Letter of Transmittal:

Curtis Jaekl

March 26th 2023

To whom it may concern,

Inclosed with this letter is a report on the ethical issues with artificial intelligence (AI). In addition to my report there is an executive summary for any shareholders wishing to get a brief insight into the report.

The report is structured with an explanation about three common concerns with AI. First being fairness, after explaining what is fairness versus unfairness in AI the solution provided is a mathematical and data based approach. Next, the report stresses the importance of transparent AI development and uses. Which brings forth a valuable idea called OPEN AI that allows for everything from the code to how a company is using their AI to be available to the public. Finally, privacy is a large issue in AI as the data to keep the current AI relevant has to be collected and up to date which raises the concern on how to maintain privacy while still collecting useful data. Addressing the issue of privacy there will be a government of Ontario agenda brought forth to shed light on the issue and important solutions.

If you have any questions or concerns please reach out to me via email gmjaekl@my.yorku.ca

Thank you for your time,

Curtis Jaekl

Executive summary:

To whom it may concern,

Included in this report is a discussion on common issues that occur in big companies when using Artificial intelligence. The main points are about fairness, transparency, and privacy. These values are important to uphold as they will ensure a morally sound company. Ensuring that you hold good practices will ensure a strong business and no ethical faults can happen. Having good practices can save money by preventing lawsuits and ensuring customer satisfaction.

Fairness is crucial to everyone, you, the business and the user. By ensuring that developers choose the right data set and represent attributes of the population correctly to fit the needs of the AI. Will result in a fair and useful AI.

Transparency, with an AI is necessary to maintain the trust of the users. Keeping them understanding and informing them. Will create a feeling of security when using your AI.

Privacy needs to be kept for everyone. This way data is secured and that the AI can continue to ethically collect data. As a result the newly collected data will keep the AI relevant.

In conclusion, maintaining good ethical practices will be good for business in many ways.

Kind regards,

Curtis Jaekl

Table of Contents:

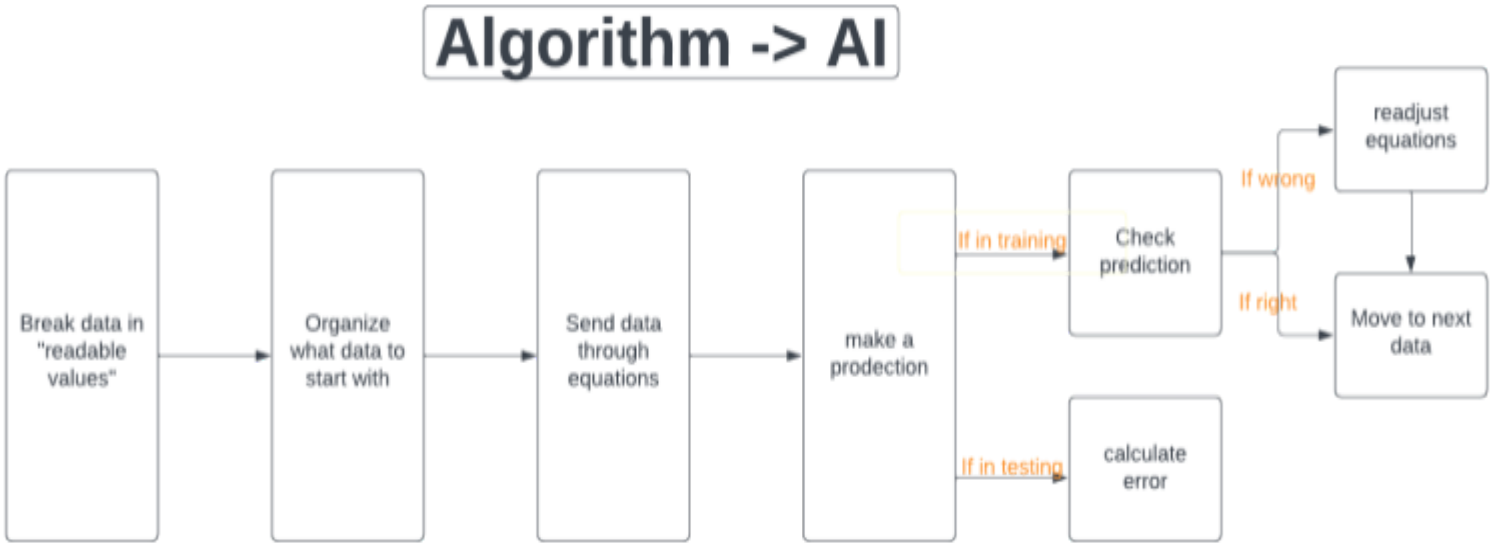
Introduction:

