HMLTWK6 – Sumbul Sultan

**Artificial Intelligence:**

*What is responsible AI?*

It focuses on ensuring the ethical, transparent and accountable use of AI technologies in a manner consistent with user expectations, organizational values and societal laws and norms. Responsible AI can guard against the use of biased data or algorithms, ensure that automated decisions are justified and explainable, and help maintain user trust and individual privacy.

An important goal of responsible AI is to reduce the risk that a minor change in an input's weight will drastically change the output of a machine learning model.

Within the context of conforming to the four tenets of corporate governance, responsible AI should be:

* Each step of the model development process should be recorded in a way that cannot be altered by humans or other programming.
* The data used to train machine models should not be biased.
* The analytic models that support an AI initiative can be adapted to changing environments without introducing bias.
* The organization deploying AI programming is sensitive to its potential impact -- both positive and negative.

*When has AI failed or been used incorrectly?*

There have been instances where artificial intelligence has failed or not been used correctly.

For example in 2018, Amazon had to axe their ai for recruitment because their engineers, trained it to be misogynistic. One of the engineers behind the ai said “they literally wanted it to be an engine, where I am going to give you 100 resume’s and you’re going to spit out the top 5 and we’ll hire those”. But the engineers realised that they actually taught the ai that male candidates were better. This happened because they trained the ai based on engineering applications and used the current engineering employees as a benchmark, which is heavily male dominated.

Another way ai is being used maliciously is through cybercrime. Hackers weaponize malware to counter attack the cyber security. For example criminals use ai to hide malicious codes in benign apps. They program the codes to executed after a certain time period the app had been installed e.g after 6 months, or when a certain amount of user download the app. They do this to maximise the impact of the attack. Keeping the codes hidden requires the application of ai models.

*What should organisations do to show they are being responsible with AI?*

Organisations should have an understanding where AI sits in the value chain and creating the right structures to ensure long-term governance by:

1. Establishing internal governance, for example by an objective review panel, that is diverse and that has the knowledge to understand the possible consequences of AI infused systems. A key success factor is leadership support and the power to hold leadership accountable.
2. Ensuring the right technical guardrails, creating quality assurance and governance to create traceability and auditability for AI systems. This is an important part of every organisation’s toolkit to allow operational and responsible Ai to scale.
3. Investing more in their own ai education and training so that all stakeholders, both internal and external, are informed of AI capabilities as well as the pitfalls.