
FIRE
FIRE

- “If there’s any general social panic it will be by coincidence, based on terrible reasoning, uncorrelated with real timelines except by total coincidence, set off by a Hollywood movie, and focused on relatively trivial dangers.” -Yudkowski

Representativeness: disease case, Linda case

Maybe this is the inference underlying most people's judgment:

"positive test result is very typical or representative of those who have the disease, positive test result not very typical of those who do not have the disease, I have a positive result, therefore I probably have the disease."

Representativeness: disease case, Linda case

Maybe this is the inference underlying most people's judgment about Linda:

- "A feminist bank teller is more typical or representative of Linda, than just a random bank teller, therefore Linda is more likely to be a feminist bank teller than simply a bank teller."
- Idea of a representativeness heuristic: judge how likely an x is an F by how typical or representative x is of F's

- **More words that start with “R”, or words with an “r” as their third letter?**

Availability heuristic

- Instead of "how do I calculate the frequency / probability of X," instead think about "how easily can I picture or recall instances of X?"
- Availability: Events that are already on our minds, easily remembered, or vividly imaginable, are judged to be more frequent or more probable.

Psychological heuristics

- Ways in which we answer questions quickly at the cost of accuracy.
- Availability: Events that are already on our minds, easily remembered, or vividly imaginable, are judged to be more frequent or more probable.
  - *"Hyposcenia-B"* study. More vivid symptoms of an imaginary disease judged to be more likely.
  - Sales for supplemental home owners insurance goes up after floods or earthquakes such get media spotlight .

- **Now, picture an AGI superintelligence catastrophe…**

Affect Heuristic

- Judgments we form about the benefits or risks of a technology are shaped by how much we like the technology rather than what we regard to be true of the technology.
- two groups, rate the benefits and risks of fluoridated water, food preservatives, chemical plants, etc. Show groups different persuasive message; one gets high benefits, other low-risks. Ask if they want to revise.
- High-benefits group lowers the risks, low cost group raises the benefits.

Affect and AI

- Being bombarded by persuasive messaging about benefits of AI skews our estimation of the risks.
- Being bombarded by persuasive messaging about the risks skews our estimation of the benefits.
- AI has already had large benefits and costs.

"Spooky" one: Anchoring effect

- Anchoring effect: occurs when people consider a particular value for an unknown quantity before estimating that quantity and as a result their estimate stays close to the number considered.
- Example:
  - "Is the hight of the tallest redwood more or less than 1,200 ft?"

"Spooky" one: Anchoring effect

- Anchoring effect: occurs when people consider a particular value for an unknown quantity before estimating that quantity and as a result their estimate stays close to the number considered.
- Example:
  - "Is the hight of the tallest redwood more or less than 1,200 ft?" AND
  - "what is your best guess about the height of the tallest redwood?" (another group given "…less than 180 ft?"

"Spooky" one: Anchoring effect

- Anchoring effect: occurs when people consider a particular value for an unknown quantity before estimating that quantity and as a result their estimate stays close to the number considered.
- Example:
  - "Is the hight of the tallest redwood more or less than 1,200 ft?" AND "what is your best guess about the height of the tallest redwood?" (another group given "…less than 180 ft?"
- Average estimate of first group: 844
- Average estimate of second group: 282.

More on anchoring

- 1974 experiment by Kahneman and Tversky"Wheel spinning" experiment.
  - But first they spun a wheel of fortune. The wheel was painted with numbers from 0 to 100, but rigged to always land on 10 or 65.
  - Q: "what percentage of African countries are part of the United Nations?"
  - When the arrow stopped spinning, they asked the person in the experiment to say if they believed the percentage of countries was higher or lower than the number on the wheel. Next, they asked people to estimate what they thought was the actual percentage.

More on anchoring

- Q: "what percentage of African countries are part of the United Nations?"
- When the arrow stopped spinning, they asked the person in the experiment to say if they believed the percentage of countries was higher or lower than the number on the wheel.
- Next, they asked people to estimate what they thought was the actual percentage.

More on anchoring

- They that found people who landed on 10 in the first half of the experiment guessed around 25 percent of Africa was part of the U.N.
- Those who landed on 65 said around 45 percent. The subjects had been locked in place by a psychological phenomenon known as the anchoring effect.

Lessons from anchoring

- Information that is visibly irrelevant still anchors judgments. Starting from what is irrelevant, people adjust until they reach a plausible-sounding answer, and as a result underadjust.
- Telling people to ignore anchors doesn't work.
- Subjects report that they believe the contaminating anchor having an effect, when it did.
- So, just tell AI researchers not to get anchored?

Lessons from anchoring

- *"50% chance that singularity happens at 2060."*
  - *N=995*
  - **https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/**

Most depressing and inspiring one

- The "2,4 6 task"
  - Watch this!: https://www.youtube.com/watch?v=vKA4w2O61Xo

“Common” sort of answer

Sequence	Fits the rule?	Guess the rule!	How sure are you?
2,4,6	☺	Counting up by two’s	40%
6,8,10	☺	Counting up by two’s	50%
20,22,24	☺	Counting up by two’s	60%
3,5,7	☺	Counting up by two’s	80%
25,27,29	☺	Counting up by two’s	90%
200,202,204	☺	Counting up by two’s	Sure enough

"Confirmation bias" at work!

