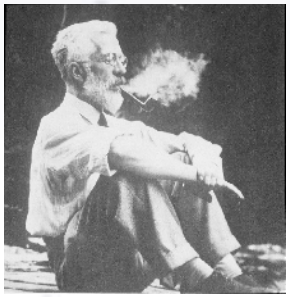
Applied ML

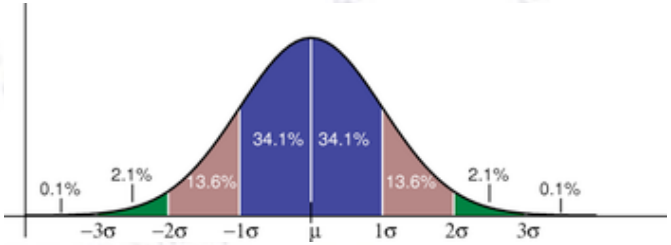
A discussion of Ethics in ML








Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

Ethics in Machine Learning

NOTE ON LIMITED WARRENTY!

I'm no expect on ethics, and the following are just thoughts for discussion.

Thanks to ATLAS colleague Savannah Theis for several cases/references used in the following.

Ethics in Machine Learning

While ML holds many opportunities, there are certainly also some pitfalls. Many of these are of technical character, but ETHICS is also one such.

In a famous case, Target (US large supermarket chain) data mined for about 25 products, that indicated that the costumer was pregnant:

> One Target employee I spoke to provided a hypothetical example. Take a fictional Target shopper named Jenny Ward, who is 23, lives in Atlanta and in March bought cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug. There's, say, an 87 percent chance that she's pregnant and that her delivery date is sometime in late August.

Ethics in Machine Learning

While ML holds many opportunities, there are certainly also some pitfalls. Many of these are of technical character, but ETHICS is also one such.

In a famous case, Target (US large supermarket chain) data mined for about 25 products, that indicated that the costumer was pregnant:

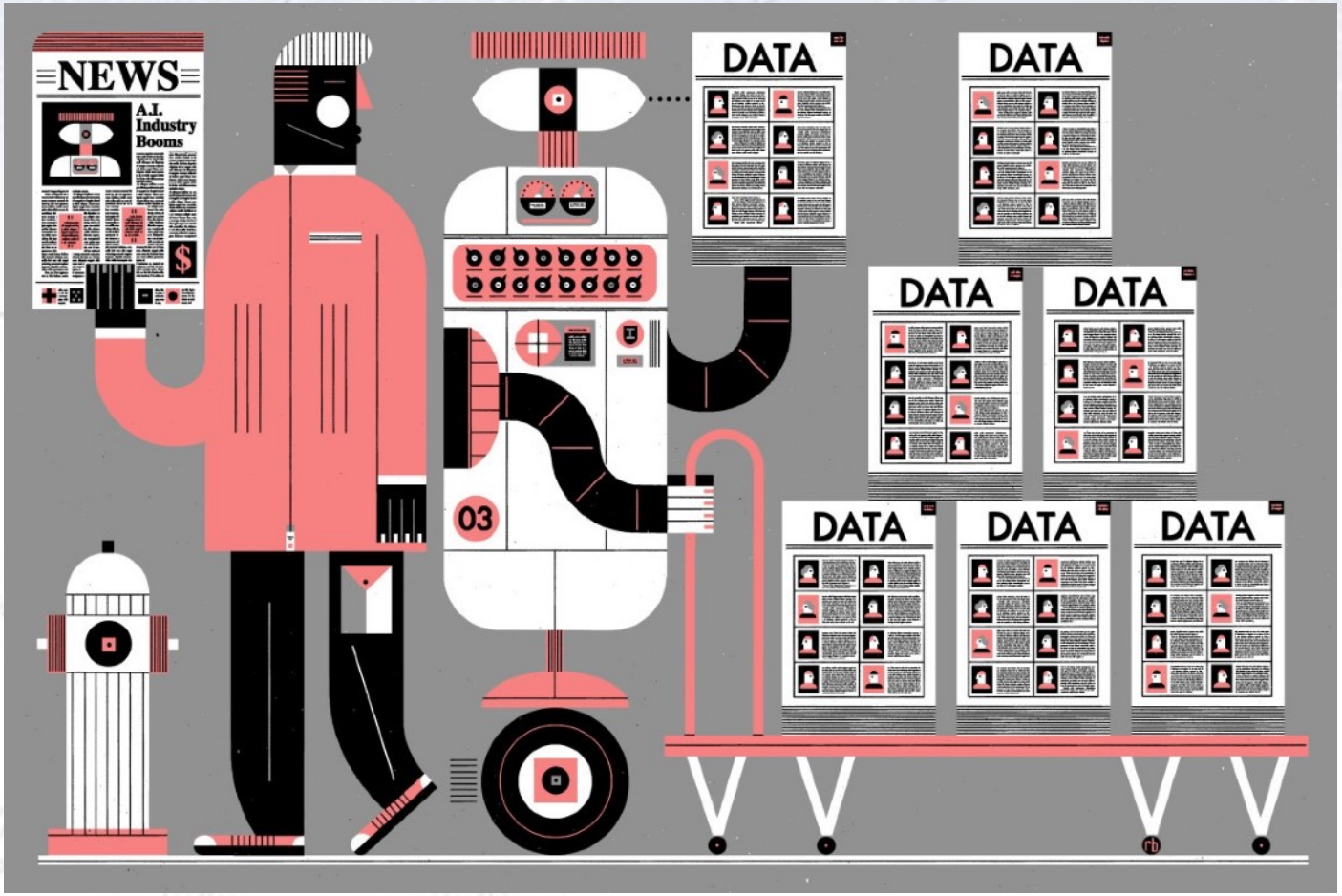
Hmmm… really? (*)