Majority of people interact with artificial intelligence (AI) more frequently than they realise. For example, the “for you page” on facebook or when applying to a job. The “for you page” uses AI to try and show you content it predicts you will enjoy. Companies use AI to go through resumes and choose the candidates. Since AI plays such an important and constant role in our lives it is important to maintain good ethical practices. The ethics should be established to create fairness for the correct audience. In addition, transparency is important for the public to understand the development and use of AI. Finally, privacy, To maintain a relevant AI it needs relevant data. Collecting data should be done in an ethical way to ensure privacy of the user is kept. Furthermore, each of these important ethical needs can be met with the correct adjustments to the technology and the laws surrounding it.

AI is the recreation of intelligence with mathematical computation that will navigate data to then establish patterns. Once recognizing patterns, AI can then make predictions with and about other known data. An algorithm is the mathematical computation that takes input from a data set. First an algorithm has to be trained with a data set. To accomplish this, a data set is broken down into values, for example, each character inside a document will be given a value. These values then go to the first step of the algorithm where values are sorted depending on how a developer structured the algorithm. This is like a human deciding what chapters to read in a

textbook when trying to learn something. Next, the values go through numerous equations to be weighed on importance. Sorting information can be interpreted as reading the textbook and taking notes. Towards the end of the algorithm the AI will send out a result. This result is then compared against the known actual result. Similar to practising homework questions to test your knowledge. If the algorithm got the answer right, it gets an opportunity to rest and goes onto the next piece of data. However, if the algorithm is wrong, there has to be a “review” which means it works backwards (back propagation) to then adjust each equation. After the training of an algorithm it is good practice to then test it by using some known information and hiding the results from the algorithm to find its percent error. Figure 1 should provide illustration on the flow of training and testing an algorithm.

Figure 1.



The current paper presents a novel demonstration in which Python and a few libraries like Numpy and Sklearn are used to create a common algorithm called a linear regression model. This linear regression model was used to predict Canada's

