ARIS PROJECT

ARTIFICIAL INTELLIGENCE SKILLS FOR ICT PROFESSIONALS

ARIS PROJECT

AI SKILLS FOR ICT PROFESSIONALS

Co-funded by the
Erasmus+ Programme
of the European Union

1 INTRODUCTORY PARAGRAPH

Artificial Intelligence is having a huge impact in our lives and, presumably, this impact will increase in the near future. Our life is affected by AI technology almost continuously since many algorithms interact with us through our cell phones or our PC internet browsers and we are using them most of the day. Such prevalence has raised social concerns about the legal and ethical implications behind all these technologies. Tech companies are also showing concerned about ethical aspects and some of them are even recruiting ethithists to drive them through their innovation and deployment processes. In this lesson we will describe some aspects that have entered the public conversation. We do not want to be exhaustive, but give the reader a glimpse of the many aspects of this conversation.

2 LECTURE NOTES

.

2.1 Motivation

Human beings have developed ethical rules that shape our society. One possible definition of **ethics** is a set of rules that we, as a society, decide that it is good to voluntarily follow because they make our world a better place. Many of these rules are quite old and universal, such as our commitment to not killing other human beings. Others are relatively new like the opposition to slavery. In many cases ethical rules end up being reflected in national laws. However, it is important to notice that ethical rules do not need to be very precise, since very often are just behavior guidelines (e.g. it is unethical lying to our friends). Also we follow them voluntarily, probably because we want to feel good and to project a good image of ourselves. On the other hand, laws must be much more precise in order to minimize different interpretations. Also the law is enforced, and we must follow it independently of whether we like it or not, because otherwise we will be punished.

An **ethical dilemma** is a situation in which a choice has to be made between two courses of action, either of which entails transgression of a moral principle or causes some harm. It is important to have some ethical standards to take the most appropriate decision.

Technology advances often create new ethically delicate situations. Because laws take some time to adapt to changing environments it is important to have ethical guidance. Presumably, what ethics dictate will be later on reflected in the law.

Artificial Intelligence is impacting our lives dramatically and there is a social claim about this progress being made following strict ethical principles. Some tech companies are hiring ethicists to help them with their business processes, AI researchers are raising some flags about how to handle the fore coming technologies and aware citizens are pushing politicians to observe and protect civil rights. All in all, the ethical aspects of AI are currently an open fascinating conversation.

This course is directed to ICT professionals and they may think that ethics does not have anything to do with them. CEOs and other executives are those making strategic decisions and those who should worry about doing things ethically. This may be true in part, but ICT professionals cannot avoid their share of responsibility. In the words of Cherri M. Pancake, the ACM's president:

"Computing professionals are the first line of defense against the misuse of technology. Our collective understanding of computing systems puts us in a position to protect sensitive information and ensure that systems integrate in ways that are appropriate, safe, and reliable. Society needs to be assured that we are committed to ethical conduct as the foundation of our work. That need has become the personal responsibility of every professional in our industry"

In this lesson we will mention some topics of ethical concern. For a more thorough analysis we urge you to visit "The Ethics of Artificial Intelligence: Issues and Initiatives", a document from the European Parliament Research Service.

2.2 Algorithmic Fairness and Algorithmic Bias

More and more software is being developed to guide **decision-making**. Machine Learning is used to learn outcomes from previous cases and assume that new cases will show similar patterns. In many domains those algorithms run with little or no human understanding and the quality of their outcome depends in a very unclear way on design decisions made by software engineers or on the quality of the data used to train the algorithm.

These algorithms for decision making can have important impact on people such as determining bank loan acceptance, granting releases on bail, job selection, etc.

It is known that humans have prejudices that affect our own decision making and we expect those taking decisions being able to be objective. Typical **prejudice variables** are race, gender, sexual orientation and, depending on the context, also age, nationality, etc.

