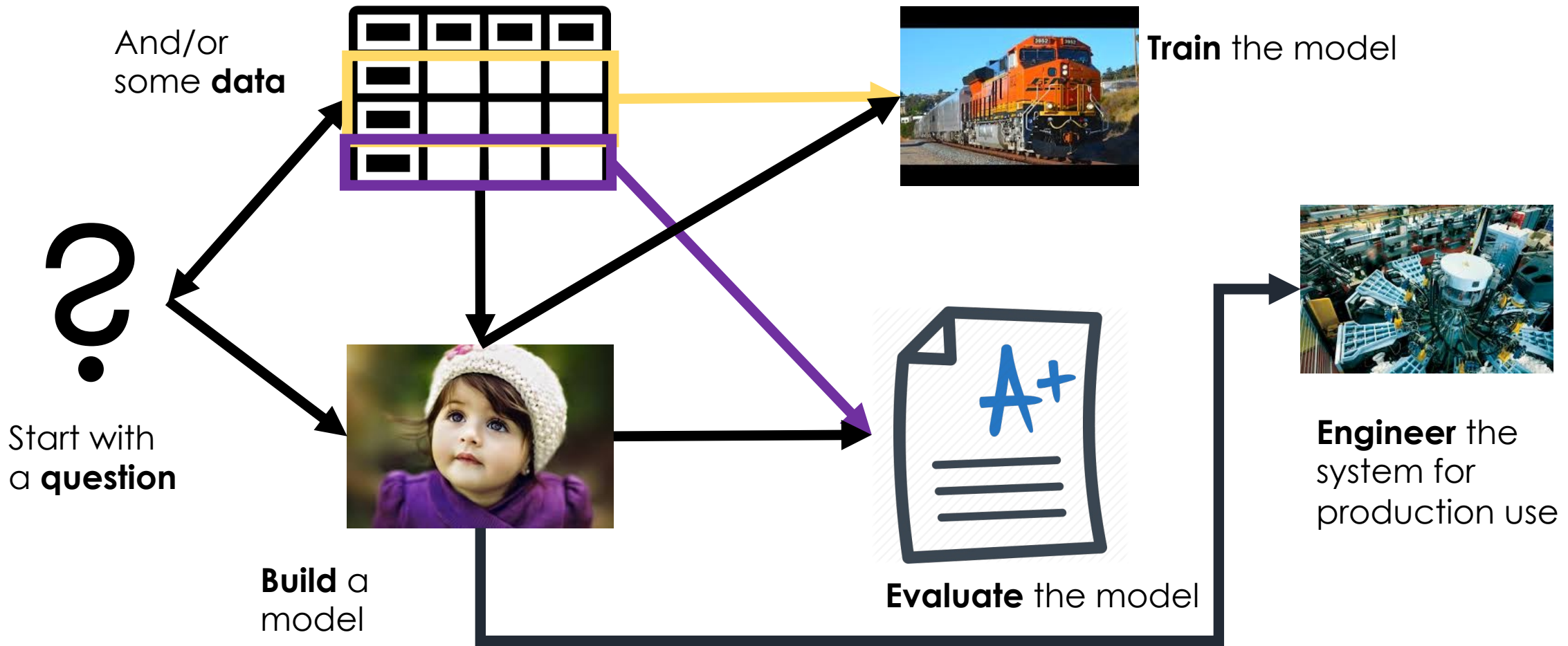
“Algorithms can be biased”



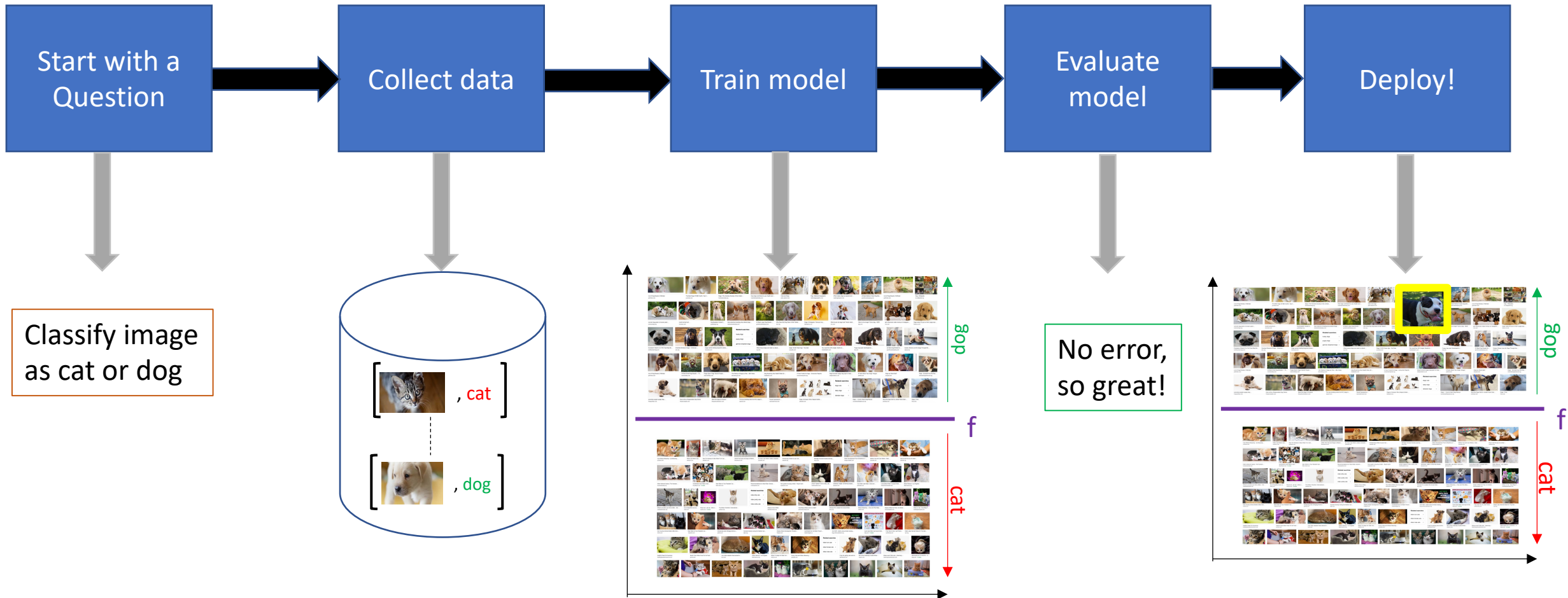
Modern Algorithm = Machine Learning (ML)