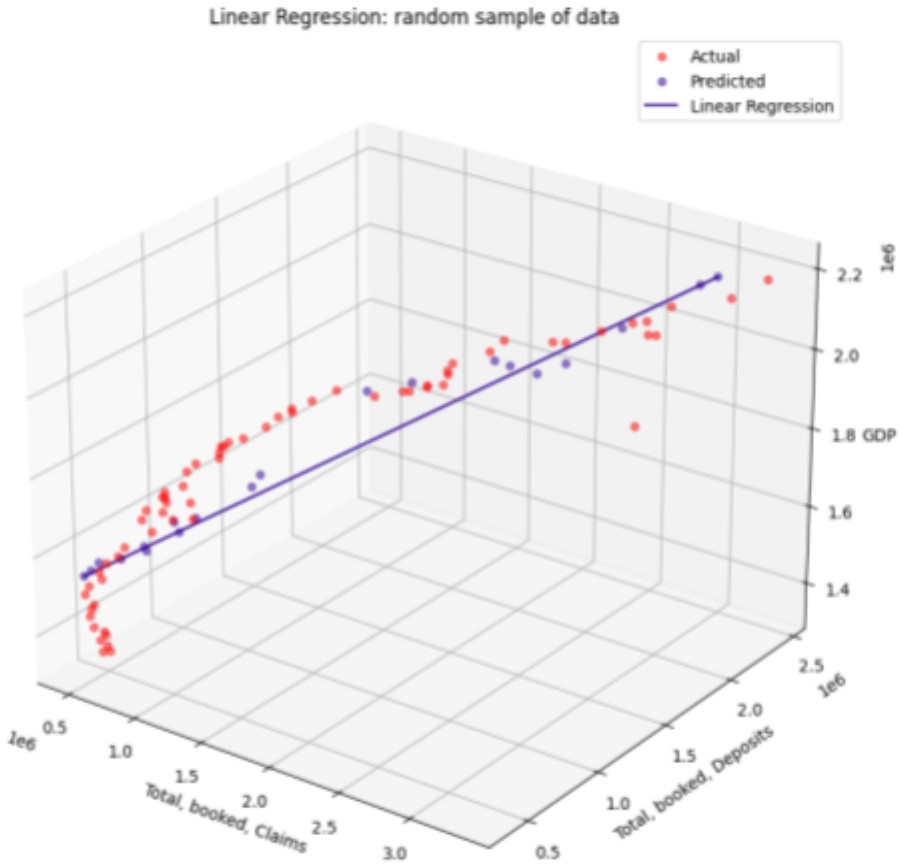
growth domestic product from the national banks deposits and claims. The deposits and claims can be taught as the debit and assets of the country. Linear regression models are motivated from the well known linear equation eq(2)

$$Y = mX + b \qquad eq(1)$$

$$Z = c_1X_1 + c_2X_2 + b \qquad eq(2)$$

The equation for the linear model is eq(2). Z is the output or the prediction, depending on if you are training, testing or using the model. C_n is the coefficient that can be adjusted to improve the accuracy of the model. X_n is the known parameter in this case “Total book, Claims” and “Total book, Deposits”. B is the intercept. For a linear regression model it is the mean value of the correct known Z values. In Figure 2 is a graph of the original values and the predicted values along with the line of regression.

Figure 2.



In Figure 2 this graph shows as claims and deposits increase so does the GDP. It is important to note the process in how the data was used. Eighty percent of the population was masked and used to train the model. Which resulted in the coefficient being [0.59472508, -0.18060489] for claims and deposits respectively. Coefficients are important as they weigh the data. If you want something to affect your prediction more, have it weighted closer to one. The intercept for the line is 1419965.3. The equation would be ‘GDP’ = 0.59472508(‘Total book, Deposits’)-0.18060489(‘Total book, Claims’)+1419965.3 where GDP, Deposits,

Claims are in the 10^6 dollars. The average error of 10000 tests for this model is 4.89% and a variance score of 0.79. Having a variance score closer to one means your data and model are relevant and good for predicting the outcome.

Introduction to fairness:

It is crucial to understand all sides of a situation to generate fairness. This being said it is difficult to generate a single definition of fairness as it may be unique to a situation. One example where it is important to establish an understanding of fairness is banks and insurance policies. Car insurance is less expensive to people who are less at risk of getting in an accident. It is well known that males 18-25 are more likely to drive carelessly than their female or other gender counterparts. Recognizing that the insurance company's job is to make money and provide insurance, the insurance company should charge the males 18-25 more than the females in the age range to ensure they don’t lose money if they were to make a claim. An example of application of AI to an insurance company is a customer could file out an application online which then can be processed immediately by an AI algorithm. The AI will then generate a policy in minutes. The algorithm would generate the policy charge with certain known information like past driving record, age, gender, type of car that is being insured. When the AI gives a policy it would

likely be more expensive on average for males 18-25 relative to the other genders. Is this fair? No. However, it is important to recognize that this age group is more likely to get in a car crash. The insurance company should cover any risk from these drivers by charging more. In conclusion, car insurance is a business. Applying AI to insure people should be on an equity basis and not an equality basis because insurance is a business. Losing money is not the goal.

