Artificial Intelligence Bias

Bias is an inherent human trait and can be reflected and embedded in everything we create, particularly when it comes to technology.

What is AI bias?

- AI Bias refers to an anomaly in the output produced by a machine learning algorithm.
- Bias in AI is when the machine gives consistently different outputs for one group of people compared to another.
- Typically, these bias outputs follow classical societal biases like race, gender, biological sex, nationality or age.
- This may be caused due to prejudiced assumptions made during the algorithm development process or prejudices in the training data.

What are the types of AI Bias?

- **Cognitive bias** - These are unconscious errors in thinking that affects individuals' judgements and decisions.
- These biases could seep into machine learning algorithms via either designers unknowingly introducing them to the model or a training data set which includes those biases.
- **Lack of complete data** - If data is not complete, it may not be representative and therefore it may include bias.
- It is also difficult to find out the factor that causes the biased output due to the 'black box effect' in AI.

> *Black box AI is any artificial intelligence system whose inputs and operations are not visible to the user or another interested party and are impenetrable.*

What could be done?

- **Blind Taste Test Mechanism** - It works by checking if the results produced by an AI system are dependent upon a specific variable such as their sex, race, economic status or sexual orientation.
- **Open-Source Data Science (OSDS)** - Opening the code to a community of developers may reduce the bias in the AI system.
- **Human-in-the-Loop systems** - It aims to do what neither a human being nor a computer can accomplish on their own.
- It leads to more accurate rare datasets and improved safety and precision.

What are the future aspirations?

- It may be impossible to fully eradicate bias in AI systems due the biases in developers and engineers that get reflected in the systems.
- In the interim, regulators and states must step up to carefully scrutinise, regulate or in some cases halt the use of AI systems which are being used to provide essential services to people.