80250993 Machine Learning
@Tsinghua University

Chapter 21
Ethics Issues in Machine Learning and AI

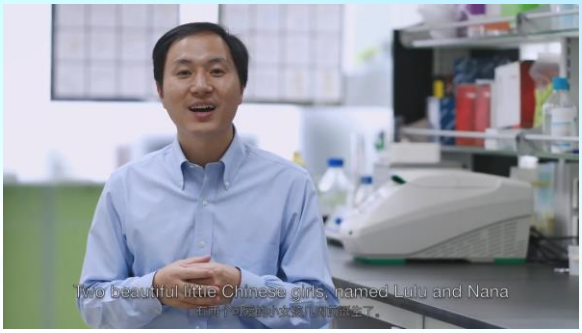
Xuegong Zhang
Dec. 23, 2021

The Ethics of Science

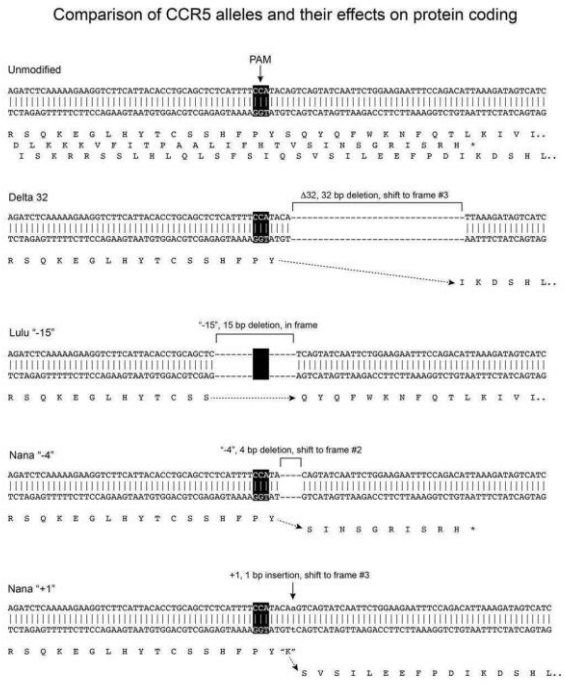
- **How does science actually benefit people?**
- Are there distinct kinds of benefits that science confers?
- Who benefits the most from science?
- Who decides what scientific questions get attention and public support?
  - Who should decide?
- Do some people, or some countries, disproportionately benefit from science at the expense of others?
  - If so, is this situation justifiable, or is it unfair?

Dr. Kevin deLaplante, Critical Thinking about Science, Unit 5: The Ethics of Science

