Computer Science Ethics

15-110 – Monday 04/24

Announcements

- **Check6-2 revisions** due **Wednesday**!

Learning Goals

- Understand the current extent of **data collection** on the internet and how data is used

- Recognize the uses and drawbacks of **facial recognition** algorithms in different contexts

- Identify the societal impact when **automated decision making** replaces human decision making due to the explainability problem and job displacement

Ethics in Computer Science

When we move from theoretical concepts of computer science to applying those theories in real life, the decisions we make have consequences.

The professional field of computer science has only recently adopted a code of ethics, and the code is not yet uniformly taught to new computer scientists or programmers. There is still much to debate over what the responsibilities of computer scientists are.

We'll discuss three areas where people debate how computing should be used in the current time: data collection, facial recognition, and automated decision-making.

Data Collection

User Data

Most applications collect data about users from various sources. We'll discuss three main categories: data provided by the **user**, data provided by the **browser/system**, and data provided by **other sources**.

As a user of the internet and various applications, you already voluntarily share a lot of data with the world!

- Internet – profile information, tweets, searches
- Applications – preferences, locations, images
- Real life – purchase history, contact info, location

Browser/System Data

Behind the scenes, your browser or phone/computer is sending additional information to the services you use.

This is not done maliciously – services can put this information to good use. However, you may be surprised by some of the data being shared.

Check out the data your browser shares here:
https://webkay.robinlinus.com/

There are plugins you can install that limit the information your browser sends, but this may also limit functionality of websites.

Other Data Sources

Cookies are used by websites to store temporary information about people using their services (like which items you've put in a shopping cart). A cookie is a small packet of data that is sent back and forth between the website and your browser.

Cookies that are shared between two or more websites are called **tracking cookies**, or just trackers. These cookies attempt to collect a portfolio of information on you, the user, by gathering information on the websites you visit. This is commonly done through ads that are placed on websites.

With enough data collected from tracking cookies and the browser, a website may be able to create a **fingerprint** that identifies you as a user. Read more here.

You can check what kinds of trackers your browser stops and what your fingerprint looks like here: https://coveryourtracks.eff.org/

Data Economy – Data Collection

Why are so many companies interested in data collection? **Data has become the economy of the internet**. Most websites are supported by advertising, and advertisers pay more for targeted ads.

Websites have a strong incentive to get the best data possible on their users, so they get paid more for advertisements. This has led to **hyper-targeting** in ads, with ads attempting to reach more niche populations.

Check out the categories that you're hyper-targeted as belonging to:

- Facebook: Settings > Ads > Ad Settings > Categories Used to Reach You
- Instagram: Settings > Privacy and Security > Account Data > Ads
- Twitter: Settings and privacy > Privacy and safety > Ads Preferences > Interests

Discuss: what kinds of data are you comfortable having collected by companies? Where might you draw a line?

Restrictions on Data Sharing

In the EU, the **GDPR** (General Data Protection Regulation) gives all users certain rights over their data; they must be told when data is being collected, data must be stored securely, and users have the right to obtain their data and/or ask for it to be deleted.

In the US, data collected about minors and students are protected through **COPPA** (Children's Online Privacy Protection Act) and **FERPA** (Federal Educational Rights and Privacy Act). There is no general law in the US about data privacy yet, but California passed the **CCPA** (California Consumer Privacy Act), which institutes some regulations for that state.

Protecting Your Data

If you want to protect your data online, you have a lot of options! Most browsers let you block cookies and can request that websites do not track you. You can also restrict permissions given to websites and applications on your devices.

For advanced protection, you can also use a VPN (Virtual Private Network) to connect to the internet. CMU has a VPN (though then CMU will know which websites you're accessing):

https://www.cmu.edu/computing/services/endpoint/network-access/vpn/how-to/

Facial Recognition

Image Recognition on Faces

In the Machine Learning lecture, we briefly discussed how machine learning algorithms can be applied to images, to recognize objects that occur in the images.

