# Find out what Responsible AI is?

The use of AI with big data poses considerable challenges to any business from both an ethical and legal perspective. For example, systems which make predictive decisions for customers using AI, may do so based on biased data, leading to customers being unfairly discriminated against. Alternatively, AI systems with big data may find and use correlations between data to make decisions which aren’t ethical and violate, in the UK discriminatory laws on grounds of gender, race, age, religion etc.

It’s therefore essential for organisations to ensure they use data in an accountable way and evidence ethical use compliant with data protection laws and mitigating against any litigation. For some, the term Artificial Intelligence can provoke thoughts of progress and productivity. For others, the outlook is less positive. Many concerns such as. To make things worse, many of these issues are unique to AI. This means existing guidelines and laws are not suitable to address them. This is where Responsible AI comes in. It aims to address these issues and create accountability for AI systems.

These concerns have brought about the call for organisations using big data and AI to implement suitable governance frameworks ensuring that they implement transparent, fair, secure, and inclusive solutions, and mitigate against explicit and implicit bias which might cause unfair outcomes against specific demographics or individuals, for example when used in employment, health care, financial or criminal justice systems. The form of governance is called ‘Responsible AI’, bringing a new focus for the already large ethical and legal teams of an abundance of corporates with big tech players taking the lead.

Over the past 4 years, Microsoft have been operationalizing responsible AI athrough a central effort led by the Aether Committee, the Office of Responsible AI (ORA) and Responsible AI Strategy in Engineering (RAISE), and developing One Engineering System (1ES)—a set of tools and systems built on Azure ML that will help customers adopt responsible AI practices and internal engineering groups implement Microsoft companywide rules for responsible AI.

Two years ago Google introduced its [**AI Principles**](https://www.blog.google/technology/ai/ai-principles/), which guide the ethical development and use of AI in our research and products. The AI principles articulate [our Responsible AI goals](https://ai.google/responsibilities/responsible-ai-practices/) around privacy, accountability, security, fairness and interpretability that whilst ‘The development of AI is creating new opportunities to improve the lives of people around the world, from business to healthcare to education. It is also raising new questions about the best way to build fairness, interpretability, privacy, and security into these systems fairness, interpretability, privacy, and security into these systems’. They even provide a responsible AI toolkit for use with Tensor Flow!

More widely, both governments and NGO’s are endevouring to bring about responsible AI through both regulatory and advisory approaches with, the world economic forum are striving to bring a global consensus on this, as discussed recently on the agenda at Davos.

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

* In 2018, Reuters reported that Amazon had been working on an AI recruiting system designed to streamline the recruitment process by reading resumes and selecting the best-qualified candidate. Unfortunately, the AI seemed to have a serious problem with women, and it emerged that the algorithm had been programmed to replicate existing hiring practices, meaning it also replicated their biases. The AI picked up on uses of “women’s” such as “women’s chess club captain” and marked the resumes down on the scoring system. Reuters learned that “In effect, Amazon’s system taught itself that male candidates were preferable.” Rather than helping to iron out the biases present in the recruitment process, the algorithm simply automated them. Amazon confirmed that they had scrapped the system, which was developed by a team at their Edinburgh office in 2014. None of the engineers who developed the algorithm wanted to be identified as having worked on it.
* In 2019, Facebook was found to be in contravention of the US constitution, by allowing its advertisers to deliberately target adverts according to gender, race and religion, all of which are protected classes under the country’s legal system. Job adverts for roles in nursing or secretarial work were suggested primarily to women, whereas job ads for janitors and taxi drivers had been shown to a higher number of men, in particular men from minority backgrounds. The algorithm learned that ads for real estate were likely to attain better engagement stats when shown to white people, resulting in them no longer being shown to other minority groups.
* Early in 2020, an Australian screenwriter announced the launch of Giggle, a social networking app that aimed to allow girls to chat in small groups, or “Giggles,” and relied on gender verification AI software to ensure that only girls were able to join. The platform was less than well received, particularly on Twitter where people drew comparisons between the software and the eugenicist practice of phrenology. The AI also automatically excluded many trans girls, meaning that if they wanted to use the app they would have to contact the makers directly to have their gender verified, which in itself entails ethical conundrums and raised questions about how sensitive the developers were to the real life application of their software. As of April 2020 the app still hasn’t been officially launched.
* Cambridge Analytica had a facebook app that harvested data of users and their friends called ‘ this is your digital life’ and sold it to clients without the users permission for political purposes. Cambridge Analytica used the data to provide analytical assistance to the 2016 presidential campaigns of [Ted Cruz](https://en.wikipedia.org/wiki/Ted_Cruz_2016_presidential_campaign) and [Donald Trump](https://en.wikipedia.org/wiki/Donald_Trump_2016_presidential_campaign).

# •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).

Organisations which breach GDPR are subject to fines by the ICO. In the case of the Cambridge Analytica scandal facebook were fined just £500k.

GDPR insists organisations opt users out of data collection by default, and actively obtain an opt in from users along with a privacy statement, in order to use their personal data.

An individuals rights around data must be supported, even when their data is used in AI, whether in the training data, used to make a prediction during deployment, and the result of the prediction itself or that might be contained in the model itself. i.e

* The [right to be informed](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-be-informed/);
* the [right of access](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/);
* the [right to erasure](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/);
* the [right to rectification](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/); and
* the [right to data portability](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/).

This means their data can be pulled from the model at any time if they so request.

Also, fully automated decision making based on data is not advised, and prohibited under certain circumstances ‘Regardless of their relative merits, automated decisions are treated differently to human decisions in data protection law. Specifically, Article 22 of the UK GDPR restricts fully automated decisions which have legal or similarly significant effects on individuals to a more limited set of lawful bases and requires certain safeguards to be in place’.

Generally it is recommended that the organisation should:

* consider the system requirements necessary to support a meaningful human review from the design phase. Particularly, the interpretability requirements and effective user-interface design to support human reviews and interventions;
* design and deliver appropriate training and support for human reviewers; and
* give staff the appropriate authority, incentives and support to address or escalate individuals’ concerns and, if necessary, override the AI system’s decision

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

Comply with the relevant data protection law, for example GDPR: <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/#howshouldweaddress>

They must be able to prove

* the data is being used fairly, lawfully and transparently.
* Users understand how their data is being used
* Their data is being held securely

Aside from looking to the data protection laws in any country, It’s key to set out principles, and a strategy and guidelines for each areas AI, ad ensure guidelines are followed in all areas of AI implementation to mitigate against litigation and any reputational damage where the company is found to have used AI unethically, even inadvertently.