Responsible Artificial Intelligence

Responsible AI is a set of guiding principles which exist for the purpose of ensuring that Artificial Intelligence (AI) are used in a manner which is fair and ethical to all and that AI systems should be transparent and secure.

According to Microsoft, Responsible AI is comprised of six key principles (Microsoft, 2022):

1) Fairness — systems should reduce existing unfairness or biases in society rather than keeping it the same or make it worse.
2) Reliability and safety — the systems should work as intended and they should be monitored to ensure that mistakes aren't being made, especially when it comes to anything which affects human lives e.g., diagnoses of diseases.
3) Privacy and security — AI requires the use of data, which includes user data, and it is vital that this isn't leaked.
4) Inclusiveness — all communities should be included when considering who the AI is designed for.
5) Transparency — the system should be open, which serves several purposes, from trustworthiness to debugging any issues that may arise. People should be able to understand the behaviour of the system and know the reasoning behind decisions that have been made.
6) Accountability — AI technology has real-world impact, which means that there is a need for people to take responsibility it, which serves as encouragement to ensure that the principles are upheld.

With these principles in place, AI can be used to bring benefit to society instead of harm. Different organisations will have different principles but they all have the same intention.

However, AI can fail and has failed, leading to a spectrum of outcomes, from when Microsoft released a chatbot, on Twitter, to interact with users leading to certain users using the opportunity to teach it controversial sentiments, which it then repeated in its Tweets (Hunt, 2016), to when a jaywalking pedestrian was killed during a trial of a self-driving Uber car, as the AI didn't recognise any pedestrian that didn't use the crosswalk as a 'person' (McCausland, 2019).

Privacy is also a concern, and the GDPR covers the legal aspects of any data of subjects in the EU. Article 22 of GDPR covers AI.

In brief, it states that subject has the right to not be a subject in any AI-based decision-making which will affect then. This does not apply if the subject and the data-controller are entering into a contract which requires the use of AI, the subject explicitly gives consent, or if the EU/EU-member State authorises it, whilst ensuring the subjects rights and legitimate interests are protected) — the very least of these measures is the right of the subject to object to any decision the AI has made. In addition to this, data used for these decisions cannot be based on categories stated in Article 9, paragraph 1, unless points (a) and (g) or paragraph 2 applies.

References

Hunt, E. (2016, March 24). *Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter*. Retrieved from The Guardian Website:

https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

Intersoft Consulting. (2022, January 30). *Art. 22 GDPR Automated individual decision-making including profiling*. Retrieved from GDPR-Info: https://gdpr-info.eu/art-22-gdpr/

McCausland, P. (2019, November 9). *Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk*. Retrieved from NBC News Website: https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

Microsoft. (2022, January 30). *Responsible AI principles from Microsoft*. Retrieved from Microsoft Website: https://www.microsoft.com/en-us/ai/responsible-ai