## S2 - Discuss the positives and negatives of automated security alert systems with regards to usable security.

Usable security focuses on designing for the user such that the security must be easy to use, within their capabilities and not require stress of mind. Usable security is important because we know humans are in the loop at all times. Even if a system becomes fully automated, humans are still involved during the development. Therefore, given that we can't take the human out of the loop it is important that we fit the task to the human, likely resulting in more psychological acceptance.

There are many positives of an automated security alert system. The first being the amount of data it can handle. Automated systems are likely to be able to process a lot more than a human can, and at a faster rate, due to the natural physical and mental limitations that humans have. Having this automation means that users are not asked to do a task outside of their capability.

Additionally, using these automated systems makes securing something a lot easier. Aside from initially learning what the alerts may mean and how to set it up, the alert system can be easy to use and doesn't require stress of mind. However, if there are a large number of alerts and signals to remember, the user may not be able to retain the long series of rules, resulting in the automated system not being as usable.

Another positive of automated systems is that they are not affected by the physical and social context in which they are operating. If a human had to manually flag up security alerts this would be arduous. Their performance would be affected by external factors, such as stress, how tired they are, or how much experience they have. If the user is busy they may not check the systems as well. In contrast, the automated system will not have these problems and will always perform to the same standard.

However, an automated security alert system may lead to alarm fatigue. The alert system may produce a high number of false positive alerts that only the human can filter out. This can be a high cognitive load and eventually the human may become blind to alerts that they see often. This is because processing a high number of false positives can push the user outside of their capability. An example of alarm fatigue we see every day is with cookie banners. Every website is compelled by law to ask the user to accept or reject the non-essential cookies. Because of the high frequency of banners a user may encounter in a day, they eventually become blind to them and accept all cookies.

Furthermore, an automated system may have biases. These may have been baked in by the developers during the development of the system and may go unnoticed. In particular, if the system is helping the user to achieve their goals, they are unlikely to pay much attention to how the system works, and are more focused on only actioning the alerts. In contrast, a human manually looking at the system may be more aware of the biases, and in particular if more people are working on the system they may be able to recognise other people's implicit biases and pull them out of the system.

Moreover, automation can result in an overreliance on the system. This in turn leads to less secure behaviour if people forget that the system they are using is not automatic. An example is automatically locking computers. A user may forget that they need to lock their computer as soon as they leave their desk, as it usually locks automatically after 10 minutes, and by the time they get back to their desk after lunch, it is always locked. They are unaware that for 10 minutes they are leaving their computer unlocked. A similar effect may be seen with automated security alert systems. Users may not be as wary of their surroundings due to overreliance, but if they visit a new place without security they may not be as cautious because they are used to relying on the system.

Overall, I believe automated security alert systems can be beneficial to users with regards to usable security.

## S3 - Twitter is rolling out a new service to allow members of the public to flag misinformation. How could this intervention go wrong? How might it be made safer?

Misinformation can be defined as the act of giving out wrong information about a topic [1]. For example, information may be spread on Twitter saying that the Covid-19 vaccine can be harmful to pregnant woman, although studies have shown that this is not the case.

Allowing members of the public to flag this misinformation results in less reliance on the social media company. This means they can focus on other aspects of their platform, such as security or rolling out new features. Additionally, it means they refrain from taking a view on controversial topics. For example, if they did not flag up the misinformation regarding the Covid-19 vaccines, users may interpret this as Twitter being against vaccines and therefore boycott the platform. Furthermore, by having less reliance on automation, people can't adapt their tweets to get through the filters as the criteria is unknown. This contrasts with spam filters on email inboxes for example, where attackers adapt their work to pass through the classifier without getting caught.

However, allowing members of the public to flag misinformation can lead to problems. One problem is that it is hard to understand why someone may have flagged a tweet for misinformation. For example, they could report a tweet simply because they do not like the person who is tweeting. Over a continuous period of time, if a person's tweets are always being flagged for misinformation they may feel their right to free speech is being violated and may also be emotionally harmed. Furthermore, they may start to identify with their label, resulting in them posting tweets containing misinformation because they know their account would be flagged up either way. To make it safer for all users, Twitter could implement a feature that when a tweet is being reported, the reporter must justify why they think this is misinformation. This means users who are spam reporting may be put off because of the increased effort to report. However, it is important to note that implementing this feature may result in genuine reporters also not reporting because they do not wish to expend the effort explaining why this is misinformation. There should also be an option for Twitter to remove labels when they are incorrect, to minimise emotional harm done to incorrectly reported users.

Another problem which may arise is conflicting opinions. A controversial tweet may be flagged up for misinformation for multiple reasons, some of which may conflict. It is difficult for Twitter to then flag a tweet as misinformation as it may be seen as taking a side as mentioned above. To make this safer, every tweet that has been flagged for contrasting opinions can be reviewed by a team. However, the team may get a lot of requests resulting in them being overworked. Alternatively, the tweets could be reviewed using an automated system. This could, however, result in some tweets not being flagged due to biases in the system that were baked in during development, due to the developers' biases but also biases in the training data.

