Responsible AI

Artificial intelligence (AI) is the development of computer systems that can perform tasks with little to no human interference, tasks that would normally require human intelligence to complete. These include visual perception, speech recognition, language translation and decision making. AI is seen by some as being beneficial for promoting productivity and progress but concerns others, however, prompting thoughts of unemployment; unfair decisions; and privacy concerns. Responsible AI aims to combat these issues and create accountability for AI systems.

Responsible AI is a framework for governing how AI is used, with the aim of ensuring that potential issues are prevented from arising. Whilst the framework will differ between companies, they all strive for the same results which is an AI system that is interpretable, fair, safe and private. As well as this, it also needs to be clear who is accountable if something goes wrong due to the AI. The level of interpretability is important when designing an AI system. Some models are easier to interpret than others and responsible AI can define how interpretable models are built or when a less interpretable model is acceptable. Fairness is important, otherwise systems could discriminate against certain groups of people because of bias in the data. Generally, the more interpretable a model is, the easier it is to correct bias, ensuring fairness. Although safety and security have been priorities in software development for some time, new issues arise when involving AI. This is because AI is not deterministic. If AI encounters an unfamiliar situation, then it can make unexpected decisions – a serious problem with self-driving cars, for example.

Privacy is an extremely important aspect of responsible AI. For this reason, personal data (such as medical history) should not be handled by an AI system. Within the EU and the UK, GDPR exists, to govern this. Article 22 of GDPR law states that AI, including profiling, cannot be used as the sole decision-maker in choices that can have legal or similarly significant impacts on individuals' rights, freedoms, and interests. When AI fails, the GDPR imposes fines that are up to 4 percent of a company's global turnover, or €20 million (whichever is greater).

Evidence of why responsible AI is important can be seen in scenarios where AI has failed. Examples of this include IBM's "Watson for Oncology", which gave out unsafe treatment recommendations; Microsoft's AI Chatbot, which became corrupted by Twitter Trolls; Apple's face ID, that was fooled by a 3D mask; and Amazon's failures with both recruitment and facial recognition (the AI recruitment system was inadvertently trained to be misogynistic, and the facial recognition matched 28 members of the US Congress with convicts).

To ensure that they are being responsible with AI, organisations should remember that they are accountable and adhere to the four points mentioned previously, confirming that their AI systems are interpretable, fair, safe and private. By doing so, they can be assured that their AI system is indeed responsible, helping them to avoid any negative implications.