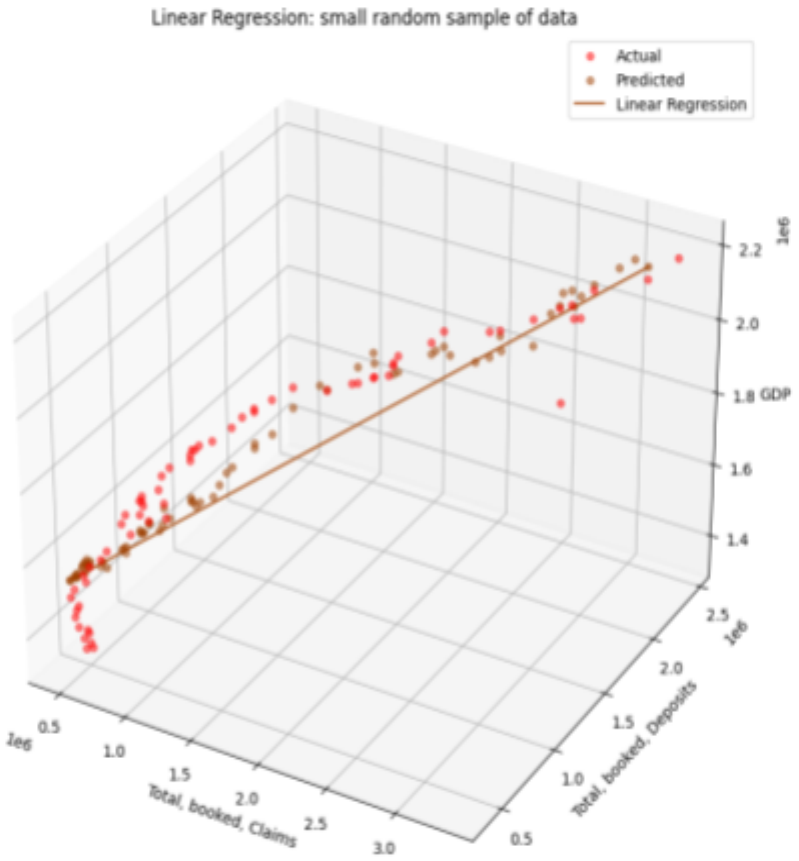
It is crucial to know when to enact fairness. Next, understanding what causes unfairness. *Addressing Bias in Artificial Intelligence in Healthcare*[1] talked about an algorithm that was to predict whether a woman would test positive for breast cancer or not. After some research into the algorithm it was more accurate when predicting results for white women then black women. This isn't fair, everyone deserves quality healthcare. The bias towards white women happened because most of the data used to train the AI was from white women. Black women were underrepresented. The point here is there is a clear bias which was from the data provided to the AI. It is intuitive that we provide great health care to everyone, so we must translate this into our AI design and emphasise that this is a time to provide fairness and represent all audience demographics equally.

One common cause of bias is poor use of data. An example of poor use of data is misrepresenting attributes or subjects in a population. To misrepresent you can overrepresent white women in the case of healthcare AI. Another way to misrepresent is to have an extremely small sample. Figure 3 shows the linear

regression model devised in this paper trained with only 15% of the data and not 80% as before.

Figure 3.

In figure 3 it is not noticeable that the model was severely under trained (due to the small sample size) and that data got misrepresented. Testing the model gave an average error of 4.74% which is slightly less than with the normal practice of