We say that an algorithm is **fair** or **unbiased** if it treats identically two different individuals that only differ in a prejudice variable. When AI algorithms started to be used to guide decision-making, it was generally believed that they were going be fair, because they are trained with cold data and they should not suffer from our social bias. However, it did not take long before it was observed that it was not the case. Some well-known examples include racist recommenders of penitentiary benefits or sexist recommenders of candidates for job vacancies.

First attempts to solve the problem simply considered the elimination of the critical variables, but it turned out that the algorithms could infer those variables with high accuracy using other seemingly inoffensive variables (e.g. in some cases you can guess with high accuracy the race or religion of a person from its zip code)

Some researchers claim that **bias may come from the programmers**, who unintentionally contaminate the algorithm through design decisions with their own prejudices. Other times it is the **poor quality of the data used to train** the algorithm what causes the bias. Because very often data is extracted from the Internet, and the amount of available data is massive, engineers tend to believe that it's a good representation of reality, but it is well known that the **internet content is highly biased** by age, gender, race and many other sensitive variables.

Many researchers are trying hard to establish the theoretical foundations in order to guarantee unbiased algorithms but there are still many difficulties to be solved. In the mean time we need to be aware of the problem and do our best to minimize it.

.

2.3 Privacy and the Surveillance Capitalism

We interact with the world wide web more and more every day, and this interaction leaves huge amounts of personal information that can be easily gathered by tech companies. Your intelligent speaker may know if you have tonsillitis, your cleaning robot may know if you

need a sofa, your smart tv may know if you are sad, your pictures manager may know if the tv in your living room is old,… and this are simple examples. From **crossing data** and using AI algorithms, it is plausible that some company has a record about you containing with high accuracy your age, sexual orientation, political ideas and even psychological profile.

While most people think that we are the customers of the tech companies whose products we use (such as social media, internet browsers, cell phone apps) some people has started to claim that **we are not the customers but the commodity** these companies make profit with. In fact, some authors claim that personal data has become the most valuable resource on earth.

Shoshana Zuboff coined the term **surveillance capitalism** to describe this new scenario with three main ingredients: aggressive data gathering with little awareness of the users, AI algorithms able to exploit these data and reduce uncertainty by discovering the probabilities of future behavior of people, ecosystem of companies willing to structure and monetize this data

There are many ethically questionable aspects behind this new business model but we will mention two:

Not many people are aware of the amount of data that they are giving away and the amount of data that can be inferred (with high accuracy) about them. Also we have very little knowledge about the "value" of our data, so there is no correlation between the value of the service that user receive and the value of the data that they give in return.

User's data is mostly used for advertisement purposes and advertisement is essentially about manipulating our behavior. While advertisement has been with us for centuries and most of the time we do not see it as unethical, we have to be aware that in our highly interconnected society we are allowing companies a much more advantage for that manipulation. It is not at the same ethical level to show adds of sport gear to somebody that the AI algorithm believes to be sports aficionado as to advertise a gambling web to somebody that the AI algorithm believes that is having financial problems

2.4 Algorithmic Accountability

.AI algorithms are increasingly being used to guide decision-making and such decisions have impact in our life. Therefore, it is reasonable that we expect some level of **accountability** on these decisions. If a company, say a bank, makes a decision such as declining a loan request, the customer has the right of having some sort of **explanation** since otherwise the decision may appear as arbitrary. Also it should be possible to **hold the company responsible** for the decisions made by its algorithms in a similar way as it is responsible for the decisions made by its employees. In order to achieve this, AI practitioners are working on how to enforce the following principles:

1. **Explanation**. AI algorithms should provide some sort of support for their decisions. The support has to be understandable by humans without having to know the algorithm details.

2. **Transparency**. Companies cannot hide behind the complexity or the proprietary nature of the algorithm to avoid giving explanations.
3. **Audit**. Algorithmic techniques, accuracy, robustness should be checked by independent third party. This requires quality and methodological standards

2.5 Prudence

There is no question about the huge success of AI systems in a variety of domains and how this success has been publicized in general media. Besides the many articles reporting accomplished success, many others speculate about the future success. Companies and researchers may be tempted to exaggerate about it. For companies it is good marketing to explain the amazing features of their current products and the even more amazing features of the products soon-to-come. For researchers it is a good strategy to guarantee funding for their work: past success gives credibility for future success. Besides, because the general audience does not understand scientific and technical writing, it is tempting to use oversimplifying metaphors that very often look to optimistic.

