what Responsible AI is?

For some, the term Artificial Intelligence can provoke thoughts of progress and productivity. For others, the outlook is less positive. Many concerns such as unfair decisions, workers being replaced, and a lack of privacy and security are valid. To make things worse, many of these issues are unique to AI. This means existing guidelines and laws are not suitable to address them. This is where Responsible AI comes in. It aims to address these issues and create accountability for AI systems.

**Why we need Responsible AI**

When we talk about AI, we usually mean a machine learning model that is used within a system to automate something. AI can be used for anything from insurance underwriting to detecting cancer. The defining characteristic is that there is limited\no human input into the decisions made by the system. This can lead to many potential issues and companies need to define a clear approach to using AI. Responsible AI is a governance framework aimed at doing exactly that.

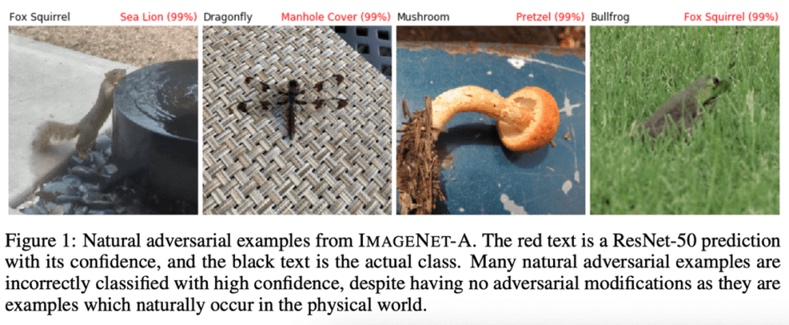
Chart

Description automatically generated

The framework can include details on what data can be collected and used, how models should be evaluated, and how to best deploy and monitor models. The framework can also define who is accountable for any negative outcomes of AI. Frameworks will differ between companies. Some will define specific approaches and others will be more open to interpretation. They all seek to achieve the same thing. That is to create AI systems that are interpretable, fair, safe and respectful of a user’s privacy.

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

Deep learning, the set of algorithms that is often used to implement AI, started its triumphal procession with the breakthrough in image recognition, also known as Computer Vision about 20 years ago. It solved earlier unsolvable task of distinguishing cats from dogs and vice-versa, and went on with more complex and demanding tasks. Now it is a common possession to believe that the computer vision is a robust and reliable technology that can hardly fail. However, a year ago, researchers from Berkeley, University of Chicago and University of Washington collected 7,500 unedited nature photos which confuse the most advanced computer vision algorithms.



**Implications of when AI fails. There is a specific article in the GDPR Law that covers this, especially with automated decision making.**

Article 22 of the GDPR states that individuals have the right not to be subject to a decision that has a legal or similar effect upon them and, that is based solely on automated decision-making (without human intervention). There are some exemptions to this right; where the use of personal data is necessary to enter into a contract, if the processing is authorised by law or if explicit consent is given by the data subject.

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

To ensure that any processing of personal data is lawful, fair and transparent, individuals should be provided with specific, clear and meaningful information about how automated decisions are being made about them. Organisations therefore need to communicate the following:

organisations should both understand the challenges and risks around AI and take these fully into account in its design and deployment. Organisation has developed a comprehensive Responsible AI Framework and Toolkit to help companies focus on and address five key dimensions when designing and deploying responsible AI applications:

**Governance**: Governance serves as an end-to-end foundation for all the other dimensions.

**Ethics and regulation**: The core goal is to help organisations develop AI that is not only compliant with applicable regulations, but is also ethical.

**Interpretability and explainability**: Provides an approach and utilities for AI-driven decisions  
to be interpretable and easily explainable by those who operate them and those who are affected by them.

**Robustness and security**: Help organisations develop AI systems that provide robust performance and are safe to use by minimising the negative impact.

**Bias and fairness**: Address the issues of bias and fairness—recognising that while there is no such thing as a decision that is fair to all parties, it is possible for organisations to design AI systems to mitigate unwanted bias and achieve decisions that are fair under a specific and clearly- communicated definition.