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 [4]. 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 [6]. 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 [5].

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 [9]. 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. BLN will also need to allow users who are subject to the decisions made by the algorithm must have an avenue or medium to correct any potential issues the algorithm is creating. This avenue to report the decisions made by the AI system is especially important if BLN decides to create and deploy their AI system without transparency about how it works.

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 [6] 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 [9].

Fundamentally, by using a responsible AI design, the AI system should not exempt or devalue 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 [9]. 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 [9]. 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 [9].

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 ensure that they make their users aware that the AI system is getting 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 [9].

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" [3]. While Microsoft also stated that "the more you chat with Tay, the smarter it gets, learning to engage people through casual and playful

conversation" [3]. 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 [3]. What happened to Microsft's Tay is something BLN need to be cautious about when implementing their chatbot for their 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 [8]. Therefore, BLN needs to be mindful that their AI system does not attribute ethnic and minority background users in a harmful way when deciding on desirable attributes on their users to match them up.

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 [1]. Therefore we can conclude that the big four tech firms have amassed too much power, and a lack of trust is getting created around them with what they are doing. So BLN needs to be mindful that they are not getting perceived as becoming untrustworthy and making sure they always stay transparent as they get bigger.

Facebook, the parent company to Whatsapp, decided to change the terms and conditions to their privacy policy to the app Whatsapp. Users perceived these changes with the terms and conditions because Whatsapp was perceived to be sharing more data with Facebook than before. The change resulted in users leaving Whatsapp and going to rival applications like Signal and Telegram due to the lack of trust in the company and the changes. However, Facebook responded that they are not doing anything different than they were before and, it was just a change in wording to be transparent about what they were doing [7]. Therefore, BLN needs to make sure, as they are holding loads of personal data about their users, that they are perceived to be looking after it and treating it correctly and not using it for their own needs that do not benefit their users.

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)

With BLN aiming to match up people based on their traits and characteristics, BLN must follow the diversity, non-discrimination and fairness requirement of the Assessment List for Transworthy Artificial Intelligence (ALTAI). ALTAI state that unfair bias must get avoided, and a fostering of diversity should be in place. Ensure that the AI systems are accessible to all, no matter their background or disability. Making sure that no one received discrimination or prejudice while using the AI system [2]. It is also vital that BLN consider ALTAI requirements for privacy and data governance, including making sure BLN follows GDPR rules [4]. ALTAI recommend that to be entirely respectful of the privacy and guidance requirement, to comply, BLN would need to have procedures in place to ensure legitimised access to data while also taking into account BLN's need for quality and the integrity of the data [2]. BLN also needs to ensure that their systems have technical robustness and safety, a suggested aspect of a trustworthy AI design, ensuring unintended harm gets prevented or minimised. Therefore, BLN needs to make sure they are building a resilient and secure system. Ensuring that all the users' details and content are kept safe and that the ai system is "safe, ensuring a fall back plan in case something goes wrong" and "being accurate, reliable and reproducible" [2].

For BLN to seem trustworthy to its users, ALTAI recommends that the AI system have transparency and accountability. To meet ALTAI's requirements, BLN will need to ensure they have systems in place that show traceability mechanisms. ALTAI recommend that the AI system getting used will need to be explained in a format that the average user will understand. BLN needs to ensure that they let their users know that they are fundamentally interacting with an AI system while informing its limitations and capabilities [2]. ALTAI framework suggests that in order for BLN to get perceived as trustworthy, BLN will need to make sure they have tools in place to make sure they have responsibility and accountability for their AI systems outputs [2]. BLN will need to ensure that they regularly audit their AI system and address any issues that arise. BLN could implement an AI Ethics Review Board, which will review the AI system and then hold the AI developer to account for the AI system's performance and accuracy. Ultimately, looking at areas that might have gone wrong and improving them regarding trustworthy design [2].

Ultimately BLN needs to make sure they empower the user when they provide suggestions on how to improve their chances of finding a partner. ALTAI

states that allowing the user "to make informed decisions and fostering their fundamental rights" [2] is essential. ALTAI refers to this trait as "human agency and oversight", a method of steps that ensures oversight on the AI system by having human-in-the-loop, human-on-the-loop, and human-in-command approaches to their systems [2]. BLN also needs to remember that they serve their users and do what is required and needed by them. Therefore, when designing their system, they should make sure they think of environmental and societal well-being. The ALTAI state that, on this matter, "AI systems should benefit all human beings, including future generations," and that is a crucial point to remember. Putting procedures in place now that will not only benefit the users now but will also benefit the potential users' many generations in the future [2].

References

[1] Ghaffary, S., and Rey, J. D. The big tech antitrust report has one big conclusion: Amazon, apple, facebook, and google are anti-competitive, 2021. Recode, Online: https://www.vox.com/recode/2020/10/6/21505027/congress-big-tech-antitrust-report-facebook-google-amazon-apple-mark-zuckerberg-jeff-bezos-tim-cook.

[2] High-Level Expert Group on Artificial Intelligence (AI HLEG). Assessment list for trustworthy artificial intelligence (altai) for self-assessment, 2020. European Commission, Online: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

[3] Hunt, E. Twitter taught microsoft's ai chatbot to be a racist asshole in less than a day, 2016. The Guardian, Online: https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.

[4] Information Commission's Office. Guide to the general data protection regulation (gdpr), 2018. UK.Gov, Online: https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation.

[5] Information Commission's Office. Data protection by design and by default, 2021. Online: https://ico.org.uk/for-organisations/accountability-framework/policies-and-procedures/data-protection-by-design-and-by-default/.

[6] Muller, B. Designing-in trust, understanding, and negotiation law-xai relection, 2021. Swansea University.

[7] Statt, N. Whatsapp clarifies it's not giving all your data to facebook after surge in signal and telegram users, 2021. The Verge, Online: https://www.theverge.com/2021/1/12/22226792/whatsapp-privacy-policy-response-signal-telegram-controversy-clarification.

[8] Vincent, J. Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech, 2018. The Verge, Online: https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai.

[9] Wyner, A. Artificial intelligence and the law - regulating ai, 2021. Swansea University.