The Machine Learning Pipeline

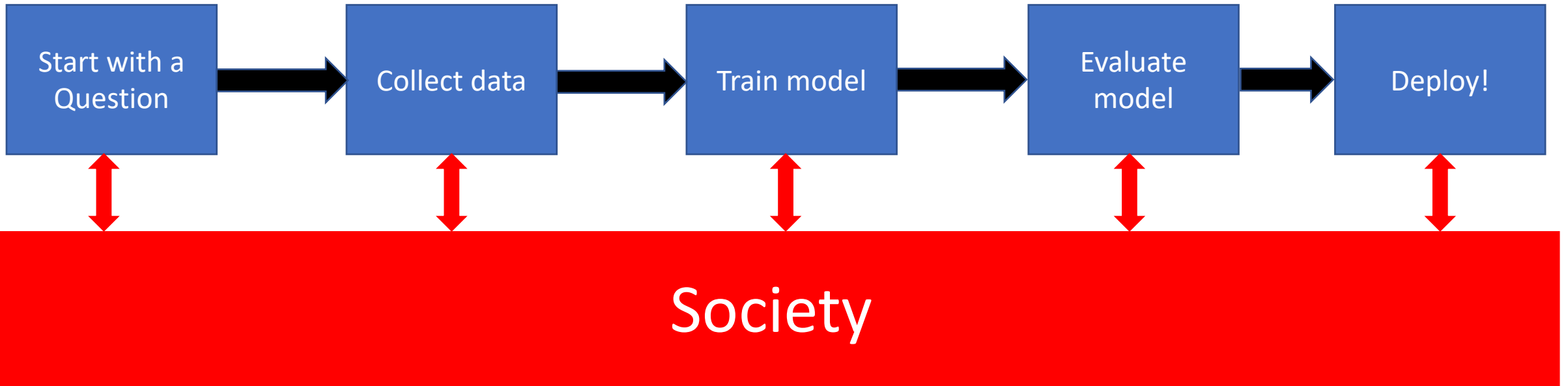


Back to cats vs. dogs

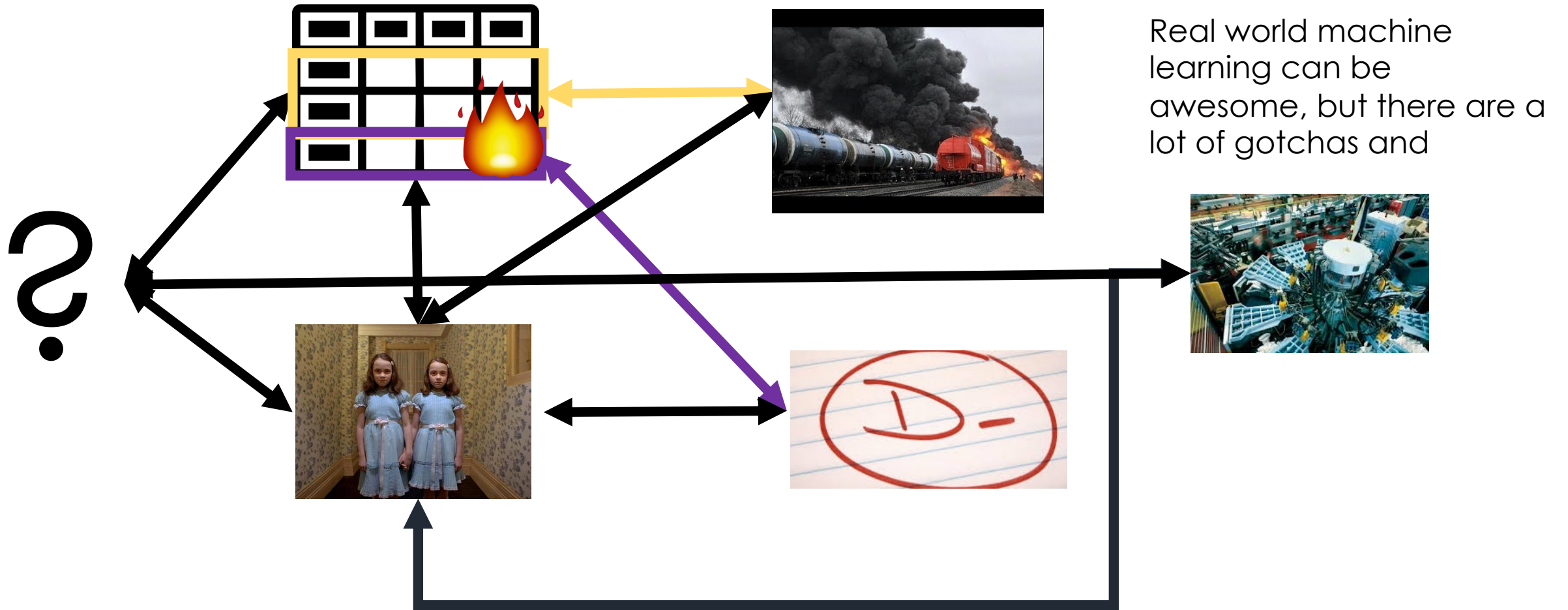




What is missing from this picture?



