TECH TALENT ACADEMY WEEK 7 HLT

What is responsible AI:

Responsible AI is the practice of designing, developing, and deploying AI with good intention to empower employees and businesses, and fairly impact customers and society—allowing companies to engender trust and scale AI with confidence

**Instances where AI has failed or has been used maliciously or incorrectly**

Tesla cars crash due to autopilot feature: Elon Musk’s Tesla Model S crashed north of Houston, killing two people. The car had missed a slight curve in the road, leading it to ram into a tree. As per preliminary investigations and witness statements, the driver’s seat was empty during the crash. As a result, it is believed that Tesla’s Autopilot or Full Self Driving (FSD) system was engaged during the crash. Tesla’s AI-based Autopilot feature can control steering, acceleration, and in some cases, braking.

However, the feature has come under increased scrutiny due to several crashes involving the vehicle. Several safety advocates have criticized Tesla for not doing enough to prevent drivers from relying heavily on its Autopilot features, or for using them in situations that the feature is not designed for.

## **Amazon’s AI recruiting tool showed bias against women**

Amazon’s AI-based experimental hiring tool had a major flaw: it was biased against women.

The model was trained to assess applications by studying resumes submitted to the company over a span of 10 years. As most of these resumes were submitted by men, the system taught itself to favor male candidates. This meant that the AI downgraded resumes with words such as “women’s” (as in the case with “women’s chess club captain”). Similarly, graduates from two all-women’s colleges were also ranked lower.

By 2015, the company recognized the tool was not evaluating applicants for various roles in a gender-neutral way, and the program was eventually disbanded.

**What should organizations do to ensure that they are being responsible with AI and the wider use of data in general**

The Data Ethics Framework guides appropriate and responsible data use in government and the wider public sector. Teams should work through the framework together throughout the process of planning, implementing, and evaluating a new project. Each part of the framework is designed to be regularly revisited throughout your project, especially when any changes are made to your data collection, storage, analysis or sharing processes. The framework is split into overarching principles and specific actions. Overarching principles are applicable throughout the entire process and underpin all actions and all aspects of the project. Specific actions will guide you through different stages of the project and provide practical considerations.