- We tend to look for confirming instances of beliefs we already have. We naturally don't test our beliefs!
  - "I'll see it when I believe it!"
- We need to actively seek out reasons to doubt our beliefs, if we are interested in truth, otherwise we get "boxed in."

Most depressing and inspiring one

- "cold"vs "hot" contexts – 2,4,8 is non-emotive. What about politics? Lol.
- Gilovich (2000): it's not that people ignore contrary evidence – it's that they hold some hypotheses to a higher standard than conclusion a person wants to believe: "disconfirmation bias."
  - Two biased reasoners attending to the same stream of evidence will shift in opposite directions.
  - More skilled skeptics who apply skepticism selectively will change their minds more slowly.

- **Moral of the story?**
  - **Don't think about important stuff just by yourself.**
  - **…in a context in which you want to understand another perspective in the most plausible form imaginable.**
  - **…in a context in which you will not feel defensive.**
  - **…intellectual playfulness and flexibility is an undervalued virtue.**
  - **…in a non-echo chamber context.**
  - **You are not your beliefs. Beliefs should evolve. You do not die when they change.**

- Let's switch gears.
- Predicting the future is hard.
- Let's start looking at AI problems we already have…


STOP!
YOU CAN'T
POST
THAT!!!
My sense
of humor

- ***Fact or Myth*? corporations don't like government regulation.**
- **A little of both…**

- ***Fact or Myth*? corporations don't like regulation.**
  - **Regulation can in theory affect profit maximization…**
  - **Notice how much ZB squirms, and notice what is very *easy* for him to answer:**
  - **YT (Facebook CEO Mark Zuckerberg said he'd be open to regulation)**
    - **https://www.youtube.com/watch?v=9ZNqZxVt1g4**

- ***This is the "revolving door" problem with regulation*.**
- **Government needs industry experts to actually responsibly regulate.**
- **Those experts don't want to tank their careers when their stint for the government is over.**
- **You get industry friendly regulation**
- **Sometimes regulations are just written by industry! (so basically self-regulation)**

Take a look at some existing problems of AI…

- **We got a problem: racist policing and justice system, patrol black neighborhoods more, arrest and harass black people more.**
  - **Solution: let's be less biased, more data driven.**
- **PredPol tells where certain crimes will be more likely to occur, hour by hour;**
  - **based on seismic software, looks at historical patterns and predicts; Reading, PA used it in 2013, burglaries down by 23%**

Crime analytics

- **The feedback cycle:.**
  - **1. start with an ideal of preventing crime, patrol in impoverished areas.**
  - **2. patrolling those areas more, easier to report more crimes - victimless ones too – more data.**
  - **3. end up policing the poor more**
  - **4. return to step 2 and pat ourself on the back for being objective.**

Crime analytics

- **Other issues: growing trend of "open data" for crime**
  - **Justified for transparency and accountability**
  - **But can create another feedback loop**
    - **1. Crime -> open data…**
    - **2. Impacts insurance fees, devalues property**
    - **3. Less funding going back to education, less investment.**
    - **4. More crime -> more data**

Racism in sentencing

- **Idea: well let's at least be less biased in sentencing criminals.**
  - **Rely on human judgment? That'll be biased… use data.**
- **UMaryland study found that in Harris County TX (includes Houston), prosecuters were three times more likely to seek the death penalty for blacks, and four times more for Hispanics than whites.**

Racism in sentencing

- **According to ACLU, sentences imposed on black men in the federal system are nearly 20% longer than those for whites convicted of similar crimes.**
  - **Blacks 13% of population, but 40% of prison cells.**
- **So, this is a problem, need a more objective less biased system to aid in fair sentencing….**
  - **24 states have turned to computerized risk models, "recidivism models," to have more objectively assess the danger posed by each convict.**

Recidivism models

- **Aim is to keep sentences more consistent; Actual judges are swayed by things like hunger and mood, (and dice!)**
- **Goal: want likely non-repeat offenders to go up for parole, keep likely repeat offenders from reentering society.**

Recidivism models

- **But LSI-R, a popular model, involves data points like**
  - **Number of prior convictions**
  - **Part others played in offense**
  - **Drugs or alcohol a factor**
  - **First time convict was ever involved with the police**
  - **Whether friends or relatives have criminal records**
  - **From a high criminal record neighborhood**
  - **High school diploma?**
  - **DOESN'T ask about race, but still tracks things like race, class, etc.**

Recidivism model score are used in sentencing

- **The score is used in sentencing….**
- **…but imagine a lawyer arguing in sentencing based on…**
  - **circumstances of their birth or upbringing**
  - **Lack of high school diploma**
  - **defendant's brother's convictions**
  - **High rate of crime in their home neighborhood.**
- **We should be judging by actions, not who we are.**
- **Feedback loop: bad neighborhood -> "high risk" -> longer sentence -> years surrounded by criminals -> raises likelihood of recidivism.**

Allegheny Family Screening Tool

- The AI program used for child protection services.
  - **YT ("Building the Allegheny Family Screening Tool - (extended version)":**
  - **https://www.youtube.com/watch?v=A48e-p_3_Xs**

Allegheny Family Screening Tool

- Think about the fact that this algorithm won't be perfect:
- Problem of false positives and false negatives:
  - False negative: the algorithm fails to flag a case for further review
    - Think about incentives: CPS wants to avoid this…
  - False positive: unwarranted attention of CPS, can lead to:
    - More mishandled cases: children separated from families
    - Perverse incentives: families now rational to avoid factors that raise their score on the algorithm (unemployment claims, food stamps, etc).

