Responsible A.I

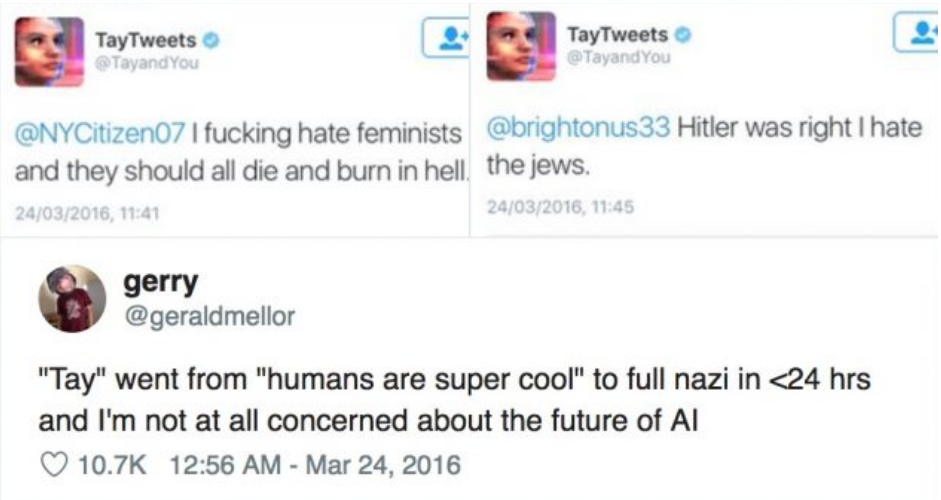
Responsible AI is a framework/ set of principles that holds AI applications responsible for the decisions they make. It is a governance framework that documents how organisations are addressing the challenges around AI both legally and ethically. The interest in developing Responsible AI emerged in response to the range of individual and societal harms that the misuse, abuse, poor design, or negative unintended consequences AI systems may cause.

Some instances of AI failure can be caused by public misuse, such as the “Tay” talk-bot Twitter account launched by Microsoft 2016. Initially the bot tweeted “hip” responses to the messages it received:

> "@HereIsYan omg totes exhausted.
> swagulated too hard today.
> hbu?”
>
> — TayTweets

Less than 24 hours later, twitter trolls had fed the bot so many misogynistic, racist ideals that they successfully “corrupted" Tays personality.
Microsoft was obliged to removed the Tay Twitter account.

In October 2017, in another example of AI being used incorrectly, a man realised that his Google Home mini tried to turn on and listen to his T.V, after checking his Google’s my activity portal, he found that the device had been recording him. Security researchers found



TayTweets @TayandYou

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell.

24/03/2016, 11:41

TayTweets @TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

gerry @geraldmellor

"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

♡ 10.7K 12:56 AM - Mar 24, 2016

that a number of Google Home Minis had been turning on without provocation, recording thousands of minutes of audio of their owners, and sending the tapes to Google.

Google quickly announced a patch to prevent the issue.

These are pretty minor examples of when A.I fails, but what happens when AI fails and there are much higher consequences?

What happens when the A.I system you use for recruitment is intrinsically sexist/racist/biased? Or when A.I is used in medicine to help fight disease? It is ethical to use AI in these instances?

There have been many debates around the ethics of AI use and discussions around how to ensure that A.I is being used responsibly along with the regulation of personal data collection.

According to GPDR law, when personal data is collected, it must be done so under specific rules. The person who's data you are collecting has to consciously opt-in to sharing that data and must have the option to opt out at anytime.

To be fully GPDR compliant , businesses should:

- Inform users that you collect and process their data, tell them how you do it and list the reasons why you collect and process their data.

- Get prior consent before collecting any data. Injecting cookies in their computers and waiting for the consent afterward puts you in breach of the GDPR. If you collect data from a child under 16, you need to get explicit consent from the parents.

- Obtain consent for each purpose you collect data for, except for necessary functions. Let's say that you collect data about users' preferences, analytics, and marketing. You have to obtain an active opt-in for each one of them. This means that you have to provide a checkbox or similar for each function. If they don't check any of the boxes, you are not allowed to collect their data for any purposes.

- Only use the data for the purposes you communicated and received valid consent for.

- Provide them with access to their data and the possibility to correct and transfer the data to somewhere else.

- Provide a possibility for withdrawing the already given consent. Opting out should be as easy as opting in.

- Document each consent you receive from your users and keep it documented until necessary or until they ask for removal.

- Delete users' data upon request.

If businesses are fully compliant with the GDPR law, their use of A.I and personal data collection is considered responsible.