For this reason, some researchers and AI practitioners recommend extra effort to also communicate the limitations of the technology. Current AI systems may work well in a majority of cases, but users must know that their full correctness has not fully checked. Also many fundamental problems of artificial intelligence are not yet solved and will require radical breakthroughs, and perhaps some issues will never be solved. This request for **prudence** appears in the **Barcelona Declaration for the proper development and usage of Artificial Intelligence**.

As an example, consider face recognition which is one of the most successful applications of AI. A recent US government study analyzed nearly 200 algorithms and observed that they have worst performance on non-white faces. Being aware of this limitation is important since such algorithms may be used law enforcement, border control, and other applications throughout society.

Another example is self-driving cars. In 2015 The Guardian predicted 10 million self-driving cars on the road by 2020, But the year is here — and the self-driving cars aren't. There are many technical and legal problems that still need to be solved. So when will they be in our roads? Some experts say that pretty soon, while others say that never.

2.11 References

Ethics of Artificial Intelligence and Robotics – Vincent Müller, The Stanford Encyclopedia of Philosophy, Summer 2020

The Ethics of Artificial Intelligence- European Parliament, 2020, ISBN: 978-92-846-5799-5 (www.europarl.europa.eu)

Policy and Investement Recommendations for Trustworthy Artificial Intelligence
ec.europa.eu/digital-single-market/en/news/policy-and-investmentrecommendations-trustworthy-artificial-intelligence

Machine Bias
(https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)

Explainer: what is surveillance capitalism and how it shapes our economy
(https://theconversation.com/explainer-what-is-surveillance-capitalism-and-how-does-it-shape-our-economy-119158)

Its 2020: where are our self-driving cars?
(https://www.vox.com/future-perfect/2020/2/14/21063487/self-driving-cars-autonomous-vehicles-waymo-cruise-uber)

Barcelona Declaration for the proper development and usage of Artificial Intelligence
(www.iiia.csic.es/barcelonadeclaration)

IEEE STANDARDS ASSOCIATION, *Ethically Aligned Design*:
standards.ieee.org/industryconnections/ec/autonomous-systems.html.

A US government study confirms most face recognition systems are racist. MIT Technology Review, dec 2020 (https://www.technologyreview.com/2019/12/20/79/ai-face-recognition-racist-us-government-nist-study/)

3 QUESTIONS AND ANSWERS

Question	Answer
What distinguishes an ethical principle from a law?	An ethical principle needs not to be precise and it is followed voluntarily. A law must be as precise as possible and it is mandatory to be followed
What is an ethical dilemma and why tech companies should worry about them?	An **ethical dilemma** is a situation in which a choice has to be made between two courses of action, either of which entails transgression of a moral principle or causes some harm. Being ethical, companies contribute to a better world.
Why ICT professionals should care about ethics?	Because they may be the first to see and those more qualified to assess potential ethical dilemmas.
What is algorithmic fairness and why it is a desirable feature?	An algorithm is fair if it treats equally individuals that only differ on prejudice variables.
Why removing from data prejudice variables does not prevent AI algorithms from being biased?	Because these variables may be inferred from other seemingly vacuous variables
What the term surveillance capitalism refers to?	A new business model in which companies gather and sell customers data, mainly for advertisement purposes. AI algorithms are used to infer features that may not be directly available in the data.
Why companies and researchers should be prudent about the performance of their present and future products?	Because they may raise expectations that cannot be fulfilled
What is algorithm accountability?	The possibility of explaining algorithm decisions and asking for responsibility of their consequences.

4 USE CASES

4.1 USE CASE 1

.

