AI and Ethics

Lecture 7 AI Respect and Human Rights

Human Rights and AI

- What are human rights, and how do they tie into the current ethical guidelines and principles of AI?
- The three rights of particular importance to AI:
  - the right to privacy,
  - security, and
  - inclusion

Human Rights

- During the COVID-19 pandemic, governments have struggled to find effective policy-making strategies for exiting lockdown in a safe way.
- According to epidemiologists, opening up society requires efficient tracking, tracing, and monitoring.
- In many cases, this has led to the utilization of various tracing and tracking apps.
- These apps have raised several concerns about privacy and security.
- Critics have seen them as the first steps towards the algorithmic surveillance of citizens

- In London, authorities decided to try something new.
- Together with scientists, they developed methods for “capturing activity over London” to better understand the city’s level of activity.
- In a project called Odysseus, authorities get information about the distribution of activities in London by combining machine learning algorithms, statistical time-series analysis, and image processing.
- This information about the activity in the streets of London can be utilized for the safe reopening of streets and for public health planning.

- In Odysseus, the data comes from a wide range of sources.
- Odysseus combines aggregated, anonymized mobile phone data, anonymized credit card transactions, satellite navigation data, and data from sensors and traffic cameras in the streets.
- This data is used to create counts of vehicles, cyclists and pedestrians, and to indicate the density and impacts of social distancing.
- Special attention is paid to the anonymization of data so that individuals cannot be identified.

- The right for a safe environment is one of these.
- Odysseus provides an example of how AI can be used in a way that respects and promotes the right to safety or to a healthy environment.
- At the same time, the project must take other rights – such as the right for privacy – into account.
- In London these concerns were taken seriously.
- To secure privacy, Odysseus is designed in a way that all the data is anonymized and individuals cannot be identified from the images taken by the traffic cameras.

- Privacy and security have raised a lot of media attention.
- They are important, but it’s necessary to consider the impact of AI on the full spectrum of fundamental human rights and freedoms, too.
- How will AI impact on the right to education and work, or for a fair trial, to fair and open elections, to freedom of speech, and to assembly and demonstration?
- And what about special groups, such as children?
- But first, let’s discuss what human rights are.

What are human rights?

- Human rights form the foundation of the current ethical guidelines and principles of AI.
- This makes human rights a fundamental component of contemporary AI ethics.
- As rights, human rights are **universal**: all humans are entitled to have them.
- One does not have to be a particular kind of person or a member of some specific community to have human rights.
- Human rights are **norms** that protect all people, everywhere from political, legal, and social abuses.

- **Civil and political rights**, such as the right to life, liberty, and property, freedom of expression, pursuit of happiness, and equality before the law
- **Social, cultural and economic rights**, including the right to participate in science and culture, the right to work, and the right to education
- The role of human rights is to protect people's ability to form, construe, and pursue their own conceptions of a worthwhile life – it's not just about the ability to live “in liberty, happiness and well-being”.

What is a human right?

- A human right is a norm which can exist on different levels:
- a shared norm of actual human moralities
- a justified moral norm supported by strong reasons
- a legal right at the national level (where it might be referred to as a “civil” or “constitutional” right)
- a legal right within international law

Universal Declaration of Human Rights

- UDHR is a document which was drafted by representatives with different legal and cultural backgrounds from all regions of the world.
- The declaration was proclaimed by the United Nations General Assembly in Paris on 10 Dec. 1948 as a common standard of achievements for all peoples and all nations.
- Conceptually, human rights are grounded in agency and autonomy.
- They have an ethical priority: if they compete with other considerations such as economic wealth, national stability or some other factor, human rights should be prioritized.

- In the context of AI, this prioritization implies the following requirements:

- AI applications that could clearly violate human rights should not be used
- AI applications that prevent people from enjoying their human rights or actively put them at risk of human rights violations should not be used

- However, human rights have certain context-sensitive properties that allow individuals to prioritize a specific human right if needed.
- Some rights are more fundamental than the others. For example, when the right to life conflicts with the right to privacy, the right to privacy will generally be outweighed.
- In recent years, privacy and security concerns have dominated the discussion on AI and human rights.
- Emerging combinations of big data analytics, surveillance technologies and developing biometric recognition methods have recently received significant media and policy attention.
- Also, the right to equality and inclusion has raised a lot of public discussion. In the next section, we'll take a brief look at these discussions.

Examples of human rights: privacy, security, and inclusion

- **Privacy**
- Privacy concerns are raised, for example, by digital records which contain information that can be used to infer sensitive attributes (age, gender or sexual orientation), preferences, or religious and political views.
- Biometric data also raises privacy concerns, as it can reveal details of physical and mental health.
- Often the real worry is not the data itself, but the way the data can be used to manipulate, affect, or harm a person.
- Ethically, privacy is related to personal autonomy and integrity. Following the principles set out by John Locke, a right to control our own personal lives has been seen as central to our autonomy.
- If that right is taken away, it violates something fundamental about our psychological and moral integrity.