> One Target employee I spoke to provided a hypothetical example. Take a fictional Target shopper named Jenny Ward, who is 23, lives in Atlanta and in March bought cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug. There's, say, an 87 percent chance that she's pregnant and that her delivery date is sometime in late August.

(*) A link to the story in Forbes Magazine can be founder by clicking this text.

Is your software racist?

Not only may ML algorithms pick out cases “a little too well”. It may also have other “features” such as being racist!



Link to Politico article

Is your software racist?

Not only may ML algorithms pick out cases "a little too well". It may also have other "features" such as being racist!

Google Translate. Translating from Turkish, the output read like a children's book out of the 1950's. The un-gendered Turkish sentence "o is a nurse" would become "she is a nurse," while "o is a doctor" would become "he is a doctor." Why? Google's Translate tool "learns" language from existing texts, often including cultural patterns regarding how men and women are described.

Microsoft Twitter chatbot. It started spewing racist posts after learning from other users on the platform.

Google's photo-recognition. In a particularly embarrassing example in 2015, a black computer programmer found that this tool labeled him and a friend as "gorillas."

Link to Politico article

ML in the real world

ML in the real world

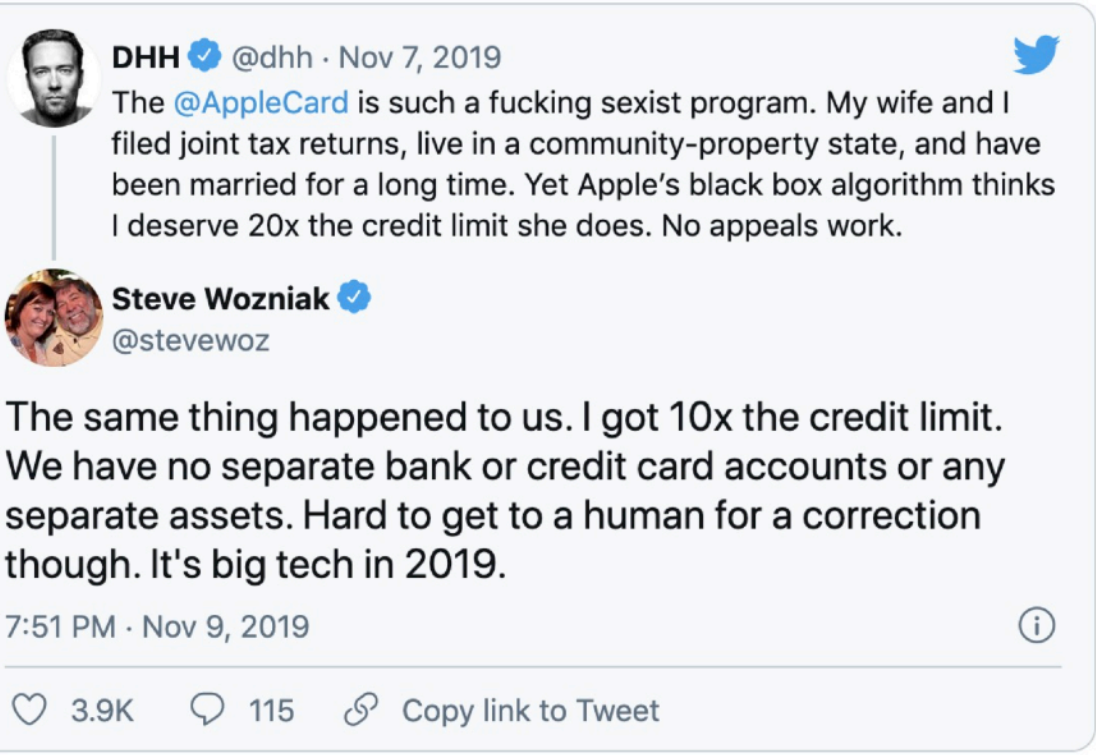
Machine Learning has entered many different aspects of society:

- **Entertainment**:
  Providing AI in games (Go, Chess, video games), generating screenplays, music and art, optimising visual effects, etc.
- **SoMe & Information**:
  Google results, ordering, translations, image captioning, news feed curation, SoMe ranking, click-bait optimisation, etc.
- **Societally**:
  Spam detection, image tagging, playlist generation, selection of commercials, GPS routing, spam detection, text prediction, self-driving cars, etc.
- **Financially**:
  Credit evaluation, loan offers, insurance rates, fraud detection, costumer ranking, stock trading, pension packages, etc.
- **Medically**:
  Cancer detection, producing treatment plans, drug discovery, hospital usage optimisation, pandemic predictions, etc.

