Possible ethics on machine learning biases and their impacts in future prospects

Marley Oliveira Bacelar
UNICAMP - Universidade Estadual de Campinas
2021

Abstract
The article states some examples for machine learning bias and ethical dilemmas that occur due to progressive improvement in machine learning based on previous studies by web research. The article evaluates the history of traditional programming to machine learning and explains how ML is implemented and how it lead to efficiency in banking, criminal justice, and medical fields. This also explains the possible bias that can occur by using the algorithmic systems in the society and ethical dilemmas regarding the ML in accordance with previously conducted studies. Finally it explains how to attain a better future with unbiased algorithmic process which will drive the society into a pleasant and fairer offers and decisions.
1. Introduction

New programs and software launching in the modern world far more efficiently than before intensively, due to the effect of Artificial intelligence and machine learning, with the aim of easing the day to day activities of the individuals. ML extensively used for the software nowadays than that of traditional programming. Traditional machine programming witness time consuming and energy wasting process while machine learning engage in self processing methodology by the machines itself. Although this seems easy, ML will cause potential bias in the society due to erroneous processing of data fed into the machine, where the later outcome going to be hazardous for the society. Bias is a complicated term with good and bad connotations in the field of algorithmic predicting machine. For phenomena with legal and ethical consequences, we should ensure the fairness of the process by providing higher attention to possible bias from the beginning of the model building. Highest algorithmic functionality is the most reasonable way to neglect biased decisions where more data ensure better model which will cause higher accuracy of the process. The improvement of ML is double edged sword, as it is being beneficial for the individuals, at the same time bring catastrophic effects which cause due engulfing the improved technology by the cyber hackers. It is possible to suggest more research under age inequality and gender disparity as there are enormous amount of research conducted under racism prejudice of ML but less amount of the others.

2. From traditional programming to machine learning

Machine learning programming can be sated as the new insertion in the 19th century for the broad programming world which is responsible for revolutionize of business for the last 10 decades, where conventional programming is also an essential part of it. The first computer
programming initiated in mid-1800’s by a famous scientist Ada Lovelace. Conventional programming includes rule coding and writing lines of coding which processed manually, where by efficiently using specific type of programming language (conventional procedural language) such as C, C++, Java, JavaScript, python. Together with this method of programming, in machine learning, the programmer creates the logic of program by inserting data according to program’s programming logic. The approach is algorithm dependent and depends on training by the insertion of particular data for a program, multiple algorithm can work as well. After feeding the computer with a set of instructions in the first place, the computer will utilize computing abilities to help human process the data rapidly and in an efficient manner. With this mode of programming, the design and development of logic is up to the program.

3. Why machine learning?

Machine learning extremely powerful in modern world task effectiveness, as they basically involve in making decisions by the machines itself without specifically programmed. Fundamentally the models utilize algorithms and semantic network. After Arthur Samuel first came up with the term Machine learning in 1952, it was revealed that Machine learning commensurate the model of brain cell interaction in 1949 by Donald Hebb further states (Foote, 2019) that the interaction mechanism in machine learning in a computer is exactly identical to the brain function as “when one cell repeatedly assist in firing another, the axon of the first cell develops synaptic knobs (or enlarges them if they already exist) in contact with the soma of the second cell. This exactly corresponds to the relationship of artificial neural networks and artificial neurons (nodes). (Hebb, 2012) Although some software engineers and data scientists engage the traditional conventional programming models, instead of deep neural networks
which power AI applications, where as traditional programming does not work most of the
times as they have limited capacity to capture information about training data.

Nowadays it is quite evident that the evolitional change is developing with accompany
of big technical companies by smoothing out the process along with assembling new machine
learning specific platforms, with end to end functionality. According to the study applications
of AI in machine learning conducted by Sumit Das and the crew there are four types of machine
learning, namely supervised learning, unsupervised learning reinforcement learning and
recommender system. Machine learning engineering takes place in three stages according to
the article by Catherine Dong in 2017 (TechCrunch Is Now a Part of Verizon Media, 2017).
Namely data processing, model building and thirdly deployment and monitoring. First and third
stages are sandwiched by the model building which is Machine algorithms where it learns to
predict input data.