A real view of the ML Pipeline



Real world machine learning can be awesome, but there are a lot of gotchas and



Slide by Kenny Joseph



Real world goal

natural language processing blog

24 August 2016

Debugging machine learning


I've been thinking, mostly in the context of teaching, about how to specifically teach debugging of machine learning. Personally I find it very helpful to break things down in terms of the usual error terms: Bayes error (how much error is there in the best possible classifier), approximation error (how much do you pay for restricting to some hypothesis class), estimation error (how much do you pay because you only have finite samples), optimization error (how much do you pay because you didn't find a global optimum to your optimization problem). I've generally found that trying to isolate errors to one of these pieces, and then debugging that piece in particular (eg., pick a better optimizer versus pick a better hypothesis class) has been useful.

For instance, my general debugging strategy involves steps like the following:

1. First, ensure that your optimizer isn't the problem. You can do this by adding "cheating" features -- a feature that correlates perfectly with the label. Make sure you can successfully overfit the training data. If not, this is probably either an optimizer problem or

my biased thoughts on the fields of natural language processing (NLP), computational linguistics (CL) and related topics (machine learning, math, funding, etc.)

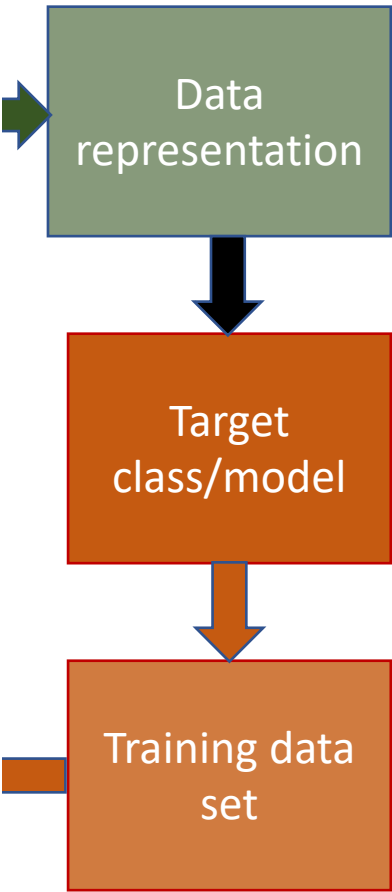
About Me



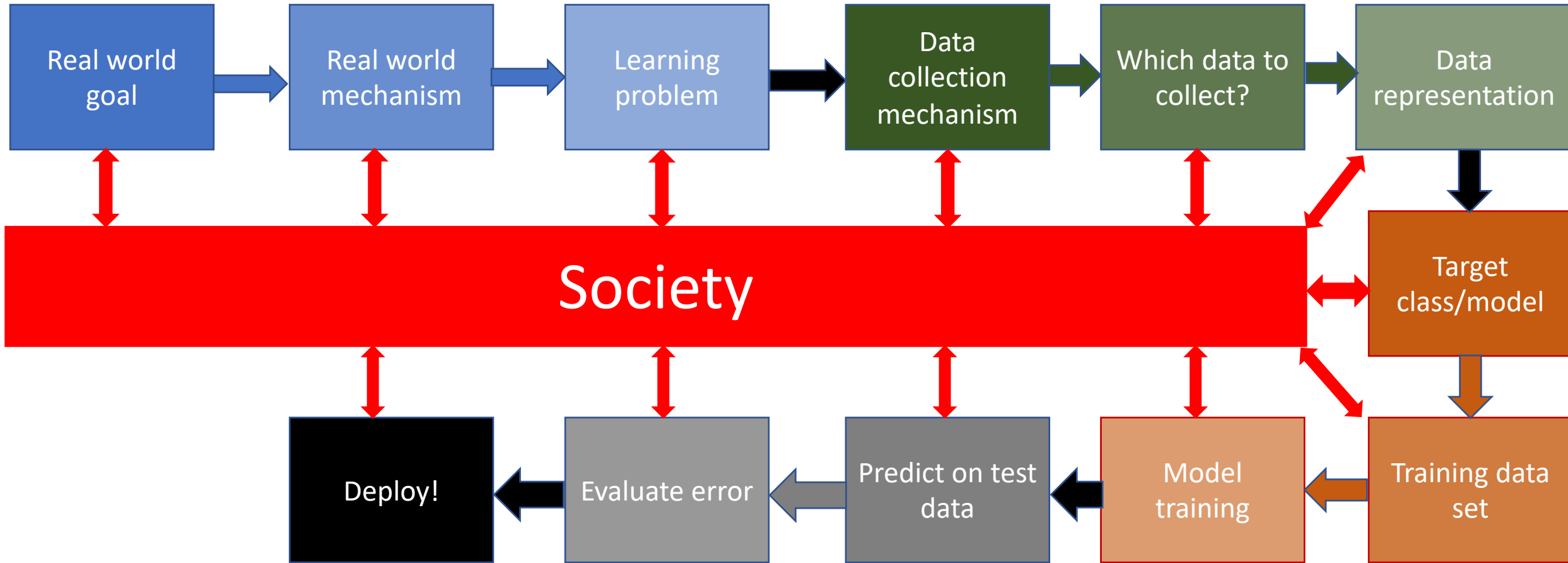
hal
View my complete profile

Labels

acl (3)
ACS (2)
advising (1)



What is missing?