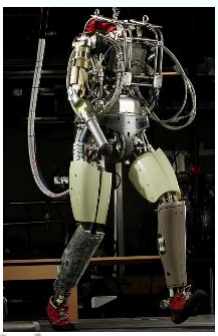
Editing Babies

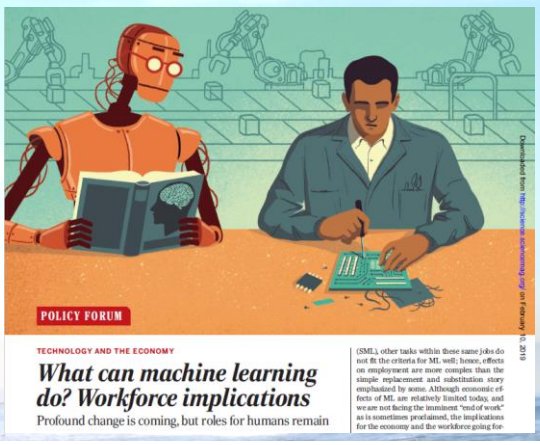


Xudong Zhang



Artificial Intelligence & Machine Learning



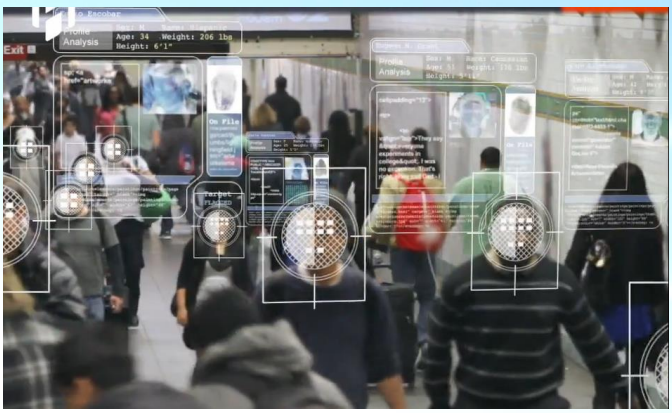


Xudong Zhang

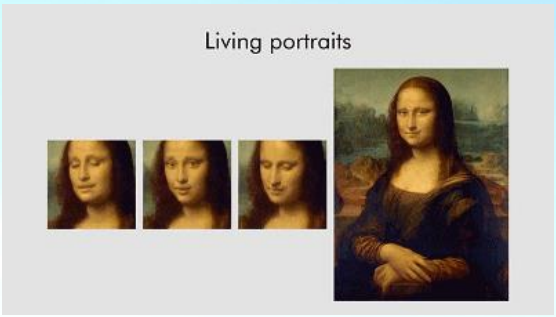
Will AI steal our jobs?

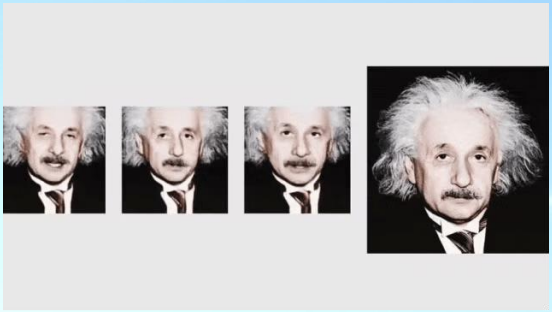
https://v.qq.com/x/page/x0601aigfrx.html

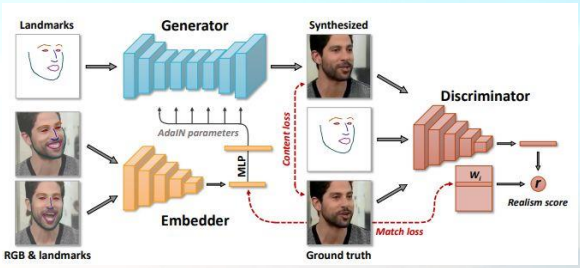
How will AI change our life?



Advanced adversarial learning







Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, Victor Lempitsky, Few-Shot Adversarial Learning of Realistic Neural Talking Head Models, arXiv:1905.08233, 2019

Will AI change our mind?

Example 1: “Google’s racist algorithm”

https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

WIRED BUSINESS CULTURE GEAR IDEAS SCIENCE MORE SIGN IN SUBSCRIBE

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn't found one.

THE VERGE TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT MORE

GOOGLE TECH ARTIFICIAL INTELLIGENCE

Google ‘fixed’ its racist algorithm by removing gorillas from its image-labeling tech

Nearly three years after the company was called out, it hasn’t gone beyond a quick workaround

By James Vincent | Jan 12, 2018, 10:35am EST

Back in 2015, software engineer Jacky Alciné pointed out that the image reco
Google Photos were classifying his black friends as “gorillas.” Google said it
mistake, apologized to Alciné, and promised to fix the problem. But, as a new
nearly three years on and Google hasn’t really fixed anything. The company h
image recognition algorithms from identifying gorillas altogether — preferring,
service rather than risk another miscategorization.

A spokesperson for Google confirmed to *Wired* that the image categories “gorilla,” “chimp,” “chimpanzee,” and “monkey” remained blocked on Google Photos after Alciné’s tweet in 2015. “Image labeling technology is still early and unfortunately it’s nowhere near perfect,” said the rep. The categories are still available on other Google services, though, including the Cloud Vision API it sells to other companies and Google Assistant.

https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

Example 2: “Google’s rightwing bias”

Google’s search algorithm appears to be systematically promoting information that is either false or slanted with an extreme rightwing bias on subjects as varied as climate change and homosexuality.

