Artificial Intelligence

### What is Responsible AI

The practice of designing AI that positively impacts people by considering 4 key principles. fairness, transparency and explainability, human-centeredness, and privacy and security.

### Instances where AI has failed? (Or been used maliciously or incorrectly).

Twitter apologised for a “racist” image cropping algorithm, users discovered the feature was automatically focussing on white faces over black ones. The company says it had tested the service for bias before it started using it, but now accepts that it didn't go far enough. There’s a vox video I really enjoyed watching here <https://youtu.be/Ok5sKLXqynQ> which goes into great detail on it. But also raises points around how America's Health insurance system and crime prediction tools.

### Implications of when AI fails

Data isn’t abstinent from history. It’s been proven that crime data in america contains some elements of racial profiling which highly affects the performance of said tool. Machine learning is highly unregulated when it comes to reviewing different demographic subgroups.

GDPR gives people the right to know how, where and why their personal data is being processed. Also gives them the option of opting out and requisiting its deletion where no longer needed and must give consent for it’s processing to begin with.

But with Twitter being so visible, how else is AI being misused. How else are companies making decisions that affect my life with a flawed system.

### What should organisations do to ensure they are being responsible with AI and the wider use of Data in general

Evaluate and assess data with an active eye for people most likely to be negatively impacted. If you know a particular marginalised group is most vulnerable in a particular situation then prioritise them in your evaluation. Be more mindful of where we implement such systems and how much human oversight which needs to be included as well.