Not the only ML+society pipeline in town

INSIDE AI

Black-Boxed Politics:

Opacity is a Choice in AI Systems



Katarzyna Szymielewicz

Follow

Jan 17 · 23 min read

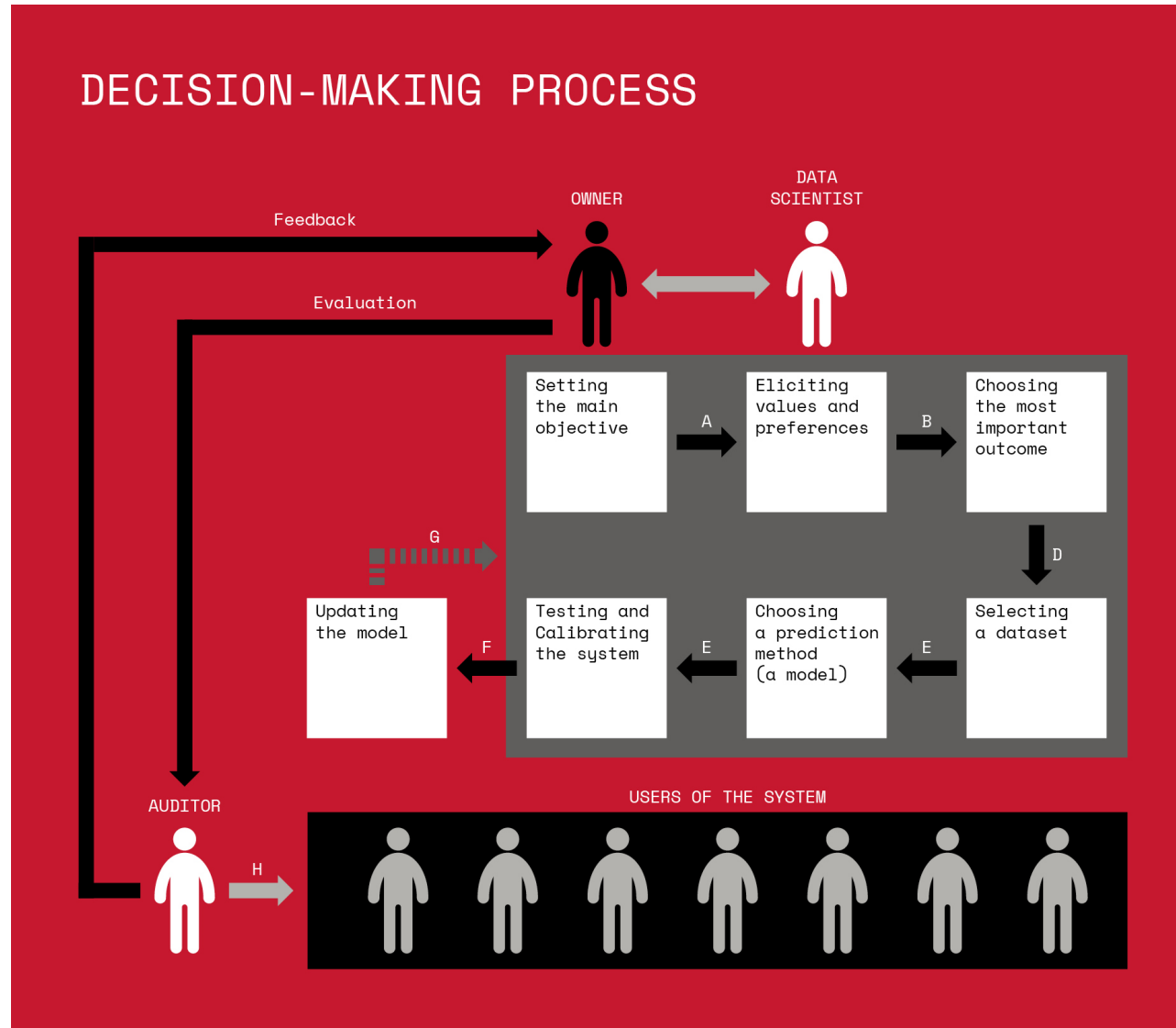


Written by: [Agata Foryciarz](#), [Daniel Leufer](#), [Katarzyna Szymielewicz](#)

Illustrations by: [Olek Modzelewski](#)

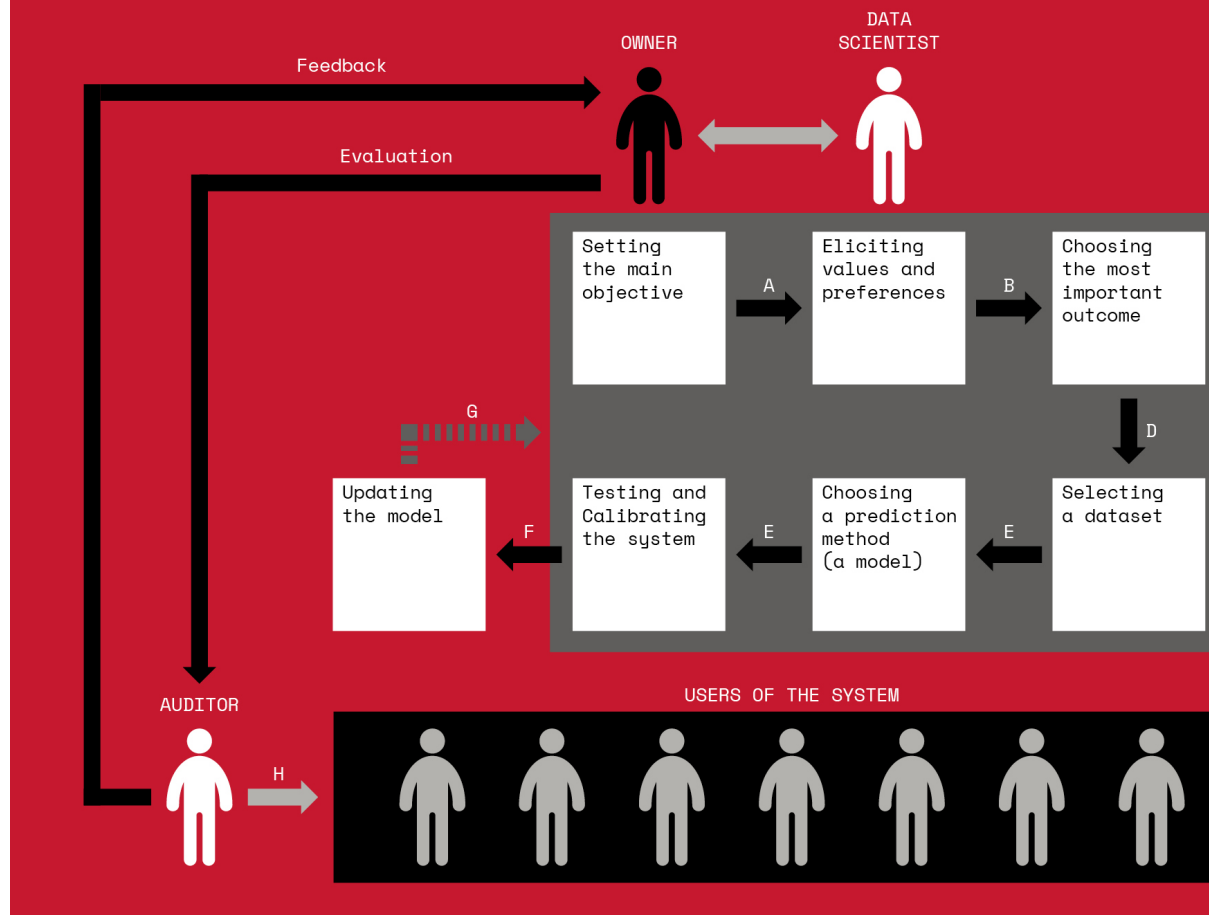
Artificial intelligence captures our imagination like almost no other technology: from fears about killer robots to dreams of a fully-automated, frictionless future. As numerous authors have documented, the idea of creating artificial, intelligent machines has entranced and scandalized people for millennia. Indeed, part of what makes the history of ‘artificial intelligence’ so fascinating is the mix of genuine scientific achievement with myth-making and outright deception

