CSCM21: Designing in Trust, Understanding, and Negotiation - Coursework: BeatLonliNess plc

Andy Gray
445348

28/04/2021

1 Based on the scenario summarised above and the aspects of responsible design learnt in the lectures, discuss issues (legal, ethical, and technological) with the business proposal of BeatLonliNess plc. (13 Marks)

In legal aspects, the BeatLonliNess (BLN) plc platform must abide by the data protection and GDPR rules that the British government and the EU have set out. These are rules to ensure that companies keep their users' data safe and secure while also holding the user's information relevant to the organisation. BLN could achieve this from the word go as they start to expand by using privacy by design method. Privacy by design ensures that privacy is a requirement in the design and development process that includes encryption, federated learning, differential privacy, access control, transparency, and finally, consent [?]. The Information Commissioner's Office (ICO) states that policies and procedures are needed to get implemented to ensure data protection issues get considered when systems, services, products and business practices involving personal data are designed and implemented. Therefore, as a result, personal data gets protected by default, ensuring that safeguarding individuals' rights. These rights include data minimisation, pseudonymisation and purpose limitation [?].

With BLN using AI-supported algorithms, they must get designed to be reliable. For the algorithms to be reliable, they would also need to carry out their tasks with high accuracy, ensuring that the generated results are what the designers expect to generate, allowing the system's logic and architecture to facilitate transparency and explainability [?]. Therefore it would be a good idea for BLN to build their algorithms with explainability within them. Making the algorithm explainable would allow the users to trust the algorithms more and see what factors impact their matchings. However, the data must not create any potential bias within the models to allow the matches to happen effectively, but removing bias is challenging to spot and remove. BLN must remove any potential bias from their datasets to ensure that no member of the platform gets discriminated against, whether it be because of their gender, race, religion, ethnicity or skin colour.

These issues also lead to BLN making sure that the algorithms and models they use are also responsible AI. For the AI to be responsible, the models will need to be transparent, trustworthy, ethical, and respecting users' privacy [?] or at the very least carry out most of them. For example, the algorithm uses common interests and individual personal characteristics like eye colour, hair colour and weight. So the algorithm needs to make sure it does not discriminate against the user for being overweight, for example, as this is one of the metrics used to calculate a potential match. However, BLN also needs to be seen as trustworthy as they will have all the user's personal information, images and videos. Therefore for people to be willing to provide this information, they need to be perceived as trustworthy. Ensuring that they are trustworthy is essential as they keep this information safe and ensure that people who should not have access to the content should not be and doing everything to prevent any data leaks. Therefore the technology must stay neutral. The AI system should not reduce the procedural and substantive requirements that are usually attached to a decision when the decision-making process gets entirely controlled by a human [?].

Fundamentally, by using a responsible AI design, the AI system should not exempt or attenuate the need for fairness. The AI system users and anyone subject to the decisions getting made by the system must have an excellent way to be able to correct and discriminatory or unfair situations that the AI has generated, whether that be through a biased or inaccurate system [?]. Therefore, BLN must carry this out with compatibility with the human agency and uphold the human rights fundamentals. Therefore they must monitor their AI systems to ensure that they attempt to mitigate potential consequences that the AI system might generate. Ensuring that they are consistent with the moral purpose of beneficence and non-maleficence [?]. So BLN must assess the social, political and environmental impacts that the system might have, especially when BLN expands to the EU. Developing and deploying a system that has taken a firm stance on a responsible AI design will reduce the risk of harm and, for any potentially unforeseen circumstances, provide strategies for any mitigating strategies to any potential risk [?].

Therefore, BLN must make awareness and educate their users on the AI system's limits. By doing this, they are ensuring that they are transparent and fair to their users. BLN should also make sure that they are making users aware that the AI system is designed to achieve specific goals set, knowledge, and experience. Additionally, that limitation will still be present, especially within the datasets used to train them. By BLN having a comprehensive approach to fairness should aim to address fairness in the AI. BLN should aim to do this by using technical experts' close engagement, including AI and social sciences. Additionally, due to the desire to expand to the EU, BLN should aim to work closely with governments and other organisations to develop their AI system and deploy it to the public within the legislation surrounding it. Ultimately, create a fair and non-discriminative system that is open and transparent with appropriate accountability principles [?].

2 Summarise relevant examples of related media coverage in the last 5 years. (7 Marks)

Microsoft in 2016 released an AI chatbot called Tay.ai. The chatbot was released onto Twitter and described by Microsoft as a "conversational understanding experiment" [?]. While Microsoft also stated that "the more you chat with Tay, the smarter it gets, learning to engage people through casual and playful conversation" [?]. However, the chatbot did not stay spirited for long. As soon as Tay launched, Twitter users' starting tweeting the bot with all sorts of misogynistic, racist, and unpleasant remarks. Therefore, this caused Tay to repeat these thoughts back to users [?].

In 2015, a software engineer Jacky Alciné discovered that Google's image recognition algorithms in Google Photos labelled his black friends as gorillas. Google said it was appalled at the mistake and promised to fix the problem. However, Google has not fixed it. They have just blocked its algorithms from identifying gorillas altogether [?].

A long-awaited report from top Democratic congressional lawmakers about the dominance of the four biggest tech giants had a clear message on Tuesday: Amazon, Apple, Facebook, and Google engage in a range of anti-competitive behaviour, and US antitrust laws need an overhaul to allow for more competition in the US internet economy [?]. — conclusion that the big four tech firms have amassed too much power.

[1] [2] [3] [4] [5]

3 Use the ETHICS GUIDELINES FOR TRUSTWORTHY AI published by AI high-level expert group of the European Commission in April 2019, in particular the TRUSTWORTHY AI ASSESSMENT LIST (p.24 of the report and standalone document), to discuss requirements for the system proposed by BeatLonliNess plc. (Links to these documents are posted with the assessment brief.) (10 Marks)

References

[1] Bruce, P., Bruce, A., and Gedeck, P. *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python.* O'Reilly Media, 2020.

[2] Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, 2019.

[3] Grus, J. *Data science from scratch: first principles with python.* O'Reilly Media, 2019.

[4] Koehrsen, W. An implementation and explanation of the random forest in python, 2018. Towards Data Science Medium, Online: https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76.

[5] Lander, J. P., and Beigelmacher, M. The essential guide to quality training data for machine learning, 2018. Cloud Factory, Online: https://www.cloudfactory.com/training-data-guide.