# Artificial Intelligence (AI)

In Data Science we process a lot of data through AI. With GDPR, it is becoming increasingly important to understand the ethics behind the data that is collected, stored, processed and evaluated.

**Your task is to (max 500 words):**

1. Find out what Responsible AI is?
2. Find instances where AI has failed? Or been used maliciously or incorrectly.
3. 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 opt-out options)
4. What should organisations do to ensure that they are being responsible with AI and the wider use of data in general?

Extra:

<https://machinelearningmastery.com/tensorflow-tutorial-deep-learning-with-tf-keras/>

# What is Responsible AI

Ref: <https://www.forbes.com/sites/forbestechcouncil/2021/09/01/six-essential-elements-of-a-responsible-ai-model/?sh=640f019456cf>

As the use of artificial intelligence or AI has expanded in business and government, we need to incorporate ethics, morals and values. Not doing so would be careless or catastrophic, however, this is a grey area - who determines what is and isn’t ethical and moral?

Moral and ethics are often used interchangeably, which means we need a framework that captures the nuances, this we call ‘responsible AI’.

**Responsible AI is composed of autonomous processes and systems that explicitly design, develop, deploy and manage cognitive methods with standards and protocols for ethics, efficacy and trustworthiness**. This includes standards for; data, algorithms, technology, human computer interaction or HCI, operations and ethics, morals and values

Responsible AI must be all of the following:

1. **Accountable**: Algorithms, attributes and correlations are open to inspection.

2. **Impartial**: Internal and external checks enable equitable application across all participants.

3. **Resilient**: Monitored and reinforced learning protocols with humans produce consistent and reliable outputs.

4. **Transparent**: Users have a direct line of sight to how data, output and decisions are used and rendered.

5. **Secure**: AI is protected from potential risks (including cyber risks) that may cause physical and digital harm.

6. **Governed**: Organisation and policies clearly determine who is responsible for data, output and decisions.

# - Instances where AI has failed, used maliciously or incorrectly

https://www.lexalytics.com/lexablog/stories-ai-failure-avoid-ai-fails-2020

## Microsoft Tay - The Racist Chatbot - (AI Used Maliciously)

Microsoft tried to create a chatbot that could respond on Twitter in a ‘casual and playful’ way. However, within 24 hours of launch, internet ‘trolls’ had corrupted it by flooding it with racist, misogynistic, and anti-semitic tweets, from which the AI was learning. Microsoft pulled the plug after attempting to make it work. They had not planned for a failure as catastrophic as this.

* 1. **Amazon’s Recruiting Tool - The Misogynistic AI Recruitment Tool - (AI Failure)**

Amazon tried to create a tool that could be given 100 CV’s and it would recommend five of them and then recruit them directly. However, what the AI had ‘learnt’ was that white and male candidates were better. By training the AI on a dataset from engineering job CV’s that were majority white males and benchmarking against ‘majority’ white male roles at Amazon.

## IBM’s Oncology Expert Advisor - Unsafe Treatment Recommendations (AI Used Incorrectly)

IBM partnered with Texas University Cancer Centre to build a tool that would eradicate cancer by enabling clinicians to uncover insights from patient databases. It turns out the tool was making mistakes with some being dangerous. The tool was learning from ‘hypothetical’ patients instead of real patient data. The tool was scrapped as it didn't have the trust of clinicians for its unsafe recommendations and the total cost was $62m!

# - Implications of when AI fails

<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-uk-gdpr-say-about-automated-decision-making-and-profiling/>

The GDPR requires consent to be opt-in. It defines consent as “freely given, specific, informed and unambiguous” given by a “clear affirmative action.” It is not acceptable to assign consent through the data subject’s silence or by supplying “pre-ticked boxes.”

The UK GDPR gives a specific right to withdraw consent. You need to tell people about their right to withdraw, and offer them easy ways to withdraw consent at any time.

## We Buy Any Car Limited

ICO fined We Buy Any Car Limited £200,000 after sending 191.4 million marketing emails and 3.6 million marketing SMS messages to individuals without fully satisfying the requirements of the soft opt in**.**

Under GDPR, organisations who fail to comply and/or suffer a data breach could face a fine. In the most serious cases, this fine could be up to 17 million euros, or 4% of a company’s annual turnover. This upper limit far exceeds the current maximum fine of £500,000 allowed under the Data Protection Act.

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

Three steps companies can take are:-

1. Establishing internal governance, e.g. 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.
2. 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.
3. 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.