AI is the housing of machine learning, where it is subcategory of AI that include rarefied
statistical techniques which allow machines to develop at tasks with experience. This category
comprise of the sub category; deep learning, which constitutes algorithms that allow software
to train itself to perform duty. (Speech and image recognition by exposing multilayered neural
networks to vast amount of data) and it is the model which sandwich the other two stages in
ML.

The first stage of machine learning involves cleaning up and formatting of enormous
amount of data which will be fed in to the model. The last stage involves classification and
monitoring of the model. Data used to train algorithms are finite and does not reflect the reality
most of the time, which will results in bias that arise from the choice of test and train data
which represent the true population. Limited data training, which is able to model and
accurately classify data is an assumption made for a fairer outcome.
Most of the companies despite AI research labs, do not depend on neural networks, instead they depend upon traditional machine learning models such as linear/logistic regression, random forests boasted decision trees, which utilize for services such as friend suggestions, as targeting, supply/demand stimulation and search results ranking. This is sometimes habitual due to rough situations experienced by training deep neural networks as they require more time to train and specifications for the process ML in the machine, such as specific hardware and GPUs). Even though deep learning permits difficult situations rather than that of traditional methods, when training and building up, it results fewer number of potential errors than that of old tools. If engineers use traditional methods with the big companies who needs to format millions of data points, the process might acquire up to several days and most of the task will be utilized for repetitive work as most companies are suburbanized and also due to cause of compound errors. Traditional unit test does not work with machine learning, as with traditional software testing, due to unpredictable nature of the results that occur at the end. Therefore instead of unit tests, engineers manually monitor dashboards and program alerts for new models, as the final outcome of machine learning is accepted to become independent of engineers, so that they can perform alone. With regards to accessing these issues, huge companies started with available resources to develop machine learning unique specific tools which will lead to attain seamless end to end machine learning platforms. (How Different Are Conventional Programming and Machine Learning? -, 2021)

4. Machine learning bias and ethics

Bias is a situation which occurs due to systemically discriminated outcomes that results by algorithm due to erroneous assumptions predicted within the machine learning stages. Mireille Hildebrandt, a lawyer and a philosopher at intersection of law and technology has extremely stated on the issue of bias and fairness in ML algorithms (Hildebrandt, 2018).
she argues that bias free machine learning does not exist and a productive bias is essential for an algorithm in order to be able to model the data and are related forecast. In accordance with no free lunch theorem (“The No Free Lunch Theorem and the Human-Machine Interface,” 1999) particular classifiers must include certain bias towards certain distributions and functions to be better at modeling those distributors, as all classifiers have the same error rate when averaged over all possible data generating distributions. The bias really depend on the data fed into the system. Even though the predictive system involves in massive aid in the field such as banding, medical educational, there are major drawbacks occur due to inaccurate training of the machine, human errors that cause while building up the model, or erroneous decisions that machines gather from various data that is fed into the machine (Lee, 2021). This process clearly delineates by a famous computer study slang, as garbage in and garbage out syndrome, where garbage in is insertion of biased data while garbage out is catastrophic outcomes due to erroneous processing.

Ethics related to machine learning is a part of ethics of AI which concerned with ensuring moral behaviors of manmade machines that utilize AI. There is no solution in solving ML bias issues. It depends on framework of the given system and the most sensible manner to neglect biased decisions while managing highest algorithmic utility. Algorithms are found to be ethically challenging due to scale of analysis and complexity of decision making, also due to uncertainty and opacity of the work done by algorithms. The effect of these are becoming increasingly problematic as algorithms are progressively rely on learning capacities. (Mittelstadt et al., 2016)

Is the new technology neutral?

