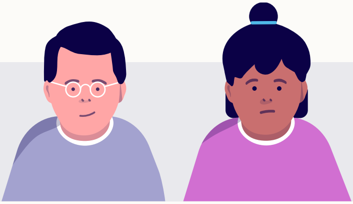
II. The societal implications of AI

In the very beginning of this course, we briefly discussed the importance of AI in today's and tomorrow's society but at that time, we could do so only to a limited extent because we hadn't introduced enough of the technical concepts and methods to ground the discussion on concrete terms.

Now that we have a better understanding of the basic concepts of AI, we are in a much better position to take part in rational discussion about the implications of already the current AI.



Implication 1: Algorithmic bias

AI, and in particular, machine learning, is being used to make important decisions in many sectors. This brings up the concept of algorithmic bias. What it means is the embedding of a tendency to discriminate according ethnicity, gender, or other factors when making decisions about job applications, bank loans, and so on.

Note

Once again, it's all about the data

The main reason for algorithmic bias is human bias in the data. For example, when a job application filtering tool is trained on decisions made by humans, the machine learning algorithm may learn to discriminate against women or individuals with a certain ethnic background. Notice that this may happen even if ethnicity or gender are excluded from the data since the algorithm will be able to exploit the information in the applicant's name or address.

Algorithmic bias isn't a hypothetical threat conceived by academic researchers. It's a real phenomenon that is already affecting people today.

Online advertising

It has been noticed that online advertisers like Google tend to display ads of lower-pay jobs to women users compared to men. Likewise, doing a search with a name that sounds African American may produce an ad for a tool for accessing criminal records, which is less likely to happen otherwise.

Social networks

Since social networks are basing their content recommendations essentially on other users' clicks, they can easily lead to magnifying existing biases even if they are very minor to start with. For example, it was observed that when searching for professionals with female first names, LinkedIn would ask the user whether they actually meant a similar male name: searching for Andrea would result in the system asking "did you mean Andrew"? If people occasionally click Andrew's profile, perhaps just out of curiosity, the system will boost Andrew even more in subsequent searches.

There are numerous other examples we could mention, and you have probably seen news stories about them. The main difficulty in the use of AI and machine learning instead of rule-based systems is their lack of transparency. Partially this is a consequence of the algorithms and the data being trade secrets that the companies are unlikely to open up for public scrutiny. And even if they did this, it may often be hard to identify the part of the algorithm or the elements of the data that lead to discriminating decisions.

Note

Transparency through regulation?

A major step towards transparency is the European General Data Protection Regulation (GDPR). It requires that all companies that either reside within the European Union or that have European customers must:

- Upon request, reveal what data they have collected about any individual (right of access)
- Delete any such data that is not required to keep with other obligations when requested to do so (right to be forgotten)
- Provide an explanation of the data processing carried out on the customer's data (right to explanation)

The last point means, in other words, that companies such as Facebook and Google, at least when providing services to European users, must explain their algorithmic decision making processes. It is, however, still unclear what exactly counts as an explanation. Does for example a decision reached by using the nearest neighbor classifier (Chapter 4) count as an explainable decision, or would the coefficients of a logistic regression classifier be better? How about deep neural networks that easily involve millions of parameters trained using terabytes of data? The discussion about the technical implementation about the explainability of decisions based on machine learning is currently intensive. In any case, the GDPR has potential to improve the transparency of AI technologies.

Implication 2: Seeing is believing — or is it?

We are used to believing what we see. When we see a leader on the TV stating that their country will engage in a trade-war with another country, or when a well-known company spokesperson announces an important business decision, we tend to trust them better than just reading about the statement second-hand from the news written by someone else.

Similarly, when we see photo evidence from a crime scene or from a demonstration of a new tech gadget, we put more weight on the evidence than on written report explaining how things look.

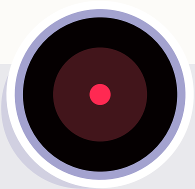
Of course, we are aware of the possibility of fabricating fake evidence. People can be put in places they never visited, with people they never met, by photoshopping. It is also possible to change the way things look by simply adjusting lighting or pulling one's stomach in in cheap before–after shots advertising the latest diet pill.