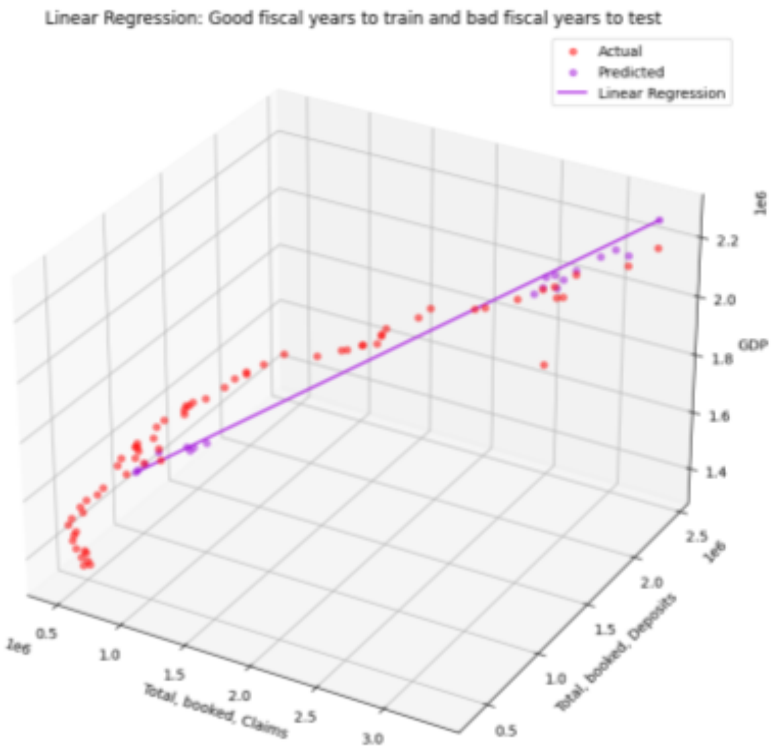
masking 80%. In addition, the variance was increased to 0.82% which is not good as

over time there could be drastic outliers that can be avoided if the model was trained with more data.

Next, Figure 4 the data that was used to train this paper's linear regression model was biased by only giving good fiscal years. Then tested with terrible fiscal years. Examples of bad years are the 2008 stock market crash, the 2020 covid-19 lockdown then recovering from the lockdown.

Figure 4.

Figure 4 shows that when given data that is inherently good or excellent to train the algorithm. It performs poorly as they are unprepared for bad situations.



Introduction to transparency:

Transparency in AI is a little more straightforward. It is important that there is transparency in AI as a whole. Some examples include the algorithm itself, or how a company may plan to use their AI, what data was used, how the collected data, etc. The list can be really long when thought about. For the record, transparency is important so that the public is well informed of what is going on and so that there are no awful things happening in the background. Information has to be public as if it is policed privately then there could be collusion that would cause further issues. However, there is the issue that if everything is made public then there is no competition between companies. To tread this fine line between ethical business and bankruptcy, and avoid scandal, a comprehensive level of public information on AI policies and biases is necessary.

Introduction to Privacy in AI:

Privacy in AI is by far the most important out of the three values discussed. It is crucial to maintain ethical privacy policies in the space of AI. The area of importance is data mining. Data mining subprograms are developed into many

applications or websites or other software. These subprograms run in real time and collect new data constantly. The data collected is significant to keeping algorithms up to date. Algorithms have to be trained with data. The more relevant the data used the better the results will be. Many companies mine data. For example, Alphabet, Google's parent company, collects data from its google accounts. Similarly, FaceBook’s owner Meta uses data from all its social media platforms and Apple from IPhone users[7]. The issue is how, and what data is being collected. In the past, there have been issues with people not being explicitly told that their data was being collected which breaches privacy[7] and shows no respect for the user.

The action to take:

The grey area called fairness and dealing with it:

As we define fairness to be continuous and not discrete, there are a few common situations that occur. AI is only black and white since on the lowest level it is literally binary. The first situation that was discussed is when is it okay to have bias whether it is intentionally programmed or accidentally through data misuse. The first example mentioned is the case of car insurance. as it would be wise for the insurance company to have bias to certain drivers that would be more likely to get into car accidents. Next, there is the unintentional bias that is not necessarily directly harmful

or malicious . However, it should still be recognized. Figure 5 is all three of the previous figures plotted on one graph.

Figure 5.

