The task

**Responsible AI**

Responsible AI is a governance framework that documents how a specific organization is addressing the challenges around artificial intelligence (AI) from both an ethical and legal point of view.

Four key principles of responsible AI:

* fairness,
* transparency and explainability
* human-centeredness
* privacy and security.

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

IBM’s “Watson for Oncology” Cancelled After $62 million and Unsafe Treatment Recommendations.

In 2013, IBM partnered with The University of Texas MD Anderson Cancer Center to develop a new “Oncology Expert Advisor” system. internal IBM documents and found that IBM’s Watson was making erroneous, downright dangerous cancer treatment advice. The Software was trained on a small number of hypothetical cancer patients, rather than real patient data. The result, medical specialists and customers identified “multiple examples of unsafe and incorrect treatment recommendations,” including one case where Watson suggested that doctors give a cancer patient with severe bleeding a drug that could worsen the bleeding.

**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)**

Article 22(1) of the UK GDPR limits the circumstances in which you can make solely automated decisions, including those based on profiling, that have a legal or similarly significant effect on individuals. “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 affects him or her”

A social security process which automatically evaluates whether an individual is entitled to benefit and how much to pay is a decision ‘based solely on automated processing’ for the purposes of Article 22(1). As well as having a legal effect, the amount of benefit received could affect a person’s livelihood or ability to buy or rent a home so this decision would also have a ‘similarly significant effect’.

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

Establishing internal governance, for example by an objective review panel, that is diverse and that has the knowledge to understand the possible consequences of AI infused systems. A key success factor is leadership support and the power to hold leadership accountable.

Ensuring the right technical guardrails, creating quality assurance and governance to create traceability and auditability for AI systems. This is an important part of every organisation’s toolkit to allow operational and responsible AI to scale.

Investing more in their own AI education and training so that all stakeholders – both internal and external – are informed of AI capabilities as well as the pitfalls.