Note

AI is taking the possibilities of fabricating evidence to a whole new level:

Face2Face is a system capable of identifying the facial expressions of a person and putting them on another person's face in a Youtube video.

Lyrebird is a tool for automatic imitation of a person's voice from a few minutes of sample recording. While the generated audio still has a notable robotic tone, it makes a pretty good impression.



Implication 3: Changing notions of privacy

It has been long known that technology companies collect a lot of information about their users. Earlier it was mainly grocery stores and other retailers that collected buying data by giving their customers loyalty cards that enable the store to associate purchases to individual customers.

Note

Unprecedented data accuracy

The accuracy of the data that tech companies such as Facebook, Google, Amazon and many others is way beyond the purchase data collected by conventional stores: in principle, it is possible to record every click, every page scroll, and the time you spend viewing any content. Websites can even access your browsing history, so that unless you use the incognito mode (or the like) after browsing for flights to Barcelona on one site, you will likely get advertisements for hotels in Barcelona.

However, as such the above kind of data logging is not yet AI. The use of AI leads new kinds of threats to our privacy, which may be harder to avoid even if you are careful about revealing your identity.

Using data analysis to identify individuals

A good example of a hard-to-avoid issue is **de-anonymization**, breaking the anonymity of data that we may have thought to be safe. The basic problem is that when we report the results of an analysis, the results may be so specific that they make it possible to learn something about individual users whose data is included in the analysis. A classic example is asking for the average salary of people born in the given year and having a specific zip code. In many cases, this could be a very small group of people, often only one person, so you'd be potentially giving data about a single person's salary.

An interesting example of a more subtle issue was pointed out by researchers at the University of Texas at Austin. They studied a public dataset made available by Netflix containing 10 million movie ratings by some 500,000 anonymous users, and showed that many of the Netflix users can actually be linked to user accounts on the Internet Movie Database because they had rated several movies on both applications. Thus the researchers were able to de-anonymize the Netflix data. While you may not think it's big deal whether someone else knows how you rated the latest *Star Wars* movie, some movies may reveal aspects of our lives (such as politics or sexuality) which we should be entitled to keep private.

Other methods of identification

A similar approach could in principle be used to match user accounts in almost any service that collects detailed data about user behaviors. Another example is typing patterns. Researchers at the University of Helsinki have demonstrated that users can be identified based on their typing patterns: the short intervals between specific keystrokes when typing text. This can mean that if someone has access to data on your typing pattern (maybe you have used their website and registered by entering your name), they can identify you the next time you use their service even if you'd refuse to identify yourself explicitly. They can also sell this information to whoever wants to buy it.

While many of the above examples have come as at least in part as surprises – otherwise they could have been avoided – there is a lot of ongoing research trying to address them. In particular, an area called differential privacy aims to develop machine learning algorithms that can guarantee that the results are sufficiently coarse to prevent reverse engineering specific data points that went into them.

Implication 4: Changing work

When an early human learned to use a sharp rock to crack open bones of dead animals to access a new source of nutrition, time and energy was released for other purposes such as fighting, finding a mate, and making more inventions. The invention of the steam engine in the 1700s tapped into an easily portable form of machine power that greatly improved the efficiency of factories as well as ships and trains. Automation has always been a path to efficiency: getting more with less. Especially since the mid 20th century, technological development has lead to a period of unprecedented progress in automation. AI is a continuation of this progress.

Each step towards better automation changes the working life. With a sharp rock, there was less need for hunting and gathering food; with the steam engine, there was less need for horses and horsemen; with the computer, there is less need for typists, manual accounting, and many other data processing (and apparently more need for watching cat videos). With AI and robotics, there is even less need for many kinds of dull, repetitive work.