Following a recent investigation by the Observer, which found that Google’s search engine prominently suggests neo-Nazi websites and antisemitic writing, the Guardian has uncovered a dozen additional examples of biased search results.

Google’s search algorithm and its autocomplete function prioritize websites that, for example, declare that climate change is a hoax, being gay is a sin, and the Sandy Hook mass shooting never happened.

https://www.theguardian.com/technology/2016/dec/16/google-autocomplete-rightwing-bias-algorithm-political-propaganda

Support The Guardian
Available for everyone, funded by readers
Contribute → Subscribe →
Sign in
The Guardian

News Opinion Sport Culture Lifestyle

World UK Environment Science Cities Global development Football Tech Business More

Google

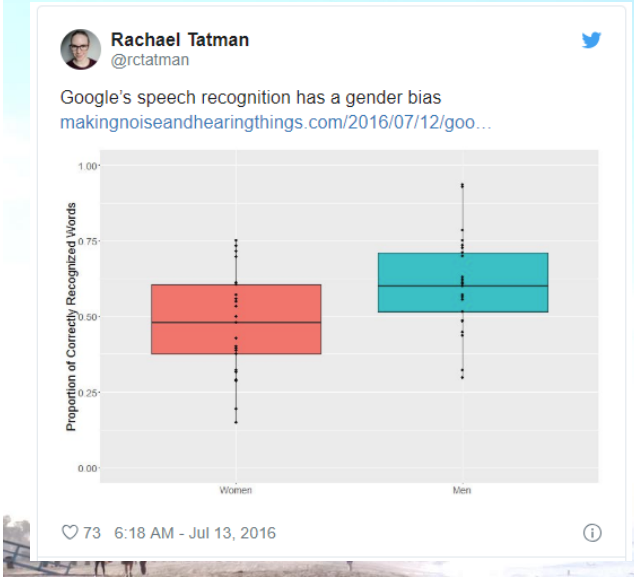
This article is more than 2 years old

How Google's search algorithm spreads false information with a rightwing bias

Search and autocomplete algorithms prioritize sites with rightwing bias, and far-right groups trick it to boost propaganda and misinformation in search rankings

▲ Manipulations by Google and third parties trying to game the system impact how search engine users perceive the world, even influencing the way they vote. Photograph: Virginia Mayo/AP

Example 3: "Google's gender bias"



The Daily Dot

Research shows gender bias in Google's voice recognition

Google can recognize men's voices better.

https://www.dailydot.com/debug/google-voice-recognition-gender-bias/

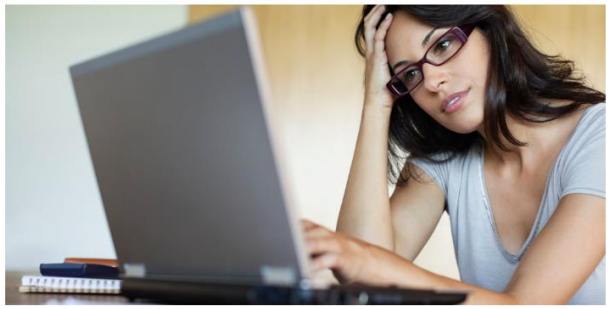


Photo via Sam Edwards/GettyImages

Selena Larson— 2016-07-15 03:17 pm

Voice recognition technology promises to make our lives easier, letting us control everything from our phones to cars to home appliances. Just talk to our tech, and it works.

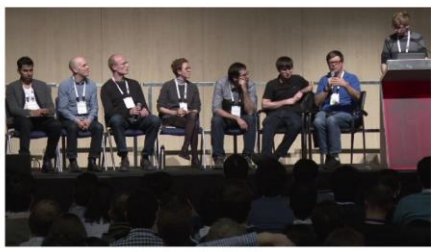
As the tech becomes more advanced, there's another issue that's not as obvious as a failure to process simple requests: Voice recognition technology doesn't recognize women's voices as well as men's.

