**Artificial Intelligent (AI) HLT**

**What is responsible AI?**

A framework that documents how a specific organisation is addressing the challenges around AI from a legal and ethical point of view. The framework focuses on accountability, fairness, safety, privacy, interpretability.

**When AI fails and has been used maliciously or incorrectly?**

AI that failed was the Microsoft Tay which was a chatbot that was created to impersonate human like interactions. The chatbot started to make racist and derogatory remarks.

Amazon also create a recruiting tool which turned sexist and chose to hire men employee based on the machine learning algorithm used in training found that majority of resumes came from men and therefore determined woman were not suitable for technical roles

Uber experience AI failure with self-driving car fatality which resulted in a loss of life because no amount of machine learning can calculate random incidents.

**Implication of when AI do fail?**

Data security is a major issue with AI failure, this is due to AI applications accessing great amounts of data to work. This can result in weak spots and vulnerabilities for hackers to target and compromise. There is also limited amount of people who can understand AI well enough to create these systems

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

Organisations need to develop responsible AI framework that ensures the following: -

* Governance, ethics and regulations, interpretability and explain ability, robustness and security and bias and fairness.
* Governance which serves as an end-to-end foundation for all other components.
* Ethics and regulations to develop AI that is compliant with regulation and meet ethical requirements.
* Interpretability and explain ability to those who operate and those who are affected by them
* Bias and fairness to ensure AI systems to diminish unwanted bias and make decisions that are fair under specific and clear communicated definition