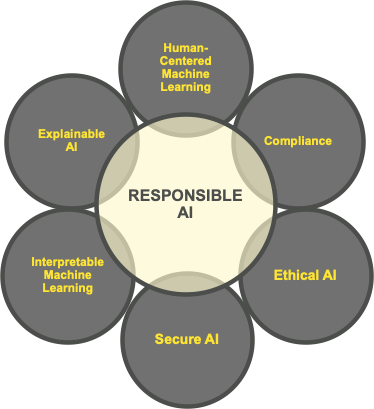
**What is responsible AI?**

Responsible Artificial Intelligence (AI) can be defined by several factors through which companies should consider in all stages of creating a machine learning model and successful and ethical AI. At present, although many major corporations have called for AI regulations, it is up to its creators to ensure that the models meet ethical, moral, and professional standards.

To the left is a diagram presenting some of the key features of responsible AI. Key to all of the points is that AI is transparent, unbiased, secure, and compliant with existing policy.

Interpretable machine learning and explainable AI come under the category of transparency. Both factors relate to how a model can be explained. This may mean explaining the inner workings of the machine learning architecture, or it may mean explaining the decision making that the AI model has made and the data and weightings that affected it.

Ethics in AI relate to the prospective bias that may exist within a chosen data set. It is important that, particularly in the training phase, the AI model is given unbiased data that does not then discriminate against particular demographics within populations.

Finally, responsible AI should be secure against potential threats which could adversely impact its decision-making capacity. In addition, all of the data used should be compliant with UK GDPR or other national data protection policy.

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

There have been a number of high-profile cases in which AI has been used irresponsibly or maliciously in recent years. Notably, cases such as Cambridge Analytica using data from Facebook to influence election, Facebook itself spreading misinformation, Uber’s fatal failures in its driverless cars, Clearview harvesting images and profiling individuals for use within law enforcement to name a few.

One such case related to criminal profiling published, and later recalled, by Harrisburg University was titled *Criminal tendency detection from facial images*. The paper suggested that AI could be used to predict if someone would be a criminal based solely on a front-facing picture of their face. The paper used data 8401 greyscale mug shots as training data and presented the machine learning process in a logical way. However, the research was discredited on the basis of the bias which existed within the data as a result of the pre-existing bias which exists within the justice system. Additionally, it would have been incredibly difficult for the creators to have explained the model and weightings in a credible and ethical fashion.

**Implications when AI fails.**

The GDPR has very specific guidelines relating to automated individual decision-making and profilingin article 22 of the data protection act. The UK GDPR also has specific regulations based on profiling and decision making when legal or other similarly significant issues may affect individuals. In order to use AI for automated decision making and profiling, the organisation must comply with the following standards/requirements that the use is: necessary for entry into or performance of a contract, authorised by domestic law, or based on the individual’s explicit consent. Special category data (ethnicity, criminal convictions, sexual orientation, etc.) can only be used in cases of substantial public interest or, again, when explicit consent is given by the individual. In addition, GDPR stipulates that companies should take steps to minimise errors and bias in the processing as well as giving individuals information and access to the information gathered about them. Failure to comply with GDPR can lead to a maximum fine of up to 10 million Euros or 2% of an organisation's global turnover.

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

There are many things which an organisation can do to ensure that they are being responsible with AI. These things are centred around the factors mentioned earlier, generally with regard to transparency, the ethical use of data, as well as keeping in mind national policies and best practice.

In order for organisations to make these assurances it is vital that they establish an internal governance model which includes guidelines on training, transparency, and accountability. This may mean sufficient training, or that the workforce is adequately qualified to present the traceability of an AI decision making process and to audit them when necessary. In addition, trained workforces and stakeholders ensure that both the advantages and pitfalls of AI are well understood, in order for the organisation to successfully harness the powerful business and industry insights it can produce. Finally, a diverse workforce may help to eradicate some of the bias that can affect the decision-making processes in AI.

<https://searchenterpriseai.techtarget.com/definition/responsible-AI>

<https://www.datanami.com/2020/04/06/brief-perspective-on-key-terms-and-ideas-in-responsible-ai/>

<https://syncedreview.com/2021/01/01/2020-in-review-10-ai-failures/>

<https://www.bbc.co.uk/news/technology-59149236>

<https://www.wired.com/story/algorithm-predicts-criminality-based-face-sparks-furor/>

<https://www.bbc.co.uk/news/technology-53165286>

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0282-4#availability-of-data-and-materials>

<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/>

<https://gdpr-info.eu/art-22-gdpr/>

<https://cloudblogs.microsoft.com/industry-blog/en-gb/cross-industry/2020/01/08/3-ways-organisations-can-use-ai-in-a-responsible-way/>