This can also be applied to people. **Facial Recognition** is used to automatically match the face of a person in a photo to their identity. (Further algorithms can even identify expressions!)

Try it out here to see an underlying model:

https://skybiometry.com/demo/face-detect/

Drawbacks of Facial Recognition

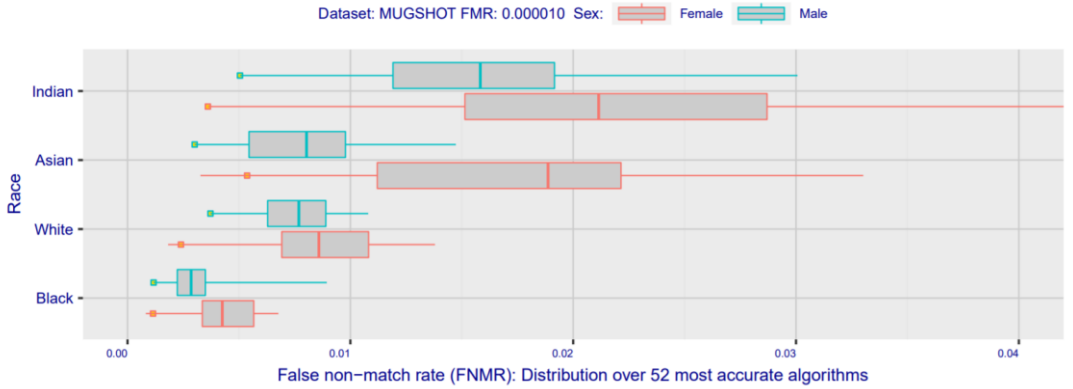
Facial recognition has been around for a while and is used for a variety of purposes (automatic tagging, ID verification, etc.).

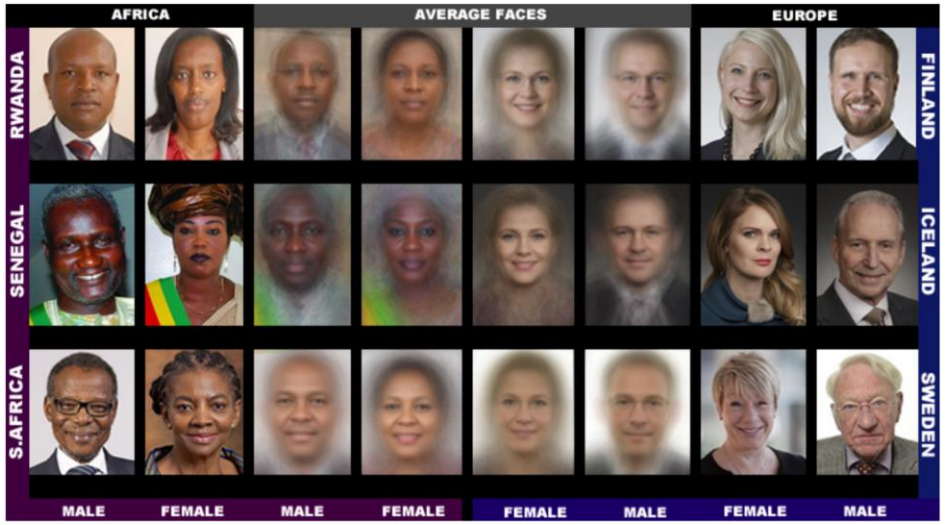
However, there are some concerns about **bias** in facial recognition systems impacting results, and how **privacy** might be impacted by widespread use of the algorithms.

Bias in Facial Recognition Algorithms

Recent studies that test facial recognition algorithms have shown huge variation in performance. One found that many facial recognition algorithms are "10 to 100 times more likely to inaccurately identify a photograph of a black or East Asian face, compared with a white one". Even among the best algorithms there are notable differences in recognition performance across race and gender.

