

# Project 1 – Ethics in A.I./D.S./M.L

Soufiane Fadel (5314E)

October 16, 2019

**Number of words : 2024**

**Topic:** The ethical issues surrounding the use of artificial intelligence (A.I.), data science (D.S.), and/or machine learning (M.L.) algorithms in the public sector, the private sector, or in academia.

New technologies often raise new moral issues. For example, the emergence of nuclear weapons put a great deal of pressure on the distinction between combatants and non-combatants that was at the heart of the just war theory formulated in the Middle Ages. New theories were needed to reinterpret the meaning of this distinction in the nuclear age. With the emergence of new machine learning techniques and the ability to use algorithms to perform tasks previously performed by humans, as well as to generate new knowledge, we are again confronted with a set of new ethical issues . These questions not only focus on the possibility of damage caused by misuse of data, but also on how to maintain confidentiality when data is sensitive, how to avoid bias in data selection, how to prevent disruptions and "Hacking" data and privacy issues. transparency in data collection, research and dissemination. Many of these issues underlie a broader question of who owns the data, who has the right to access it and under what conditions. There are currently no agreed answers to these questions. Nevertheless, it is extremely important to confront them and try to define common ethical rules. When agreement can not be reached, it is important to consider the conflicting values and specifically articulate the assumptions underlying the work in different models. An interesting illustration is the equity debate in models predicting the risk of recidivism among black and white accused in Broward County, Florida. Should a risk score be: equally accurate in predicting the likelihood of recidivism for members of different racial groups; assume that members of different groups have the same chances of wrongly predicting that they will reoffend; or suppose that the failure of prediction of recidivism occurs at the same rate in all groups. Recent work has shown that it would be impossible to satisfy all three criteria at the same time; If you meet two people, you will not respect the third. We must therefore decide which aspects of equity are the most important. technologies such as AI , Big Data and machine learning are evolving more rapidly, the rapid development of these technologies poses ethical, philosophical and moral problems in several areas.and a lots of futuristic fears exist about the handling of personal data or the use of machines for military or professional purposes , it's true that these intelligent machine systems are making our life easier and better but the ethical issues surrounding the use of these technologies is still a common thought that arises when we read about any significant advancements in those technologies . there are ethical questions associated with exploration algorithms and also privacy concerns associated with the use of private data and what should be private vs not , what are some ethical issues regarding machine learning ? how far can we go with big data ? And how far are we willing to delegate our decisions to the machine?

Machine learning, also known as statistical learning has recently made progress ,Its applications are multiple: search engines, image and speech recognition, translation ,automatic, conversational agents for example. They begin to emerge in sectors such as health, energy, transport, education, commerce and banking. The success of machine learning is based on increasing computing, storage and data processing capabilities (Big data), and building systems capable of . learning, abstraction and reasoning is one of the goals of researchers in artificial intelligence. To do this, the algorithms learning methods use different statistical methods and one of the major problems of these algorithms are : Some times they are just blackboxes where the data goes in,(inputs) and decisions come out (output) but we don't know how these decisions are obtained from the input so the ethical question here is how humans can trust and work effectively in cooperation with robots and AI driven systems when AI makes decisions the no human can explain ? So