In 2014 Google developed an AI system to review job applicant's resumes and automatically select top candidates. The system would rank candidates from one to five stars. In 2015 the company realized that the system was not ranking in a gender-neutral way.

The reason was that the system was trained observing patterns with employees over the last 10-year period. Most came from men, since the tech industry is largely dominated by men. Thus, the system taught itself that male candidates were better and penalized resumes including the word "women". Even if the gender was not explicitly included in the data, it could be inferred easily from many resumes. For example the algorithm would end up penalizing candidates reporting having participated in cooking clubs (mainly feminine) or having studied in all-women colleges.

In 2016 the company abandoned the project because they could not easily fix its bias.

4.2 USE CASE 2

In 2014 Facebook and Cornell University's Department of Communication and Information Science made a study entitled "Experimental evidence of massive-scale emotional contagion through Social Networks" in which they influenced the emotions of more than 600,000 Facebook users. Facebook researchers directly manipulated Facebook users' news feeds to display differing amounts of positive and negative posts in order to determine whether their subsequent posts were affected by the posts they were viewing.

Facebook allowe the scientists (both internal and those from Cornell) access to the huge amount of data. The Facebook researchers performing the data-collection did not seek explicit informed consent from any participants which is standard practice in current internet research ethics procedures.

5 PRACTICAL EXERCISES

5.1 EXERCISE

Companies' income depends on how much they sell. If their products last too long it may be against their own interest. There are two ways to make customers buy products more frequently: to innovate so the customer wants to get the newest, more advance, model; to design products so that they decrease their performance after some time, so the customer needs to buy a new product. The second strategy is called programmed obsolescence. Identify potential cases of programmed obsolescence in the ICT world and discuss its ethical implications.

5.2 EXERCISE

Personalized marketing is becoming more and more popular, mainly because old-fashioned marketing was extremely inefficient. Most countries have laws protecting vulnerable (e.g. kids) people from old-fashioned marketing. Identify potential cases of unethical personalized marketing and discuss if it seems possible to write laws preventing them. As an example, psychologists say that a depressed teen-girl is likely to buy make up if suggested. What would you think if a music provider would sell data of teenager users that suddenly transition from listening to cheerful to sad music?

5.3 EXERCISE

At least 21 US states use AI systems to grade essays written by high-school students. Their advantage is that it alleviate the work burden of already burn-out teachers an also that they allow instant and unlimited feed back to students that can use them 24/7. As an ICT professional, what do you think about these systems? Can you anticipate any ethical dilemma in their use? You can find some answers and much more information here:
https://www.vox.com/recode/2019/10/20/20921354/ai-algorithms-essay-writing

6 MULTIPLE CHOICE QUESTIONS

1. Why society wants tech companies to follow ethical standards:
    a. Because otherwise they may be fined by governments and go bankrupt
    b. Because it makes the world a better place
    c. Because otherwise nobody will want to work there
2. Why ICT profesionals should care about ethics:
    a. Because otherwise they may break the law and be imprisoned
    b. Because they may be the first ones identifying ethical dilemmas
    c. Because it makes them feel better among their family and friends
3. Which of the following is not a prejudice-variable:
    a. Income
    b. Sexual orientation
    c. Religion
4. Which of the following actions a Company can do to improve the accountability of its algorithms:
    a. Protect its liability claiming that the algorithm is complex and its decisions cannot be explained
    b. Write a legal disclaimer for any consequence of the algorithm
    c. Ask an independent third party to analyze the algorithm
5. What of the following actions is in line with the request of prudence made in the Barcelona Declaration for the proper development and usage of Artificial Intelligence:
    a. Companies should exaggerate the features of their products in order to increase the enthusiasm of customers which will be good for their business and consequently to global economy
    b. Researchers must give the most optimistic versión of the results of their reseach, because it is good that funding agencies have confidence and allocate resources to the development of AI technology
    c. AI has already shown great success in the last decade. A prudent, responsible and accurate description of its systems is the best way to have society trusting AI and using it for a better fairer world.