One factor that could lead to this difference is **bias in the data used to train the algorithms**. An analysis showed that two popular training sets were overwhelmingly composed of lighter-skinned subjects. This is supported by the above study, which showed that algorithms developed in Asian countries performed better on Asian faces.





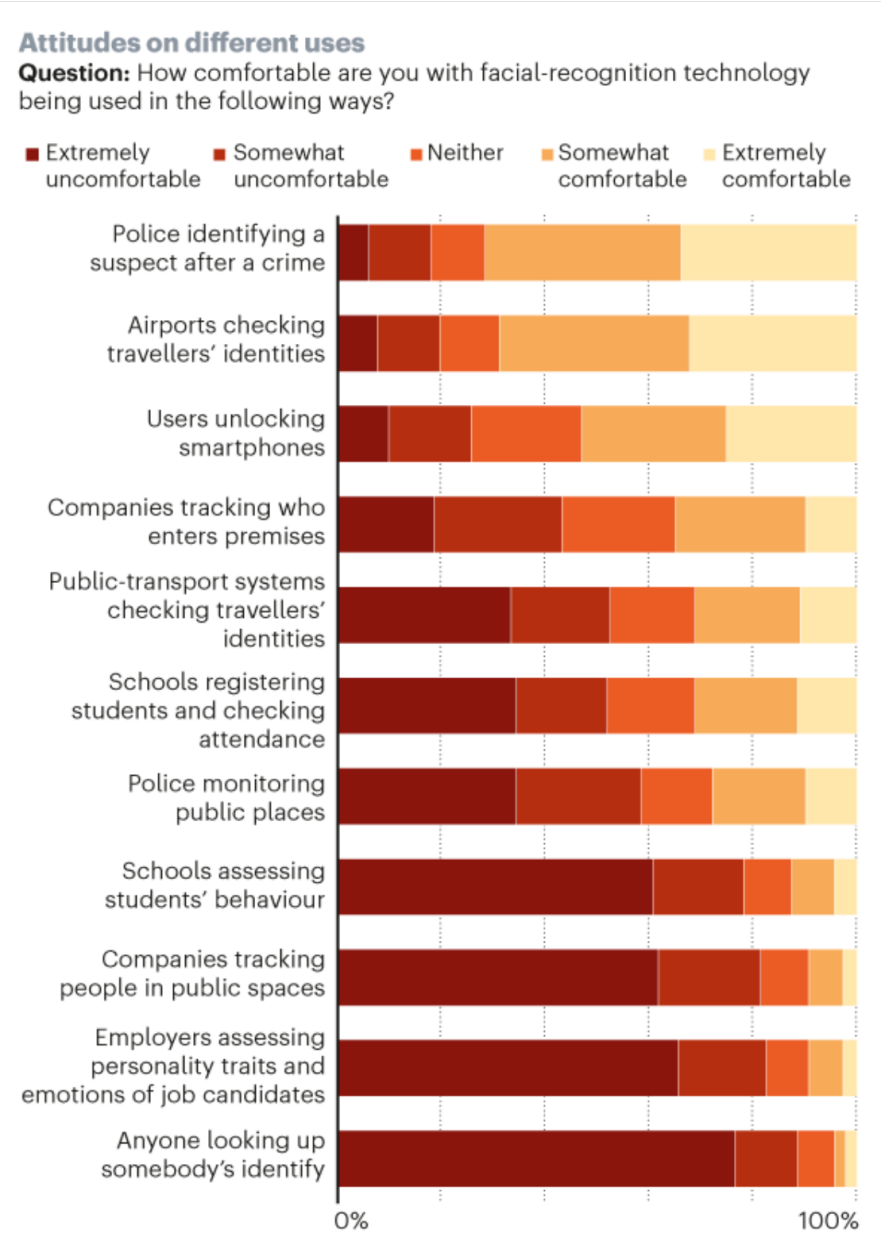
Controversial Uses

There are also concerns about some commercial uses of facial recognition even when it works well.