Not the only ML+society pipeline in town



A walkthrough

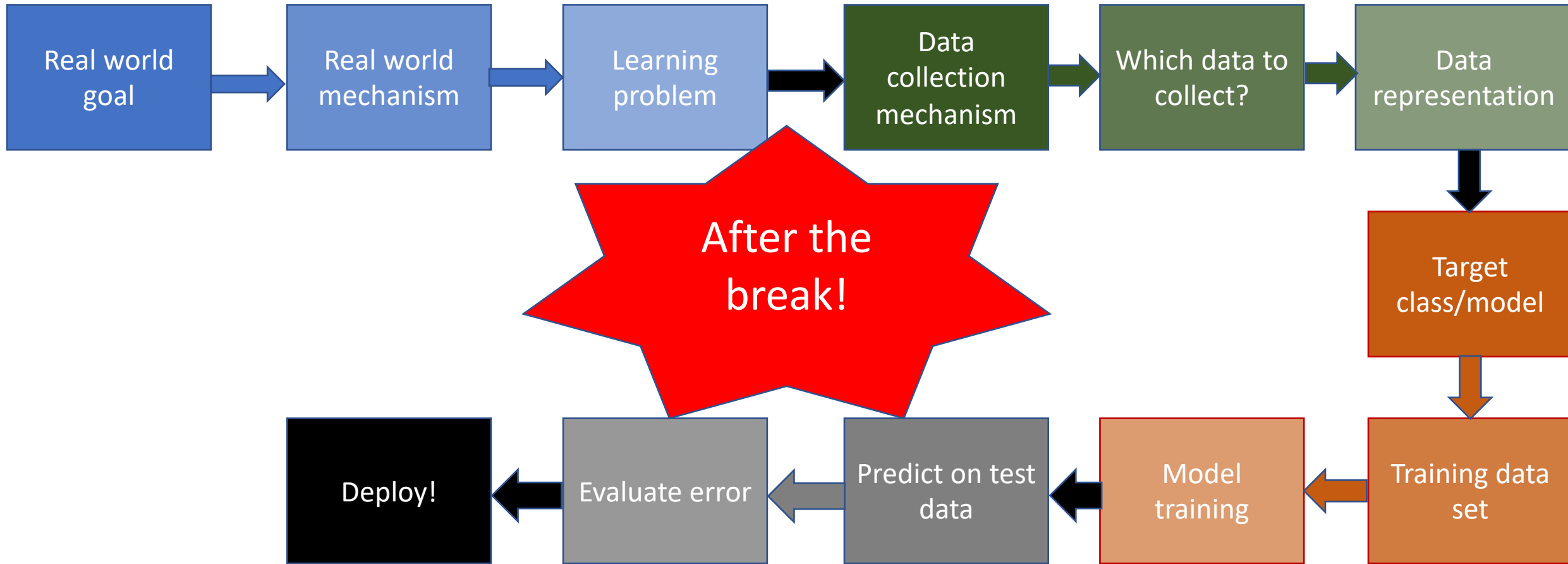
DECISION-MAKING PROCESS



HUMAN DECISIONS THAT SHAPE AN AI SYSTEM

Steps	Examples
Setting the main objective [A]	<p>This decision, made by the owner of the system, is likely to be formulated in business or political language, and describes a process or decision that a model will help to measure about or improve. From a data scientist's perspective, such a framing leaves scope for various interpretations. This is just the beginning of a longer conversation.</p> <p>We want to support patients with the greatest health needs by enrolling them in dedicated medical support programs.</p> <p>— or —</p> <p>We want to identify patients with medical conditions for which treatment is currently under-resourced in our hospital system (eg. waiting disorders).</p>
Eliciting values and preferences [B]	<p>This stage can involve conversations with various stakeholders who will be involved with the automated decision support system to understand their needs. Those needs will be balanced against various limitations (budget restrictions, constraints that arise from interpreting the main objective mathematically), as well as potentially competing objectives between stakeholders.</p> <p>The hospital management is eager to use the resources to realize existing services to cut spending, while doctors argue that the program should supplement existing programs.</p> <p>— or —</p> <p>A patient advocacy group wants the resources to be allocated to the underfunded existing disease treatment program, while physicians insist the program should be used primarily to support diabetes and the elderly.</p>
Choosing the most important outcome ("optimisation logic") [C]	<p>In the world of data science, one cannot expect the system to achieve multiple diverse objectives at once - rather, a single outcome has to be defined.</p> <p>We want only patients threatened with the most serious conditions to be eligible for access of public resources) - we will prioritize precision (ensuring all identified individuals fit our criteria).</p> <p>— or —</p> <p>We want to identify all patients who could benefit, so that no one is unfairly excluded - we will prioritize recall (ensuring we find all individuals that may fit our criteria).</p>
Selecting a dataset [D]	<p>In the real world, access to high-quality and carefully collected data is limited. At this stage our data scientist will have to compose - from everything that was available (e.g. purchases from data brokers) and carefully collected - a dataset that will be comprehensive enough to construct an accurate prediction model.</p> <p>Private electronic health records of the hospital's patients (including existing codes and medications prescribed - relevant to the task at hand, but there are privacy and data quality concerns).</p> <p>— or —</p> <p>Public datasets from other hospitals and the national health services, which include data for a more diverse patient population, but with less detail.</p>
Choosing a prediction method (a model) [E]	<p>Knowing what data is available for training, our data scientist is now able to choose the prediction method that will perform best against the chosen indicators of success. These indicators, such as the mathematical interpretation of the outcome, have been set in previous steps.</p> <p>We use: Logistic regression - well established in statistics, does not require making statistical assumptions, but will require additional work in preparing data.</p> <p>Loss function: Standard cross-entropy loss with L1 regularization (pushing the model to drop the least important elements of input consistency).</p> <p>— or —</p> <p>We use: A neural network - can sometimes produce more accurate predictions, requires less data pre-processing, but is not easily interpretable, and can generate counterintuitive predictions that are hard to trace in some cases.</p> <p>Loss function: Standard cross-entropy loss modified to include a fairness constraint, ensuring equal rates of correct predictions for men and women.</p>