Example 4: from NIPS to NeurIPS

https://medium.com/@therese.koch1/nips-ai-conference-to-continue-laughing-about-nipples-at-the-expense-of-women-in-tech-8c0fa74b1ec4

NIPS AI Conference to Continue Laughing about Nipples at the Expense of Women in Tech

Therese Koch Follow
Oct 26, 2018 · 4 min read



Panel discussion on Adversarial Training at NIPS 2016.

This past Tuesday, the Neural Information Processing Systems conference announced that after a survey of former attendees, they would not be changing their name. If you're quick with acronyms, you have likely already

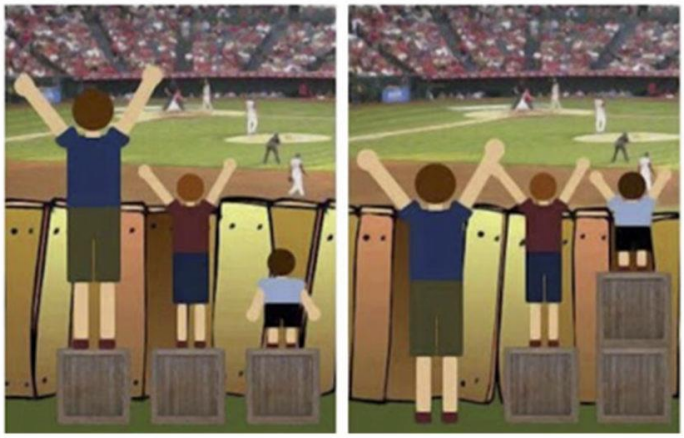
This past Tuesday, the Neural Information Processing Systems conference announced that after a survey of former attendees, they would not be changing their name. If you're quick with acronyms, you have likely already figured out why this possibility was being discussed in the first place. NIPS is one of the biggest and most prestigious artificial intelligence conference in the world. It has been held annually since 1987, and will be hosted from December 3rd to 8th in Montreal, Canada this year.

The sexual connotations of this name haven't been a secret for many years but certain AI researchers, particularly women, started voicing their concerns more publicly surrounding last year's conference in Long Beach, California. In a field already so strongly dominated by men, many women feel uncomfortable with this name, which often elicits crude jokes and opens the door for harassment. For instance, the audience cheered Elon Musk as he joked about tits and nips in a keynote talk last year. The AI company Dessa (formerly DeepLearni.ng) was promoting t-shirts with the slogan 'My NIPS are np-hard' — Their self proclaimed 'diverse' team of 28 employees includes only 4 women.

So when the conference finally announced they would consider a name change, it seemed like a step in the right direction. The survey they conducted, however, was not. As a conference on what is essentially fancy statistics, you might expect them to have a better grasp on concepts like sampling bias.

Of the 2270 former participants who responded to the survey, 1881 were men and only 294 were female (95 chose not to disclose their gender)

