# Effects of Parameter Changes in a Simple Genetic Algorithm and Relevant Ethical Issues

**Suleima Abbara - 19020111**

## 1 INTRODUCTION

Artificial Intelligence is a term that is being used more often, and is implemented in many areas and fields, however, the lack of sufficient guidance on how to use machine learning is causing a lot of overlooked consequences of misuse. Multiple examples of such are discussed later, followed by the implementation and testing of two problems in a simplified evolutionary algorithm. The algorithm is meant to find the best solution for each problem, with parameters that control the output and its values being varied to project the effects they have. The equations will be tested in multiple runs, outputting the values of the generations in tables in a text file, as well as the use of graphs for better visual aids. With each run, the range of change of the parameters will be tweaked to get closer to the optimum solution.

## 2 BACKGROUND RESEARCH

AI in Social Media: Manipulation of Behaviour:

In recent years, the world of social media has witnessed the viral spread of misconceptions, conspiracy theories, and political propaganda on many AI driven platforms. Common examples include the 2016 US presidential elections being affected by the online spread of false political information (Walch, 2021), the misbelief that the Earth is flat (Landrum, 2021), and perhaps most recently is the fabrications and distorted stories on many social media platforms during the COVID-19 pandemic (Cinelli, 2020).

With 3.78 billion people using social media in 2021 (Statistica, 2021), 2.85 billion monthly active users on Facebook, and 211 million daily active Twitter users (Snider, 2021), it is concerning knowing the amount of data those platforms collect from their users, claiming that they are needed for the algorithm to personalize users' experience, however, this also makes users easy targets for unsupported information; scientific, political, or otherwise, potentially creating "filter bubbles*" of false information. The use of fake news in order to manipulate the behaviour and opinion of entire groups is easily achievable with modern use of AI. (Kaplan, 2020)

It is clear that AI can be harmful to communities with the absence of implied ethical rules, therefore AI systems need a framework with a clear set of guidelines. As social media is globally used, a proposed solution must be of global scale, to have regulators internationally and governments create the ethical rules, and set implications of consequences that penalise companies and developers collecting and using unconsented personal data for hidden purposes like manipulation of behaviour. As for combating the spread of misinformation, social media platforms have adopted several means to combat this issue, such as YouTube removing misleading content about the COVID-19 vaccines (YouTube, 2021), while Instagram has adopted third-party fact-checkers, labelling posts which are rated "false" or "partly false", and reducing their visibility. (Instagram, 2019)

*Filter bubble: a situation in which someone only hears or sees news and information that supports what they already believe and like, especially a situation created on the internet as a result of algorithms (=sets of rules) that choose the results of someone's searches. (Cambridge Dictionary | https://dictionary.cambridge.org/dictionary/english/filter-bubble )

AI in employment:

Looking at the way the economic system works, it is mainly dependent on growth and productivity to increase wealth. However, by using AI, a company would need less employees and smaller teams to accomplish tasks, meaning the revenue will be given to less people, as the wealth increases with increased productivity. Therefore, those who own AI driven companies will get most of the wealth. Currently, AI runs in an unregulated environment where the responsibility of consequences is often lost and unclear. Since AI algorithms are easily adapted and spread, it means that changes on the market are faster and more impactful, as those algorithms replace human thought and information processing (Muller, 2021). Therefore, the transition to an era run by AI will cost jobs, so who bears the responsibility of this cost? And who is responsible of the distribution of the benefits of AI?

Furthermore, the industries that are adapting technology and automation (such as AI) showed a decline in the wages of workers who perform predictable and routine work, as their tasks are easy to replace. The NBER estimates that the 50% to 70% of changes of the US wage structure is related to this reason (Acemoglu et al, 2021), as companies replaced humans that did routine work. On the other hand, the increase of productivity and wealth creates more opportunities and more jobs, and there is no denying that human labour will still be needed, such as companies like Amazon hiring more people during the Covid-19 pandemic despite investing to use technology in its warehouses to cut down on costs. (Amazon, 2020) that doesn't mean that the main issue isn't there, as income inequality and job loss are still prominent effects of technology and the rapid use of AI needs to be addressed in clear ethical guidelines to define which tasks cannot be replaced by AI, which ones can be assisted by AI, and which ones can be completely overtaken by it without having drastic consequences.