Case: Apple Pay

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sex

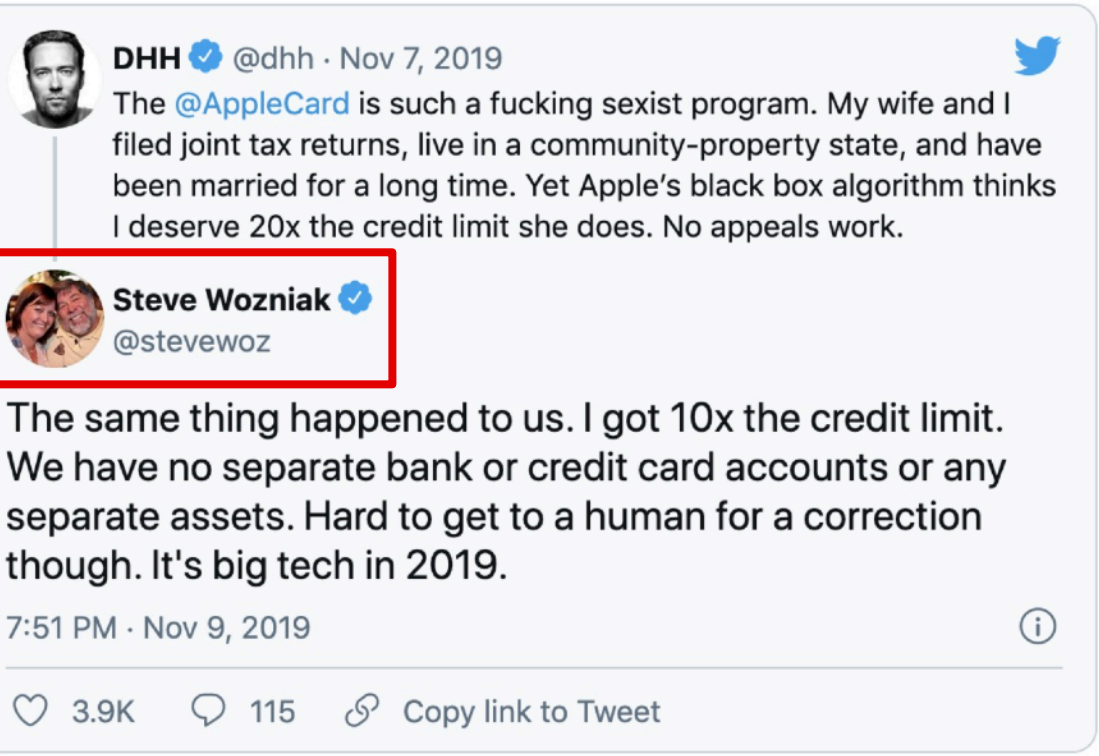


Apple Card Discrimination (NY Times)

Case: Apple Pay

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sex



Apple Card Discrimination (NY Times)

Case: Apple Pay

Apple Pay Card gave higher (or any) credit limits to **men!**

Why do you think?

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sex

DHH @dhh · Nov 7, 2019

The @AppleCard is such a fucking sexist program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

Steve Wozniak @stevewoz

The same thing happened to us. I got 10x the credit limit. We have no separate bank or credit card accounts or any separate assets. Hard to get to a human for a correction though. It's big tech in 2019.

7:51 PM · Nov 9, 2019

3.9K 115 Copy link to Tweet

Apple Card Discrimination (NY Times)

Case: Apple Pay

Apple Pay Card gave higher (or any) credit limits to **men!**

Why do you think?

Possible causes:

- Models trained on historical data, and may therefore reflect a past society.