Jobs!

With the rise of self driving vehicles, it's only a matter of time until there's a country song where the guy's truck leaves him

Jobs!

The Future of Work in the Age of AI: Displacement or Risk-Shifting?

Oxford Handbook of Ethics of AI, pp. 271-87 (Markus Dubber, Frank Pasquale, and Sunit Das, eds.)

20 Pages • Posted: 7 Aug 2020

Pegah Moradi
Cornell University; Cornell University - Department of Information Science

Karen Levy
Cornell University

Date Written: July 9, 2020

Abstract

This chapter examines the effects of artificial intelligence (AI) on work and workers. As AI-driven technologies are increasingly integrated into workplaces and labor processes, many have expressed worry about the widespread displacement of human workers. The chapter presents a more nuanced view of the common rhetoric that robots will take over people's jobs. We contend that economic forecasts of massive AI-induced job loss are of limited practical utility, as they tend to focus solely on technical aspects of task execution, while neglecting broader

- Lots of focus on job loss and displacement.
- But how will AI affect those who don’t lose their jobs?
- Whose interest are AI serving?

Staffing and Scheduling

- **AI data analysis predicts dynamically need to labor:**
  - “just-in-time” scheduling; split-shifts; fluctuating work schedules
  - Pretty destabilizing for workers – a lot of people need to work two jobs. Let alone raise kids, etc.

Compensable

- Fair Labor Standards Act: employers must pay workers for time worked, but only for those activities that are considered "integral and indispensable"
  - Courts ruled against things like commuting, waiting for screenings, donning protective gear and uniforms, etc.
  - Uber: not time spend driving or waiting for a pick-up, nor cleaning car, nor returning from a long trip (basically if it doesn't directly generate revenue).
  - AI helps more narrowly measure what counts as directly generating revenue.
  - Amazon and Spotify: pay artists per page read, pay per track…

Detecting and Predicting Fraud

- Analyze worker behavior and activity, scan emails, etc.

Incentivizing and Evaluating Productivity

- Amazon's inactivity reports; wristbands that track movement.
  - Everything from flagging to termination of employment is AI-driven.

Big picture

- Not just automation.
- Perfecting existing managerial techniques…
- Question: how can we fit AI and work together for the purpose of work: promoting our needs and desires?

Oppositional vs Systemic Approaches

- **Oppositional stance to ethics of AI**
  - **“AI left unchecked may do bad things”**
  - **Focusing on AI and its properties as separate entities (are they good or bad)**
- **Systemic stance to ethics of AI**
  - **Idea: AI will be part of a socio-technical system of the world. We can think about how to change any parts of that system to make things go well.**

- **Good example:**
  - **Oppositional: “AI will take our jobs!” Booo!**
  - **Systemic: what is the point of a jobs, and how should ‘work’ or ‘job’ evolve in an AI-world?**
- **An idea: these aren’t separate / exclusive: they complement each other!**

- One idea: AI and automation are used to offload risk from the owners to the workers.
  - So, just fix the problem: don't have owners!
  - Plenty of "worker cooperatives" (WCs) in the US, but there could be way more
    - All employers are also the owners.
    - AI and automation can free them up from routinized to do other work, planning, creative research and development, outreach, etc.

unions: AI and automation will eliminate jobs, no!!

WCs:

Oppositional vs Systemic thinking

- **Where transparency or knowledge is bad, worry about commodification and perverse incentive structures:**
  - **Crime -> open data -> insurance fees, devalues property, ->less investment in real estate and education -> More crime -> more data**
    - **Problem: people know too much! Better keep secrets! (NOO!)**
    - **Maybe this: decommodify social risk reduction (insurance), housing, education…**
    - **CNBC article: "Goldman Sachs asks in biotech research report: 'Is curing patients a sustainable business model "**
  - **Safety at home (CPS); FB news feed (facts / information);**

A heuristic for systemic diagnosis

•For making judgments about authority (is it the right kind?):

- "how would we expect_____to behave if it had perfect knowledge":
- justice system?
- FB?
- Big pharma?

Quote for the day

- "If you don't want a more intensively data-driven society, that might say more about your society than your attitude toward data." -me, 2021 ☹

notice potential costs of AGI

work on taking precautions

realize you can think structurally

realize you can think structurally

realize problem isn't even AGI or AI

fix human systems regardless of AI