- Many have proposed the principle that people should have control over their own data – and that data concerning them should not be allowed to be used to harm or discriminate against them.
- According to some, this right to have “full control over one’s own data” should be a human right.
- But what, exactly, is your “own data”?
- Is it the raw data, or the collected and analyzed data?
- If the data is used for secondary purposes, is it still your data? Or, as Wachter and Mittelstadt (2019) remark,
- does the content of inferences that can be drawn from your data belong to your “own data”?

- Wachter and Mittelstadt (2019) propose that the right for the control of your own data should be reformulated as a right for the “right to reasonable inferences”.
- According to them, it is crucial that we can also control the “high-risk inferences” that can be made about us through big data analytics.
- These inferences are privacy-invasive or reputation-damaging, or have low verifiability (in the sense of being predictive or opinion-based) while being used for important decisions.

General Data Protection Regulation (GDPR)

- The General Data Protection Regulation (GDPR) is a legal framework.
- It sets guidelines for the collection and processing of personal data from individuals who live in the European Union.
- The GDPR's aim is to give individuals control over their personal data.
- Any information that relates to an individual who can be directly or indirectly identified is “personal data”.
- This includes names, social security numbers and email addresses. Location information, biometric data, ethnicity, gender, web cookies, and political or religious beliefs can also be personal data.
- Pseudonymous data (data that does not directly identify an individual but can be connected to them) can also fall under the definition if it’s easy to individuate someone from it.

- The data subject must give specific, unambiguous consent to process the data.
- Consents must be “freely given, specific, informed and unambiguous.”
- Data subjects can withdraw previously given consent whenever they want.
- Children under 13 can only give consent with permission from their parent.
- The GDPR recognizes several privacy rights for data subjects. Their aim is to give individuals more control over the data.

GDPR Privacy Rights

The right to be informed

- A person must be told about the use of their personal data

The right of access

- It should be explained how someone's personal data is used

The right to rectification

- A person has the right to be forgotten and the data deleted

The right to restrict processing

- A person can deny the use of their personal data

Data Protection Principles in GDPR

- **Lawfulness, fairness and transparency**: Processing must be lawful, fair, and transparent to the data subject
- **Purpose limitation**: You must process data for the legitimate purposes specified explicitly to the data subject when you collected it
- **Data minimization**: You should collect and process only as much data as absolutely necessary for the purposes specified
- **Accuracy**: You must keep personal data accurate and up to date

Data Protection Principles in GDPR….

- **Storage limitation**: You may only store personally identifying data for as long as necessary for the specified purpose
- **Integrity and confidentiality**: Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (for example by using encryption)
- **Accountability**: The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles

How to protect privacy: data anonymization methods

- The GDPR permits organisations to collect anonymized data without consent, use it for any purpose, and store it for an indefinite time – as long as organisations remove all identifiers from the data.

Generalization

- A method that deliberately removes some of the data to make it less identifiable.
- Data can be modified into a set of ranges or a broad area with appropriate boundaries.
- You can remove, for example, the street address while including the information about the town name.
- In this way, you can eliminate some of the identifiers while retaining a degree of data accuracy.

How to protect privacy: data anonymization methods

Pseudonymization

- A data management and de-identification method that replaces private identifiers – names, ID-codes – with fake identifiers or pseudonyms,
- For eg., replacing the identifier “Shanti Devi” with “Saara”.
- Pseudonymization preserves statistical accuracy and data integrity.
- The modified data can be used while still protecting data privacy.

- **Synthetic data**
- A method for using created artificial datasets instead of altering the original dataset.
- The process involves creating statistical models based on patterns found in the original dataset.
- One can use standard deviations, medians, linear regression or other statistical techniques to generate the synthetic data.

- Data-anonymization can be challenging.
- There are also methods for “de-anonymization”.
- De-anonymization methods attempt to re-identify encrypted or obscured information.
- For example, cross-reference anonymized information with other available data in order to identify a person, group, or transaction.

Safety and security

- The right to safety is a norm protecting individuals from physical, social and emotional harms, including accidents and malfunctions.
- Security means safety from malicious and intentional threats.
- As a right, safety creates a moral obligation to design our products, laws and environment in such a way that safety can be protected even in unconventional circumstances or impairments.

AI as an existential threat

- The conversation around AI as an existential threat takes a highly speculative and future-oriented stance towards artificial intelligence.
- It focuses on asking what kind of threats to humanity are posed by AI systems if they become too complex to control.
- However, the plausibility of a future of super-intelligent AI has been called into question, both by philosophers and technologists.
- As things stand, there is no reason to assume that superintelligence will emerge from developing contemporary algorithmic methods.