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sex

DHH @dhh · Nov 7, 2019

The @AppleCard is such a fucking sexist program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

Steve Wozniak
@stevewoz

The same thing happened to us. I got 10x the credit limit. We have no separate bank or credit card accounts or any separate assets. Hard to get to a human for a correction though. It's big tech in 2019.

7:51 PM · Nov 9, 2019

3.9K 115 Copy link to Tweet

Note that removing class labels (i.e. gender) from training data doesn't force a fair outcome! Why?

Apple Card Discrimination (NY Times)

Case: Apple Pay

Apple Pay Card gave higher (or any) credit limits to **men!**

Why do you think?

Possible causes:

- Models trained on historical data, and may therefore reflect a past society.

Note that removing class labels (i.e. gender) from training data doesn't force a fair outcome! Why?

- Other variables may correlate and hence reveal gender.

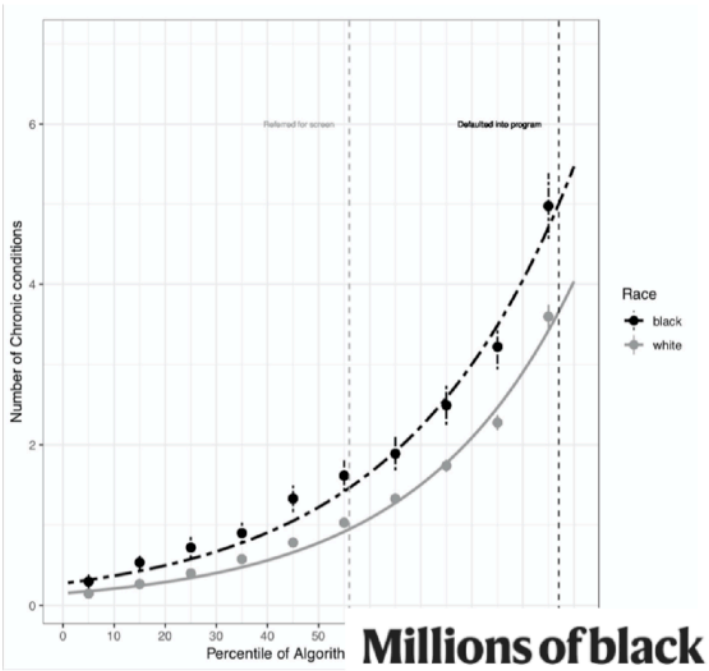
Can you think of ways to solve this problem/bias?