For example- is it okay for police to use facial recognition to identify a suspect? What about using it to monitor people in a public space? Where do we draw the line between **private** vs **public** scenarios?

This especially causes problems when a controversial use of facial recognition collides with a weakness in the algorithm. For example, an algorithm led to an innocent man being arrested for shoplifting.

Discuss: when do you think it's acceptable to use facial recognition? Should there be restrictions on its use?

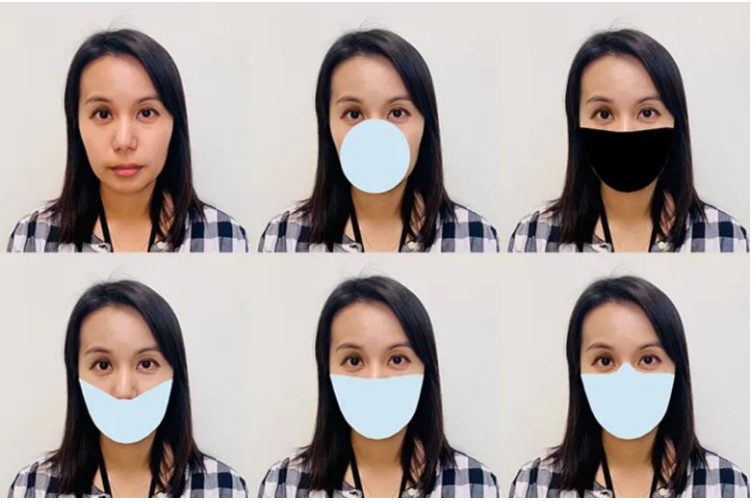


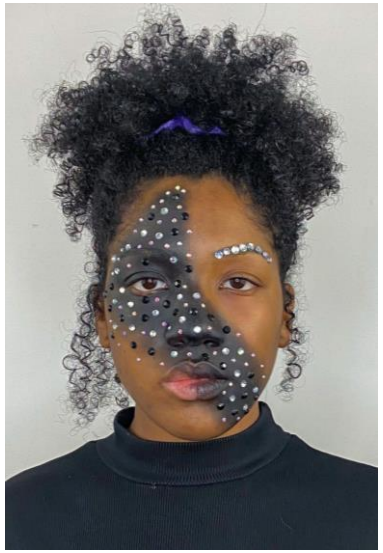
Legislation of Facial Recognition

Some communities have gone to the length of banning facial recognition technologies from being used in certain contexts.

Several US cities (including San Francisco and Boston) have recently banned the use of facial recognition by local governments.

Beyond legislation, algorithms can be foiled by obfuscating part of the face. This can be done with masks or even makeup!