## 3 EXPERIMENTATIONS

The evolutionary algorithm used to test the two minimisation fitness functions is a simplified version of a genetic algorithm that can be represented with some steps. The first step is creating a population of individuals with a string of randomised values as the genes. A population array will include individuals, which are arrays of randomly generated float numbers between the MIN and MAX of the fitness function. That, is the first generation. That generation is then evaluated using the fitness function. In a number of loops, called generations, the selection process starts; by randomly selecting 2 individuals from the population, comparing the two individuals and choosing the one with the smaller fitness value, and add it to the offspring array. Once finished, the temporary array will hold only copies of the individuals in the population. Now the crossover process begins by taking the first two individuals inside, choosing a random point in their genes, and swapping the values after that random point between the two selected parents, and moving on the next two individuals, repeating the same steps. Then, to start the mutation, the algorithm adds or subtracts a small random value (mutation step) from a gene at a specified rate (mutation rate). At this point, the offspring is created. the offspring array is then passed through the fitness function again, followed by the algorithm finding the worst individual in offspring, and switching it with the best individual of the population. Once this process has looped over multiple generations, a graph will show the values reducing over the specified generations.
This report is meant to highlight the effects of changing the parameter values in while searching for the solution, so the identified parameters in the used algorithm are:

- P = number of people in a population.
- N = number of genes in an individual
- G = number of generations
- MUTSTEP = the value of which the gene will be altered (MS)
- MUTRATE = the probability of the mutation event occurring (MR)

In the given problems, (x) is the gene, (d) is the number of genes, and (i) is the number of the referred gene in the individual. The algorithm runs 10 times and averaging the results to get a more accurate estimate of each alteration since it's a stochastic algorithm.

First Problem:

$$f(X) = \frac{1}{2}\sum_{i=1}^{d}(x_i^4 - 16x_i^2 + 5x_i)\, where\, -5 \leq x \leq 5$$

This fitness function loops through the genes of an individual, performing some calculations and adding them together then dividing it by half. To start the experimentation, some parameters are set to be able to compare the values properly, so the set parameters are N=20, and G=100. MS and MR will be at 0.1 as a start. An initial run of the code for a population of 50 at (MS=0.1 , MR=0.1) shows the best solution moving from an average of -430 to -700 (figure 1). The averages in the first generations vary greatly from the averages in the final generations. Changing the population to a higher value (P = 100) resulted in a lower best solution, reaching closer to -740 (figure 2). Changing the population back to 50, then increasing N to 30, gave a very low solution value (nearly -1000) (figure 3), which allows to assume that the higher the number of genes, or the higher the population is (the bigger the scale), the better. It is the same with increasing the generations to 200 (N = 20, P = 100), the best value reached -750 (figure 4), however the higher the generation means the longer it takes to reach the best solution.