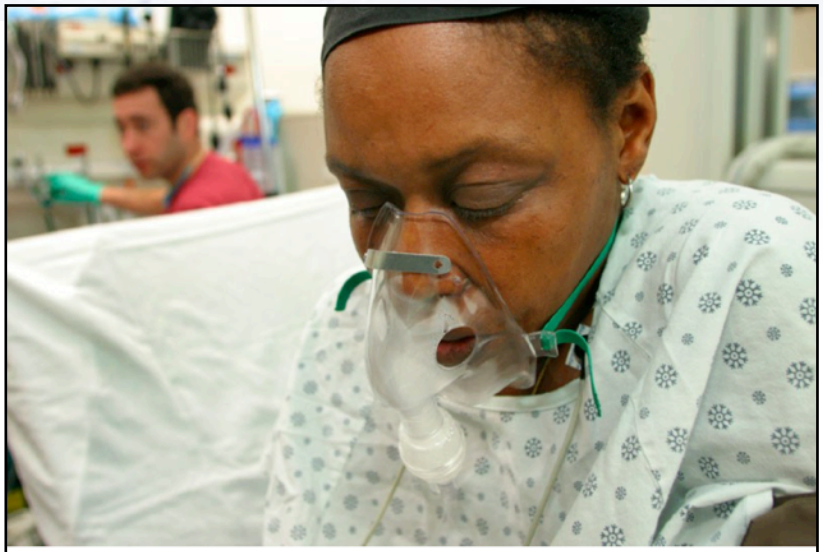
Apple Card Discrimination (NY Times)

Case: US health care

Healthcare risk assessment under-estimates disease severity in African American patients.

- Healthcare spending in the previous year was weighted.
- Ignoring broader context/domain knowledge can be devastating.





Black people with complex medical needs were less likely than equally ill white people to be referred to programmes that provide more personalized care. Credit: Ed Kashi/VII/Redux/eyevine

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

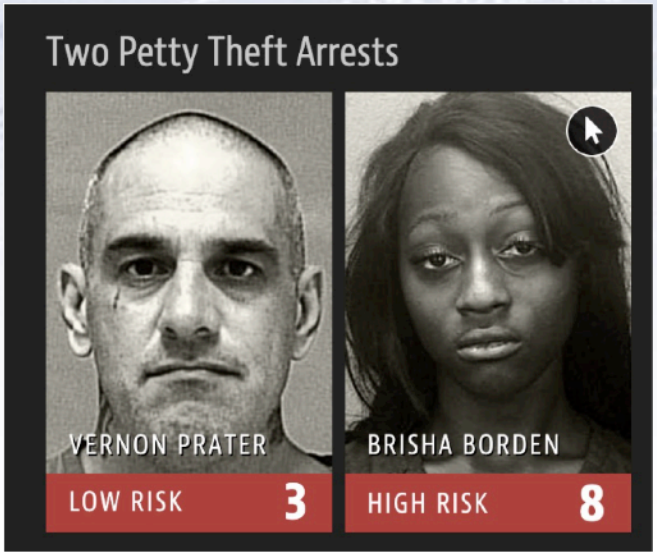
Black people affected by algorithms (Nature)

Case: US criminal bias

COMPAS Recidivism prediction tool predicts higher risk scores for minorities:

- Race is not an explicit factor in the score: based on survey questions and criminal records
- But there is historical bias in which communities are policed and who is sentenced
- Known relationship between socioeconomic status and petty crime (all crimes are considered in the model, training data not shared).

Overall accuracy was considered but not accuracy across classes and severities



Minority bias in criminal risk assessments

What to do about it?

ML should be Scientific

ML is (especially in society) facing a reproducibility crisis.
Designing a **good** ML model is like running a scientific experiment:
We don't know apriori what will work best!

Apply a scientific approach:

Step	Example
1. Set the research goal.	I want to predict how heavy traffic will be on a given day.
2. Make a hypothesis.	I think the weather forecast is an informative signal.
3. Collect the data.	Collect historical traffic data and weather on each day.
4. Test your hypothesis.	Train a model using this data.
5. Analyze your results.	Is this model better than existing systems? *
6. Reach a conclusion.	I should (not) use this model to make predictions, because of X, Y, and Z.
7. Refine hypothesis and repeat.	Time of year could be a helpful signal.