Wide spread of AI and ML technology progressively engulf the world over the past decades, which has resulted in spreading into new applications. The safety and fairness
restrictions therefore reaching the nadir by eroding the effectiveness and productivity. It is exclusively assumed that AI technology and ML is neutral by some people, but the reality is much far away from the situation. This is extremely possible and acceptable as ML algorithms invented by people itself and who all have bias. Therefore in order to fully expose the fruitful outcome with AI and ML, we need to unclouded the effective process by examining possible bias in the first place. In order to proceed this, examination of the origin and cause of the bias is extremely necessary. The bias possibly can occur due to two ultimate reasons. The first ultimate reason for the bias is the data that insert in order to be learnt by the machine. To state an example, (Alexander, 2018) when one intrude Google to search professional haircuts, Google has witnessed more racial and gender diversity. This case has been highlighted originally by twitter users back in 2016. The reason for the above phenomenon is crystal clear that Google results reflects editorial decisions that captured from articles in Google which prioritize white men. The way of algorithmic model development could be stated as the most possible cause for second mode to originate. This is human errors that occur throughout the development of the process and insertion of biased data within the training period. In order to site an example, charging of penalties to Facebook Company for excluding people from viewing certain advertisements based on traits that are on protected characteristics such as race, national origin and religion could be stated. (Wagner, 2019)

**Machine learning and criminal justice**

In order to judge whether to release or hold up defendants awaiting trial, judges must consider defendants flight risk or possibility to black sliding. Machine learning majorly support in this field so as to achieve decisions by considering recidivism. To state a genuine example, the system COMPAS which used in Broward County, Florida in pretrial hearings could be mentioned (Rieland, 2018). The system has invented to neglect any human bias arise. This
functions by employing 137 questions in a mode of questioner to apply a risk score for each individual. The outcome risk score value then used by judges in pretrial bearings to conclude if an individual can be released or detain in jail until their trial. The technique considered here is that if the machine can predict accurately who is the most probable in witnessing additional crime in the future. But the process could be fairer if it is taking sentencing decisions of the hands of humans and their biases. This will merely keep penal system cost at a low position by choosing people who should keep out of jails that is unlikely to commit future crimes. Machine prediction of suspicions history using algorithmic models are derived from data on the past cases, information on judge decisions and information on recidivism. Although this seems more accurate and confident, there are specifications which we have to be careful when considering support from machine learning systems as they inherit bias preferably through data. Therefore it is acceptable that machine learning cause various ethical dilemmas. Age at main crime, the year of crime, sex, nationality, sentence, previous number of crimes, whether probation was given or not are the features related demographics of the above scenario. By using the COMPAS system a defended Eric Loomis sentenced to six years in prison due to high risk status in performing crime in the future. However after the challenging by the defendant’s lawyers on the methodology of the algorithm, an investigation on the system detected that this program resulted in significant racial disparities. There was almost twice the probability of black defendants at future criminals than that of white defendants. The article ‘AI is now used to predict crime, but is it biased?’ by R. Reiland that published by Smithsonian magazine has declared that this system works as black boxes and the author questioned the transparency of the algorithms used in and how they function.

In the future, if the systemic function still going to be used by the courts without any development or correction it will be catastrophic by majorly affecting the future due to jailing
good-youth and releasing worst-youth that needed to rehabilitate, (disregarding racism) which will result in poisoned society subsequently.

**Financial sector and Machine Learning**

AI and ML dominating financial sector progressively by taking over mundane repetitive paper works to themselves and allowing the workers to perform in an outstanding and efficient manner which leads to profitable outcome. These unprecedented level of automation will lead to many benefits such as improved productivity, personalized customer service, more precise risk assessments. Despite all these ML in fraud detection and prevention could be stated as the top benefit. Unlike pure rule based software, AI based solutions smartly derive correlations in fraudulent patterns. Types of banks fraud that present nowadays are credit or debit card fraud, mortgage fraud, document forgery. ML is already being used to make or assist decisions in banking sector such as credit ratings, loan approvals. Since machine learning involves in predicting circumstances based on past events, they fail to predict results based upon behavior that has not been statistically measured. The biases arise in this sector due to minority population poorly represented in a data set. Therefore the AI should be rational, fair and dispassionate where algorithms replace judges, corporate leadership, loan officers, mortgage brokers which will enable fair outcome driven poorly.

