## HLTWK7 – Responsible AI

*Your task is to:*

*• Find out what Responsible AI is?*

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

*• 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).*

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

Responsible AI is about designing and deploying AI with good intention to empower and fairly impact society. Therefore, allowing companies to grow and scale AI with confidence. Responsible AI allows businesses to flourish by

* Minimising unintended bias – algorithms to be built on underlying data that is unbiased and a representative of people.
* Ensuring transparency – which helps build trust with employees and customers to develop explainable AI across processes and functions.
* Protecting the privacy and security of data – makes sure that personal or sensitive data is never used unethically.

A recent example when AI failed, was with Microsoft’s Chat bot, Tay. Engineers had built the bot with data from the public domain, whilst it was built in the lab, it had undertaken cleaning and filtering process to pull out relevant information. However, AI is built by learning, and in doing so, when it went out live to the public, the bot was collecting information put in by the public, mostly incorporate and offensive remarks, in turn the bot was relaying the offensive remarks without the filtration it was trained to do.

GDPR is the most impactful law globally to ensure data is a regulated in the market. Data is the key source of AI. When AI processed information about an individual or profiling an individual for decision making, the GDPR laws apply. Article 22 has specific provisions around profiling and automated decisions which are very narrow in it applicability, however all profiling is regulated by the rest f GDPR, including consent, the right to objection, to access meaningful information, right to opt out in the interest of fairness and transparency. Any failure to comply with the GDPR laws, could face a fine or up to 4% of a company’s turnover or 17 million euros depending on which is higher.

An important goal of responsible AI is to reduce the risk that a minor change in an input weight will drastically change the output of a machine learning model. To be considered as a responsible as an AI company, companies must build with resources according to company wide standards, which mandates the use of:

* Sharing code repositories
* Approved model architecture
* Sanctioned variables
* Stability standards for active ML, to make sure that the AI programme works as intended.
* 