HLT WEEK 8 AI ( Gokhan Koyunlu)

1. Find out what Responsible AI is?

AI and the machine learning models that support it should be comprehensive, explainable, ethical and efficient.

* Comprehensiveness – comprehensive AI has clearly defined testing and governance criteria to prevent machine learning from being hacked easily.
* explainable AI is programmed to describe its purpose, rationale and decision-making process in a way that can be understood by the average end user.
* Ethical AI initiatives have processes in place to seek out and eliminate bias in machine learning models.
* Efficient AI is able to run continually and respond quickly to changes in the operational environment.

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

## AI fails to do image recognition

Researchers from Berkeley, University of Chicago and University of Washington collected 7,500 unedited nature photos which confuse the most advanced computer vision algorithms.

## b) AI that hated humans

*“Hitler was right to hate the jews”* - said Tay, a Microsoft most advanced chatbot after 24 hours of ‘learning’ from interactions with humans. The idea was to create a slang-laden chatbot, that would bring a new level of machine-human conversation quality. But it turned out to be *“a robot parrot with an internet connection”* as it was described in the Verge

The chatbot was built on top of AI technology stack developed in the company, but harsh reality seems to have spoiled the innocent AI worldview: a good example of how data can destroy an AI-model built in a ‘clean’ lab condition without immunity to harmful influence from the outside.

## c) AI to fight cancer could kill patients

Another failure cost 62 million dollars, which were spent by IBM to build an AI system that would help fight cancer. But again, the result was rather disappointing:

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 out options).

The complications of AI In addition, suppliers of AI solutions could be liable if they fail to exercise reasonable skill and care in performing their responsibilities: such as**defects in AI technology performance arising from faulty implementation or software engineering design or build flaws**.

The vast scope of GDPR has raised fresh challenges — chief among them is the complex interaction between AI and the GDPR. In particular, this shines a spotlight on Article 22, which concerns automated profiling and decision-making, where the incorrect use of personal data can have huge ramifications for the individuals concerned.

The problem is that existing AI system logic takes automated decisions without user consent. Since data is the engine behind AI, Article 22 impacts every industry hoping to leverage the power of technology to drive efficiencies through automated means.

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

Transparency , Accountability , Fairnes