**Plans and Progress:**

Broad ranges of tools are integrated into this project, and work together to achieve a goal of autonomy and strengthened Cybersecurity. The edge given to this project is a use of the modern, ultimately still-developing Machine Learning technology: Artificial Intelligence with the function of self-learning from provided and accessible data.

Largely, the objective of this project is to eliminate the need for updating the software, to limit storage use on a system by integrating the software to a database, and to provide an adequate, learning security system against viruses.

This progress report is to provide information to current research headway and ideas, as well as conceptualizing future plans and issues that may arise. Given the group’s current knowledge, we rely on conceptualization for the future creation of this tool and the abilities required to make that happen.

The project arose in a combination of some of the group members’ individual designs about their own projects. These had roots in databases and security measures, both of which are central to the current project. Relevancy to today’s technological climate was also considered while approaching the choice- Cybersecurity is a main concern, given that new ways are being advanced upon to both defend and attack systems, in line with technology’s constant progression.

Cybersecurity plays an important role in society by employing different sources and people across all cultures, as well as the latest technology. Its implementation is in high demand, as technology improvement will bring an unfortunate increase in cyber-threat. Hence, it is essential to ensure the safeguard of civil liberties and wrights worldwide through applied regulations, even if it could imply some privacy loss. In this, antivirus software plays a major role in keeping systems safe from attacks. Cyber-attacks exist with the aim to disrupt and put at risk our society’s wellbeing by imposing a global de-centralised treat.

When combining a breadth of knowledge that is able to be stored on a database with the ability to combat viruses, a formidable defense is created. Integrating self-learning into this defense can form new level in managing this consistent and ever-growing threat. In line to the present state of antivirus software in the increasingly cyber world, this project holds the purpose of removing the concerns of both constant scanning of systems and refreshing of software.

To fulfill this purpose, Machine Learning plays a key part in this project for the reason of human error, performance constancy, and persistent monitoring. The ever-adapting nature of this software commits to this idea of a stronger, more prevalent security. Machine learning is a branch of artificial intelligence, and has a sole purpose: to analyse large amounts of data, ideally without any form of human intervention. Thus, the automation aspect of this project arises.

The method this project intends to use is known as pattern recognition. Upon making an analysis of large amounts of data and identifying designs within it, this method produces analytical model building. What’s appealing about this is that the intelligence builds the more it encounters- the more it learns, the stronger it becomes.

In combining machine learning and security, the constant updating of antivirus clients is eliminated- instead, the system learns as it goes along. Additionally, it solves the issue of what antivirus brand are appropriate for, rather than having a program that scans for viruses that it knows about, it learns about threats based upon prior experience. This deems it superior in purposefulness over current software due to this factor of self-learning, where the owner would not need to worry about threats attacking their system if they haven’t updated or purchased new antivirus software.

Machine Learning was the major change from the original plan. Initial feedback for this project suggested the exploration of Machine Learning rather than broadly using Artificial Intelligence (AI) as a descriptor for the system’s operating. Python and R are common languages to create algorithms in Machine Learning. The data to be analysed by this software is added to by predictive patters, and will use dates of major incidents, type, typical targets, and possibly more aspects. These large amounts of information is to be stored about the viruses encountered.

Here, a database is necessary, although it provides its own problems for systems that are to remotely access it. The major hurdle would be to figure out how to code a system in order to learn from information that is stored elsewhere. The antivirus would need to fetch and analyse data and compare it to the supposed infiltrator to determine if it is a threat, what the threat could do, and what its targets within the system would be. Standard Query Language (SQL) is a requirement for making pathways to and from the database. However, the location of an AI such as the Machine Learning mechanism involved in this project would be within the database, rather than in the software. This is because in today’s database technology, Python or R coding is also captured by SQL server-stored procedures. In this way the usual procedure for storage can transformed into an AI-stored one.

Programming pattern learning into this AI (within the area of Machine Learning) can go multiple ways. The algorithms that can be created are dependent on various aspects of the data’s intended output, and on the ability to work off of trained data only; that is, if the data has been inputted into the AI (supervised), or whether it can determine new patterns from encountered data (unsupervised). This project will operate on an unsupervised learning basis, in order to continue picking up new viruses.

Algorithms to approach the data (viruses) this project is to encounter are narrowed when considering it supervised or unsupervised- even more so when taking in to account the nature of the data, and the information desired to be extrapolated from it. Principal Component Analysis (PCA) procedure is statistical, and widely used in exploratory data analysis as well as in predicative models, both of which are relevant to the countermeasures this project is aiming for in terms of security. To prevent future attacks, this element of prediction is necessary. As an unsupervised regression algorithm, it recognises the variables in data and attempts to draw relationships from it. By making use of all aspects of the data [dates, types, targets, etc], a PCA algorithm could prove to be invaluable to the prediction of the manner of future attacks.

In terms of the categorisation of the viruses and providing a suitable countermeasure, Clustering has many types of useful unsupervised algorithms when it comes to statistical analyses, pattern recognition, and information retrieval, all of which are relevant to the project. An ensemble learning algorithm (supervised, and so not able to pick up new information) could also have use in increasing the implementation of a multitude of other algorithms.

Although only in mock-up stages for what the program will look like, there is confidence in the theory that this project will be able to be produced smoothly. With technology in its current state, there are few limits on a project such as this when enough time, work, and expertise is put into it. This project is still in the theoretic stage as, for the group, it as an ambitious project. Yet, it is interesting to research and learn about- and would be a useful tool if it were to become a reality. Already, the environment of Cybersecurity is growing and incorporating more and more AI aspects. Threats to security are upgrading into new and more maladaptive attacks on systems as well- thus, a large database of all encountered viruses is not only practical but valuable for making use of this information for proactive analysis as well.