Safety in AI

- Safety in AI is the practical question of designing systems which behave in a safe and predictable manner.
- As AI systems are integrated into ever-widening areas of life, it becomes more important that the systems are well designed to account for the complexity of the world.
- A very practical and already existing example of this is lane guard technology, which uses machine learning to prevent cars from veering outside of their lanes.
- Machine learning researchers have found that some lane detection algorithms are quite easy to confuse with rogue road markings, causing the car to veer off the road by following the fake lane markings.

- One could argue that the right to safety obligates technology producers to account for these kinds of scenarios:
- the fact that the environment was not ideal does not excuse the system malfunctioning.
- This feature is called **robustness** – the capacity of the system to work predictably under new and unpredictable circumstances.

- The ethically – and legally – significant question is “what are the acceptable limits to robustness?”
- It is conceivable that there are a set of circumstances so incredible that even if the system’s safety cannot be assured,
- We can concede that “nobody could have realistically seen that coming”.
- Where this limit is, though, is a difficult problem, and definitely not one that is exclusive to AI or even technology.

Case: Caged pavements – AI safety and environmental uncertainty

- A difficult problem for autonomous vehicles is the complex unpredictability of the urban traffic environment.
- While AI-driven vehicles are constantly being developed to include better ways to model their surroundings, even a small group of individuals – all performing their own movement goals within a shared space – will create a constellation that is difficult to predict.
- When technical solutions in the cars are too far off, another way to approach the issue is to contain the uncertainty in the environment.

- In a New York Times column, Eric A. Taub proposed a solution: by enclosing pavements in cages, with traffic-light-synced gates at crossings,
- So that the complex traffic environment is simplified to become more understandable to autonomous vehicles and therefore safer.
- However, this safety comes at an obvious cost: limiting the freedom of pedestrians, and a redistribution of accountability.
- This means we should look at the intersecting limits of the right of safety vs freedom.
- Which one is more important?

- A further interesting line of thought that can be traced here is the criminality of what in the US is called “jaywalking”, or
- Walking across the road at locations without zebra crossings.
- The concept of jaywalking did not exist until the roads were reconceptualized with motor vehicles as the primary users.
- How comparable is this to the thought of caging pavements?

Producing safety with AI

- Can AI make the world safer?
- Can AI make the world feel safer?
- And safer for whom?
- Robotization can provide an example of this concept in practice.
- The work of handling hazardous materials or
- working in hazardous environments can be delegated to robots, protecting the health of human (or animal) workers.

- AI-powered surveillance is used in many domains: in public spaces, in law-enforcement work through predictive policing, and in domestic life through products like Amazon’s Ring.
- “Still, the social implications of being recorded have not changed: when we walk into a store, we generally expect that the presence of cameras won’t affect us.
- We expect that our movements will be recorded, and we might feel self-conscious if we notice a camera, especially if we were doing anything that we feel might attract attention.
- But unless something dramatic occurs, we generally understand that the videos in which we appear are unlikely to be scrutinized or monitored.”

- Constant surveillance produces **“chilling effects”**.
- That is, the awareness that our actions are constantly watched limits our true freedom to act in the world.
- Imagine that whenever you leave your house, you are tailed by two police officers.
- They never interact with you, just follow ten meters behind you.
- You will probably feel unsettled and unable to go about your day as you normally would.
- In this way, safety is sometimes at odds with personal freedom and privacy.

A safe and healthy environment: AI and climate change

- Safety also means the right to a safe and healthy environment.
- Nowadays, this right is threatened by climate change.
- The effects of climate change are already visible – storms, droughts, fires, and flooding have become more common, more frequent and more devastating.
- Global ecosystems are changing.
- They all impact the environment on which our existence depends.
- The report on climate change, 2018 estimated that the world will face catastrophic consequences unless global greenhouse gas (GHG) emissions are eliminated within thirty years.

- AI could be a powerful tool for tackling climate change.
- It can be used as a resource for monitoring, understanding, and predicting the consequences of climate change.
- AI can accelerate the development of more ecologically sustainable societies.
- It can be used to design green cities, environment-friendly transportation, to reduce the ecological impact of industry, and
- Design equipment that can help study and maintain the diversity of ecosystems.

- Many potential problems are associated with the deployment of AI – for instance, innovations that seek to reduce greenhouse gas emissions may actually increase energy consumption and emissions.
- Given the data and resource-intensive character of contemporary AI, the technology itself still struggles with energy consumption and carbon footprint.
- One must also pay attention to the environmental impact of raw material extraction for supporting the manufacturing of AI technologies, which can be significant.

Summary

- Safety plays into AI technologies in multiple different ways.
- These all raise questions about the balancing of normative values: while calls to make “AI for good” sound promising,
- In practice the enactment of rights and normative values in technological systems often collides with the many conflicting interests and deep injustices existing in the world.
- When evaluating safety, it is then important to evaluate what other rights intersect in practice and ask, “safety for whom?”

Thank you