Testing and Calibrating the system [F]	<p>The model has to be tested, using training data. Our data scientist will now look at the errors that have occurred. Errors can be summarized by multiple metrics. At this stage data scientists choose metrics that are most relevant to evaluate the model. Based on test results, they decide whether errors are acceptable (i.e. how harmful it would be if a certain type of error happened after deployment).</p> <p>Are there any groups that are only classified as high-risk after analyzing much higher need than other groups?</p> <p>Based on our owners, is the system fulfilling its desired goals to what extent?</p> <p>Asking the following questions:</p> <ul style="list-style-type: none"> How often is our system correct? How often is it wrong? If it is wrong, what is the reason for the error? Are there any groups that are only classified as high-risk after analyzing much higher need than other groups? Based on our owners, is the system fulfilling its desired goals to what extent?
Updating the model [G]	<p>If the test results are not satisfactory, this is the signal to redesign one or more components of the system. The decision regarding which component to update and how is most often controlled by the data scientist alone.</p> <p>Let's add data from another hospital to include patients with more diverse socioeconomic backgrounds.</p> <p>— or —</p> <p>Let's change our most preferred outcome. Let's add mathematical constraints in order to change model behavior.</p> <p>— or —</p> <p>Let's rethink the main objective!</p>
Evaluating before deployment [H]	<p>The decision to evaluate the system in real-life conditions before its deployment is not obligatory. In the absence of population, it is usually data scientist (rather than system owners) who decide whether and what tests to run, based on their own assumptions regarding what the system should do, or shouldn't do.</p> <p>By contrast, in more mature engineering fields, such as civil engineering, there is a well-defined set of required tests and measurements that have to be reported for a system.</p> <p>Asking the following questions:</p> <ul style="list-style-type: none"> Does the model perform as well in the real world as it did when built in our data? How does the system perform with regard to various groups of people (eg. different genders, socioeconomic status, age)? Are there errors that the user sees which we did not catch? Is there any undesirable behavior we haven't anticipated? How does the model compare to human decisions? How is it different?

A walkthrough



Have you heard of COMPAS?

COMPAS (software)

From Wikipedia, the free encyclopedia

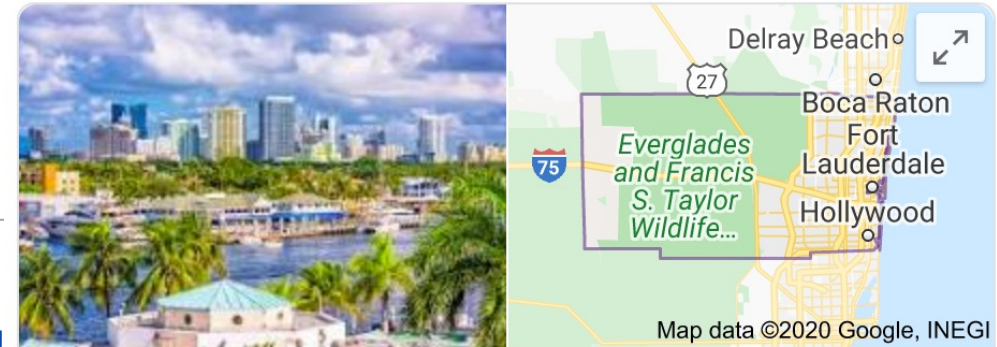
COMPAS, an acronym for Correctional Offender Management Profiling for Alternative Sanctions, is a [case management software](#) used by [U.S. courts](#) to assess the likelihood of a [defendant](#) becoming a [recidivist](#).^{[1][2]}

COMPAS has been used by the U.S. states of New York, Wisconsin, California, Florida's [Broward County](#), and oth

Contents [\[hide\]](#)

- [Risk Assessment](#)
- [Critiques and legal rulings](#)
- [Accuracy](#)
- [Further reading](#)
- [See also](#)
- [References](#)

