Algorithmic bias, robotics and legal liability

1. Define algorithmic bias. (1)

2. Explain the causes of algorithmic bias (4)

3. What steps can programmers take to avoid introducing algorithmic bias into AI applications? (3)

4. Who should be legally liable (responsible) if an autonomous (self-driving) vehicle harms a human? The manufacturer, the user, the programmer? (2)

5. Robots may collect and use personal data.
   They also may have to make decisions that affect human safety. Eg. self-driving vehicles.
   They may also be used in military situations.
   List some of the ethical issues surrounding the use of robots in these scenarios (6)

ANSWERS

1. **Algorithmic bias** is when a computer program produces **unfair or discriminatory outcomes** due to **biases** in the data it was trained on or the way it was designed. These biases can lead to unintended consequences and impact people differently based on factors like race, gender, or socioeconomic status. For example, if software is choosing which students to interview for a university place or choosing which people to interview for a job.

2. Algorithmic bias may be caused by bias within the training data used to train the algorithm. If training data reflects societal inequalities (e.g., predominantly white male job applicants), the algorithm may continue those biases. It also might be caused by the internal biases of the programmers who are responsible for developing the algorithm.

3. Programmers should ensure that training data is wide-ranging and represents the entire population. Eg. if developing a facial recognition app, ensure that people of a wide range of ethnicities are included. Programs should be reviewed (audited) by external teams of people to check that unintended bias hasn't been introduced by the programmers or the training data.

4. This is complex and there is no correct answer because all parties may be partly responsible for an outcome. Autonomous vehicles may encounter unpredictable situations that they have not specifically been programmed for. Data breaches affecting robots or AI may be the fault of the human using the system or a fault with the manufacture. Each case would need to be looked at individually.

5. **Personal Data Collection by Robots**: **privacy** (user consent, data **minimisation** – only collecting the bare minimum of data) and **consent** (people opt in not opt out.); **transparency** (Users should understand they are talking to or using a robot, why & how they use the data collected);

   Self-driving cars: impact on jobs, **safety issues** especially in unexpected situations;

   Lethal Autonomous Weapons (LAWS): issues around **accountability** – who is responsible if things go wrong?; should LAWS be making decisions about the life and death of humans? Potential **algorithmic bias** in the programming