A study conducted at UK Berkley in 2018 revealed that machine learning based interest ratings on Latin/African Americans is 6-9 basis points higher that of face-to-face decisions, which eventually unclouded the factor that existence of discrimination in machine learning. Furthermore, the earlier studies conducted in 1990 in Boston on data loan applications, with additional borrower data collected via survey by the federal reserve bank of Boston, revealed that with the credit history and loan value ratio, the minority experienced major rejection (28%) than that of white applicants rejection (20%) with maintaining identical property values and
personal characteristics conditions on them. Another recent study in 2015 by Cheng, Ling and Liu to compare mortgage interest rates for minority and non-minority borrowers, found that black borrowers on average pay approximately 29 basis points more than white borrowers with the difference larger for young borrowers with low education, subprime borrowers and women. (Bartlett et al., 2017) These factors basically enhance the racism bias exist in financial institutions. Basically the development must take care when building a model that will not inherit bias with human processes. In order to build up a fair model, the developers must consider fairness and ethical issues at the beginning. If the initial data set feeding to the machine is biased, the training will eventually biased too. According to the Jay Budzik on Zest finance (Azulay, 2019), the history of discriminating decisions in banking industry together with lack of transparency is the primary to blame for the bias. There are two major categories that can be seen in financial technology innovation. The credit scoring systems which take data inputs about a person such as spending behavior, location and age and will provide output score for them. Thereby conclude to accept or reject the credit card or loan application. This can be place into first category. Second category delineates customer point of interaction with a financial institution. Nowadays this process takes place predominantly via smart phones which will provide conclusion on application or provide some personalized recommendations which is done through machine learning.

There are various ethical challenges in relation to banking in using AI. It is extremely clear that banking customers tend to provide all richer data sets they own to organizations without any awareness. Using the social, financial and personnel data provided to the banks, along with the digital behavior of the customers, the bankers tend to provide advertisements according to the customer browsing history on the products and services interested in. therefore it is clearer that it is ethically unrespectable to manipulate the spending habits of people by the bankers.
In rejection of a loan application, the world before AI used to explain the reason for the conclusion, while in this era, AI systems just spew out the decision without any explanation or feedback. Here, the managers fail to explain the outcomes as the process is not merely by them as it is ML predicted. Therefore knowing algorithmic model extremely well by the related persons is quite necessary. Another ethical issue is implication is with cyber security. Through a robust password protection and user authentication. AI engage in a great war. On the other hand, AI could be used with malicious purposes as well. Automation can also lead to lesser awareness by the customers on behaviors of spending which will result in limitless online shopping, where one can cause serious financial damage. In order to improve the above ethical problems in the future, there are certain considerations that need to take into account. Everyone has the right in protection of their personnel data. The terms and conditions in a process that is concerned should be stated clearly to the users. There should be improvements in providing terms and conditions to the users in future in clearer manner about the data they providing. Explainable algorithmic model should be established in order to be transparent the relevant process to all the users which will thereby crack open the AI black box.

### Medical clinics and Machine Learning

In a world with a progressively growing and aging population, the health is the major valuable asset and a blessing to each individual. In order to improve diagnosis, treatment selection, health system efficiency, machine learning is used progressively in the modern medical field. In the healthcare field the patient’s diagnosis method and treatment selection vitality depend upon the patient’s history. The more clinical data history gathered, the more complicated data analysis will result. Therefore machine learning is extremely useful in facilitating the process, and save time. The mechanism also facilitating in earlier diagnosis of disease or recommend optimally individualize treatment plans. Even though the innovations
provide vast number of benefits, there are some challenges with the emergence of ML in the field. Where careful attention is necessary.