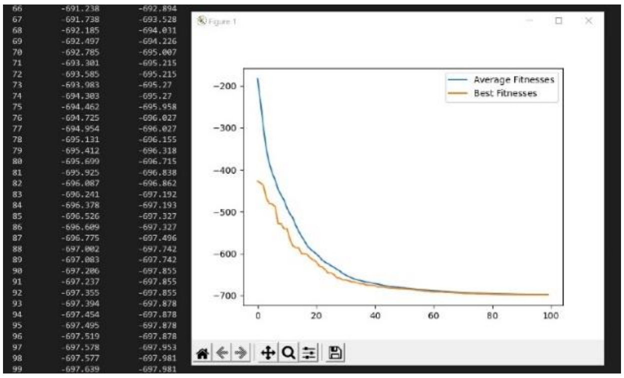


**Figure 1**

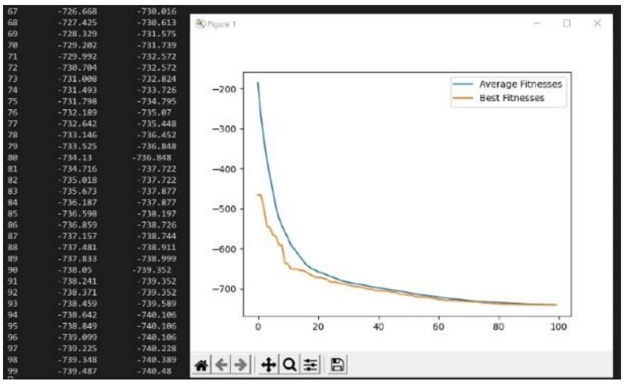


**Figure 2**

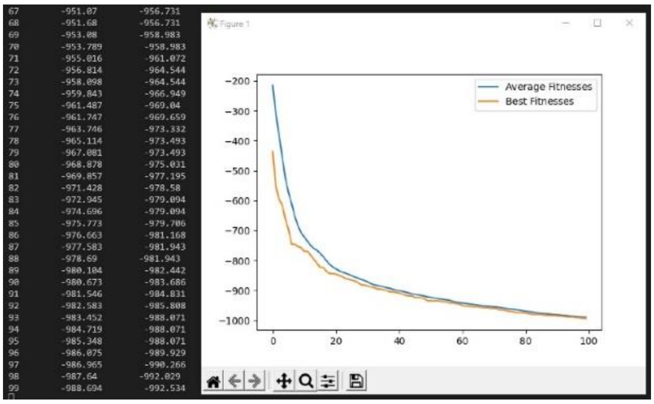


**Figure 3**

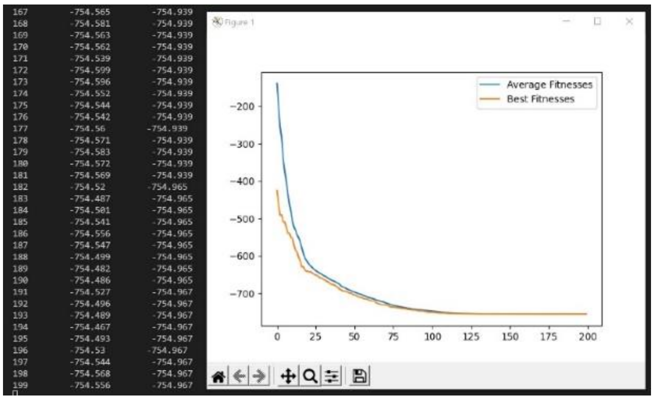


**Figure 4**

The next parameter changes will be in the mutation step (MS) and the mutation rate (MR). They will change from 0 to 1, by adding 0.1 in each run, producing the main table for the first problem (Table 1 in appendix A) put in a surface graph (figure 5), which holds the averages of the best solutions of 10 runs for each parameter change. (P = 200, G = 100, N = 20).

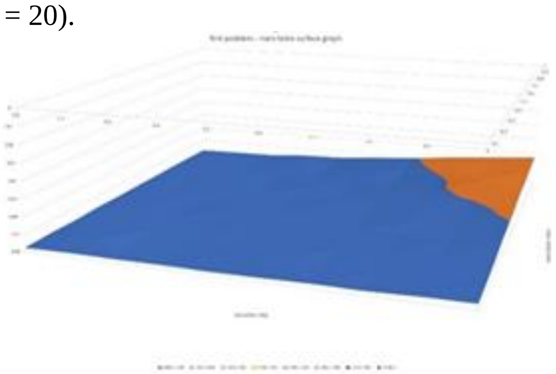

**Figure 5**

When looking at the best solutions in (table 1, appendix A.), it is noticed that the best values (highlighted in dark yellow) are scattered, mostly concentrated when MS is small. Big values are focused at the bottom right of the table, when MS and MR are both closer to 1. The smallest values seem to be in two dips, the first one is when MS is 0.1 with most MR values, and the second one is when MS is between (0.3 , 0.8), while MR is between (0.1 , 0.2). However, the lowest value is (-779.606) when (MS = 0.5) and (MR = 0.1). Attempting to reach an even better solution than this one, the algorithm is run again with the previous MR and MS, outputting a table (figure 6 for graph & table 2 in appendix a.) which shows the optimum solution reaching -780 twice.

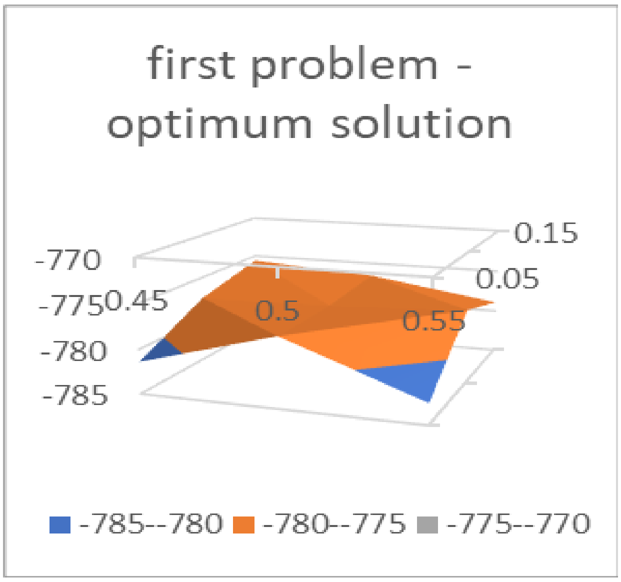


**Figure 6**

Second Problem:

$$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{d} i(2x_i^2 - x_{i-1})^2 \ where - 10 \leq x \leq 10$$

