Responsible Artificial Intelligence refers to a best practice governance framework aimed at ensuring that organisations account for ethical and legal considerations when launching any AI initiatives, (Gillis, 2021). The main aim of the framework is to outline an organisation’s commitment to ensure that its use of Artificial Intelligence is non-discriminatory in any way. The Institute for Ethical AI and Machine Learning lists the following eight principles of responsible machine learning development:

* Human Augmentation
* Bias Evaluation
* Explainability by justification
* Reproducible Operations
* Displacement Strategy
* Practical Accuracy
* Trust by Privacy
* Data Risk Awareness

However, the governing standards are developed by individual organisations and therefore tend to vary from organisation to organisation. They are typically developed by data scientists, software developers or chief analytics officers in accordance with their organisation’s models. For instance, Microsoft lists the following six as their Responsible Artificial Intelligence principles:

* Fairness
* Reliability and Safety
* Privacy and Security
* Inclusiveness
* Transparency
* Accountability

Whilst on the other hand Google believes that Artificial Intelligence should:

* Be socially beneficial.
* Avoid creating or reinforcing unfair bias.
* Be built and tested for safety.
* Be accountable to people.
* Incorporate privacy design principles.
* Uphold high standards of scientific excellence.
* Be made available for uses that accord with these principles.

There are some overall similarities in most of the basic principles with degrees of variation here and there within the industry but there is yet to be any regulation to specifically police these standards. This can be concerning in situations when Artificial Intelligence fails or has undesirable consequences whether intentional or not. One prominent example of Artificial Intelligence failing was the autonomous Uber vehicle that struck and killed a pedestrian in Arizona in March 2018. The organisation was not held accountable for the death of the pedestrian despite the federal regulators’ findings of Uber’s technology being flawed and insufficient.

Another example is that of Amazon’s facial recognition technology which was found to be flawed as it wrongfully identified some prominent professional athletes as criminals. Amazon had been pitching for this technology to be used by law enforcement in their criminal investigations therefore the implications for this would have been huge, had the technology been given a green light. This was in direct violation of paragraph 1 of Article 22 of the General Data Protection Regulation which stipulates that, “The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her”.

Overall, organisations should ensure that Data Protection is at the forefront of any design to ensure that the individual rights of any potential automated decision making are within the confines of transparent, lawful and non-discriminatory processes that are also tested for safety and reliability.