**Artificial Intelligence**

Responsible AI is a governance framework that documents how a specific organisation is addressing the challenges around artificial intelligence from both an ethical and legal point of view. Resolving ambiguity for where responsibility lies if something goes wrong is an important driver for responsible AI initiatives.

Responsible AI should include the following qualities and principles:

* Explainable
* Monitorable
* Reproducible
* Secure
* Human-centered
* Unbiased
* Justifiable

Organisations should follow the Data Ethics Framework provided on the gov.uk website. The specific actions are as follows:

1. Define and understand public benefit and user need
2. Involve diverse expertise
3. Comply with the law
4. Review the quality and limitations of the data
5. Evaluate and consider wider policy implications

Within each point there are sub points to also consider. The website advises to score the project out of 5 for each point. If the score is 3 or below in any of the principles this could indicate the need for additional checks and potential changes to make to the project more ethical.

One instance where AI has failed to function in the desired way was for Amazon who wanted to automate their recruitment to speed up the candidate selection process for thousands of vacancies Amazon has:

“They literally wanted it to be an engine where I’m going to give you 100 resumes, it will spit out the top five, and we’ll hire those.” – amazon engineer

It turned out to be a PR fiasco, the system turned out to be misogynistic, giving preference to white males. Most probably, the training data, that was used to build the model, was unbalanced that resulted in candidate selection bias.

AI is being used maliciously in numerous ways one being AI-Supported password guessing. Cybercriminals are employing ML to improve algorithms for guessing users’ passwords. More traditional approaches, such as HashCat and John the Ripper already exist and compare different variations to the password hash in order to successfully identify the password that corresponds to the hash. With the use of neural networks and GANs, however, cybercriminals would be able to analyse vast password datasets and generate password variations that fit the statistical distribution. This will lead to more accurate and targeted password guessed and higher chances of profit.

For “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).” I have found article 22 in the GDPR Law however I cannot seem to find where it covers implications of when AI fails – not sure if I have found the wrong article?

Source:

* <https://www.immuniweb.com/blog/top-10-failures-of-ai.html>
* <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/exploiting-ai-how-cybercriminals-misuse-abuse-ai-and-ml>