Responsible AI

There are 8 principles of responsible AI.

1. Producers of ML models must commit to assessing the human impact of incorrect predictions, and where possible, put humans in place to review the process
2. Monitor bias, and develop systems that remove it from the process
3. Improve transparency and explainability of systems to all involved.
4. Make processes reproducible, and therefore easier to monitor and refine
5. Ensure practical accuracy
6. Work to communicate impacts, so that workers are protected when processes are automated.
7. Develop and maintain systems of privacy between stakeholders
8. Ensure data and learning model security while ML systems are being developed.

ethical.institute

Failing to consider these principles can lead to situations where human ethics and laws are breached.

A real world AI failure

A real-life example of AI failure is a machine learning questionnaire algorithm known as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). It was used to determine the risk of whether an arrestee would be a repeat offender.
A study showed that this algorithm was racially-biased despite never asking about race. Individuals were asked questions that modelled existing social inequalities, and minorities, particularly black people, were more likely to be labelled “high-risk” repeat offenders.

A case such as this shows that machine learning algorithms are only as effective as the data they are trained on. More importantly, it would be in conflict with GDPR rules where they apply. According to article 22 of the GDPR law:

"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

The use of automated decision making where legal effects are carried out, is dependent on the explicit consent of the individual affected. However, even where this is the case, the transparency and explainability of the system should be evident (XAI).

How companies can meet AI responsibilities:

- Use a human centred approach to design and deployment. Involve subjects from the outset, telling them what kind of information is being gathered, why and how it is

being used. It needs to be tested and monitored early on, with a sample of people from diverse backgrounds (depending on what personal data is available).
- Once the model is deployed, it needs to be continuously monitored, allowing a space or facility for feedback, so that mistakes can be rectified.
- The raw data should be examined thoroughly by a number of people, to eliminate the possibility of inaccuracies or bias.
- The limitations of the AI model must be clear, and communicated to the user, which will improve feedback, and ensure that insights gained are of high quality. This will ensure that the AI will improve in line with human use. It cannot imply it can do anything beyond its capabilities.

- Continuously test the AI model, and continue to monitor and update it. This understands that many factors can develop over time to render the AI inaccurate, unusable or physically/mentally harmful.