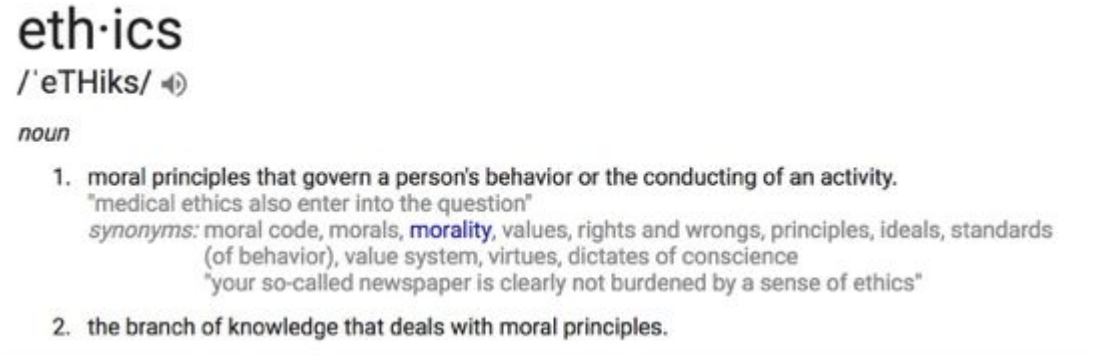
CSE 519: Data Science

Steven Skiena
Stony Brook University

Lecture 24: Human-centric Data Science

Ethics and AI



eth·ics
/ˈeTHiks/

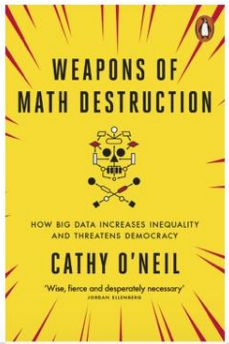
noun

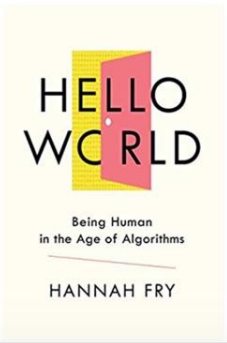
1. moral principles that govern a person's behavior or the conducting of an activity.
   "medical ethics also enter into the question"
   synonyms: moral code, morals, morality, values, rights and wrongs, principles, ideals, standards (of behavior), value system, virtues, dictates of conscience
   "your so-called newspaper is clearly not burdened by a sense of ethics"
2. the branch of knowledge that deals with moral principles.

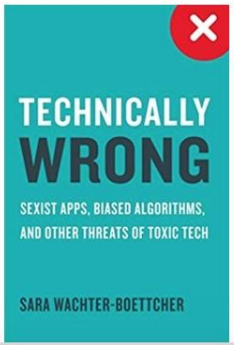
I think the definition gets right that ethics governs people's behavior.
People build and apply technology, so the ethical demands are upon those who build and use AI systems.
Moral principles are often in tension with each other, and different people can reasonably hold to different beliefs and standards.

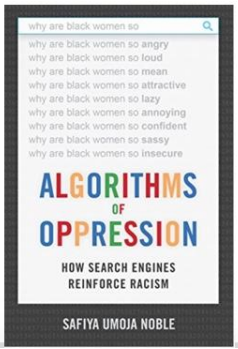
Ethical Concerns about ML Models

- Economic dislocation: will these machines take jobs from people?
- Privacy concerns: is too much of my private information being given to machines to build these models?
- Bias concerns: are these models learning the wrong thing from the training data?
- Agency issues: are people adequately in the decision loop to step in when models go astray?









Economic Dislocation

- Many blue collar jobs are threatened: taxi/truck drivers, security guards,…
- Many classes of white collar work are threatened: lawyers, translation/transcription, education, pathology/medicine…
- This double threat makes it hard to know what to train for!
- But: future jobs will involve working with AI systems to do better work. This suggests the need for a broad general education: communication, technology, and empathy.
- But: unemployment is under 4%, seemingly impossible 10 years ago.
- But: technology has always destroyed jobs, yet generated new and better jobs, in ways impossible to predict.
- The world will always change rapidly and you will have to adapt to it.

Privacy Concerns

- The power of Big Data is such that large companies gather private information on a scale many people find threatening: Google, Facebook, Amazon/Alexa.
- Face identification means ubiquitous video surveillance.
- Europe does a much better job protecting data privacy by law than the United States.
- More subtle concerns about what inferences are being made about you: e.g. models that predict whether you are gay or pregnant from observable data.

