HLT Week 7

Find out what Responsible AI is?

Responsible AI is a governance framework that legally and ethically governs how an organization is addressing the challenges around artificial intelligence. It helps identifies who is responsible if something goes wrong which is an important incentive for responsible AI initiatives.

Find instances where AI has failed? Or been used maliciously or incorrectly.

AI can be used for “deepfakes” which is an abuse of AI that manipulates audio and visual content. This can be used to spread disinformation on the internet as it can make fake news seem legitimate as you can almost make anyone say or do anything.

An example of this is when a UK-based energy firm was conned into transferring £200,000 to a Hungarian bank account after a malicious individual used deepfake audio technology to impersonate the voice of the firm’s CEO in order to authorize the payments

Implications of when AI fails. There is a specific article in the GDPR Law that covers this, especially with automated decision making. (opt in and out options).

When AI fails it can cause casing of profiling. When using automated decision making it can cause people to be treated unfairly and stereotyped.

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

1. Establishing internal governance. A key success factor is leadership support and the power to hold leadership accountable.
2. Having technical guardrails to create traceability and auditability for AI systems.
3. Investing more in their own AI education and training so that all stakeholders – both internal and external – are informed of AI capabilities as well as the pitfalls.