**Practical: Artificial Intelligence (AI)**

**Responsible AI**

The last few years have brought to light many cases of failure of companies implementing AI without adequately considering or analysing the risks of their models.

Responsible AI is a governance framework that documents how a particular organization is addressing the challenges surrounding artificial intelligence from both an ethical and legal standpoint Resolving ambiguity about where responsibility lies if something goes wrong is a key driver for responsible AI initiatives.

As of this writing, the development of fair, trustworthy AI standards is at the discretion of data scientists and software developers who write and implement AI algorithm models of a particular organisation. This means that the steps needed to prevent discrimination and ensure transparency vary from company to company.

Just as ITIL has provided a common framework for delivering IT services, responsible AI proponents hope that a widely adopted governance framework for AI best practices will help its organizations globally to make sure their AI programming is human-centred, interpretable and explainable

What are the principles of responsible AI?

AI and the machine learning models that support it should be comprehensive, explainable, ethical and efficient.

* Comprehensiveness – comprehensive AI has clearly defined testing and governance criteria to prevent machine learning from being hacked easily.
* explainable AI is programmed to describe its purpose, rationale and decision-making process in a way that can be understood by the average end-user.
* Ethical AI initiatives have processes in place to seek out and eliminate bias in machine learning models.

Efficient AI can run continually and respond quickly to changes in the operational environment

**Instances Where AI has failed**

#### **AI despised women:**

Amazon wanted to automate its hiring process to expedite the selection of candidates for the thousands of job openings they have. Everything ended up being a public relations disaster since the system turned out to be sexist, favouring white guys. The training data used to create the model was most likely imbalanced, resulting in candidate selection bias. This is also another example of AI Failures.

**Implications of when AI fails & GDPR**

AI is not explicitly mentioned in the GPDR, but many provisions in the GDPR are relevant to AI, and some are indeed challenged by the new ways of processing personal data that are enabled by AI. There is indeed a tension between the traditional data protection principles – purpose limitation, data minimisation, and the special treatment of ‘sensitive data, the limitation on automated decisions – and the full deployment of the power The latter entails the collection of vast quantities of data concerning individuals and their social relations and processing such data for purposes that were not fully determined at the time of collection. However, there are ways to interpret, apply, and develop the data protection principles that are consistent with the beneficial uses of AI and big data

**Steps to take for responsible AI**

The best solution for organisations to be responsible with AI is to ensure the responsible building and application of AI by taking care to confirm that AI outputs are fair, that new levels of personalization do not translate into discrimination, that data acquisition and use do not occur at the expense of consumer privacy, and that their organization's balance system performance with transparency into how AI systems make their predictions. It may seem logical to delegate these concerns to data-science leaders and teams, not politicians, since they are the experts when it comes to understanding how AI works.