**Week 9: Artificial Intelligence**

**Find out what Responsible AI is?**

The practice of designing AI that positively impacts people by considering fairness, reliability & safety, privacy & security, inclusiveness, transparency and accountability.

**Find instances where AI has failed? Or been used maliciously or incorrectly.**

A) AI that, intentionally or not, reinforces existing bias within it’s training data. E.g. identifying potential offenders using US prison data will identify more black people due to historic racism within the prison system.  
B) Specific goals can lead to outputs that are harmful e.g. A chatbot tasked with reducing doctor workload suggested suicidal patients should kill themselves. Does technically meet the goal but without any safeguards in place. (https://www.theregister.com/2020/10/28/gpt3\_medical\_chatbot\_experiment/)  
Or social media generally using algorithms to increase engagement for purely financial purposes with no consideration to the longterm harm it can cause (addiction, radicalisation, emotional/developmental problems)  
C) AI used for nefarious purposes e.g. Cambridge Analytica influencing election outcomes by targeting swing voters with adverts tailored to invoke emotional responses. Voters are no longer making informed decisions and democracy could boil down to whoever has the most amount of money gets to decide the outcome. True democracy becomes impossible and the population is divided against each other with little consideration for the middle ground.

**Implications of when AI fails. There is a specific article in the GDPR Law that covers this, especially with automated decision making. (opt in and out options).**

GDPR gives people the right to know how, where and why their personal data is processed as well as opting out and requesting it’s deletion where no longer needed and must give explicit consent for it’s processing to begin with.

As automated decision making becomes more common peoples rights need to be protected by ensuring people are aware of the where, what, why, how of their data being processed, need a simple way to opt out and must be able to challenge the decision being made. Credit scores being an example, if an automated system is flagged as low rated impacting their ability to access finance, they need to be able to easily see why and challenge the decisions being made.

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

Continually monitor, review and take action to maintain compliance with legal frameworks / best practice as a starting point but also go further to ensure a holistic view of the AI is taken to safeguard all stakeholders internal and external to the business. Ensure training so that all parties involved with it’s implementation can make informed decisions that have considered all angles and potential impact.