* Including how certain you are!

ML Hypothesis

Your ML hypothesis is a combination of the model you want to build and the pattern you want to explore:

- An algorithm can distinguish between normal and cancerous brain scans based only on pixel values.
- A model can simulate tau lepton decays within a defined margin of uncertainty.

ML Hypothesis

Your ML hypothesis is a combination of the model you want to build and the pattern you want to explore:

- An algorithm can distinguish between normal and cancerous brain scans based only on pixel values.
- A model can simulate tau lepton decays within a defined margin of uncertainty.

Questions to consider, as you construct your hypothesis:

- What specifically do I want my model to be able to do?
- What is the ideal outcome/use case of my experiment?
- How will I define success (proving hypothesis) or failure (reject hypothesis)?
- What kinds of outputs do I need the model to make and how will I use them?

ML Hypothesis

Your ML hypothesis is a combination of the model you want to build and the pattern you want to explore:

- An algorithm can distinguish between normal and cancerous brain scans based only on pixel values.
- A model can simulate tau lepton decays within a defined margin of uncertainty.

Questions to consider, as you construct your hypothesis:

- What specifically do I want my model to be able to do?
- What is the ideal outcome/use case of my experiment?
- How will I define success (proving hypothesis) or failure (reject hypothesis)?
- What kinds of outputs do I need the model to make and how will I use them?

All components need to be quantifiable and measurable:

- What are your input features and how are they represented?
- How do you quantify how well the model is doing?
- What metric(s) can you use to compare different models? Are they biased?

Setting up your model

Your model is never better than your training data, so consider if you have thought of the following points:

How much data is available and does each entry have the same information?

Do you have examples of all data classes/ranges?

How expensive is it to create/collect more data or labels?

Are the available labels related to the decision you want to make?

Is there noise in your label creation or distribution?

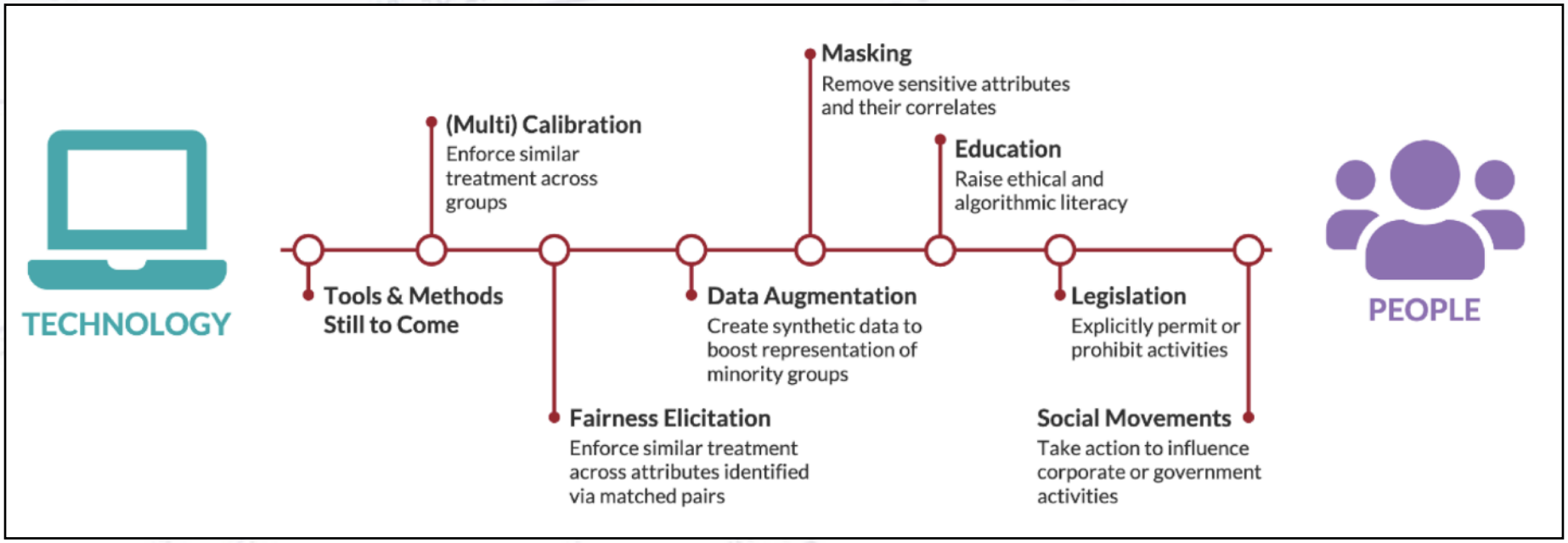
Are classes and inputs balanced and normalized?