EQUALITY vs. EQUITY



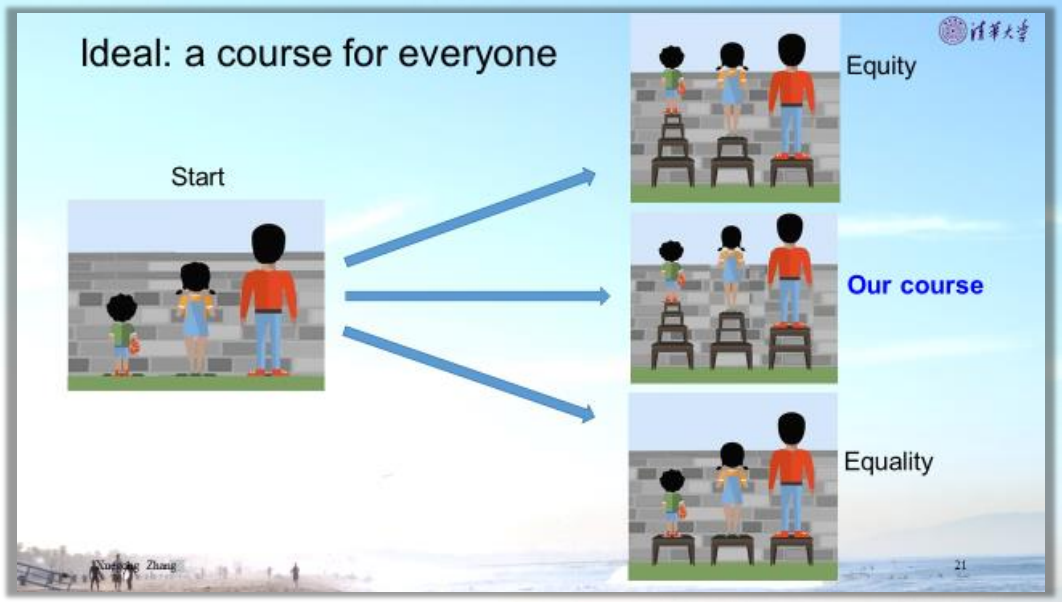
Equality = Sameness

GIVING EVERYONE THE SAME THING → It only works if everyone starts from the same place

Equity = Fairness

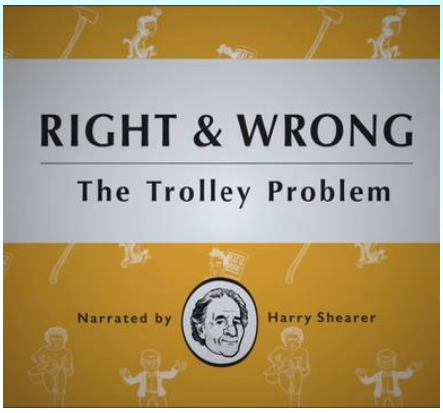
ACCESS TO THE SAME OPPORTUNITIES → We must first ensure equity before we can enjoy equality

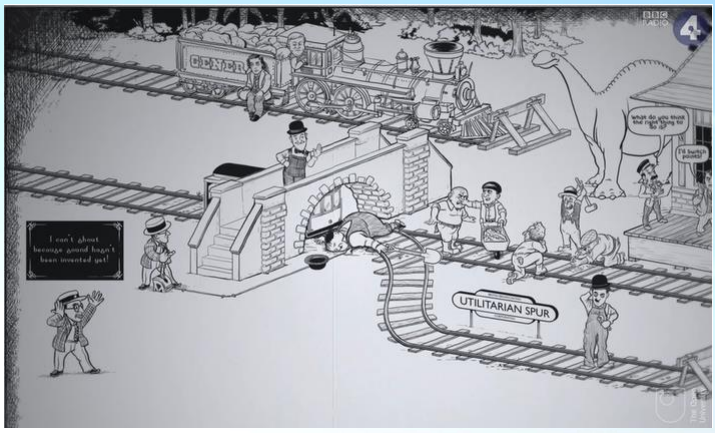
Equity image credit: Please note, this image was adapted from an image adapted by the City of Portland, Oregon, Office of Equity and Human Rights from the original graphic: http://indianfunnypicture.com/img/2013/01/Equality-Doesnt-Means-Justice-Facebook-Pics.jpg



The basic question of ethics

Moral Reasoning: What is the right thing to do?





Moral Reasoning: What is the right thing to do?

From *Illustrated Police News*, 20 September 1884.

Moral Reasoning: What is the right thing to do?

- Consequentialist
  - Locates morality in the consequences of an act
    - To maximize the utility (the overall happiness)
- Categorial
  - Locates morality in certain duties and rights

Utilitarianism
功利主义

Deontology
道义论

From *Illustrated Police News*, 20 September 1884.