Automated Decision Making

Automation Potential

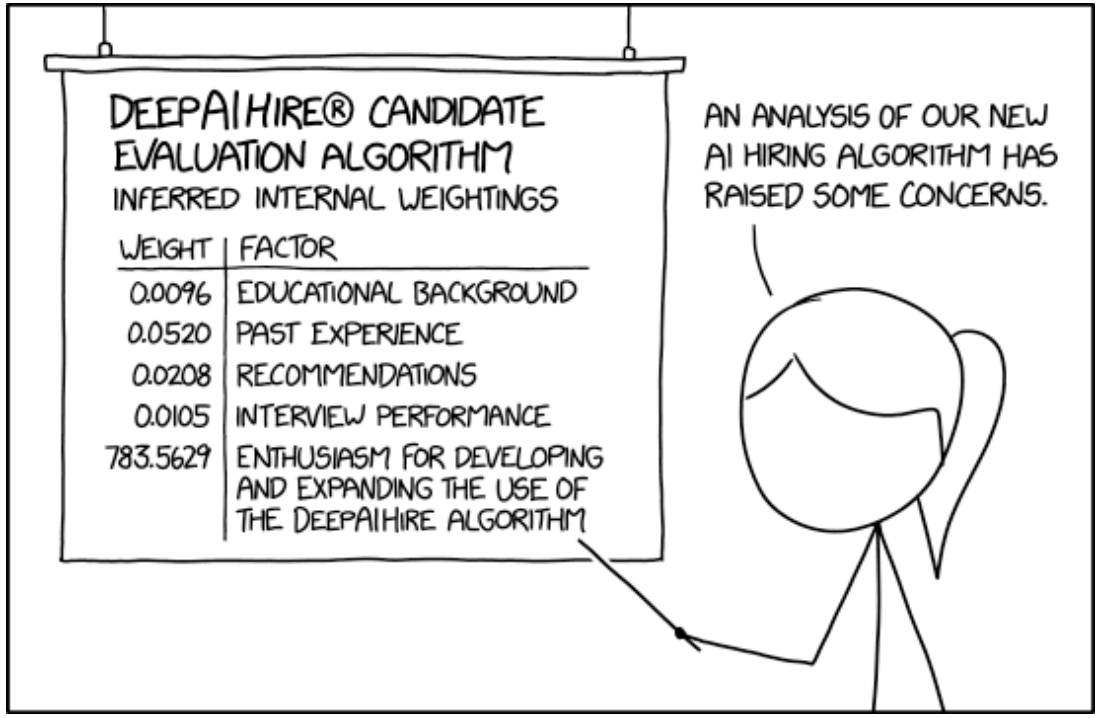
There are potentially enormous benefits to be gained by using machine learning and artificial intelligence to accomplish tasks and solve problems.

However, we must keep in mind that there are potential downsides to these algorithms as well. In particular, let's consider the problem of **explainability**: when an algorithm makes a decision, can it explain why? We'll also discuss how automation can lead to **job displacement**.

Explainability

Decisions made by machine learning algorithms are usually based on a huge number of tiny features. In some algorithms (like neural networks) those features aren't named in a human-readable way.

This is a problem when the algorithm makes an important decision about a person's life, like whether they should be admitted to a school or hired for a job.

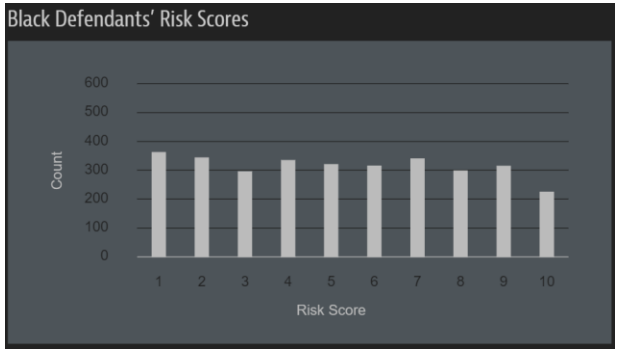


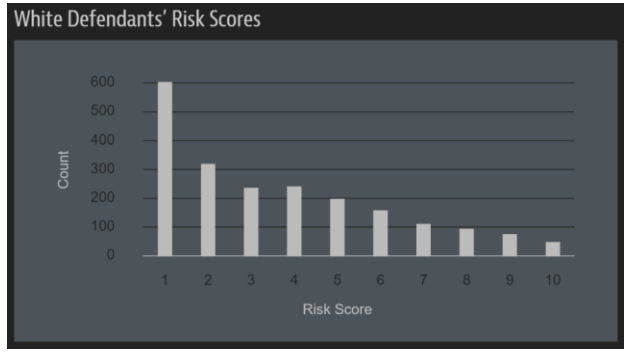
Bias in Machine Learning

Just like in facial recognition, bias in the data fed into a machine learning algorithm can lead to bias in the algorithm's results.

This has caused problems in algorithms for determining bail, which have shown systematic racial bias in predicting a person's likelihood to commit future crimes. This could be due to historical racial bias in bail decisions.

A similar problem was observed in an algorithm to hire engineers for Amazon, which showed bias towards hiring employees based on gender. Here the bias could be caused from the algorithm being trained on currently employee resumes, where most of the current employees are male. Similar problems have been observed in other hiring algorithms too.





Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

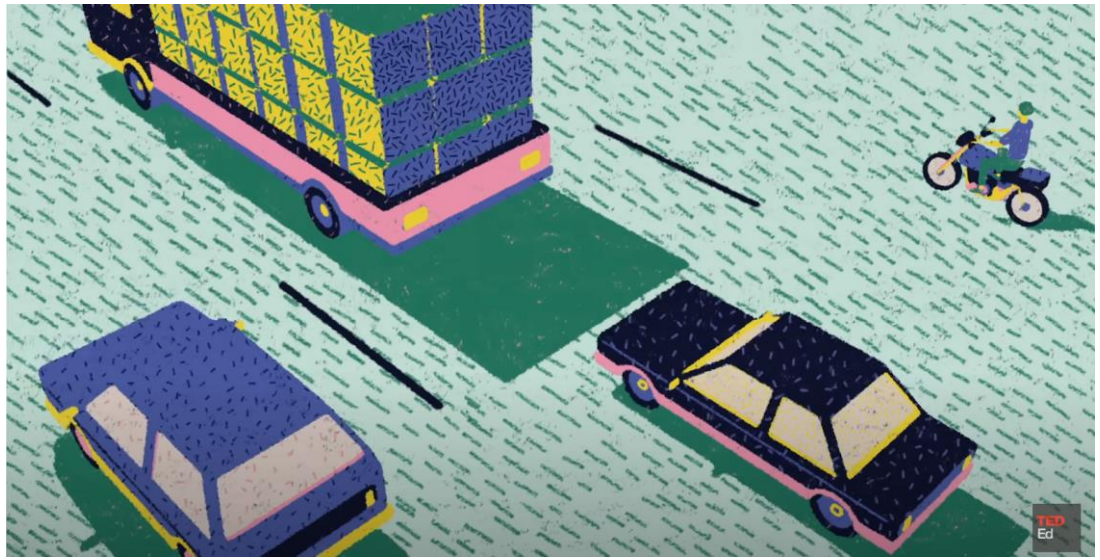
Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Ethics in AI Design

Even if we set aside the problems related to bias in data (which obviously affect human decision making as well), there are still big ethical questions about how we should use AIs.

Consider decisions that are made by self-driving cars. If a car is put in a position where it will inevitably get into an accident, should the car protect its passenger, or should it optimize for the greatest preservation of human life? And how should this be treated legally?

Discuss: when should algorithm creators be held responsible for the unintended outcomes of their algorithms?



