**Homework Week 6**

Artificial Intelligence (AI)

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

***What is Responsible AI is?***

Artificial intelligence (AI) aims to replicate human intelligence in machines/technology. Artificial Intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans.

Because AI can very easily become uncontrollable, it is important that structures and regulations are in place to help control and manage AI. This theory birthed 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.

Businesses develop responsible AI strategies and principles based on the values of the organisations as well as cultivating responsible AI-ready culture throughout their businesses. They set principles from implementation to governance with practices, tools, and technologies built on multidisciplinary research, shared learning, and leading innovation.

Principles set around responsible AI include:

* Fairness - AI systems should treat all people fairly
* Privacy - AI systems should respect privacy
* Security - AI systems should be secure
* Reliability - AI systems should perform reliably
* Safety - AI systems should perform safely
* Transparency - AI systems should be understandable
* Accountability - People should be accountable for AI systems
* Inclusiveness - AI systems should empower everyone and engage people

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

Artificial intelligence (AI) is doing a lot of good and will continue to provide many benefits for our modern world, but along with the good, there will inevitably be negative consequences.

AI has been used incorrectly or maliciously by terrorists through the expansion of autonomous drones and the introduction of robotic swarms.

It has also been used maliciously by hackers through nefarious acts such as phishing and delivery of viruses/malware to software and computers.

There is a story online about when AI has failed –

***When Alexa Became A Party Animal – Copied from https://analyticsindiamag.com/top-5-epic-artificial-intelligence-fails/***

*Amazon’s smart assistant Alexa is definitely one of those smart devices that most people want to have in their home. And why not, Alexa makes our life better (or create problems?).*

*In 2017, a German man Oliver Haberstroh witnessed something really weird with the world-renown AI Alexa. When Haberstroh was away from home, Alexa started playing music on its own at 1:50 AM and turned the volume so up that the entire neighbourhood woke and even ended up calling the cops, who broke into the house to disconnect Alexa. Well, the cops did install a lock while leaving.*

***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).***

When AI fails especially in some business domains for example in healthcare or government, it can pose serious risks to life and wellbeing.

Businesses can also lose lots of money and investments.

Failed AI can also result in loss of data and exposure to threat.

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

Google has recommended the following principles to help with responsible AI:

* Use a human-centred design approach
* Identify multiple metrics to assess training and monitoring
* When possible, directly examine your raw data
* Understand the limitations of your dataset and model
* Test, Test, Test
* Continue to monitor and update the system after deployment

Hence businesses/organisations should ensure that:

* AI benefits people and the planet by driving inclusive growth, sustainable development and well-being.
* AI systems are designed in a way that respects the rule of law, human rights, democratic values and diversity, and include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
* There is transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
* AI systems functions in a robust, secure and safe way throughout their life cycles and potential risks are continually assessed and managed.
* Developing, deploying or operating AI systems are held accountable for their proper functioning in line with the above principles.

**THE 3 CHALLENGES IN AI**

***Time:***

Time is of the essence if AI is to develop and become a very successful tool for organisations/businesses and the whole world as a whole. AI will require time for development, enough time for legislation that will ensure that AI is practised responsibly and also time for talents to develop and for the general public to fully trust in AI.

***Talent:***

Not many organisations/businesses are willing to put money into development of people skills for AI, and the subject is not well represented in industry-focused education and training curricula.

Even though the demand for the study of computer sciences, mathematics and all relevant studies for the knowledge of AI have risen over time, there are still not enough people to enable every business or organisation to unleash their vision of machine-powered progress on the world. Just as in other areas of science and technology there is a skills shortage – simply not enough people who know how to operate machines which think and learn for themselves.

Because AI is an emerging technology, there are few who possess the skills or training necessary for [artificial intelligence](https://www.tiempodev.com/blog/7-things-software-leaders-should-know-about-artificial-intelligence) development. Because this is a significant problem in the software development industry, many companies will need to allocate additional budget towards artificial intelligence development training, or the hiring of artificial intelligence development specialists.

***Trust:***

AI is perceived as a black box – people don’t feel comfortable when they don’t understand how the decision was made. “For example algorithms used by banks are mainly linear maths and it’s pretty easy to explain the path from the input to the output – ‘I denied your mortgage application because, you don’t have a job, or whatever…”

“With multi-layer neural networks, the average human doesn’t understand, so now we’re making predictions based on things that people don’t understand and that’s going to make people uncomfortable.”

Hence, somewhere AI has not been able to create trust among people. And the only solution that seems to this problem is to let people see that this technology really works. However, the reality is somewhat different. And it shows that there is a lot of opportunities to make things better by having predictions that are more accurate.

This raises problems of government overstep. Suppose, a part of the regulation tells that citizens may have the right to ask an explanation for decisions that are made about them with the help of Artificial Intelligence.