# HLT 6: Practical: Artificial Intelligence (AI)

1. Responsible AI aims to embed principles, values and fairness at the heart of Ai development and implementation. Microsoft, for example, have chosen six principles: Fairness, Reliability & Safety, Inclusiveness, Privacy & Security, Accountability and Transparency.

Increased use of technology means increased possibility for it to impose on privacy and human rights. It is now more difficult -- and more important -- to work to detect biases, identify blind spots in technologies and embed responsibility at the core of AI.

1. Instances of where Ai has failed / been used maliciously or incorrectly include many loan applications. For example, a Finnish man in a rural area was denied credit, [purely based on statistical, not personal information](https://algorithmwatch.org/en/finnish-credit-score-ruling-raises-questions-about-discrimination-and-how-to-avoid-it/). Had he been a Swiss speaking woman, he would have been approved.

Other examples: Amazon’s recruiting tool was [biased against women](https://www.reuters.com/article/amazon-com-jobs-automation/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idINKCN1MK0AH?edition-redirect=in), Microsoft’s [racist chatbot](https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation), and false facial recognition matches and [poor levels of accuracy](https://www.bbc.co.uk/news/technology-50865437) in non-white faces.

1. Implications of when AI fails

Since there are so many parties involved in an AI system, it is difficult to establish liability when AI goes wrong. For example, if a self-driving car were to crash into a pedestrian crossing the road, where would the liability lie? Is it with the driver? The car manufacturer? The software engineers? The regulators for approving the vehicle for use on roads?

AI failures have wide-reaching implications and I believe current legislation and public awareness of these issues is insufficient. Returning to the loan example, there is financial and reputational harm done to the individual denied the loan, as well as infringement to human rights. There is reputational damage to the company using the software, and erosion of trust in them and the wider system. Long term damage caused by loss of trust, potentially making people less likely to trust other AI systems they encounter.

Companies using AI systems are at risk of falling foul of the law. UK GDPR gives people the right not to be subject to solely automated decision-making processes that have a legal or otherwise similarly significant effect on their lives. This includes for example, whether to interview a candidate based solely on an online aptitude test. The following risks are highlighted:

* Profiling is often invisible to individuals.
* People might not expect their personal information to be used in this way, and not understand how the process works or how it can affect them.
* The decisions taken may lead to significant adverse effects for some people.

1. What should organisations do?

Organisations need to plan responses to potential AI incidents. They need to be able to identify and contain the harms that AI could be causing. Other best practices include keeping an up-to-date inventory of all AI systems, monitoring systems for anomalous behaviour, implementing AI-specific preventative security measures, thoroughly documenting all AI and ML systems, and keeping staff up to date with AI ethics-related issues, including translating academic thought into language accessible to engineers and sales managers.