The moral side of murder

What is the right thing to do?

r/Coronavirus
Posted by u/knightlyostrich • 7h

The Italian Society of Anesthesia, Resuscitation and Intensive Care is considering setting an age limit to access to intensive care, prioritizing those who have more years to live and better chances of survival

Europe

Xuegong Zhang

最热评论 百岁老人新冠肺炎出院新闻的跟评

用户1693604827 美国加利福尼亚州
该救活的没救活，不该救的却依然活着。
3月8日23:54 赞100 回复

云飞扬：就冲这句话，看看美国式的教育多失败。连起码的人性都没了。
34分钟前 已赞19 回复

用户5102445090：回复@用户5856924770：有限的资源下就是有该救和不该救的区分，本质是科学选择还是政治选择，任何行为都是有目的性的，没有无缘无故的爱。
49分钟前 赞2 回复

涸水之鱼：你确实不该被救，早点去死吧，孤儿
57分钟前 赞6 回复

真不好取名儿：除你以外，谁都该救活！
今天07:40 赞7 回复

用户5856924770：什么叫不该救的却依然活着，你这话太恶毒了
今天01:03 赞31 回复

麒麟兔 上海
恭喜！这百岁老人有福气！如此高寿！免疫力还这么强！！惊叹！
3月8日23:43 赞73 回复

云飞扬：前几天还有一个一百零一岁的，已经回家了。
33分钟前 赞2 回复

Infectious Disease > COVID-19

COVID-19 Triage: Who Lives, Who Dies, Who Decides?

— When push comes to shove, special committees could take emotional burden off individual clinicians

by Molly Walker, Associate Editor, MedPage Today March 23, 2020

Triage committees deciding which patients get ventilators in a potential ventilator shortage during the COVID-19 coronavirus outbreak in the U.S. may serve as a valuable buffer to protect individual clinicians from emotional damage from making the decision themselves, experts argued.

Committees made up of experts with no clinical responsibilities for patient care can ensure these decisions are based on which patients are most likely to benefit, and may help to spare clinicians from crippling emotional distress, argued Robert Truog, MD, of Harvard Medical School in Boston, and colleagues.

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

The Toughest Triage — Allocating Ventilators in a Pandemic

Robert D. Truog, M.D., Christine Mitchell, R.N., and George Q. Daley, M.D., Ph.D.

The Covid-19 pandemic has led to severe shortages of many essential goods and services, from hand sanitizers and N-95 masks to ICU beds and ventilators. Although rationing is

during which they can be saved. And when the machine is withdrawn from patients who are fully ventilator-dependent, they will usually die within minutes. Un-

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

Fair Allocation of Scarce Medical Resources in the Time of Covid-19

Ezekiel J. Emanuel, M.D., Ph.D., Govind Persad, J.D., Ph.D., Ross Upshur, M.D., Beatriz Thome, M.D., M.P.H., Ph.D., Michael Parker, Ph.D., Aaron Glickman, B.A., Cathy Zhang, B.A., Connor Boyle, B.A., Maxwell Smith, Ph.D., and James P. Phillips, M.D.

March 24, 2020 01:18 PM

Deciding who lives and who dies

Dr. Morhaf Al Achkar

TWEET SHARE SHARE EMAIL

Dr. Morhaf Al Achkar is an assistant professor of family medicine at the University of Washington in Seattle.

Modern Healthcare is providing COVID-19 coverage for free as a public service and a show of appreciation for the frontline workers in this battle against the epidemic. To support this essential journalism, please subscribe here.

I could soon be the physician following a policy that determines who would be denied medical care. At the same time, I could be one of those forbidden care if I needed it.

Medical leaders in Washington state quietly debated a plan to decide who gets care when hospitals fill up. Not many details are out, but the arguments echo a similar discussion in Italy, where an intensive-care unit protocol withheld life-saving care from certain people. The rejected were those older than 80 or who had a Charlson comorbidity index of 5 or more. With my diagnosis of stage IV lung cancer, I score a 6!

When I read the news, I was morally troubled, enraged and mortified.

I am in the same boat as many colleagues who have health issues or are older and could be asked to return from retirement or work accommodation to help out. Are we asking individuals to risk their lives, but will refuse them treatment if they get sick?

I am not familiar with empirical, objective evidence to support setting a threshold for who should or should not receive care as a way to improve outcomes for a community. Research to answer such an empirical question would have been unethical to start with. Using such a strategy also misuses predictive tools.

