**Module 10** **Societal Impact of AI**

**Outline**

**Overview**

**Segment 1: Broad view of AI impacts on human society**

* **Positive sides of AI impact**

* **Negative sides of AI impact**

**Segment 2: What is ethics in AI?**

**Segment 3: Ethical issues in spreading misinformation**

**Segment 4: Ethical issues of bias and fairness**

**Segment 5: Steps toward positive impacts of AI**

**Overview**

This module focuses on societal impact of AI that are both urgent and practical. AI is not what often science fiction portrays, in fact, the past decade has been continually transformed the world due to AI advancement. There are several benefits and risks of AI that have big influence on many aspects of human life.

This module will give you a broad view of AI impacts on various domains and show potential of AI technologies to serve human society in both positive and negative ways. Hence, many concerns raise around the advantages and disadvantages of future impacts of AI on human society. For instance, job automation, human-like robot assistants, education and learning transformation, and intelligent hiring or resource allocation systems are highly influenced by AI development. Since, ML-empowered AI systems can learn about environment and mimic human, many jobs have already taken by AI systems which can outperform human tasks amazingly like transcribers in hospitals and financial analysts.

Therefore, it is important to distinguish between what are myths about advanced AI and what are real and important about AI safety applications. Most popular mythical worries like “AI turning evil” or “AI turning conscious” are rejected by counter-myth researchers because of the consciousness misconception. So, another module aim is to help you differ mythical theory from the current facts about AI. This module suggests you to discover more fields of your interest and observe how AI changed them in-depth positively, so does if it can act like a double-edged sword.

Beside major impacts of AI, this module gives you a definition of data and AI ethics that cover a broad range of topics, many of them making headlines daily, and causing harm to real people. There is a huge variation of different points of view considering ethical AI including but not limited to: justice and human rights, disinformation, robots, civil society, law, policy, and surveillance. This module narrows down of huge amount of variation in tech ethics and investigates the legal but not ethical consequences of AI and then consider bias and misinformation across AI technologies and its impact on society.

There is a ton of courses and articles in tech ethics area including ethical issues in computing and technology. In a handy resource, there is a spreadsheet of collection of many syllabi about AI technology ethics which can be found [here](https://cfiesler.medium.com/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18). This module can be studied independently and there are no prerequisites to follow along. However, it can be integrated with CS 5820 or CS4910. It will provide useful information about societal impact of AI, as well as encourage learners to ask questions about unjust bias, surveillance, disinformation and analyze AI-powered algorithms to identify potential risks.

**Desired Learning Goals:**

Broad view of topical issues that affect human societies now and in the foreseeable future. In depth-understanding of one issue (or one group of related issues) that can contribute to meaningful debate.

**Segment 1: Broad view of current and future ethics issue**

First, we address the positive side of changing world because of many AI developers believe that AI impacts will mostly be good.

**Positive Sides of AI Impact**

AI can dramatically improve the efficiencies of workplaces and can outperform of many human tasks. The new AI technologies, such as intelligent systems might help us eradicate war, disease, and poverty, and so the creation of superintelligence technology might be the biggest event in human history. With better monitoring and diagnostic capabilities, artificial intelligence can dramatically influence healthcare industry. By improving the operations of healthcare facilities and medical organizations, and enhancing disease diagnosis, AI can reduce operating costs and save money.

Also, countless productivity in autonomous transportation and improving traffic congestion issues are remarkable positive impacts of interweaving AI in human society. Most importantly, AI helped people from all over the world to access knowledge, credit, and other benefits of global society and was a big hand of inequality and poverty reduction.

Therefore, there is no doubt that AI brought many beneficiaries into human society. But it is always worth discovering disadvantages of new arrival in order to minimize any possible disaster.

**Negative Sides of AI Impact**

Although AI undoubtedly have a major positive impact on human life from doing exhausting works and increasing efficiency and comfort in many dimensions of daily life, the threat and negative effects of widespread AI technologies are being other major concerns of people and human rights activist. Some of the global concerns are as follow.

1. **AI Impacts on Future Jobs:**

The World Economic Forum (WEF) concluded in a report that a next generation of smart machines, empowered by state-of-the-art models of AI and ML could potentially replace a large proportion of existing jobs. Workers in industries ranging from healthcare to agriculture and industrial can all expect to see changes in hiring due to AI. According to a recent story in Becker’s Hospital Review, AI will [take over](https://www.beckershospitalreview.com/finance/how-ai-can-transform-hospital-revenue-cycle-management-5-thoughts.html?utm_content=98715676&utm_medium=social&utm_source=linkedin&hss_channel=lcp-11444741) certain healthcare tasks related to revenue cycle management and likely increase economic inequality.

McKinsey Global Institute’s latest report assesses the number and types of jobs that might be created under different scenarios through 2030 and compares that to the jobs that could be lost to automation.