Note

A history of finding new things to do

In the past, every time one kind of work has been automated, people have found new kinds to replace it. The new kinds of work are less repetitive and routine, and more variable and creative. The issue with the current rate of advance of AI and other technologies is that during the career of an individual, the change in the working life might be greater than ever before. It is conceivable that some jobs such as driving a truck or a taxi, may disappear within a few years' time span. Such an abrupt change could lead to mass unemployment as people don't have time to train themselves for other kinds of work.

The most important preventive action to avoid huge societal issues such as this is to help young people obtain a wide-ranging education. This that provides a basis for pursuing many different jobs and which isn't in high risk of becoming obsolete in the near future.

It is equally important to support life-long learning and learning at work, because there are going to be few of us who will do the same job throughout their entire career. Cutting the hours per week would help offer work for more people, but the laws of economics tend to push people to work more rather than less unless public policy regulating the amount of work is introduced.

Because we can't predict the future of AI, predicting the rate and extent of this development is extremely hard. There have been some estimates about the extent of job automation, ranging up to 47% of US jobs being at risk reported by researchers at the University of Oxford. The exact numbers such as these – 47%, not 45% or 49% –, the complicated-sounding study designs used to get them, and the top universities that report them tend to make the estimates sounds very reliable and precise (recall the point about estimating life expectancy using a linear model based on a limited amount of data). The illusion of accuracy to one percentage is a fallacy. The above number, for example, is based on looking at a large number of job descriptions – perhaps licking the tip of your finger and putting it up to feel the wind – and using subjective grounds to decide which tasks are likely to be automated. It is understandable that people don't take the trouble to read a 79 page report that includes statements such as "the task model assumes for tractability an aggregate, constant-returns to-scale, Cobb-Douglas production function." However, if you don't, then you should remain somewhat sceptical about the conclusions too. The real value in this kind of analysis is that it suggests which kinds of jobs are more likely to be at risk, not in the actual numbers such as 47%. The tragedy is that the headlines reporting that "nearly half of US jobs at risk of computerization" are remembered and the rest is not.

So what are then the tasks that are more likely to be automated. There are some clear signs concerning this that we can already observe:

- Autonomous robotics solutions such as self-driving vehicles, including cars, drones and boats or ferries, are just at the verge of major commercial applications. The safety of autonomous cars is hard to estimate, but the statistics suggests that it is probably not yet quite at the required level (the level of an average human driver). However, the progress has been incredibly fast and it is accelerating due to the increasing amount of available data.
- Customer-service applications such as helpdesks can be automated in a very cost-effective fashion. Currently the quality of service is not always to be cheered, the bottle-necks being language processing (the system not being able to recognize spoken language or to parse the grammar) and the logic and reasoning required to provide the actual service. However, working applications in constrained domains (such as making restaurant or haircut reservations) sprout up constantly.

For one thing, it is hard to tell how soon we’ll have safe and reliable self-driving cars and other solutions that can replace human work. In addition to this, we mustn’t forget that a truck or taxi driver doesn’t only turn a wheel: they are also responsible for making sure the vehicle operates correctly, they handle the goods and negotiate with customers, they guarantee the safety of their cargo and passengers, and take care of a multitude of other tasks that may be much harder to automate than the actual driving.

As with earlier technological advances, there will also be new work that is created because of AI. It is likely that in the future, a larger fraction of the workforce will focus on research and development, and tasks that require creativity and human-to-human interaction. If you'd like to read more on this topic, see for example Abhinav Suri's nice essay on *Artificial Intelligence and the Rise of Economic Inequality*.

Unanswered

Exercise 24: Implications of AI

What kind of articles (in newspapers and magazines or other popular science outlets such as blogs, ...) are being written about AI - and do you think they are realistic? Do an online search about AI related to one of your interests. **Choose one of the articles and analyze it.**

1. Mention the **title of the article** along with its author and where it was published (as a URL if applicable) in your answer.
2. Explain the central idea in the article **in your own words** using about a paragraph of text (multiple sentences.)
3. Based on your understanding, how accurate are the AI-related statements in the article? **Explain your answer.** Are the implications (if any) realistic? **Explain why or why not.**

Your answer

Words: 0

Submit

Next section

III. Summary →