Bias Concerns

- Models trained on racially/gender biased data learn these biases.
- Examples: Amazon's resume screener, Google's image search.
- It is hard to build accurate classifiers for small/minority classes.
- Models for rating teachers, positioning policemen, assigning bail, and granting parole are often opaque and proprietary, and not adequately evaluated for performance or bias.
- But: people are also biased, and it should be easier to evaluate algorithms than people. Biased AI is usually accidental.
- But: fairness is often hard to define: e.g. do you want the system that minimizes errors globally or equalizes all groups?

Agency Issues

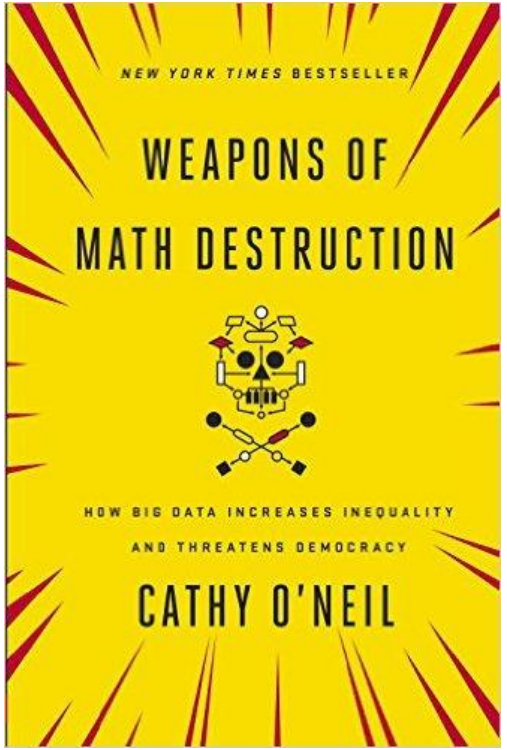
- There is no imminent danger of machines becoming intelligent and taking over the world!
- The real problem is people trusting AI models too much.
- Should algorithms be permitted to fire weapons in combat?
- When is a self-driving car safe enough to drive?
- Are mistakes correctable? Who do you complain to when a program eliminates your resume before a human ever sees it?
- Are models correctable? Are processes in place which continually evaluate models and improve them with time?
- Ultimately these are human decisions, and we must pass judgment on how to use AI or any other technology.

Weapons of Math Destruction

It is important to understand the potential harm data-driven models can cause.

Correlation is not causation, but models so-trained can trigger actions and feedback mechanisms, resulting in self-fulfilling prophesies.

It is important for us as data scientists to think about societal issues in a constructive way.



Properties of a Good Model

- It uses relevant data, not just data that happens to be available.
- It is transparent, making clear why it is making its decisions.
- There is a clear measure of success, and an embedded feedback mechanism to evaluate and learn from it.

Teacher Ratings: A Terrible Model

School systems are firing teachers based on the test-scores of their students:

- Relevant or available data?
- Transparent? Enough to see the statistical significance issues from small samples.
- Feedback/evaluation? Are you firing bad teachers or those with bad students?

Societal / Ethical Issues in Big Data

- Integrity in communications and modeling
- Transparency and data ownership
- Model-driven bias and filters
- Maintaining the security of large data sets
- Maintaining privacy in aggregated data

Integrity in Communications and Modeling

Fight the temptation to inflate your results, because you know the model and audience:

- Present p-values, not just correlations.
- Don’t cherry-pick to present the best results.
- Use visualization to reveal, not conceal.
- Make clear the limitations of your models to non-technical audiences.

Transparency and Ownership

- Does your organization follow data retention/use policies?
- To what extent do users own the data they have generated?
- Can data errors propagate? Are there mechanism for correction, like credit data?
- Is data provenance maintained?

Model-Driven Bias and Filters

Learning algorithms pick up biases from biased training data:

- Will your search engine show better job opportunities to men than women?
- Are predatory ads shown to poor people?
- Do news filters reinforce political polarization?

Maintaining Security of Big Data

There are ethical responsibilities to encrypt and delete data to avoid security breaches:

- Making 100 million people change their password costs 190 man-years of effort.
- Released data on addresses, ID numbers, and accounts persist for many years/forever.
- Can anyone get/keep life insurance?

Maintaining Privacy in Aggregate

People are often identifiable even if their names, address, and ID number are deleted.

- The AOL search engine data release.
- Destinations/tipping habits of celebrities, by correlating GPS locations of trips with paparazzi photos.

Data scientists must strive to be responsible.

Aggregating Expertise

No person/program has all the answers.

Social media and other new technologies have made it easier to collect and aggregate opinions on a massive scale.

But how can we separate the wisdom of crowds from the cry of the rabble?

Francis Galton's Ox

At a livestock fair in 1906, villagers were invited to guess the weight of a ox.