For this problem, the parameters will be set at P=200, N=20, G=300. The approach will be similar to the previous problem, running the code 10 times and averaging the results, then changing the mutation rate and mutation step, both values moving from (0) to (1) by adding (0.1) with each run. But first, the produced output from an initial run (MS and MR set as 0.1) shows the drastic fall of best solution and average solutions from the first generation to the final one. The average best shows to move from nearly 300,000 to 7 within 300 generations (table 3, appendix A.).

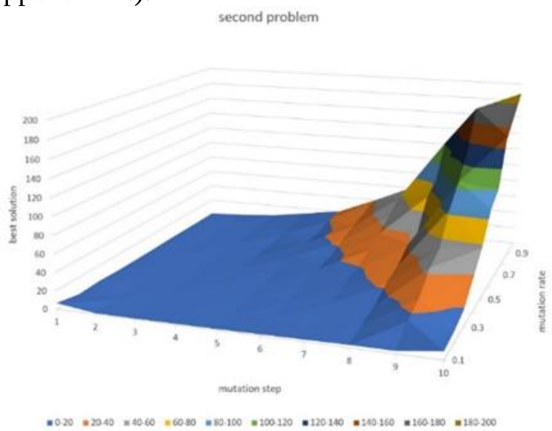


**Figure 7**

The second equation’s main table (table 4 appendix A.) shows the change of the average best of the final generation in each alteration. Generally, the closer the parameters are to (1), the less effective the mutation becomes. As the surface graph shows (figure 7), the highest values of the best solutions are after the (0.5) in both parameters. The highest solution (188) being when both mutation rate and step are at (1).

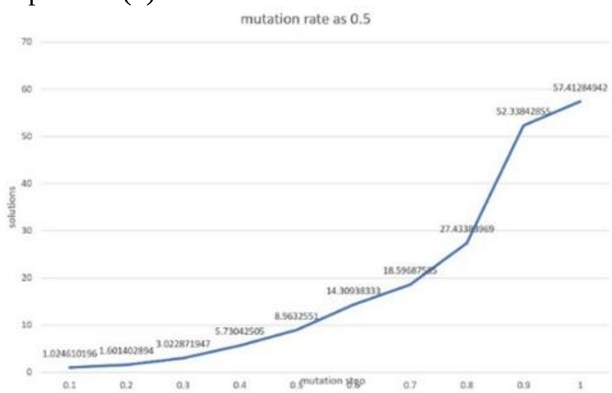


**Figure 8**

If the value of mutation rate is set at (0.5) as an example, it is noticeable that the change of mutation step affects the solution values in a steady gradual increase (figure 8). But when keeping the mutation step set as (0.5), the mutation rate is more variable in its effect on the solutions (figure 9).

**Figure 9**

Looking back at (table 4 appendix A.), the best solutions are highlighted, mostly concentrated at the smaller values of the parameters. Mutation step appears to have the most effect when it's at (0.1), with most solutions being below (2.0), therefore, the next test will run smaller mutation step and mutation rate values to see if the solutions will be closer to the optimum solution of (0).

Taking the solution of (MS = 0.1) and (MR = 0.6) (figure 10 & table 5 appendix A.), the output showed that MS being too small (0.02) begins to become ineffective in finding the best solution, reaching to above (500) at the MR of (0.56). As for the optimum solution, the main table (table 4 appendix A.) shows the smallest solution value is (0.8) at the parameters (MS = 0.3) and (MR = 0.1), however, the value of (0.8) is reached 4 times when mutation step is between (0.06) and (0.1), and the mutation rate between (0.52) and (0.58).