Are there patterns in your data you don't want the model to exploit?

Many sources of problems

What we have just discussed is only one part of the problems (albeit, a significant one).

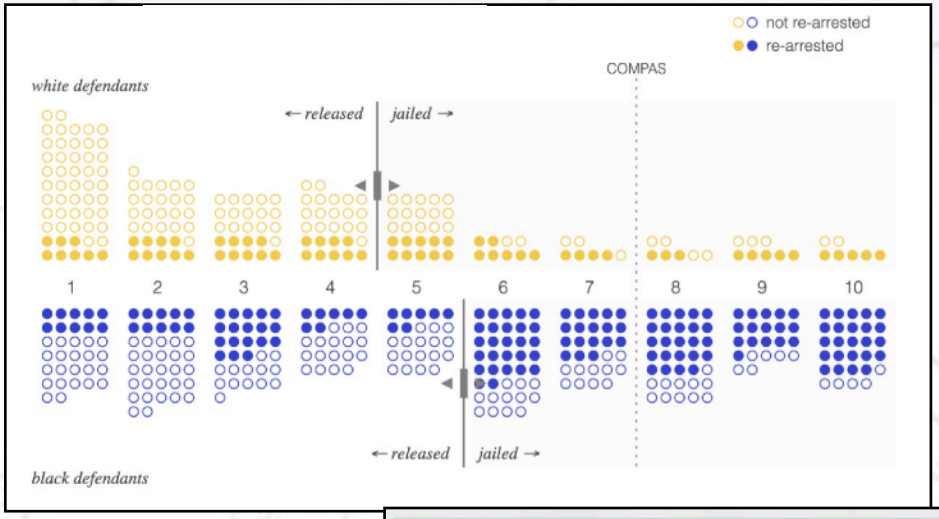
But the problem(s) are not purely mathematical problems, and many different methods and people are needed to address them.

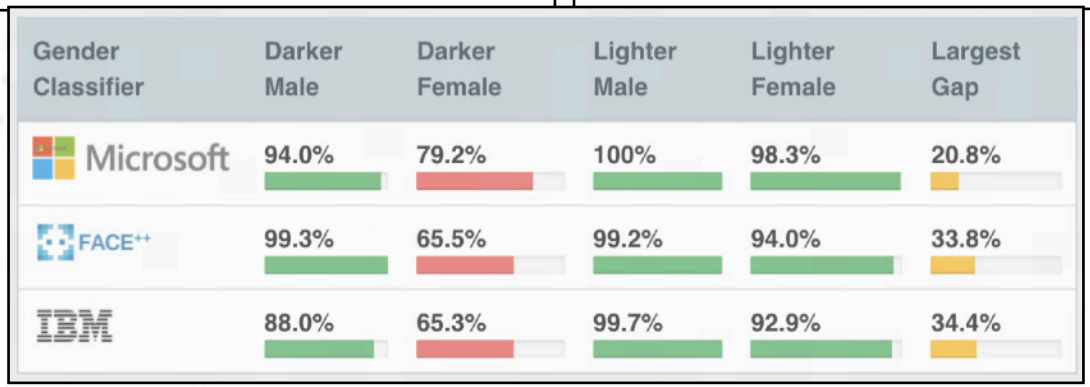


Fighting algorithmic bias

ML researchers measured the bias in several companies commercial facial recognition algorithms:

- This led some companies to modify their algorithms or suspend their facial recognition sales all together.