Galton observed that none of the almost 800 observers guessed the correct weight (1,178 pounds).

Yet the average guess was amazingly close: 1,179 pounds!

Wisdom of Crowds: Penny Demo

How many pennies do I have in this jar?

Watch the video!

Wisdom of Crowds: Rules of Demo

How many pennies do I have in this jar?

Ten of you write your opinions on cards
Ten more tell me them one by one.

Who wants to bet?

Independent Guesses

537, 556, 600, 636, 1200, 1250, 2350, 3000, 5000, 11000, 15000

Median: 1250, Mean: 3739, Actual: 1879

The median is closer than any other guess.

Conditioned Guesses

A second group of students made guesses after seeing the first group's answers:

750, 750, 1000, 1000, 1000, 1250, 1400, 1770, 1800, 3500, 4000, 5000

Median: 1325, Mean: 1935, Actual: 1879

Showing other people's guesses substantially conditioned the distribution (no outliers)

Financial Stakes

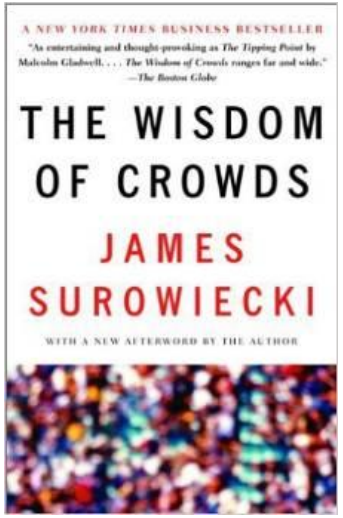
Allowing people to bet on the outcome yielded two wagers:

1500, 2000 Actual: 1879

People willing to bet their own money on an event are by definition more confident in their selection.

When is the Crowd Wise?

- When the opinions are independent: this avoids groupthink.
- When the crowd consists of people with diverse knowledge / methods.
- When the problem is in a domain where the crowd does not need specialized knowledge.
- Opinions can be fairly aggregated.



Diversities of Opinion

Crowds only add information when there is disagreement.

A committee with perfectly correlated experts contributes nothing more than any one of them.

Be an incomparable element on the partial order of life.

Mechanisms for Aggregation

On numerical aggregation, the mean or median work if errors are symmetrically distributed.

But which is more robust: mean or median?

On classification problems, voting is the basic aggregation mechanism.

But should all votes be treated equally?

The Condorcet Jury Theorem

If the probability of each voter being correct is *p>0.5*, the probability of a majority of voters being correct *P(n)* is >p.

$$P_n = \sum_{h=(n+1)/2}^{n} [n!/h!(n-h)!]p^h(1-p)^{n-h}$$

For p=0.51, a jury of 101 members is right 57%, P(1001)=0.73 and P(10001)=0.9999.

As *n* gets large enough *P(n)* approaches *1*.

Arrow's Impossibility Theorem

No election system for summing permutations of preferences as votes satisfies:

- Each voter is allowed any ranking they want.
- The relative order of any subset of candidates is independent of all others.
- If all voters like $a > b$, a wins ahead of b.
- No single voter can dictate rankings.

Crowdsourcing Services

Services like Amazon Turk and Crowdflower provide the opportunity to hire large numbers of people for small amounts of piece work.

In life, you generally get what you pay for.

Getting people to do your bidding requires both incentives and clear instructions.

amazonmechanical turk
Artificial Artificial Intelligence

Sign In

Your Account | HITs | Qualifications

442,955 HITs available now

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing that pay at least $ 0.00 for which you are qualified require Master Qualification GO

All HITs

1-10 of 3194 Results

Sort by: HITs Available (most first) GO!

Show all details | Hide all details

1 2 3 4 5 › Next » Last

Is this Category Valid for the Groupon Deal? — View a HIT in this group

Requester: Relevance Groupon
HIT Expiration Date: Jul 9, 2014 (6 days 4 hours)
Time Allotted: 5 minutes
Reward: $0.04
HITs Available: 78236

Mark all the cells in the 10 images — View a HIT in this group

Requester: Cell Counting
HIT Expiration Date: Jul 6, 2014 (3 days 9 hours)
Time Allotted: 25 minutes
Reward: $0.01
HITs Available: 33091

Extract purchased items from a shopping receipt — View a HIT in this group

Requester: Jon Brelig
HIT Expiration Date: Jul 10, 2014 (6 days 23 hours)
Time Allotted: 2 hours
Reward: $0.08
HITs Available: 26691

Geo Result Relevance-Sun Jun 08 08:20:10 PDT 2014 — View a HIT in this group

Requester: Amazon Requester Inc.
HIT Expiration Date: Jul 9, 2014 (5 days 14 hours)
Time Allotted: 60 minutes
Reward: $0.00
HITs Available: 23052