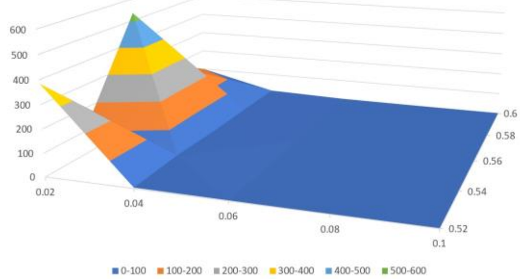

**Figure 10**

Another test run focusing on the solution value in the main table (table 4 appendix a.) of (0.8) (MS = 0.3, MR = 0.1) shows the following values (figure 11):

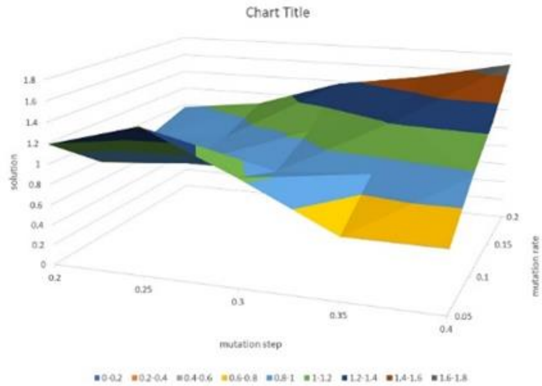

**Figure 11**

the optimum solution so far equals to (0.59) (table 6 appendix a), which appears when the MS is between (0.35) and (0.4), while MR is at (0.05). Going back to the main set parameters, while keeping in mind the values of MS and MR that give the lowest solution, testing those numbers with a bigger population would result in a better solution, same as the previous problem, with the optimum solution found is at (0.52), at (P = 500), (MS = 0.35), (MR = 0.05), and (G = 300) (Figure 12).

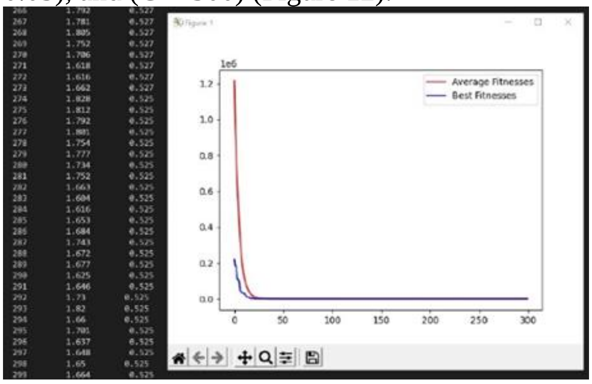

**Figure 12**

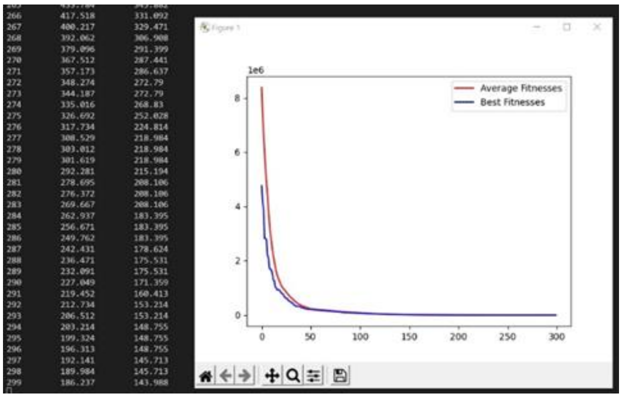

**Figure 13**

As for N, increasing N in this equation actually makes the solutions go higher, as it adds up positive numbers, so the higher the N, the higher the value. So, if (MS=0.35), (MR=0.05), (G = 300), (P=200), and (N = 50), the best solution at a test run would be back in the hundreds (186) (figure 13), therefore, when N is changed to 10, the best solution reaches (0.917) (figure 14). Although it is a low value, it is still higher than the lowest achieved so far. However, if the MR is changed to (0.005), the optimum solution easily reaches (0.09) (figure 15), which is the lowest value reached for this problem.