This intervention may also go wrong when a tweet or tweets about a topic become very controversial and receive public uproar. Although Twitter may have been trying to supress

the misinformation, because of the public uproar this misinformation is being spread. This is an example of the Streisand effect. An example of this is when Donald Trump suggested that people should inject themselves with disinfectant to kill Covid-19. This was clear misinformation but because of the public attention it received, more people became aware of the idea [2].

Another problem with flagging misinformation is that users may move on from Twitter to a platform that allows them to freely share their opinion, without the fear of being flagged for misinformation. A similar effect was seen when Donald Trump was banned from Twitter and many alt-right users moved to other platforms [3].

Overall, I believe that this service can benefit the users of Twitter, although many precautions will need to be put in place.

References:
[1] https://www.oxfordlearnersdictionaries.com/definition/english/misinformation
[2] https://www.bbc.co.uk/news/world-us-canada-52407177
[3] https://qz.com/1617824/twitter-facebook-bans-send-alt-right-to-gab-and-telegram/

## L1 - “Professionalisation of cybercrime has made everything better for cybercriminals” – Discuss.

Over the last 20 years, cybercrime has seen a rise in sophistication, commercialisation and organisation. Although some techniques have stayed the same, for example Kane Gamble used the same social engineering techniques in 2015 that Kevin Mitnick published in a book in 2003, cybercrime has seen an increase in professionalisation, partly due to a heightened level of communication and partly due to increasing defences.

An example of professionalisation of cybercrime is the emergence of the online hacktivist movement called *Anonymous* that was founded in 2003 [1]. They are structured as a swarm, meaning there are no headquarters, no leaders and no single agenda. As a group they are able to commit cybercrime at a large scale very publicly. An example of a recent attack is a Distributed Denial of Service (DDoS) attack on the Minneapolis Police Department during the Black Lives Matter movement in June 2020. This was a large-scale attack which is unlikely to be possible by one lone cybercriminal.

This increase in community allows cybercriminals to share tips and raise the complexity of their attacks. This in turn raises the commercialisation of cybercrime. Cybercriminals can become specialised on a particular area of cybercrime. They can then be hired out by a core group of cybercriminals to deliver the attack. For example, if a group want to attack a Microsoft product but they do not have the skills or knowledge, they may hire out a specialist who previously worked for Microsoft and therefore is very aware of the vulnerabilities in the platform, thus driving commercialisation.

As well as driving commercialisation, the professionalisation means that less technical cybercriminals can also get involved. Before, a cybercriminal would usually have to do everything themselves, from finding their victims, to planning and delivering the attack, and then recovering their costs, which would require a high technical aptitude. In contrast, lay users who want to be involved in cybercrime can now easily take part. For example, a key part of illegal trading are cashiers. These are people who extract funds from accounts given the details. This role does not require any technical knowledge, meaning it is easy for anyone willing to take large risks to get involved.

Another advantage for cybercriminals is that something widely used is unlikely to end even after being shut down. For example, Silk Road was a popular darknet market that got shut down in 2013. But one month later, a new version of Silk Road appeared with the same functionality. This meant the trading and cybercrime could continue. This contrasts with a lone cybercriminal being shut down. They would likely be caught and charged, and their business would not be able to continue unlike Silk Road [2].

However, the professionalisation is not all positive for cybercriminals. Because of the anonymous nature and the ease to join a group, it can be hard to identify who is genuine and wants to commit the same cybercrime and who is either an uncover investigator or a criminal who is invested for their own good, rather than working towards a shared goal. Additionally, the large scale of operations means it can be easier to get caught. For example, because lay

users can also join the anonymous networks, they may inadvertently share information that can lead investigators to take down the network. This cannot be controlled because there is no way to control who is part of the network.

Moreover, some cybercriminals may not want to commit cybercrime for the money. They may just want to show their power and destroy the reputation of some companies. This is increasingly difficult as the cybercriminal industry becomes more sophisticated, meaning it is difficult for lone hackers to break through and operate at the same scale as large groups. Furthermore, they may not wish to be associated with a group. For example, as mentioned above, *Anonymous* do not have a single agenda. This may make a cybercriminal reluctant to join this group because they cannot identify with any core aims, and although may agree with some of the attacks taking place, they may not wish to be associated with all the attacks taking place.

Another reason the professionalisation may not have made everything better for a cybercriminal is because they are frequently surrounded by other cybercriminals. This links back to the social learning theory that most of our behaviour is learnt from those around us. If a cybercriminal is always surrounded by other cybercriminals, they may acknowledge that what they are doing is bad but may fail to realise that cybercrime is abnormal, making it harder for them to leave the industry, even if they want to. A reason they may want to leave the industry, is because of the stereotypes that have been created of a cybercriminal, which has been promoted by the professionalisation of cybercrime.

Overall, I do not believe that the professionalisation has made everything better for all cybercriminals. I think it has had a positive effect on cybercriminals who would like to work in a group on large-scale attacks, but has had a negative effect on cybercriminals who like to work alone and are motivated not by money but by the chance to ruin a company.

References:
[1] https://en.wikipedia.org/wiki/Anonymous_(hacker_group)
[2] https://www.swansea.ac.uk/media/Silk-Road-After-being-closed-twice,-can-the-brand-ever-%C3%A2%C2%80%C2%98rise-again%C3%A2%C2%80%C2%99.pdf