Risk Assessment [\[edit\]](#)



Broward County

County in Florida

Broward County is a county in southeastern Florida, US. According to a 2018 census report, the county had a population of 1,951,260, making it the second-most populous county in the state of Florida and the 17th-most populous county in the United States. The county seat is Fort Lauderdale. [Wikipedia](#)

Incorporated cities: 24

Population: 1.936 million (2017)

Mayor: [Mark D. Bogen](#)

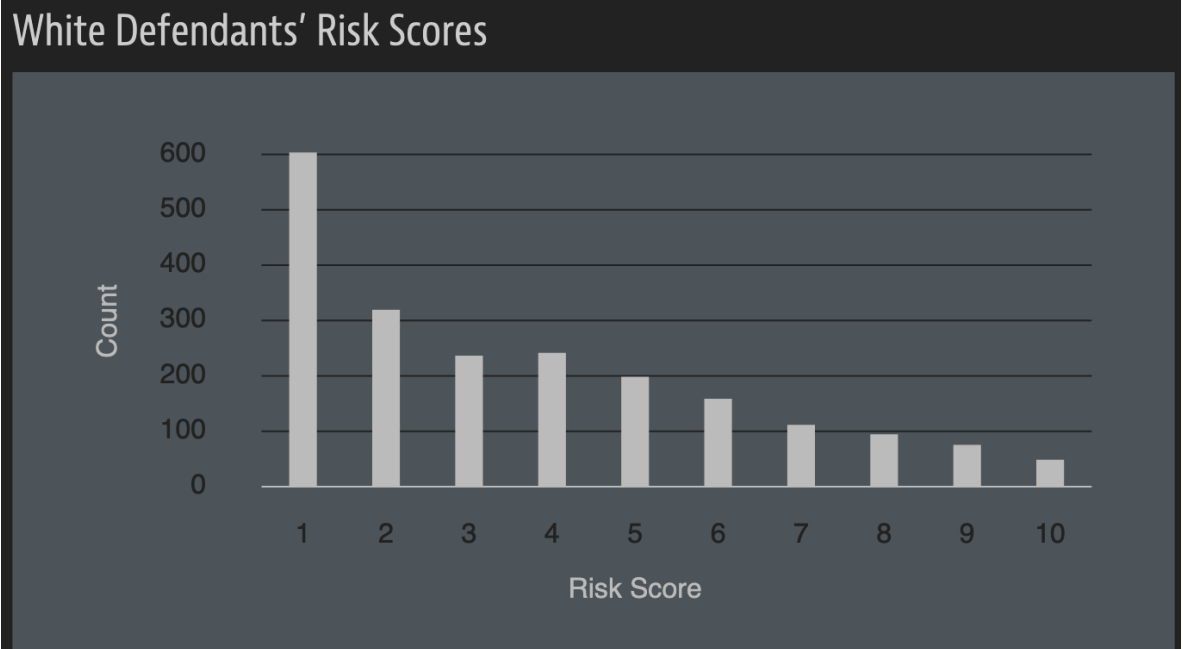
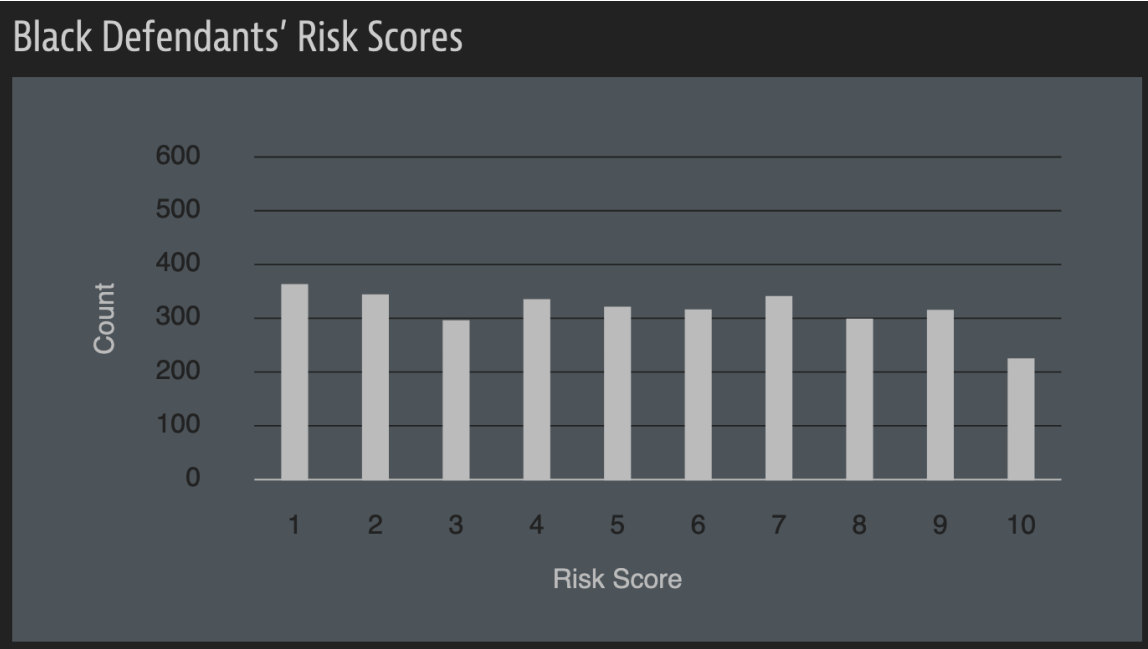
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

A sample of their result



False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.”

