**Understanding Artificial Intelligence (AI)**

*What is responsible AI?*

Responsible AI is the governance that outlines how an organisation addresses the challenges surrounding the implementation of AI from both a legal and ethical position. Responsible AI at its core is a set of principles where one of the key focuses of is to resolve issues of responsibility if any issues arise. Responsible AI was developed as a response to the range of harm that could be caused, to individuals and whole societies, if AI were misused, poorly developed, abused or due to unforeseen outcomes.

*Find real life instances where AI has failed or been used maliciously/ incorrectly.*

One of the most notable AI failures this year is the racial biased of both *twitter* and *Zoom.* Both involved racial biased against black people with twitter AI favouring white faces when cropping images and cutting out black faces. While Zoom completely removed faces of black people when virtual backgrounds were used.

*Amazon Alexia* is another AI that has had a few issues over the years in Germany back in 2017 police were called to an apartment due to a loud house party. On arrival they found nothing but an Alexia that enjoys playing loud music spontaneously when the residents are out. A more costly issue arose when a 6-year-old girl ordered £170 worth of dolls and accessories simply by asking Alexia for them. The story was picked up by local news reporters and as the new caster stated that he loved how the girl just said ‘Alexia get me a dolls house’ resulted in Alexia in several viewers homes attempting to purchase dolls houses.

A more serious AI failure occurred when Facebook was used by Cambridge Analytica to harvest over 50 million profiles, without consent, in the US and the UK to build software designed to identify and target voters with the expressed intent of influencing the voter’s choices at the ballot boxes, for both the Trump election and the EU referendum.

*What are the implications of AI failures?*

AI failures vary in severity depending on the AI’s function from the simple missteps of the racial bias facial recognition systems previously mentioned to the breaches in data handling and privacy that can occur. The development of the GDPR and the opt in/out method of data sharing has been implemented to allow people the right to withdraw consent and ensure organisation must gain consent of users to share certain data. Although more serious implications are possible the over reliance on AI can lead to fatalities as such as the death of Joshua Brown in 2016 when his tesla failed to identify a white tractor-trailer crossing the highway, whilst in ‘auto-pilot’. This is as with the racial biased a classic example of how AI is only as good as the data it’s been given. Which has two main limitations firstly, is there enough data to be a successful in a real-life situation where the AI will encounter thousands of different situations. In both situations if the AI had been given more data the issues would be less likely to occur. The second limitation being is the data pre-exposed to a bias, the data needs to be wide and varied to make it a true representation in real life settings, there are many instances of chatbots becoming hateful and single minded based on the information they have been given.

*What should organisations do to ensure that they are being responsible with AI and the wider use of data in general?*

Organisations should develop an internal governing body that understands the implications of AI but at the same time is filled with diverse people that are supported by the leadership of the organisation and can hold the leaders accountable for any shortfalls.

Organisations should understand the need for traceable audits of AI decision making for both quality assurance and to allow for them to address where issues arise. I know there is a ‘black box’ situation where organisations do not wish to expose the inner workings of their AI or the data is so complex it cannot truly be explained how the AI comes to decisions, however there should be a final level of decision checking to ensure the way the AI is reaching the decisions is not just efficient but ethically and legally sound.

Organisations should opt to use a diverse workforce when creating AI as this can go far in combatting bias within the AI as they arise and be able to identify them before the AI is released into society.