**HLT WK7**

In Data Science we process a lot of 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.

Your task is to:

**•Find out what Responsible AI is?**

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.

Responsible 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.

The four key principles of AI are: fairness, transparency and explainability, human-centeredness, and privacy and security.

**•Find instances where AI has failed? Or been used maliciously or incorrectly.**

**•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).**

There are further protections where automated decision making takes place – i.e., where an artificially intelligent system is solely responsible for a decision that has legal effects or significantly affect a data subject. This reflects the common-sense expectation that important decisions, for example whether to offer someone a job or provide a mortgage, should not be entirely delegated to a machine.

Under the GDPR, this type of automated decision making can only take place in the following situations:

* Human involvement – If a human is involved in the decision-making process, it will not be a decision based solely on automated processing. However, that involvement would have to be meaningful and substantive. It must be more than just rubber-stamping the machine’s decision.
* Consent – Automated decision making is permitted where the individual has provided explicit consent. While this sounds like an attractive option, the GDPR places a very high threshold on consent and this will only be valid where the relevant decision-making process has been clearly explained and agreed to.
* Performance of contract – Automated decision making is also permitted where it is necessary for the performance of a contract or in order to enter into a contract. An example might be carrying out credit checks on a new customer or considering whether to offer someone a job.
* Authorised by law – Finally, automated decision making processing is permitted where it is authorised by law.

Even where automated decisions are permitted, you must put suitable safeguards in place to protect the individual’s interests. This means notifying the individual (see below) and giving them the right to a human evaluation of the decision and to contest the decision

<https://www.linklaters.com/en/insights/blogs/digilinks/ai-and-the-gdpr-regulating-the-minds-of-machines>

Five mistakes made by AI in the history of humanity, from intelligent devices to chatbots to self-driving cars include:

1. AI struggles for Image Recognition

In 2015, Google learned this the hard way when it recently launched the image recognition feature in its google photos application, powered by artificial intelligence and neural networks. The image recognition features of google photos are designed to identify specific objects or specific people in the given images.But machines can make mistakes, can they? In the case of google photos, a user got offended when an image of his two black friends was tagged as “GORILLAS,” He decided to take this matter to Twitter, and Goole apologized.

2. AI in Military Services creates ethical dilemmas among People

In the past few years, scholars from the field of AI and ML have taken part in dozens of conferences and talks dedicated explicitly to the ethics and dangers of AI systems’ future. One such example is The White House, which has released its report on this issue, and even Stephan Hawking has his concerns regarding the same.

A statement from a person who is like a rock star in this field, “Peter Asaro” stated that, in some areas like the zones having all military forces removed like in between North Korea and South Korea, semi-autonomous weapons like sentinel guns that lock onto a target with no human intervention, are already deployed.

According to Peter Asaro, he said that “It’s important to realize that targeting a weapon is a moral act and choosing to pull the trigger to engage that weapon is another moral act. These are two crucial facts that we should not have made fully autonomous.

3. Smart Devices debate existential dilemmas

What is the meaning of your life, why do we continue to live, who are we and why are we here, what Is our purpose in life? You might think about these questions. To clarify things, these are some of the existential questions recently debated by two adjacent google home devices, which, for your knowledge, are driven by artificial intelligence and machine learning technology.

This happened in Jan 2017 in a live streaming service offered by Twitch. A debate was set up by putting two google home smart speakers next to each other in front of a camera, and things got weird very soon.At a point, they got into a heated debate about whether they were humans or robots. Things did not end here; the public posted several questions, and insults were exchanged like a manipulative bunch of metal.

4. Microsoft chatbot “Tay” gives spouting abusive epithets on Twitter

Microsoft ran into a significant public dispute back in the spring of 2016 when its Twitter chatbot “Tay,” which uses AI at its core, started tweeting some random and abusive epithets and also made some Nazi comments like “Hitler was right” or “9/11 was an inside job”.

According to Microsoft, Tay repeated mechanically offensive statements made by other humans trying to provoke Tay. Tay runs on top of artificial intelligence and adaptive algorithms, making Tay blending in other relevant data and phrases.

In a public press conference, Microsoft said, “ The more you chat with Tay, the smarter she gets. Maybe not so much,” and Tay was taken offline after 16 hours

5. Uber Self Driving Car ran red lights during real-world testing

Uber is one of the most popular transportation services globally in the 21st century. Still, uber had to go through rough times in late 2016 when uber conducted a test on its self-driving cars in San Francisco without approval from California state regulations.

When resulting documents showed that Uber’s autonomous vehicle ran six red lights in the city during their test ride, the situation got out of hand, but there was a driver behind the wheel to take over if something went wrong.

According to Uber’s statement in a public press conference, the traffic rules violation resulted from driver error. Still, internal documents later revealed that at least one vehicle was driving itself when it ran a red light at a busy crosswalk in daylight.

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

3 steps to ensure AI is served in a responsible way

So, where do we go from here to ensure that AI is serving our society in a healthy and 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