**What is Responsible Artificial Intelligence (AI)?**

Responsible Artificial Intelligence (AI) is the practice of designing, developing, and deploying AI with good intention to empower employees and businesses, and fairly impact customers and society—allowing companies to engender trust and scale AI with confidence.

When we talk about AI, we usually mean a machine learning model that is used within a system to automate something. For example, a self-driving car can take images from sensors.

An important goal of responsible AI is to reduce the risk that a minor change in an input's weight will drastically change the output of a machine learning model. Within the context of conforming to the four tenets of corporate governance, responsible AI should be:

* Each step of the model development process should be recorded in a way that cannot be altered by humans or other programming.
* The data used to train machine models should not be biased.
* The analytic models that support an AI initiative can be adapted to changing environments without [introducing bias](https://www.techtarget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-AI-bias).
* The organization deploying AI programming is sensitive to its potential impact -- both positive and negative.

**Find instances where AI has failed?**

**Microsoft’s AI Chatbot Tay**

With chatbots becoming popular across social networks, Microsoft launched its version for Twitter users in March 2016. Monikered ‘Tay’, it was programmed to have casual conversations in the language of a typical millennial.

According to the company, Tay leveraged AI to learn from these interactions to hold better conversations in the future. However, the Twitter chatbot had to be taken down less than 24 hours post its launch.

Targeting its vulnerabilities, trolls on the microblogging website manipulated Tay into making deeply sexist and racist statements.

Following this debacle, Peter Lee, Microsoft’s corporate VP for AI and research issued a public apology, which stated that the company took “full responsibility for not seeing this possibility ahead of time.”

**There is a specific article in the GDPR Law that covers this, especially with automated decision making. (opt in and out options).**

The GDPR (Article 22(1)) imposes legal requirements on whoever uses the AI system for profiling and/or automated decision-making purposes, even if they acquired the system from a third party. These requirements include:

* Fairness, which includes preventing individuals from being discriminated against;
* Transparency towards individuals, including meaningful information about the logic involved in the AI system; and
* The right to human intervention, enabling the individual to challenge the automated decision.

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

There are few steps to ensure AI is served in a responsible way:

Organisations must think of AI technology in a holistic way – understanding where AI sits in the value chain and creating the right structures to ensure long-term governance by:

* 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 audit ability 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.