Graphical user interface, diagram

Description automatically generated

A picture containing graphical user interface

Description automatically generated

1. **AI and Robotics:**

The main debates around any AI smarten devices are privacy and surveillance, manipulation of behavior, data ethics, biased decision making, humanoid robot interaction, and autonomous systems. For example, in creating decision-making processes that constrain and limit opportunities for human participation, it will often be impossible for the affected person to know how the system came to this output, i.e., the system is “opaque” to that person. Bias in decision systems and data sets is exacerbated by this opacity.

The use of robots in health care has raised a number of concerns for the future of de-humanized care. The growing proportion of elderly people in society, together with recent advances in robotics, makes the use of robots in elder care increasingly likely. One reason why the issue of care has come to the fore is namely that with longer lifespan people will need more care, and that it will not be possible to attract more humans to caring professions.

Despite the possible benefits, six main ethical concerns associated with:

(1) the potential reduction in the amount of human contact

(2) an increase in the feelings of objectification and loss of control

(3) a loss of privacy

(4) a loss of personal liberty

(5) deception and infantilization

(6) the circumstances in which elderly people should be allowed to control robots.

1. **AI Impacts on Hiring and Recruitment**

The volume of applicants has rapidly increased over the past few years, while human resources (HR) cannot consider all the application in a reasonable time. However, potential repercussions of new technologies on the privacy of job candidates are not hidden. There is far less information about the new generation of recruiting tools in pre-hire assessment. As a result, it is not always clear what they assess, whether their underlying hypotheses are valid, or why they may be expected to predict job candidates’ performance.

For example, social medias will use user’s intelligence and personalities for appearing hiring jobs on their platforms. But is it ethical to mine the private data for hiring purposes when users use such apps only for different purposes?

1. **Deviation from AI’s goals**

When an AI system is programmed to do something devastating, means it is deviated from its original goals. Lethal autonomous weapon systems (LAWS), are programmed to fight in a battle. In this case, AI might become a risk when they are extremely difficult to turn off due to their nature for being robust against enemies.

On the other side, when an AI system is programmed to do something positive, but the model may behave in an undesirable way to achieve its goal. For example, if you ask an autonomous car to take you to the airport as fast as possible, it might get you there chased by helicopters, doing not what you wanted but literally what you asked for. Or even using private user’s profiles in a platform in order to augmenting a face recognition training data.

**Segment 2: What is ethics?**

Perhaps the main question in this module is “What is ethics?”. Many people would like to reply to this question in their point of view which can be influence by their feeling, religious beliefs, law requirement, or society acceptance. In fact, it is more complicated and none of them can be a definition of ethics. Based on [Markkula Center for Applied Ethics](https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/), ethics consists of two things. First, it refers to well-founded standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, or fairness.

Second, ethics refers to the study and development of one's ethical standards. Ethics also means, then, the continuous effort of constantly examine and study our own moral beliefs and our moral conduct, and striving to ensure that we, live up to standards that are reasonable and solidly based.

**Segment 3: Ethical issues in spreading misinformation**

**Case study and real-world examples of disinformation**

**1. Coronavirus myths on Social Media**

Fears about the role of misinformation could play in many areas like failing chatbots and deepfakes that can be used to harass minorities or cause unfair resource allocation. Widespread disinformation indicates the complexity of data ethics issues which involves human society, bad actors or even governments in broader ways.

Back to the beginning of the Coronavirus pandemic, several false information tweeted on Twitter to not to use hand sanitizer. It was retweeted and liked a quarter of a million times. This type of a plague of disinformation illustrates how social media could worsen the outbreak by encouraging counterproductive actions. You can see countless conspiracy theories that spread in the first few weeks of pandemic [here](https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-disinformation-spread).

Graphical user interface, text, application, email

Description automatically generated

Hopefully, twitter, Facebook and other social media platforms acted against coronavirus misinformation by directing users to the official sources and updated news by reputable organizations.

[Mike Caulfield in WSU](https://infodemic.blog/) has been working on different approaches to teach students to distinguish online information and misinformation. The educational website of learning the skills that will make a dramatic difference in your ability to sort fact from fiction on the web was a result of his effort to bring the word out to the rest of the world.

He sums it up with the acronym SIFT:

1. Stop.
2. Investigate the source.
3. Find better coverage.
4. Trace claims, quotes, and media to the original context.

Each of these steps comes with a couple of go-to “moves,” such as hovering over the bio of a Twitter user before retweeting, or searching for a URL on Wikipedia before you actually visit it.

**2. Content Creation and Filtering**

The term Deepfake originates from people who shared manipulated pornographic video using face-swapping technology. Today, Deepfake refers to usage of ML and especially DNN models to generate highly realistic video or audio content designed to deceive. Instances of fake content creation abound. For example, [a group of researchers](https://www.washington.edu/news/2017/07/11/lip-syncing-obama-new-tools-turn-audio-clips-into-realistic-video/) transformed audio clips of a former US president into a lip-synced video clip. That means the system has the potential to put other people’s words into someone’s mouth. Further, Deepfake can create non-existent unique faces of one’s choosing, such as age, perspective, mood, and ethnicity, to mimic the photograph of a real person.