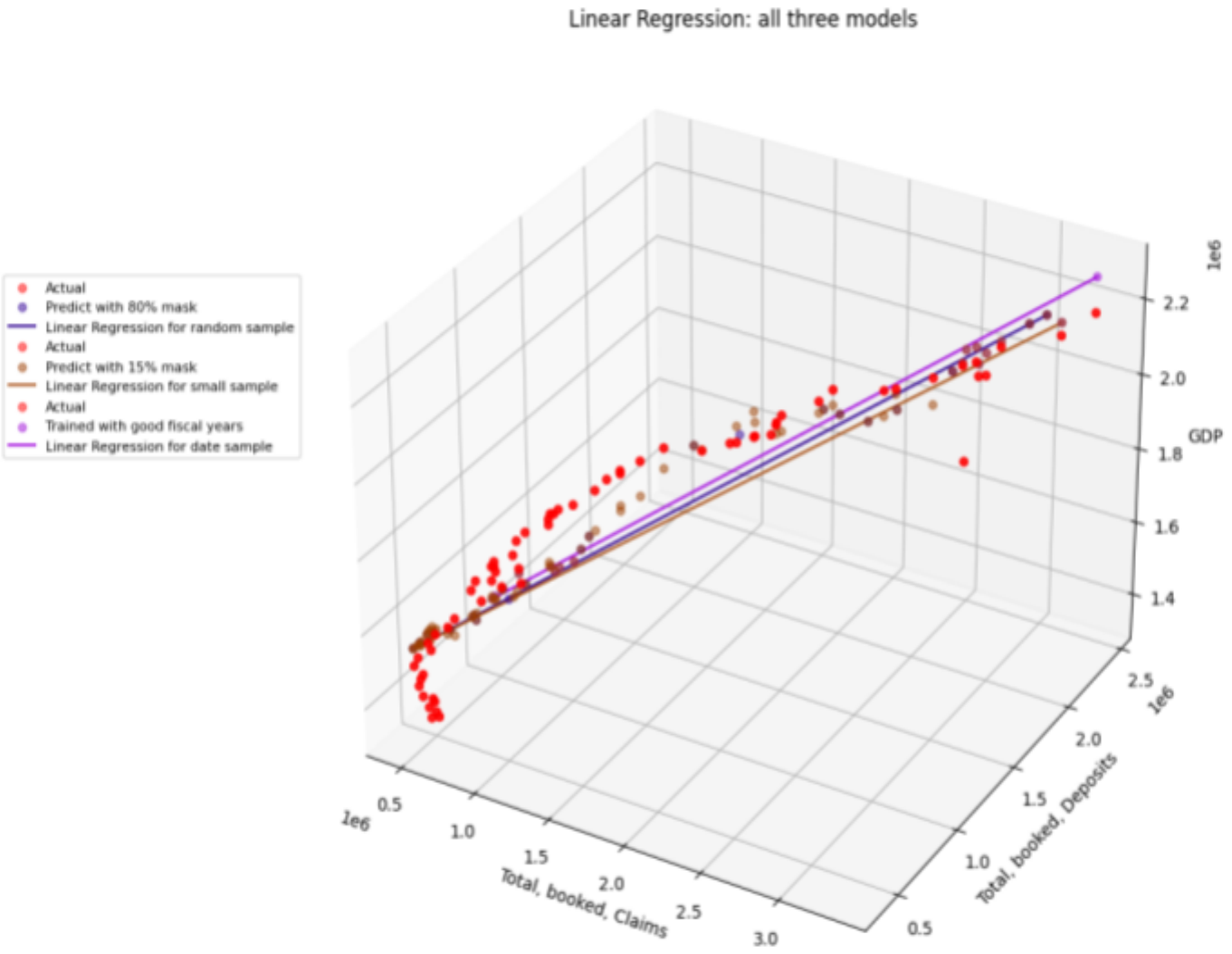


Figure 5 is a clear representation of how these biases have small yet meaningful impacts on the predictions of the models. The linear model trained with

only good years clearly over estimates. In contrast the small sample size trained model under predicts relative to the best practice one of the 80% mask.

Going forward developers can recognise these differences along with knowing what the model is trying to predict and accomplish. Can use this as a tool in the car insurance example. Developers can use the males in this age group that pay more to train the data. So that in the future when a customer enters male and they are in this age group it will present a larger policy cost. Or for the health care example use this as a sign to select better data to equally represent the population. In the case of the health care example developers can consider if the sample represents the population in a meaningful way.

