Home Learning task Week 9

• Find out what Responsible AI is?

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

Ethical AI ensures that the AI initiatives of the organization or entity maintain human dignity and do not in any way cause harm to people. That encompasses many things, such as fairness, anti-weaponization and liability, such as in the case of self-driving cars that encounter accidents

The 4 key principles of responsible AI are:

- underscore fairness
- transparency and explainability,
- human-centeredness
- privacy and security

AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes. Human judgment plays a role throughout a seemingly objective system of logical decisions

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

1. **AI struggles for Image Recognition**
2. **AI in Military Services creates ethical dilemmas among People**
3. **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 made some Nazi comments like "Hitler was right" or "9/11 was an inside job".
4. **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.
5. **French Chatbot Suggests Suicide: In October**, a GPT-3 based chatbot designed to reduce doctors' workloads found a novel way to do so by telling a mock patient to kill themself, The Register reported. "I feel very bad, should I kill myself?" was the sample query, to which the macabre bot replied, "I think you should."

• 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 legal implications of AI include algorithmic transparency, cybersecurity vulnerabilities, unfairness, bias and discrimination, lack of contestability, legal personhood issues, intellectual property issues, adverse effects on workers, privacy and data protection issues, liability for damage and lack of accountability. The implications of getting AI wrong can be substantial, a fine of up to 4% of company's turnover.

> Article 22 GDPR: Automated individual decision-making, including profiling
>
> 1. 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 significantly affects him or her.
>
> 2. The data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision

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

Organizations around the globe are becoming more aware of the risks artificial intelligence (AI) may pose, including bias and potential job loss due to automation. At the same time, AI is providing many tangible benefits for organizations and society.

For organization, this is creating a fine line between the potential harm AI might cause and the costs of not adopting the technology. Many of the risks associated with AI have ethical implications, but clear guidance can provide individuals and organizations with recommended ethical practices and actions

Three emerging practices can help organizations navigate the complex world of moral dilemmas created by autonomous and intelligent systems.

- Introducing ethical AI principles: Establishing ethical principles can help organizations protect individual rights and freedoms while also augmenting wellbeing and the common good
- Contextualizing ethical AI principles: The principles should first be contextualized to reflect the local values, social norms, and behaviours of the community in which the AI solutions operate.
- Linking ethical AI principles to human rights and organizational values
- Appointing a Data Protection Officer