CM50266 – APPLIED DATA SCIENCE
CASE STUDY – DATA PROTECTION

1. Excluding accountability, the UK General Data Protection Regulation (GDPR) defines six other key principles as follows (Data protection principles under the UK GDPR, 2018):
    - The lawfulness, fairness and transparency principle states that personal data must be processed lawfully, fairly and in a transparent manner in relation to the data subject. In other words, one must establish a lawful basis in terms of personal data collection and usage, while also ensuring that this data usage avoids violating any other laws. Additionally, data usage must not be damaging, unanticipated or deceptive with respect to those concerned (i.e. data usage must be fair). Lastly, the word transparency implies that one must be clear and truthful with those related to any personal data used, with regards to exactly how their data is being used.
    - The purpose limitation principle states that any personal data must be collected for a specifically, explicitly and legitimately defined purpose or objective. Therefore, one must abide by documentation and transparency obligations, ensuring to record these purposes as well as notify persons about these purposes from the initial point of data collection. Outside the parameters of initially defined purposes, data must not be processed any further. However, there is one exception where further processing for archival purposes in the public interest, scientific or historical research purposes or statistical purposes, is permitted.
    - The data minimisation principle states that any personal data processed must be adequate so as to sufficiently achieve the initially stated purpose, relevant so as

to be rationally linked to the purpose and limited to what is necessary for successful processing .

- The accuracy principle states that personal data collected must be accurate and up-to-date, with a clear source and status being available at all times. One must always contemplate about any challenges regarding the accuracy of the data as well as determine whether periodical updates of the data must be carried out. Importantly, any inconsistencies in personal data which has been processed, must be quickly erased or rectified without any delays.
- The storage limitation principle states that stored personal data must not be kept for any period of time exceeding the time necessary for personal data processing within the defined purpose(s). It is therefore essential to undergo periodical reviews of the stored data, in order to delete or anonymise it, once it becomes unnecessary in terms of one's purpose. The aforementioned exception seen in the purpose limitation principle also applies here, where stored personal data may be kept for longer periods, so as to serve the same purposes after processing.
- The integrity and confidentiality principle states that any stored data must be protected by established security measures as appropriate. This protection must therefore prevent against unsanctioned or illicit data processing as well as fortuitous loss, destruction or damage of said data.

2. Firstly, it should be noted that the US does not have a central federal level data protection law, such as the GDPR, and thus, there are some shortcomings in terms of the US regulations (Green, 2021). Such is observed in the provided situation, as users are indeed given the option of opting out of personal data processing for marketing purposes, however, there is no mention of being allowed to rectify any inconsistencies in processed data. This is legal in the US as there is no explicit law targeted to such (Jehl and Friel, 2018), however, this is a clear violation of the accuracy principle and thus, must be implemented in the UK version of the website, in order to comply with the GDPR.

3. The accountability principle states that one must accept responsibility for processing any personal data collected whilst demonstrating compliance with the aforementioned GDPR principles. When considering the various elements which attribute to the accountability principle, the company will be required to consider several actions in order to implement the new features. Firstly, the company will need to communicate to individuals/users the necessary information regarding its data privacy program, data processing procedures and personal rights via applicable policies. In other words, there should be transparency with regards to how an individual's personal data will be used when the new features are implemented. This should not only be mentioned when a user first signs up but should also be easily accessible via dashboards or portals on the website itself (Centre for Information Policy Leadership, 2018). Secondly, the company should perform data protection impact assessments as it regards to any risks the new features may create. Through this risk assessment, the company must be able to successfully carry out risk mitigation as it pertains to individuals' data and their rights (Centre for Information Policy Leadership, 2018).

4. As stated by the Article 29 Working Party (advisory body on data protection and privacy), “profiling often is used to make predictions about individuals based on inferences drawn from the individual or a group of statistically similar individuals.” (Association of National Advertisers, 2018.). With this in mind, the individualised recommendation system would therefore fall under the characteristic of profiling, which is an GDPR related issue if not handled appropriately. This profiling is permitted under the GDPR, once a lawful basis for it, such as individual consent, has been established (GDPRhub, 2021.). Hence, if the company desires to implement the new recommendation system, not only will they need to provide sufficiently transparent and comprehensive information regarding the profiling and the resulting decision (the new recommendation system) to potential data subjects, but they will also need to gain explicit consent to do so from same.