Anthony W. Flores

California State University, Bakersfield

Kristin Bechtel

Crime and Justice Institute at CRJ

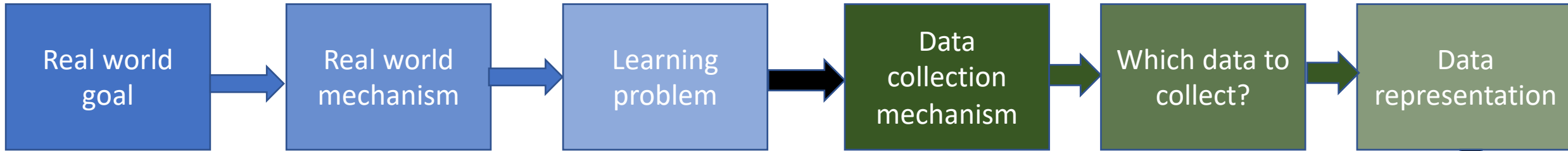
Christopher T. Lowenkamp

Administrative Office of the United States Courts

Probation and Pretrial Services Office



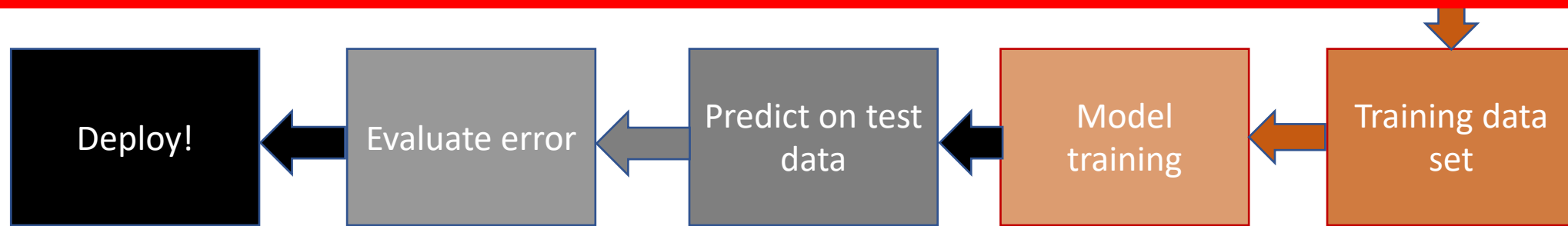
A walkthrough



The Problem

Imagine a situation where the creator of COMPAS had access to the [COMPAS dataset](#). In particular, you are in the team that wants to predict recidivism based on the [COMPAS dataset](#). How would you go about doing it?

Well, let's just walk through the ML pipeline to see how you would go about doing this.



Real world goal

Real world
goal

Real world goal

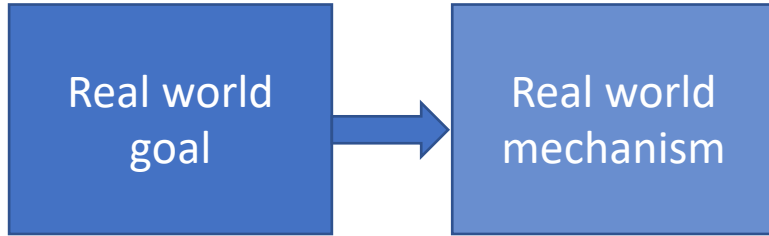
Reduce crime in society.

The Problem

Imagine a situation where the creator of COMPAS had access to the [COMPAS dataset](#). In particular, you are in the team that wants to predict recidivism based on the [COMPAS dataset](#). How would you go about doing it?

Well, let's just walk through the ML pipeline to see how you would go about doing this.

Real world mechanism



Real world mechanism

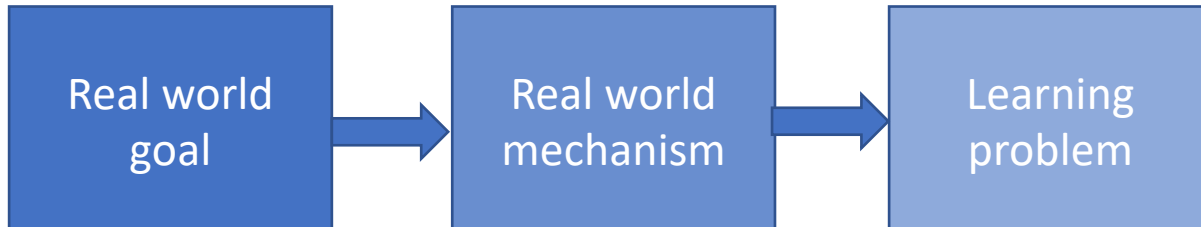
Based on some studies (or not!), your superiors decided that repeat offenders contribute most to crime. This in turn they decided would mean that if one could identify who would commit a crime again in the future, then one could use this information when making judgment on the current crime. Thus, they decided they wanted a system that can identify folks who will re-offend in the future and then promptly handed off the problem to your group to solve it.

The Problem

Imagine a situation where the creator of COMPAS had access to the [COMPAS dataset](#). In particular, you are in the team that wants to predict recidivism based on the [COMPAS dataset](#). How would you go about doing it?

Well, let's just walk through the ML pipeline to see how you would go about doing this.

Learning problem



Learning problem

Your group decides on the simplest learning problem: given a defendant *predict* if they will re-offend or not (in other words you are doing *binary classification* (binary because you are "labeling" defendants as either going to re-offend or not going to re-offend and you are doing classification because you are putting people into the two bins-- i.e. giving them a binary label and hence assigning them a "class."

There is another related option (which is what [COMPAS](#): instead of assigning defendants to two scores: they assign a score from 1 (being least likely to re-offend) to 10 (most likely to defend). This range of score (rather than a binary classification) could potentially be more useful to the end user of your system.

However, for our discussion (and indeed for most of the rest of the course), we will focus on binary classification.

The Problem

Imagine a situation where the creator of COMPAS had access to the [COMPAS dataset](#). In particular, you are in the team that wants to predict recidivism based on the [COMPAS dataset](#). How would you go about doing it?

Well, let's just walk through the ML pipeline to see how you would go about doing this.