we will not just rely on a mysterious blackbox because some times these boxes could make bad decisions because of Algorithms have been built-in biases because they are created by individuals who have conscious or unconscious preferences that may go undiscovered until the algorithms are used and also the data used produces biased models which can be discriminatory and harmful to humans . machine learning is based on some statistical models and statistics are persuasive. So much so that people, organisations, and whole countries base some of their most important decisions on organized data. But any set of statistics might have something lurking inside it that can turn the results completely upside down and one of the famous examples of statistical biases is the racial bias against African Americans the problem is that platforms like Facebook, YouTube, and Twitter are banking on developing artificial intelligence technology to help stop the spread of hateful speech on their networks the idea is that complex algorithms that use natural language processing will signal racist or violent speech faster and better than human beings. Doing this effectively is more urgent than ever in the light of recent mass shootings and violence related to online hate speech, but two new studies show, however, that AI trained to identify hate speech may actually magnify racial prejudice. In one study, researchers found that the main models of AI for the treatment of hate speech were 1.5 times more likely to report tweets as offensive or hateful when they were written by African Americans, and 2.2 times more likely to report tweets written Afro-American English (which is commonly spoken by blacks at United States). Another study found similar widespread evidence of racist bias against black discourse in five sets of widely used academic data used to study hate speech, which totaled about 155,800 publications on Twitter. The influence of bias is present in many other types of data. For example, a simple application of Machine Learning, in which computers outperform human capabilities, is loan approval. Financial institutions use historical data to form their algorithms over millions of records, allowing them to capture data models that best identify the characteristics of mortgage applicants. The problem is that these algorithms learn too much , suppose mr santiago is a brilliant graduate student from a dangerous and poor neighborhood but mr Santiago has a clean credit record and he has decided to apply for a loan to buy his first property , and according to all the criteria traditionally used by banks to assess the solvency of a person, his application must be approved. however the algorithm has another plan, it was learned that applicants from this same poor area had just had their applications rejected in recent years because of poor credit records, insufficient disposable income and so one So in this case the algorithm has discriminated racially against him but how this algorithm could be racist when the bank claim that of course this is not possible, as the algorithms have been intentionally blinded to the race of the applicants , However, the problem is that while some algorithms like DecisionTrees may enable an auditor to discover if the address information was used in a way to penalize applicants who were born or previously resided in predominantly poverty-stricken areas, other algorithms like Deep Learning are much more sophisticated and tend to be a “black box” to external inspection, and it may prove almost impossible to understand why, or even how, a certain decision has been taken. Another example is when the computer teaches itself English it becomes prejudiced against black Americans and women. Using data ”scraped” from the web called the ”Common Crawl”, a corpus containing approximately 840 billion words, it shows that machines can learn word associations from written texts and that these associations mirror those learned by humans, such as pleasantness and flowers or unpleasantness and insects. So far so good. But it also shows that machine learning absorbs stereotyped biases as easily as any other - for example, associations between female names and family or male names and career. From bad to worse, the researchers found that names associated with being European American were significantly more easily associated with pleasant than unpleasant terms, compared to some African American names. A computer builds its vocabulary using frequency data, or how often terms appear together. So it found that on the internet, African-American names are more likely to be surrounded by words that connote unpleasantness. Is that because African Americans are unpleasant? Of course not. It’s because people on the internet write and say horrible things

Another major moral issue that surrounds the using of machine learning and data science is the problem of privacy. Everyone (business, government, university) agrees to recognize Big Data as a great potential for boosting business, innovation, and the progress of society. In fact, with the proper use of big data, the economic intelligence and efficiency of the private and public sectors can be improved by developing new applications that exploit new social and economic opportunities that transform society. Big data can be used to identify correlations and to define general trends to better understand a market and its evolution, but they can also be processed in order to analyze the behavior of individuals very exhaustively. The same algorithms and analytical tools used by a distributor or banker, for example, to understand the interests, desires of individuals, and determine what they can sell to him, can be used by a government, political parties, businesses private security to calculate (and often miscalculate) if you can be a threat (eg the S cards), now

or in the future. Even if it is not the megadonnées that are in themselves worrying, but well the uses of the information that can be made, the multiplication of the sensors that takes place via the development of the Internet of Things (GPS, connected watches, activity plotters, home automation ...), considerably increases the data flows and the possibilities of pierce many spaces that were previously private. Specific examples of data collection may not seem important on their own, but when aggregated, they can create a complete image of a person that can be extremely harmful to those affected, especially in the hands of others. not allowed. Today, the risk that people lose control of their own data becomes maximum.

One final Serious problem is the impact of AI in the workforce. A year ago, the figure seemed huge. By 2021, the growing robotization and increased use of artificial intelligence and automation were to eliminate as many as 5 million jobs in the world's top 15 economies, including Canada's. There is no revolution without breaking eggs and the fourth industrial revolution should be no exception. This massive job loss forecast was formulated in January 2016 by the World Economic Forum Research Center following a survey of the world's 15 largest economies, which account for 65% of the workforce. global work. The study predicted that the increasing automation of economic activity would result in the loss of 7.1 million jobs, mostly among white-collar workers, while a significant number of administrative tasks would now be performed by machines. intelligent. In return, these job losses should be partially offset by the creation of 2.1 million new jobs related to the increased use of artificial intelligence (AI) in the fields of computer science, mathematics, computer science This is a net loss of 5 million jobs.

*"I find it somewhat ironic that many conservatives in the United States are adamant about securing the border against immigrants who will likely take jobs that few Americans want, while at the same time expressing little concern that the virtual border is left completely open to higher-skill workers who take jobs that Americans definitely do want."*

— **Martin Ford**, Rise of the Robots: Technology and the Threat of a Jobless Future

## References

- [1] <https://quantela.com/blog/?p=243>
- [2] <https://medium.com/humane-ai/ethics-in-machine-learning-54a71a75875c>
- [3] <https://towardsdatascience.com/machine-learning-and-discrimination-2ed1a8b01038>
- [4] <https://time.com/5520558/artificial-intelligence-racial-gender-bias/>