**What is Responsible AI?**

Responsible AI is a governance framework. The framework documents how specific organisations are addressing the ethical and legal challenges around the subject of AI. Within the framework would be an outline of where responsibility lies if something goes wrong. As there are currently no AI regulations then the standards and best practicies adopted will vary from organisation to organisation.

According to SearchEnterpriseAI (<https://searchenterpriseai.techtarget.com/>) the principles of responsible AI are:

* Comprehensive
* Explainable
* Ethical
* Efficient

**Where has AI failed? (used maliciously or incorrectly)**

Stories taken from 2018 in Review: 10 AI failures. (<https://medium.com>)

**IBM Watson comes up short in healthcare**

“This product is a piece of shit” wrote a doctor at Florida’s Jupiter Hospital regarding IBM’s flagship AI program Watson,[according to internal documents obtained by Stat](https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/). Originally a question-answering machine, IBM has been exploring Watson’s AI capabilities across a broad range of applications and processes, including healthcare. In 2013 IBM developed Watson’s first commercial application for cancer treatment recommendation, and the company has secured a number of key partnerships with hospitals and research centres over the past five years. But Watson AI Health has not impressed doctors. Some complained it gave wrong recommendations on cancer treatments that could cause severe and even fatal consequences.

After spending years on the project without significant advancements, IBM is reportedly downsizing Watson Health and laying off more than half the division’s staff.

**Chinese billionaire’s face identified as jaywalker**

Traffic police in major Chinese cities are using AI to address jaywalking. They deploy smart cameras using facial recognition techniques at intersections to detect and identify jaywalkers, whose partially obscured names and faces then show up on a public display screen.

The AI system in the southern port city of Ningbo however recently embarrassed itself when it falsely “recognized” a photo of Chinese billionaire Mingzhu Dong on an ad on a passing bus as a jaywalker. The mistake went viral on Chinese social media and Ningbo police apologized. Dong was unfazed, posting on Weibo: “This is a trivial matter. Safe travel is more important.”

CloudWalk Deep Learning Researcher Xiang Zhou told Synced the algorithm’s lack of live detection was the likely problem. “Live detection at this distance is challenging, recognizing an image as a real person is pretty common now.”

**Implications of when AI fails**

Using the stories above of when AI has failed, we can see that when AI fails the implications can have a detrimental consequences.

Using IBM Watson as my first case study, despite gaining key partnerships that system failed to impress Doctors who complained it gave wrong recommendations on cancer treatments that could cause severe and even fatal consequences.

The facial recognition cameras used in China showed how AI could not distinguish between a live face and a digital image and incorrectly identified someone. On this occasion the effects were only minor however this could cause problems with people being accused of things they hadn’t done. In 2019 Amazon’s Rekognition software (facial recognition) identified 26 New England (USA) professional athletes as criminals. The misclassifications were an embarrassment for Amazon, which has marketed Rekognition to police agencies for use in their investigations.

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

As the scope and use if AI grows, any organisations using this technologu need to ensure they focus upon responsible and eithcal use (rather than ethical only). As with many other standards with regards to the use of data, responsible AI should capture the wider concepts and approaches that will outline shared responsibility. All AI should be developed with the protection of the public and also innovation in mind. The rise in AI development has outlined the urgency for business leaders to ensure “they get it right”. Kate Rosenshine (<https://cloudblogs.microsoft.com/>) outlines the following 3 steps on Microsoft Industry Blogs:

**3 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 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.