This is not the story we want to leave for history. And who said that an order from a health authority takes the moral burden off your shoulders? Have we forgiven the doctors in Nazi Germany who experimented with vulnerable patients? We humans carry moral responsibility for our actions. If anything, blindly following an unjust order doubles the burden. Worse than doing what is unjust is not standing up to advocate for the vulnerable. What will be remembered is that we pacified our consciences with a piece of paper we called a "policy."

If we think we cannot save everyone, let's invite people to have conversations about death and dying. Patients and their primary-care doctors should discuss advanced directives. The patient can sign a do not resuscitate order. People could even embrace death with dignity if they live in a state that allows it.

I can make the choice to not live and forfeit my right to care. But that right cannot be taken from me. Age or health conditions cannot alter a person's entitlement.

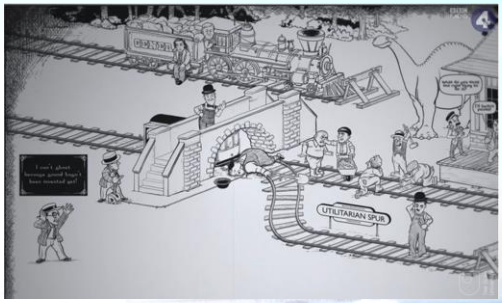
We can trust doctors' abilities to make the right moral decision, and we can give them the authority and support in so doing. In today's hyper-complex context, medical doctors should be competent to manage, case-by-case and situation-by-situation.

Yes, it will be a difficult time. When a decision has to be made between two lives, we regret having to make the decision, and we express our deep sadness. We should not make such unfortunate decisions a norm, and we should not write a policy to make it OK. It is not OK, and it will never be.

Why are these issues relevant to Machine Learning?

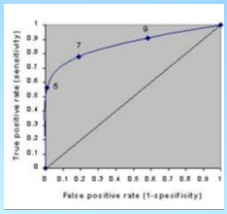
D

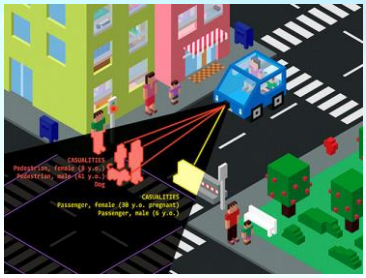
- Usually we do not meet such dilemma in everyday life,
- but we do need to code preferences/choices for such dilemmas in the designing of learning machines!

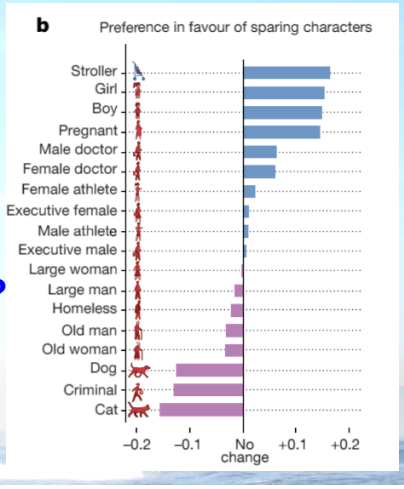


Relevance in ML?

- Are those only hypothetical dilemma?
- Are those only issues for decision-makers?







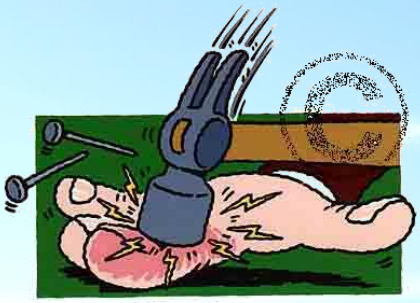
- **How do we design the objective functions?**

Xuegong Zhang

Final words

This is a broad field for great accomplishments.

-- Zedong Mao



If all you have is a hammer, everything looks like a nail.

-- Abraham Maslow, *The Psychology of Science*, 1966

Thanks to all TAs!

Qiuchen Meng（孟秋辰）

Jiaqi Li（李嘉骐）

Xi Xi（席熙）

Yixin Chen（陈奕鑫）

Minsheng Hao（郝敏升）

Welcome to my course

《科学规范与表达》

Scientific Ethics and Exposition

in the spring semester

(70250431, mixed teaching in Chinese & English)