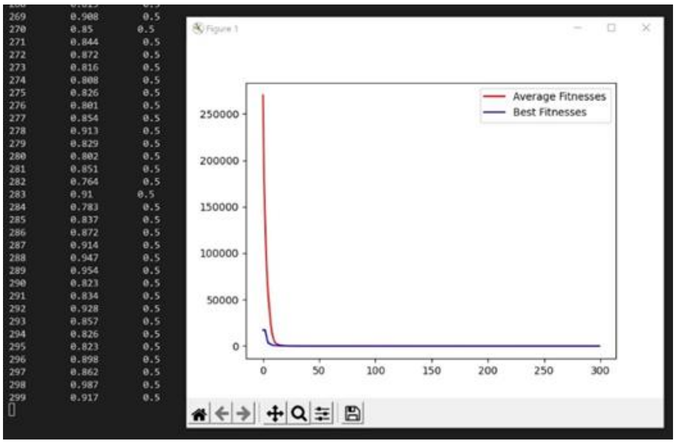

**Figure 14**

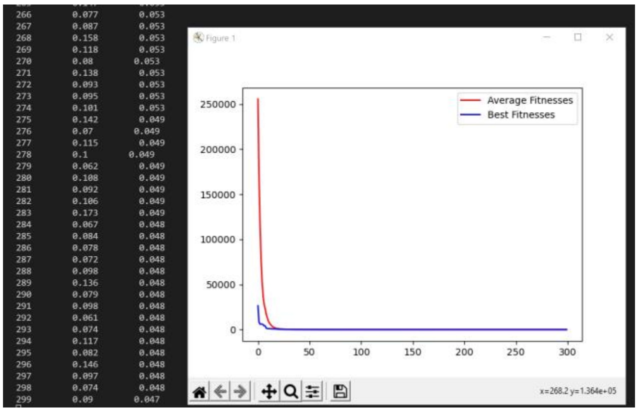

**Figure 15**

The change in the values in reaching the optimum solution greatly depends on the set parameters, which includes the amount of data, as well as the way the date is processed in the genetic algorithm. However, the effects of N, relies a great deal on the kind of numbers the problem produces, if it's negative numbers, the bigger the N, the lower the values. If it's positive numbers, the higher the N, the higher the values. The other parameters can depend a lot on the randomized data, which is why it is best to average over multiple runs.

## 4 CONCLUSIONS

As most genetic algorithms, this algorithm displays the important role that parameters play in finding solutions, and could define the search process. A different approach to solving minimisation test functions is by setting the wanted value as the condition to end the repetition loop. In a super computer, this genetic algorithm could perform better by making it change the parameters range if the output isn't what's desired after a certain time of running, by saving the best output from previous runs, and changing the parameters range to match those found, and running the program again, making the change to be of smaller values.

## REFERENCES


- Walch, K. | Forbes (2019) Ethical Concerns of AI [Online]. Available from: https://www.forbes.com/sites/cognitiveworld/2020/12/29/ethical-concerns-of-ai/?sh=2550f74a23a8 [Accessed 05 December 2021]
- Landrum, A.R., Olshansky, A. and Richards, O. (2019) Differential Susceptibility to Misleading Flat Earth Arguments on YouTube. Media Psychology [online]. 24 (1), pp. 136-165. [Accessed 05 December 2021].
- Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F. and Scala, A. (2020) The Covid-19 Social Media Infodemic. Scientific Reports [online]. 10 [Accessed 06 December 2021].
- Statistica (2021) Number of social network users worldwide from 2017 to 2025. [Online] Available from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [Accessed 5 December 2021].
- Snider, C. | Chris Snider Designs (2021) Social Media Statistics. [Online] Available from: https://chrissniderdesign.com/blog/resources/social-media-statistics/ [Accessed 05 December 2021]
- Kaplan, A. (2020) Digital Transformation in Media & Society [online] Chapter 9: Artificial Intelligence, social media, and fake news: is this the end of democracy? Paris, France: Rector & Dean, Escp Business School. [Accessed 05 December 2021].
- YouTube (2021) Managing harmful vaccine content on YouTube [Online] Available from: https://blog.youtube/news-and-events/managing-harmful-vaccine-content-youtube/ [Accessed 06 December 2021]
- Instagram (2019) Combatting Misinformation on Instagram [Online] Available from: https://about.instagram.com/blog/announcements/combatting-misinformation-on-instagram [Accessed 06 December 2021]

- Muller, Vincent | the Stanford Encyclopedia of Philosophy (2021) Ethics of Artificial Intelligence and Robotics [Online] Available at: https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/ [Accessed 14 December 2021]
- Acemoglu, A. and Restrepo, P. (2021) Tasks, Automation, and the Rise in US Wage Inequality. NBER Working Paper Series [online]. [Accessed 12 December 2021].
- Amazon (2020) Investor Relations [Online] Available from: https://ir.aboutamazon.com/news-release/news-release-details/2020/Amazon.com-Announces-Third-Quarter-Results/default.aspx [Accessed 14 December 2021]

**Appendix A – Tables: provided as a file in the folder.**
**Appendix B – Genetic algorithm: provided as a file in the folder.**