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%

Ethics discussion

Ethics in Machine Learning

It is CERTAINLY a good idea to think about the implications of using ML in ones work. Most likely it is perfectly fine (research, production, medicin, etc.), but sometimes it is less straight forward (banking, government).

I've discussed a few cases, and here are some hypothetical cases:

Housing prices:
Would it be OK for banks and/or ministry of tax to evaluate your residence based on variables including the description from the last time it was on sale?

Banks:
Would it be OK to ask for a meeting with a costumer, knowing that the financial distress seen in the bank is most likely because of a coming divorce?

Schools:
Would it be OK to ask a student for a "chat", if some ML indicated that the student was about to drop out?

Discussion

In the break-out sessions, please discuss ML ethics cases. The following might be questions that inspire (provoke?) thoughts and discussion:

1) In which ways do you think that ML will affect the ethics in medicin?

2) What changes in legislation do you think, that our the "ML world" warrants?
   - Databases
   - Collection of data
   - Surveillance
   - DNA samples
   - Browsing history

3) Do you think that ML will transform the way wars are fought? (to some extend, this has already happened).

4) What impact do you dream that ML will have on society, and is this match a good ethical standard (whatever that is)?

Resources & References

The following are resources and references that might be interesting/useful:

- AI now Institute
- Data & Society
- Berkman Klien Center
- Stanford Center for Human-Centered AI
- Montreal AI Ethics Institute
- Oxford Future of Humanity Institute
- Alan Turing Institute
- Algorithmic Justice League
- Data for Black Lives
- Resistance AI

The discussion of ethics in ML is wide ranging and multi-facetted:
Some want justice, others better medical care, social freedom, etc.

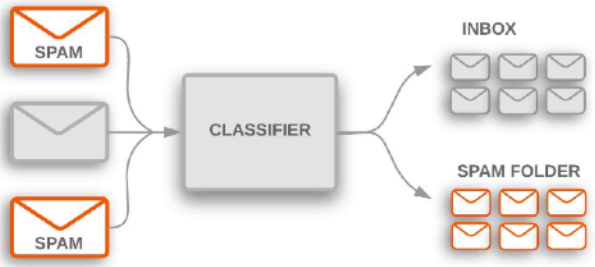
For most parts, ML simply provides a continual improvement of essentially all parts of our society. The trick (and above discussion) lies in directing the course of this development.

Bonus Slides

What can ML do?

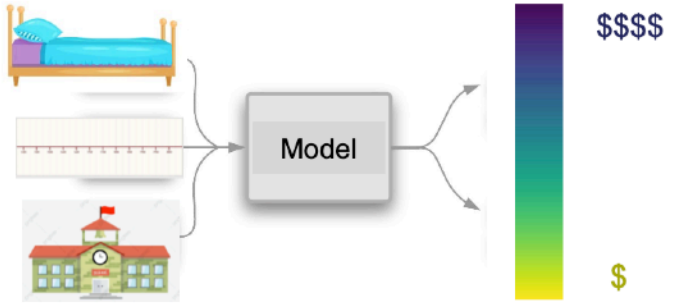
Classification:

Predict a class label for an input



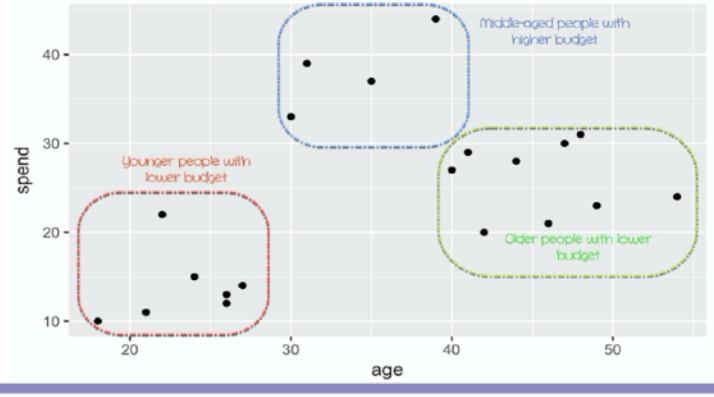
Regression:

Predict a continuous variable



Clustering:

Group similar inputs



Generation:

Construct new data within pattern