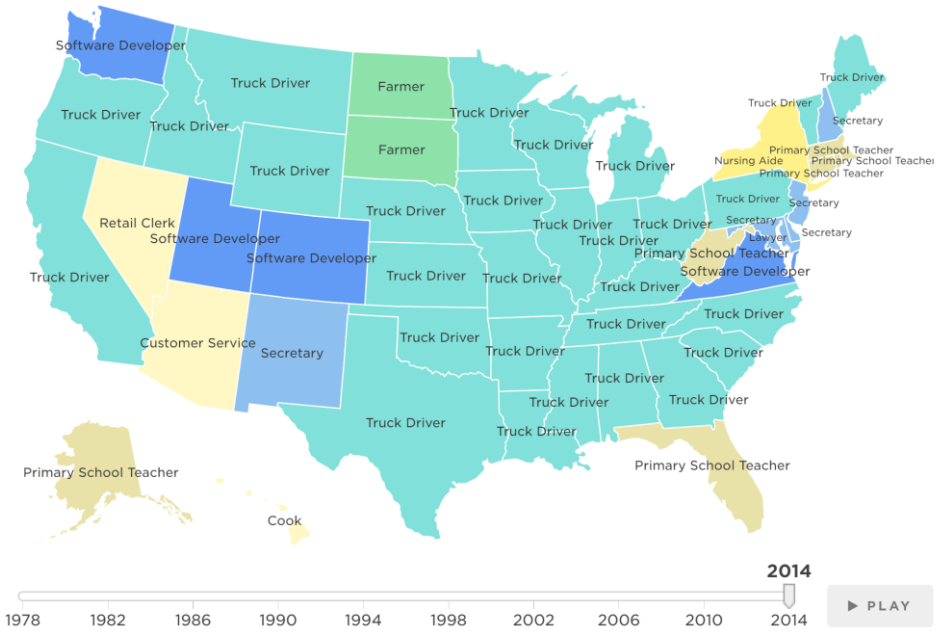
Job Displacement

Even if we set aside the question of explainability and responsibility in automated decision making, there are still practical consequences to consider. Automation will certainly affect the world economy – it's affecting it already.

Automation isn't a new process. Since the Industrial Revolution, humans have been finding new methods to solve problems that require less human labor. This means that the skills expected of workers are constantly changing.

Check out the most common jobs in the United States over the past forty years here:

www.npr.org/sections/money/2015/02/05/382664837/map-the-most-common-job-in-every-state



