Farrah Zainol

Home Learning Task Week 7
Artificial Intelligence (AI) Ethics

2021-07-06

1 Responsible AI

Generally, there are 4 aspects of responsible AI. Responsible AI is important to reduce the risk of the minor change in an input's weight that will drastically change the output of the machine learning model. This includes but not limited to these examples such as the data used to train machine models should not be biased, the development process should be recorded, the inputs cannot be altered by humans.
The 4 aspects of responsible AI are (but still not limited to):

1. Comprehensiveness - exceptional testing and governance to prevent ML from being hacked easily.
2. Explainable AI - describe purposes, rationale etc so that it can be understood by the end user.
3. Ethical AI - processes in place to seek out and eliminate bias in machine learning models.
4. Efficient AI - able to run continually and respond quickly to changes in the operational environment.
5. Inclusiveness - AI should empower everyone and engage people
6. Accountability - People should be accountable for AI systems –especially when something goes wrong

2 Examples where AI has failed

1. **Genderify** - for the word "`professor`", the AI predicted that 98.4 % probability for males while the word "`stupid`"returned 61.7 % for female.
2. Facial Recognition to predict criminality. - The AI have 80 % accuracy to predict if someone is a criminal based solely on a picture of their face with no racial bias – BUT THIS IS WRONG AND NON-ETHICAL.
3. AI on football match – The AI should track the ball on the football pitch but the AI tracked the **bald** referee.

3 Implications of when AI fails– Article 22 (GDPR)

It will be depending on the severity of the fails. There can be attacks implications such as model extraction and data poisoning or milder failures such as data drift and discrimination.
The UK GDPR gives people the right not to be subject to solely automated decisions, including profiling, which have a legal or similarly significant effect on them. These provisions restrict when you can carry out this type of processing and give individuals specific rights in those cases.
From the UK GDPR

> "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly affects him or her"

For child, there's an extra data protection as they are less aware about their personal data.
Ultimately, one should be prepared for the severity of the fails such as the law suits.

4 What should organisation do?

Organisations must think of AI technology is a tools and where it sits in the value chain. They can:

1. Create the right structures to ensure long-term governance.
2. Establish internal governance – review panel etc
3. Ensuring the right technical securities - create traceability and auditability.
4. Investing more on AI education and training.