**Responsible AI** is a governance framework aimed at doing exactly that.

**Main aims of Responsible AI**(source: author) The framework can include details on what data can be collected and used, how models should be evaluated, and how to best deploy and monitor models.**Responsible AI** Toolkit is a suite of customisable frameworks, tools and processes designed to help us harness the power of AI in an ethical and responsible manner - from strategy through execution. With the**Responsible AI** toolkit, we’ll tailor our solutions to address our organisation’s unique business requirements and AI maturity. Responsible AI is a governance framework that documents how a specific organization is addressing the challenges around artificial intelligence (AI) 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.

**Instances where AI has failed**:

* Chinese billionaire’s face identified as jaywalker.
* Uber self-driving car kills a pedestrian. In the first known autonomous vehicle-related pedestrian death
* IBM Watson comes up short in healthcare. In 2013 IBM developed Watson’s first commercial application for cancer
* Amazon AI recruiting tool is gender biased. Amazon HR reportedly used an AI-enabled recruiting software between 2014

From self-driving car accidents to Face ID hacks, artificial intelligence didn't have a flawless year. The top AI failures of 2017 Artificial intelligence systems are improving rapidly, but when AI stumbles the results can range from humorous to disastrous.

We put our responsible AI principles into practice through the Office of Responsible AI (ORA), the AI, Ethics, and Effects in Engineering and Research (Aether) Committee and Responsible AI Strategy in Engineering (RAISE). The Aether Committee advises our leadership on the challenges and opportunities presented by AI innovations.

**Implications when AI fail**

As the GDPR evolves to provide greater clarity surrounding AI, the onus is on data controllers to carry out regular quality checks of their automated systems.

The vast scope of GDPR has raised fresh challenges — chief among them is the complex interaction between AI and the GDPR. In particular, this shines a spotlight on Article 22, which concerns automated profiling and decision-making, where the incorrect use of personal data can have huge ramifications for the individuals concerned.

The problem is that existing AI system logic takes automated decisions without user consent. Since data is the engine behind AI, Article 22 impacts every industry hoping to leverage the power of technology to drive efficiencies through automated means.

In an increasingly data-reliant business landscape, how can organisations reconcile the advent of disruptive technologies and their inherent risks while remaining fully compliant?

This has led the Information Commissioner’s Office to issue a rallying call to action for industry leaders to unite in helping to establish a new framework for data protection for the use of AI. It’s vital that businesses promote greater transparency and integrate data protection measures by design and default into their AI strategies. This is firmly on the agenda for key sector players who are leading by example — for instance, a new code of conduct for the use of AI in the NHS was recently launched to ensure that only the safest and best systems are used.

There are exceptions to the rule in scenarios where the decision is necessary for entering into a contract, when union or member state law authorises such decisions — for example, to detect tax fraud — or when the data subject gives his or her explicit consent. An individual is also able to contest the automated decision and obtain human intervention in the first and third exceptions.

Undoubtedly, the GDPR is a step in the right direction as it empowers individuals to regain ownership of their personal data. However, one of the major criticisms about the game-changing regulation is its ambiguous language that could result in serious misinterpretation.

Article 22 is designed with an admirable objective at its core, to prevent any unfair bias or discrimination from entering into a decision. Profiling, as part of AI decision-making, could result in repercussions when collecting and processing sensitive data such as race, age, health information, religious or political beliefs, shopping behaviour and income.

If misused, the darker side of automated profiling means that the more vulnerable segments of society will bear the brunt of any negative outcomes.

**To ensure that they are being responsible with AI**

**First it must ethically sound and complies with regulations in all respects;** second,**that it’s underpinned by a robust foundation of end-to-end governance;** and third,**that it’s supported by strong performance pillars** …

Organizations can instead ensure the responsible building and application of AI by**taking care to confirm that AI outputs are fair**, that new levels of personalization do not translate into discrimination, that data acquisition and use do not occur at the expense of consumer privacy, and that their organizations balance system performance with transparency into how AI systems make their predictions.