In order to ensure trust and transparency of using the models, there must be a careful attention to related ethical issues. Patients are usually not aware when physicians use computer-based decisions aids for their health problems and also they are rarely able to provide the source of physician’s judgment. These facts defining raise a questions ethically as if ML predictive algorithms are well known by physicians or any enough awareness practices are taken into consideration by the patients who are being treated fundamentally on ML. A careful examination in between patient and ML decisions system need to take into consideration as the problem with one individual most probably is different to the other individual body type, as well as due to the resources allocation at the root of the algorithmic model building set by the hospital administration rather than the physician. Privacy of the medical history and rich data is the next main concern. The patient should always be alerted regarding their data and how they could be used to formulate or improve predictive algorithms and those predictive tools might play a role in their care. The clinicians must need proper education regarding ML in order to communicate information about algorithm models decision based treatment recommendations which includes details about nature, risks and benefits of particular treatment options.

In 2016 researchers in Hidelberg (The Importance of Nuance, 2021), Germany use algorithms based on computer visioning technique in order to recognize melanomas based on clinical images. They trained algorithm with more than 100000 input of images of skin lesions labeled as malignant or benign. The algorithm predicted 95% of melanomas and 82.5% of moles, but the involvement of white skinned images on training was reported to be 95%. Regarding the evidences, Dutchen pointed out major problem “if the model was implements in boarder context, would miss skin cancers in patient of color? Would it mistake darker skin
Discussion paper

tones for lesions and over diagnose cancers instead? Or would it perform well?” therefore it is clear that outcomes of the false positive and false negative results going to be predominant which will lead to mistrust in ML which is a potentially efficient clinical tool, by doctors and patients. Machine learning used by hospital administrator in order to reduce the length of stay by patients. This is done by training algorithms so as to identify which patient most likely to discharge early. The above example was stated by Marshall Chin, professor of healthcare ethics at the unit of Chicago medicine, who subsequently pointed out a problem as “if you live in a poor neighborhood or a predominantly African-American neighborhood, you were more likely to have the longer length of stay. So, the algorithm would have led to the paradoxical result of the hospital providing additional case management resources to a predominantly white, more educated, more affluent population to get them out of the hospital earlier, instead of to a more socially at-risk population who really should be the ones that receive more help” (Wood, 2018). The most recent racial disparity occurred during Covid-19 season which showed pandemic far more deadly around black and Latino communities than which communities in United States (Community, Work, and School, 2020).

Prevent bias and enlighten fairer future with AL and ML.

In order to set a process free from bias, the root cause for the preconceptions should be identified. Prejudice that currently in the process should be revealed regarding past and current algorithmic practices taking place in a particular company by questioning openly and honestly by the results from investigating the model and potential outcomes of the process. If any bias identifies, the companies could block them by removing any unnecessary data or else by removing components of input data set. Instead providing any acknowledgment on race or gender, a manager in a bank could address the zip codes, type of the car, or first names in addressing late payments or defaults at initial stage of building the model, could be stated as
Discussion paper

an excellent example for the situation. Therefore this will offer customer with much fairer
opportunities in payment plans. “It is always safe to assume that bias exists in all data. The
question is how to identify it and remove from the model” (Oliver Wyman, n.d.).

Allowing machines to train by itself by throwing enormous amount of data will cause
machine to run wild by having unconscious bias and also will be much harder to explain the
performance of machine learning models. This could be overcome successfully by
implementing traditional methods as well as sophisticated challenging machine learning
models in assessing data which will be more precise. Therefore it will be able to confirm if
much complex machine learning method is more accurate in comparison to the traditional
method. This will eventually enable to verify machine learning tool’s balance between
transparency and sophistication.