Geo Result Relevance-Fri Jun 27 10:50:25 PDT 2014 — View a HIT in this group

Requester: Amazon Requester Inc.
HIT Expiration Date: Jul 27, 2014 (3 weeks 3 days)
Time Allotted: 60 minutes
Reward: $0.00
HITs Available: 16493

Inv_B_2 — View a HIT in this group

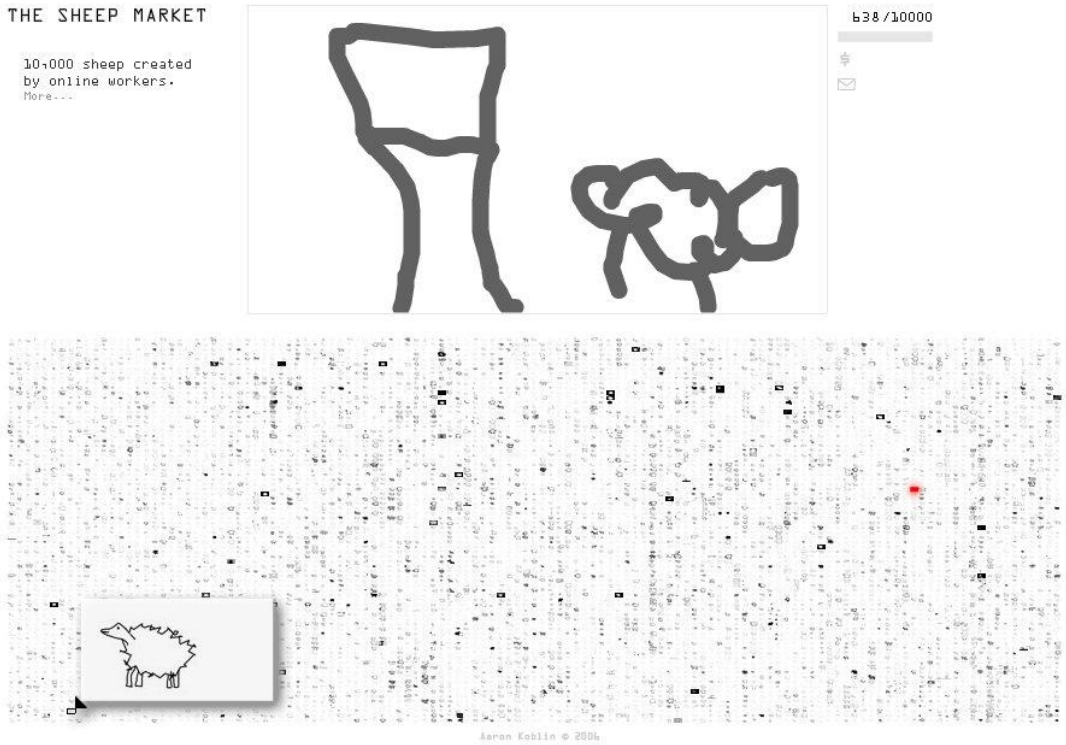
Requester: rohzit0d
HIT Expiration Date: Jul 17, 2014 (2 weeks)
Time Allotted: 48 minutes
Reward: $0.00
HITs Available: 11727

Good Uses of Turkers

- Measuring aspects of human perception (e.g. what name would you call this color?)
- Language understanding tasks (read cans for the blind).
- Obtaining training data for machine learning classifiers.
- Creative efforts (blog posts/reviews or 10,000 sheep drawings, see www.sheepmarket.com)
- Economic/psychological experiments
- Reviewing transcripts for accuracy

Sheepmarket.com

What creative endeavors can you think of that people will do for $0.02 each?



Bad Uses of Turkers

- Any task which requires advanced training.
- Any task you cannot specify clearly.
- Any task where you cannot perform some type of validation as to whether they are doing a good job or not.
- Any illegal task, or one too inhuman to subject people to (IRB ethical standards)

Applications for Turkers? (Projects)

- Miss Universe?
- Movie gross?
- Baby weight?
- Art auction price?
- Snow on Christmas?
- Super Bowl / College Champion?
- Ghoul Pool?
- Future Gold / Oil Price?

Gamification

Make things fun so that people work for free!

Motivational techniques include:

- leader boards
- points
- badges
- leader boards
- progress bars

Games with a Purpose (GWAP)

- CAPTCHAs for OCR
- ESP game for labeling image components
- Psychological/IQ testing in games/apps
- FoldIt game for predicting protein structures

Keys to success include:

- Being playable enough to become popular
- Abstracting technicality as scoring functions