5. Under the GDPR, each and every data subject reserves the right for their personal data to be erased upon request. In this case, this is being followed by the company whenever a user decides to close their account, by attempting to anonymising the data. However, to fully complete the process of data anonymisation under the GDPR, it must be impossible to determine the identity of the user via reconstructive methods. This is rather difficult when considering the exposed data or what remains after anonymisation, as it can be paired with other available information, therefore making it easy for the identity of data subjects to be deduced (Blair, Lewis, Campbell and Catanzaro, 2019). This is true with the provided case as ratings and reviews will still be available, making it possible for re-identification to take place. Hence, deleting personal data alone is insufficient here and does not achieve anonymisation in line with the GDPR, but rather, is a case of pseudonymisation.

6. This process violates the GDPR as the avatars are being generated by the system based on additional information previously provided by users. By using users' information, there a direct link is created between the avatar and the respective individual, therefore allowing for an individual's identity to be potentially exposed, which is a clear breach of the integrity and confidentiality principle (GDPR Principle 6: Integrity And Confidentiality, 2020).

7. Instead of using user information to directly generate the avatars, the website can implement a pre-defined collection of avatars from which users can choose from when signing up (Judin, 2018). Instead of using a cartoon face, avatars should be more arbitrary in nature whilst still being relatable to users. For example, a user may decide to choose their favourite animal, food or hobby as their avatar. In the case where any personal data is hashed, it is also possible to implement a unique avatar generator system, based on this hashed data. For example, GitHub's interface implements "simple 5×5 'pixel' sprite" avatars or 'identicons' generated using a hash of users' IDs (Long, 2013), which is unique to each user yet does not leak any user details. Thus, with this generalisation in mind, in both cases personal user identification is avoided and the GDPR principles are upheld.

Reference List

Association of National Advertisers. 2018. *Advertising and the GDPR's Requirements on Automated Decision-Making and Profiling*. [ebook]. Available at: <https://www.ana.net/miccontent/show/id/ii-reed-smith-gdpr-automated-decision-making> [Accessed 10 January 2022].

Blair, T., Lewis, M., Campbell, P. and Catanzaro, V., 2019. *The eData Guide to GDPR: Anonymization and Pseudonymization Under the GDPR*. [online] JD Supra. Available at: <https://www.jdsupra.com/legalnews/the-edata-guide-to-gdpr-anonymization-95239/> [Accessed 10 January 2022].

Centre for Information Policy Leadership. 2018. *The Case for Accountability: How it Enables Effective Data Protection and Trust in the Digital Society*. [ebook]. Available at: <https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf> [Accessed 10 January 2022].

GDPRhub. 2021. *Article 22 GDPR*. [online] Available at: <https://gdprhub.eu/Article_22_GDPR> [Accessed 10 January 2022].

Green, A., 2021. *Complete Guide to Privacy Laws in the US | Varonis.* [online] Varonis.com. Available at: <https://www.varonis.com/blog/us-privacy-laws> [Accessed 10 January 2022].

Incorporated.Zone. 2020. *GDPR Principle 6: Integrity And Confidentiality*. [online] Available at: <https://incorporated.zone/gdpr-principle-6-integrity-and-confidentiality/> [Accessed 10 January 2022].

Jehl, L. and Friel, A., 2018. *CCPA and GDPR Comparison chart.* [ebook] Available at: <https://www.bakerlaw.com/webfiles/Privacy/2018/Articles/CCPA-GDPR-Chart.pdf> [Accessed 10 January 2022].

Judin, J., 2018. Avatars, identicons, and hash visualization. [Blog] Jussi Judin's weblog, Available at: <https://barro.github.io/2018/02/avatars-identicons-and-hash-visualization/> [Accessed 10 January 2022].

Long, J., 2013. Identicons!. [Blog] *The GitHub Blog*, Available at: <https://github.blog/2013-08-14-identicons/> [Accessed 10 January 2022].

Nibusinessinfo.co.uk. 2018. *Data protection principles under the UK GDPR.* [online] Available at: <https://www.nibusinessinfo.co.uk/content/data-protection-principles-under-uk-gdpr> [Accessed 9 January 2022].