Artificial Intelligence

1. What is Responsible AI

AI systems will make decisions that have ethical ground and consequences. Responsible AI ensures that development processes take into account social and ethical implications of AI as it integrates and replaces systems and social structures. Principles for Responsible AI:

* Accountability
* Responsibility
* Transparency and Explainability
* Ethical purpose and societal benefit
* Fairness and non-discrimination
* Safety, privacy and reliability

Look at options and choices, document them and ensure sufficient regulations are in place to monitor and control which will drive transformation for better solutions, fairness, human-focussed, privacy and security.

1. Instances where AI has failed or been used incorrectly/maliciously in 2020:

* Facial recognition technology is the standard for mobile payment systems and banking services in China. In a video a 94 year old grandmother is seen being lifted up by her son in order to reach a facial recognition camera to activate her social security card at a bank.
* The introduction of AI ball tracking cameras promised to make live coverage cost effective but glitches have been off putting. Caledonia FC football club used the AI Ball tracking camera which confused the ball with the referee’s ball head especially when its view was obscured by players or shadows.
* After 5 years Walmart ended it’s contract with Bossa Nova Robotics which made robots that that scanned shelves for inventory. Walmart were hoping that the technology could help to reduce labour costs and increase sales by making sure products are kept in stock. 500 robots had been deployed in more than 4700 stores but Walmart ended the partnership as it found different simpler solutions that proved just as useful i.e. human workers.
* Starsky Robotics shut down the San Francisco based autonomous truck company in 2020 even though they had previously received more than $20 million in funding and achieved many driverless firsts. Their approach combined self-driving software on highways and remote monitoring and control by human drivers for the first and last mile. Self-driving problems remain too difficult to solve hence their closure.
* In France a chatbot designed to reduce doctors workload responded to a query – “I feel very bad, should I kill myself” by replying “I think you should”. Nabla, the chatbot creator wisely concluded that the erratic and unpredictable nature of the software’s responses made it inappropriate for interacting with patients in the real world.

1. Implications when AI fails (GDPR – automated decision making (opt in/out)

AI incidents can be attacks or failures. This includes potential violation of privacy and security i.e. stealing of data, incorrect predictions which could cause great harm if left unaddressed and unaccounted for. Therefore AI failures don’t require an external attacker. If the model generates false negative predictions i.e. tumour detection, organisations could combine automated imaging results with activities like follow up radiologist reviews or blood tests to catch any potentially incorrect predictions by automated decisions and even improve the combined human and machine efforts.

Under GDPR, AI can’t be used as the sole decision maker in choices that have a legal or significant impact on individuals rights, freedom and interests i.e. loan application. There are exceptions i.e. law authorises such decisions, or when individual gives his consent. Also, there needs to be a mechanism that requires an action to opt in as opposed to pre-ticked boxes. If organisations want to use the data for another purpose, consent needs to be obtained again. Individuals can withdraw consent at any time which means their data needs to be removed.

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

AI depends on large volumes of digitally stored data. Many AI applications process personal data. Personal data may contribute to the data sets used to train machine learning systems, namely to build algorithmic models. Such models can be applied to personal data, to make inferences concerning particular individuals. In 2018 the EU introduced a set data protection principles that regulates entities that process personal data of EU residents/citizens. The use of personal data has implications for privacy, data protection and the associated rights for individuals. Individual rights have been strengthened under GDPR when stricter rules apply to the collection and use of personal data. Organisations need to be more transparent, more accountable for what they do with the personal data. Appropriate processes need to be in place to deal with decisions based on automated processing. Appropriate security measures need to be in place to safeguard personal data and it should only be kept for as long as necessary. Organisations also need to be careful as machine learning algorithms have the potential to make decisions that are discriminatory, erroneous and unjustified which has implications for the data protection principle of accuracy i.e. inaccurate predictions based on biased profiling. Forbes India explains the inherent bias in data in that AI systems are only good or bad as the data they are trained on. Bad data is often laced with racial, gender or ethnic biases. If the bias in the algorithms go undetected it could lead to unethical and unfair consequences.

Organisations outsourcing analytics to companies specialising in AI and machine learning need to consider carefully who has control over the processing of any personal data as this will have implications for compliance and liability. A contract needs to be in place with clear instructions about how the data can be used and the specific purposes for its processing.

AI can be a major contributor to growth. However, businesses need to have protocols in place for implementing AI in a safe and ethical way. Organisations that fail to comply or suffer a data breach could face serious fines under GDPR and Data Protection Act.

Investigate the 3 Challenges in AI:

* Time

It takes time to deliver an AI strategy. AI is being implemented across all industries. This can save time and money and gain a competitive edge. But it takes time to do this properly i.e. determine the outcomes i.e. translate the high-level goals of the company into a business problem and subsequently determine the outcome. Companies must also identify metrics that measure its success. It takes time to develop and deploy machine learning models and applications.

* Talent

There is currently an AI skills gap. This is predicted to get even worse as the number of AI technology companies and job openings are growing. Even companies using readymade software and solutions need to ensure that they have sufficiently skilled employees to deploy and manage the systems and interpret the results correctly. Lufthansa and IBM combined their expertise to overcome this. The UK National AI strategy 2021 recommends a commitment to an ongoing 10 year programme of high level AI skill building to bridge the skills gap.

* Trust

AI is a relatively new technology and complex. Therefore people who are not trained in it can be suspicious of it. They are often seen as black boxes. Being able to describe the models decisions adequately, having good documentation and eliminating bias from the results are key considerations for companies to instil trust in AI.

Also, people have mixed feelings towards AI and fear it may make their jobs obsolete or irrelevant. For large scale adoption of AI across an organisation, buy-in is needed along with support and integration across multiple business processes, IT systems and stakeholders

The UK National AI strategy 2021 recommends ensuring public trust through public scrutiny.