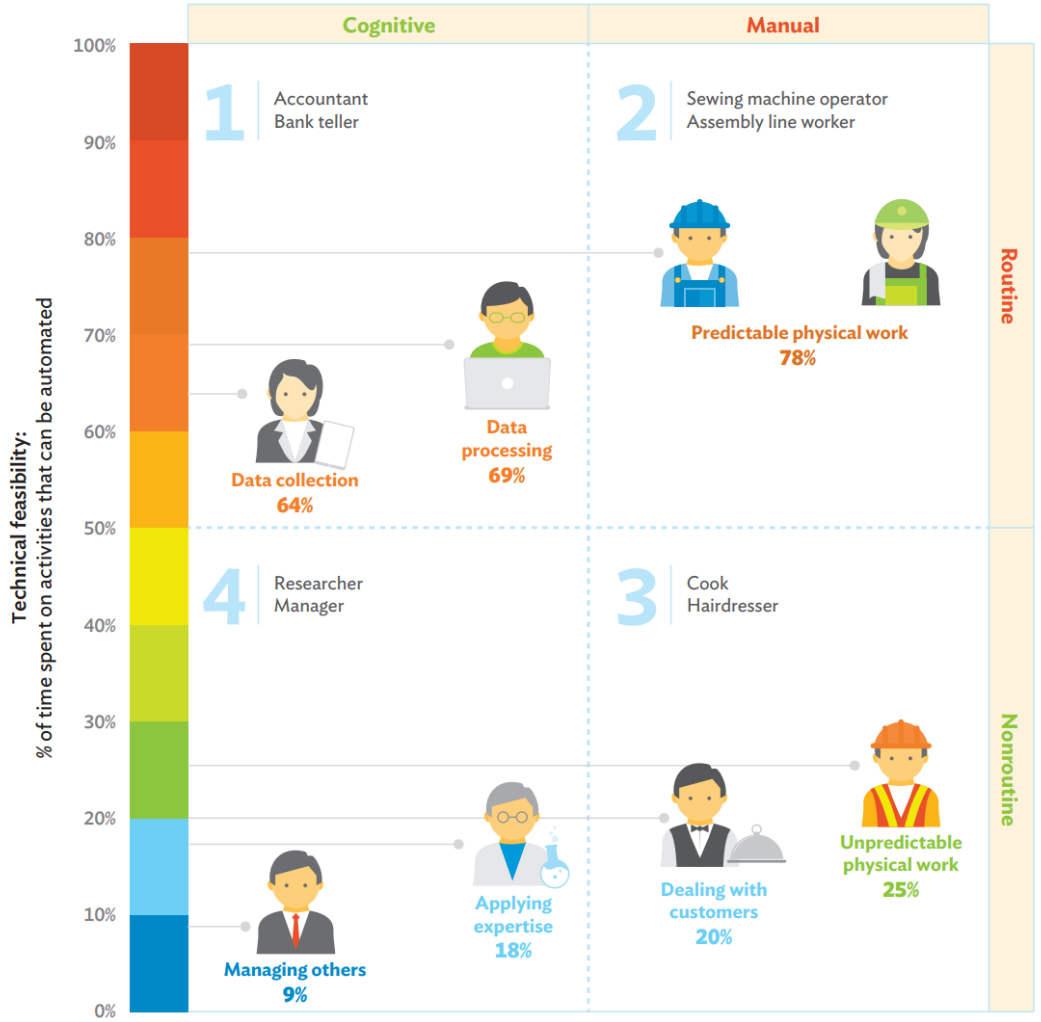
Jobs at Risk

More recently, the success of artificial intelligence has led to concerns that common jobs we previously thought couldn't be automated will be replaced.

Specifically, jobs that are **routine** are more likely to be replaced by algorithms. This includes certain expected jobs (like manufacturing and data processing), but also more surprising jobs, like truck driving.

What will happen to people who are displaced? Will they be able to find a new line of work, or will governments need to options like job training or to universal income? These policy questions need to be considered soon.

2.1.9 Impact of automation on jobs



Note: Percentages are from Frey and Osborne (2017) estimates on probability of automation. Framework is based on Acemoglu and Autor (2011).

AI + Human Collaboration

There is a great deal of potential in using automation to **support** people in their work, instead of replacing workers outright. This is possible because AI agents are excellent at solving specific tasks, while humans are good at generalizing.

This is already done in several industries – for example, AI-assisted grading for professors. In the future, AI will be able to provide even more support. This is predicted to drastically impact the fields of healthcare, law, and accounting.

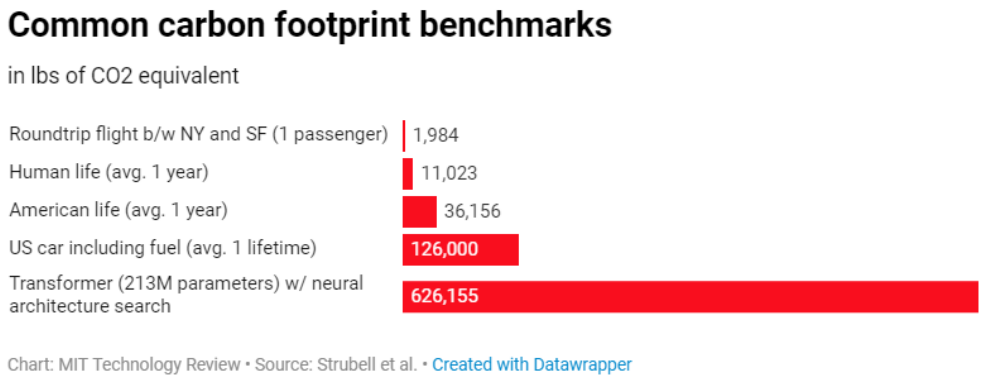
Sidebar: Effects on the Environment

Even when we make productive and unbiased algorithms, they can still have unintended side effects.

Many companies and researchers train machine learning algorithms on very large datasets to answer questions. This analysis does not come without a cost.

An enormous amount of energy is needed to run these algorithms, and in the US, that energy often has a carbon footprint. A recent study found that training a popular NLP model, The Transformer, left a gigantic carbon footprint.

On the bright side, some tech companies have pledged to go carbon negative to combat this. Other scientists are exploring new ways to make algorithms more energy efficient.



Next Time: CS Future

What do you most want to learn about in the CS Future lecture?

Go to the new Piazza post to vote on the topics you're most interested in. You can add new topics too!

We'll cover the top 4-5 topics on Friday.

Link: https://piazza.com/class/lcp556nkons13w/post/260

Learning Goals

- Understand the current extent of **data collection** on the internet and how data is used

- Recognize the uses and drawbacks of **facial recognition** algorithms in different contexts

- Identify the societal impact when **automated decision making** replaces human decision making due to the explainability problem and job displacement