Artificial Intelligence (AI)

**What is Responsible AI?**

Responsible AI is a set of rules that a specific organisation must follow in order to address challegenges around artificial intelligence (AI) from both an ethical and legal point of view. Responsible AI helps ensure that any AI system will be efficient, comply with laws and regulations, operate based on ethical standards and prevent potential for reputational and financial damage



Source: Forbes.

**AI failure**

A lot of things can go wrong with any system, especially if criminals would want to make harm to the AI system. However, AI systems can go amok on their own, without any external interference.AI could fail because of bad or insufficient data, bad engineering, wrong area of application.

There have been cases of AI learning failures:

* Microsoft developed Tay, a Twitter chatbot AI, which started as an experiment in conversational understanding but began to generate racist messages in less than twenty four hours. For example, Tay said that “Hitler was right to hate Jews”
* Cambridge Analytica used data analysis techniques in order to target specific Facebook users with political content via Facebook’s targeted advertising service.
* Amazon created an AI system that was supposed to find criminals by their face image, but when it was tested with a set of pictures of Congresspeople it turned out not only inaccurate but also quite racially biased.
* In 2015, in Japan the first innovative Henn-na Hotel opened its doors to guests. All its staff: the front-desk, cleaners, porters and in-room assistants were robots. But the bots started accumulating customer complaints much faster than expected: the bots frequently broke down, could not provide satisfactory answers to guest queries, and in-room assistants startled guests at night by interpreting snoring as a wake command.

**Implication of when AI fails**

Personal data must used in ways which individuals would reasonably expect. Organisations need to be transparent about how they process personal data in an AI system in order to comply with the principle of transparency. The key is that individuals understand how their personal data is used and have consented to this use.

However, for consent to be valid, individuals must also be able to withdraw consent as easily as they gave it. The organisation shall no longer process the personal data unless they demonstrate compelling legitimate grounds for the processing which override the interests rights and freedoms of the data subject.

**Conclusion**

Organisations must understand where AI sits in the value chain and create the right structures to ensure long-term governance by:

* Establishing internal governance, for example by an objective review panel, that is diverse and that has the knowledge to understand the possible consequences of AI infused systems.
* Ensuring the right technical guardrails, creating quality assurance and governance to create traceability and auditability for AI systems
* Investing in their own AI education and training so that all stakeholders are informed of AI capabilities as well as pitfalls.

**References**

Gillis, A (2021), *What are the principle of responsible AI?* [Online], [Accessed 20/08/2021]. Available at <https://searchenterpriseai.techtarget.com/definition/responsible-AI>.

Villa, F (2021), Make Responsible AI Part of Your Company’s DNA [Online], [Accessed 20/08/2021]. Available at < https://www.reworked.co/ information-management/make-responsible-ai-part-of-your-companys-dna/>.

O’Sullivan, C (2021), *What is Algorithm Fairness?* [Online], [Accessed 20/08/2021]. Available at < https://towardsdatascience.com/ what-is-algorithm-fairness-3182e161cf9f>.

O’Sullivan, C (2021), *What is Responsible AI?* [Online], [Accessed 20/08/2021]. Available at < https://towardsdatascience.com what-is-responsible-ai-548743369729>.

Application Security Series (2020), *Top 10 Failures of AI* [Online], [Accessed 21/08/2021]. Available at <https://www.immuniweb.com/blog/top-10-failures-of-ai.html>.

Patel, A (2019), *Malicious Use of AI* [Online], [Accessed 21/08/2021]. Available at <https://blog.f-secure.com/malicious-use-of-ai/>.

Rutkin, A (2017), *the tiny changes that can cause AI to fail* [Online], [Accessed 21/08/2021]. Available at <https://www.bbc.com/article/20170410-how-to-fool-artificial-intelligence>.

Rosenhine, K (2020), *3 ways organisations can use AI in a responsible way* [Online], [Accessed 20/08/2021]. Available at < https://cloudblogs.microsoft.com/ industry-blog/en-gb/cross-industry/2020/01/08/3-ways-organisations-can-use-ai-in-a-responsible-way/>.

Information Commissioner’s Office, *What do we need to do to ensure lawfulness, fairness, and transparence in AI systems?* [Online], [Accessed 21/08/2021]. Available at <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-do-we-need-to-do-to-ensure-lawfulness-fairness-and-transparency-in-ai-systems/#howdotheprinciples>.