Selecting meaningful samples will improve the overall performance of the algorithm. Finally the obvious is bias that does directly cause issues to a group or an individual. An example of this would be, banks in the United states when giving a loan use a unique variable that affects whether you get approved for a loan or not. The independent variable is where the person lives. Essentially where someone lives affects how likely they are to get approved for a loan or not. Currently people can apply for loans online and in person which both go through an AI to determine their eligibility. People apply for loans for many different reasons, some to buy a new boat, some to buy a new house, or want to upgrade their business. For people that live in a neighbourhood that might not be looked upon as a good or safe might be applying for a loan. If they were to be approved for a loan they could use it to move away from the bad neighbourhood. If they were to get denied this loan they may continue to be stuck in a cycle of not being able to escape their neighbourhood and continue to get denied for loans. To sum up, people apply for loans. Their applications go through an

AI to get approved or not. If they apply to improve their way of life or move from a bad neighbourhood. If they get denied the opportunity to move out of their neighbourhood then the cycle repeats. This is a tougher one to be discussed. There are two arguments to be made. Similar to the car insurance company it is the bank's business model to make money. It would not be wise to be lending money to people who are more likely to not pay it back, that is just common sense. On the other hand the AI does not know each individual's story. So it is unfair to deny people the opportunity to create a better life for themselves[6]. The question then begins to emerge: how can we stop this bias from affecting people in a negative way. Once again it is hard for an algorithm to assess an individual circumstance. It is even more difficult for an AI to genuinely believe a human if they would or would not pay back a loan. So how can we accomplish AI fairness? There is a simple solution of just not using it. This does not work as AI does such a good job at finding patterns that humans may not detect. One idea is to notice when the bias is caused by a data set. Then using a correct for the situation sample from a population. Another idea is adjusting the algorithm behind it. For instance, if we adjust a coefficient in the algorithm to not have such a big weight on the neighbourhood or even take this variable out entirely. This can simply be done by reprogramming it on a lower level. More specifically readjust the equation for the computer to compute outcomes. In conclusion, bias is not the fault of the AI, but by the algorithm developed by humans. It is our job to use trial and error to figure out an algorithm that weights certain variables less. To provide a fair opportunity to improve their life.

Transparency issues and a common sense solution:

Constantly there is miss communication between companies and the public. They might not want to expose their secret recipe to stay ahead of the competition. Or don’t want to reveal a cpu design to stay faster and better than the opposition. However, transparency with AI is crucial. It is important for the public to know and understand how it works to ensure their privacy, just and of course they're safe. On numerous occasions big companies like facebook or google have miseused ai or misled the public about their AI designs[5]. Transparency is not a new concept. It is about being open. Which use leads us to an important technological solution. Again it is not a new idea. OPEN AI(OA)[3]. OA this is as straightforward as the title suggests. Let their AI be open for the public to review. To take it a step further alongside releasing the code, provide a detailed report about the maths of the algorithm, and how the company intends on using it. The code and the reports allow for a more wide range of people to be able to review the work. IN addition, if it is public any one for that matter would have access to it. There are some counter arguments here to also consider. The first one is: Doesn’t this defeat the point of capitalism? It doesn't allow one company to keep a crucial advantage it might have over another. But this argument is weak. For there are laws like copyright and patents. There can be patents for intellectual property, and technology so AI can fit in nicely there. In addition, there are laws preventing monopolies. The Sherman Act to be specific. Having OA allows there to be no monopolies. Going back to OA, it is not new as many companies already use the ideology, examples, GITHub and open source software that allows people to upload and share code or use other people's code for

their own needs. This can be extended to corporations. Bringing forth new ideas is that OA allows for the continual growth of the community as people can learn from the existing code or algorithms, and build on existing ones. In conclusion, Transparency is a constant struggle with big corporations. However, it is important to have it as it will keep users safe and allow for new growth in that space. Transparency will not hinder the current global business model in technology.

