

## Compulsory HLT

### Practical : Artificial Intelligence (AI)

In Data Science we process a lot data through AI. With the GDPR, it is becoming increasingly important to understand the ethics behind the data that is collected, stored, processed and evaluated.

#### **Your task is to:**

- Find out what Responsible AI is?
- Find instances where AI has failed? Or been used maliciously or incorrectly.
- 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).
- What should organisations do to ensure that they are being responsible with AI and the wider use of data in general?
- Maximum 500 words.

#### **ANSWER BELOW**

##### Responsible AI is:

- “Governance framework that documents how the organisation is addressing the challenges around AI from both an ethical and legal point of view. Resolve ambiguity of WHERE responsibility lies if something goes wrong..”

[www.techtarget.com](http://www.techtarget.com)

- Weighing the challenges and opportunities presented by AI innovations, setting out rules and governance processes across the whole organisations to ensure advancement of AI is driven by ethical principles that put people first, such as Fairness, Reliability & Safety, Privacy & Security, Inclusiveness, Transparency, Accountability.

##### Instances AI has failed in

The concerns with AI are safety, human replacement, AI's lack of awareness, privacy concerns, bias, and AI's lack of human dignity.

The 3 instances AI has failed are -

1. Microsoft's most advanced chatbot, being not immune to profane and hate languages that 24 hours chatting with humans on Twitter had ruined it's innocent worldview and started spewing hateful speech.
2. AI developed to aid battle against cancer had an utter failure as it gave multiple dangerous and erroneous therapy suggestions.
3. FaceID on iPhones could be bypassed with a 3D mask of the owner's face with 2D eyes

##### Implications when AI fails

1. AI's learning system is through the training data its creators have provided or it has data mined itself. GIGO (Garbage in garbage out) concept of when the input data is not good enough or not statistically proportionate representation of the real world data, the output data will also be not good enough. Of tons of different types of data, how can we know the data that is provided as training data are lacking, statistically representative or biased?
2. AI's models decay overtime as well, as patterns in the new data drift away from patterns learned in training data, the accuracy of its output will start falling, resulting in the fall in the quality of the AI system results over time.
3. The complexity of the AI systems makes it extremely difficult to carry out audits, debugging, and understanding what went wrong almost impossible.

#### What should organisations do to ensure they are being responsible with AI

1. Documentation of the intended uses of the AI and where it is not intended to be used. The risks identified and mitigations in place for specific uses. As there is potential for AI to be repurposed for other uses, the original developers need to capture assumptions and tradeoff decisions.
2. Accountability to be tied to AI's impact. With more potential to influence people's behaviour and livelihood, the more careful our use of AI should be and our governance need be.
3. As our assumptions shape AI, there is no such thing as a neutral, impartial or unbiased AI. Hence, organisations should strive for diversity in teammates' backgrounds and experiences, multidisciplinary perspectives, and be open and responsive with any raised issues and provide documentations for all the assumptions that went into the AI system.

Total words: 483 words