In late 2019, a study created a Deepfake bot that generated indistinguishable Deepfake text from real submissions for federal public comment website. Even a human classifier was not better than random guessing in discerning bot submissions from real comments.

There are many ethical concerns around malicious applications of Deepfake from politics and fake news to impersonation. Although researchers are striving to create tools to detect deepfakes, the harms will likely be merely slowed, rather than stopped. Racial and cultural bias could also occur due to prejudice in the training dataset for these tools.

**3. Recommendation Systems Discredit Social Media**

Guillaume Chaslot the former Google engineer and founder of the [AlgoTransparency](https://www.algotransparency.org/) to provide transparency on YouTube’s algorithms. It’s mission is to expose the impact of the most influential algorithms. Google explains why AI is designed to maximize watch time: “If viewers are watching more YouTube, it signals to us that they’re happier with the content they’ve found. It means that creators are attracting more engaged audiences. It also opens up more opportunities to generate revenue for our partners.”

On the other side, AI amplify resentment against other media and can influence content creators. Any smart AI that optimizes engagement with itself will have a tendency to discourage engagement to other channels.

**Segment 4: Ethical Issues of Bias and Fairness**

The bias and fairness issue in machine learning can be classified as the potential sources of unwanted consequences. A study provides a framework that partitions sources of biased data in machine learning into five distinct categories spanning the data generation and machine learning pipeline. Representation bias, Evaluation bias, Historical bias, Measurement bias, Aggregation bias. The fact is that machine learning can strength bias and even reproduce them.

**1. Failures in Facial Recognition Technologies**

Perhaps one of the most scandalous series of failing AI technology is the potential weakness of facial recognition system which has demonstrated prejudice against women. [Joy Buolamwini](http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf), a researcher at the Massachusetts Institute of Technology Media Lab, during her PhD program on how computer detect and recognize people’s faces, figured out IBM face recognition system was working well on lighter skinned face but not so well to detect her own face. She tested four famous and available demos softwares including IBM, Face++, Microsoft, and Megvii (created by a Chinese company); two of those did not detect her face and two others misgendered her. She found out all companies product performed better on males than women and also performed better on lighter subjects that darker ones, amongst all of them IBM had the largest racial and sexual biases with 34% error rate.

Graphical user interface, application

Description automatically generated

**2.** **Recidivism Algorithm in Predictive Policing**

Even when race and gender are not inputs, machine learning will find latent variables. Major predictive policing algorithm is fundamentally flawed. Because historical bias is a structural issue with the first step of the data generation process and can exist even given perfect sampling and feature selection.

Table

Description automatically generated

**3.** **Exclude older workers, minorities, women from recruiting advertisement**

Graphical user interface, text, application, email

Description automatically generated

**Segment 5: Steps toward positive impacts of AI**

There is no kind of a general solution for biased algorithms, but still there are some helpful and practical ways to reduce and tame the biased systems. Analyze every project that we work and answer critical questions about AI such as,

1. should we even be doing this? (sometimes the answers is to not build it, for example facial feature discovery for ethnicity recognition or Gaydar program from Stanford university)
2. What biases is in the data? (recognizing all data and understand data and how they are gathered)
3. Can the code and data be audited?
4. What are the error rates for different subgroups?
5. What is the accuracy of a simple rule-based alternative? (sometimes a random guess solution or a simple linear regression can do better than a complex model)
6. How diverse is the team that built it?
7. What processes are in place to handle appeals or mistakes? (examine and evaluate the model before lunching to make sure no negative consequences happen)

Another effort for documenting dataset is “[Datasheets for Datasets](https://arxiv.org/abs/1803.09010)” which proposes that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses. Datasheets for datasets will facilitate better communication between dataset creators and dataset consumers and encourage the machine learning community to prioritize transparency and accountability.

**Motivation for dataset creation.**

* Why was the dataset created?
* What tasks could the dataset be used for?
* Has dataset been used for any tasks already?
* Who funded the creation of the dataset?

**Dataset composition.**

* What are the instances?
* Are relationships between instances made explicit in the data?
* How many instances there?
* Who was involved in the data collection process?
* Over what time frame was the data collected?
* How was the data associated with each instance acquired?

**Dataset maintenance**

* Who is supporting, hosting, or maintaining the dataset?
* Will dataset be updated?
* How will updates be communicated?

**Legal and ethical considerations**.

* If the dataset to people or was generated by people, were they informed about the data collection?
* If it relates to people, were they told what the dataset would be used for and did they consent?
* If it related to people, could this dataset expose people to harm or legal action?
* If it relates to people, does it unfairly advantage or disadvantage a particular social group?
* If it relates to people, were they provided with privacy guarantees?

References

1. <https://www.mckinsey.com/~/media/mckinsey/industries/public%20and%20social%20sector/our%20insights/what%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/mgi%20jobs%20lost-jobs%20gained_report_december%202017.pdf>
2. <https://link.springer.com/article/10.1007/s10676-010-9234-6>
3. <https://plato.stanford.edu/entries/ethics-ai/#Intr>
4. <https://depts.washington.edu/hints/articles/robotic_pets_2.pdf>