Privacy and the big agenda:

Privacy has to be addressed as not many people realise the issues surrounding how data is currently being collected and used. First, to apply an understanding AI needs to be trained and it has to be trained using data that has known outcomes, which means it is from the past. Or if it is some kind of adaptive model like instagrams advertising AI, it tries to show you ads that you are mostly to click on. The data collected for this is in real time as there are no known results. It simply uses stuff you clicked on last to judge what you click in the future, which makes it impossible for anyone to predict the future with one hundred percent accuracy. For the other type of learning it is given old data sets with known answers and the AI works as described before. In both cases the data is “mined”. Data mining refers to the collection of data for the specific use of profiting from weather that is directly or indirectly to train companies algorithms[7]. It is no secret that this is happening however it is important to get it right, to maintain privacy. The main issues

are these companies leak information that they have collected about people and they might be collecting extremely private information without your permission, for example health information. The main issue might be the security of information. However this is not a report about cybersecurity. The solution has to come from within the AI itself to follow the trend of this report. The important solutions are having no collection of data be the default option rather than having data collect unknowingly. Next there is the transparency, keep it open about what is being collected from there ai. Finally that is a positive sum gain, meaning that the data collected also helps the users too. For example, if the data collected is used to give ads based on users interest it can also give ads to help them, for example, an ad for suicide prevention[4].

The results:

There are three important areas of ethics in AI. First being fairness. It is truly difficult for a computer which operates in ones and zeros to understand the complexity of humans which makes fairness left to be enforced by humans. Data bias can create unfairness and is it up to the developer to understand the right time to use bias. The car example is the right time to use bias or the author's algorithm trying to illustrate bias. The wrong time to use bias is when trying to equally represent a population. AI in health care should be created equal for the population, after all

healthcare is available to everyone here in canada. In addition, it is crucial to seek out areas of improvement for fairness as bias can happen unintentionally. Second, transparency, it is not a new idea, it has to be revisited thoroughly to ensure that companies have quality assurance that will ensure the safety of the public. Thirdly, privacy issues with data mining. Data mining is necessary to keep a good user experience and to ensure that algorithms are kept relevant. But collecting data should be done ethically and with the user's privacy in mind. To reiterate the solutions of fairness. Knowing the situation and context of the use of AI there is a time to use bias and time to not. Understanding the area it is up to the developer to choose the correct samples to train the AI. This can be paired with manipulating coefficients to have different weights on the data to generate more equitable predictions.

Next, transparency needs to be given from big tech companies. The idea of open AI allows for everyone from a common user to an expert in the field to have more confidence in their user experience. Finally, for privacy the user should be able to choose whether to share their data or not. Along with their personal information not being intruded upon. Once the user consents to giving their data it needs to be kept private and safe.

References:

1. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. *JAMA.* 2019;322(24):2377–2378. doi:10.1001/jama.2019.18058

2. Fredrik Heintz, Transparency in artificial intelligence, Volume 9, Issue 2, May, 2020, 10.14763/2020.2.1469

3. CHI EA '20, Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, April, 2020 Pages 1–8https://doi.org/10.1145/3334480.3383051

4. Ann Cavoukian Ph.D, Privacy by Design, Information & Privacy Commissioner, Ontario, Canada,

2011,https://www.ipc.on.ca/wp-content/uploads/Resources/PbDReport.pdf

5. Jain, P., Gyanchandani, M. & Khare, N. Big data privacy: a technological perspective and review. *J Big Data* 3, 25 (2016). https://doi.org/10.1186/s40537-016-0059-y

6. Feuerriegel, S., Dolata, M. & Schwabe, G. Fair AI. *Bus Inf Syst Eng* 62, 379–384 (2020). https://doi.org/10.1007/s12599-020-00650-3

7. WSDM '19: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, January 2019, Pages 840–841, https://doi.org/10.1145/3289600.3291384

Appendix

Appendix A

Table 1. Screenshots of My code

For a link to the Github for the code. Click here.