Machine learning will be trained to perform independent, despite it is ‘must factor’ to
train the models will new data sets as the model operating is constantly changing, and also it is
a huge necessity to implement careful planning to avoid unintended biases in the future process.
If negative bias going to retain in the process, it will impact the effectiveness of the machines
decisions in the future. Otherwise, creators risk undercutting machine learning’s positive
advantages by building models with a based mind of their own.

Conclusion
Together with human bias created in the root of building algorithmic models, it is
understandable that machine makes biased decisions on their own by categorizing data fed into
the machine. Therefore it is extremely important to identify the particular bias before
introducing the process to work for the society. It is clear with the examples and explanations
provided for the bias occurred in recent years that how the efficiency of ML will destroyed,
even though it is providing great deal of service with the day to day work load. Catastrophic outcomes will be prominent and predominantly destructing the whole society progressively will be efficient in the future generations as internet and cyber performance will be vastly increased.

References

Alexander, L. (2018, July 2). Do Google’s “unprofessional hair” results show it is racist? The Guardian. https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist

Azulay, D. (2019, December 13). How Lenders Can Win More Business with Machine Learning – with Jay Budzik of Zest AI. Emerj. https://emerj.com/ai-podcast-interviews/how-lenders-can-win-machine-learning/

Bartlett, R. P., Morse, A., Stanton, R., & Wallace, N. (2017). Consumer Lending Discrimination in the FinTech Era. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.3063448

Community, Work, and School. (2020, February 11). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/index.html

Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. International Journal of Computer Applications, 115(9), 31–41. https://doi.org/10.5120/20182-2402
Foote, K. D. (2019, March 13). *A Brief History of Machine Learning*. DATAVERSITY.  
https://www.dataversity.net/a-brief-history-of-machine-learning/

_How different are Conventional Programming and Machine Learning? _.-. (2021, March 5).  
Tech Blogger. https://contenteratechspace.com/how-different-are-conventional-programming-and-machine-learning/#:%7E:text=Machine%20learning%20is%20also%20known%20as%20augmented%20analytics.,are%20fed%20to%20the%20algorithm%2C%20creating%20the%20program.

Lee, E. (2021). How do we build trust in machine learning models? *Available at SSRN 3822437*.

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine Learning in Banking Risk Management: A Literature Review. *Risks*, 7(1), 29.  
https://doi.org/10.3390/risks7010029
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Oliver Wyman. (n.d.). *The Risk Of Machine-Learning Bias (And How To Prevent It).* https://www.oliverwyman.com/our-expertise/insights/2018/dec/risk-journal-vol-8/rethinking-tactics/the-risk-of-machine-learning-bias-and-how-to-prevent-it.html

ProPublica. (2020, February 29). *Machine Bias.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Rieland, R. (2018, March 5). *Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?* Smithsonian Magazine. https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/

*TechCrunch is now a part of Verizon Media.* (2017, August 8). Extra Crunch. https://techcrunch.com/2017/08/08/the-evolution-of-machine-learning/

*The Importance of Nuance.* (2021). Harvard Medicine Magazine. https://hms.harvard.edu/magazine/artificial-intelligence/importance-nuance

The no free lunch theorem and the human-machine interface. (1999). *IEEE Control Systems, 19*(3), 8–10. https://doi.org/10.1109/37.768535
Wagner, K. (2019, March 28). *HUD says Facebook’s ad business leads to housing discrimination.* Vox. https://www.vox.com/2019/3/28/18285782/facebook-hud-lawsuit-charges-housing-discrimination-ad-targeting

Watney, C. (2018, May 9). *It’s time for our justice system to embrace artificial intelligence.* Brookings. https://www.brookings.edu/blog/techtank/2017/07/20/its-time-for-our-justice-system-to-embrace-artificial-intelligence/

Wood, M. (2018, December 2). *How to make software algorithms for health care fair.* UChicago Medicine. https://www.uchicagomedicine.org/forefront/patient-care-articles/how-to-make-software-algorithms-for-health-care-fair-and-equal-for-everyone