**Case Study 1 – Data Protection**

CM50266 Applied Data Science

Ans 1. The seven data privacy principles are:

I. Any processing of personal data should be fair and warranted by a reason which falls under the six lawful bases that businesses have to process EU citizens data. The data practices should be shared with data owners in clear, concise and sensible language.
II. The purpose for collecting personal data should be clearly stated, and the collection and processing should be limited until the specific purpose is completed.
III. The collection of personal data should only be limited to the minimum requirements of the organization to achieve its processing purposes.
IV. There should be practices followed to ensure the collection of personal data is accurate and updated if required. Additionally, there should be reasonable steps taken to ensure personal data which is inaccurate should be erased or rectified.
V. Collected personal data should be stored only until it is no longer a requirement for the stated processing purposes. After this period the data should either be deleted or anonymized to prevent identification of data subjects.
VI. There should be practices followed to ensure the personal data is protected from unauthorized or unlawful processing and accidental loss, destruction, or damage. (GDPR, 2018)

Ans 2. In the current structure of ACME Review Inc, there is no mention of explicit consent taken by the company to generate targeted advertisements. While there is a mention of users have the ability to opt-out of the use of the data to select the adverts shown, this means that targeted advertising is enabled without the user's permission. Hence the company will have to explicitly ask for consent to share data with target advertisements once a user arrives on the website page, additionally the company will have to share the list of advertisement companies with whom they share data. This is required to comply with GDPR principles of transparency. (Andreou et al., 2018)

Ans 3. For implementing the ability to generate car recommendations, ACME Review Inc will have to explicitly ask for permission to use the personal data and car ratings to generate a personalized profile. As car recommendations are not covered under their existing purpose,

hence they will have to ask for consent. (Panteli, 2019). Additionally, as the avatars are generated from personal data and there is no consent asked for it, hence they should auto-generate a blank face cartoon and let the users modify it in any manner possible. Though as the avatar modifications can serve as personal data, there should be explicit permissions to use that data from the users. The users should also have the right to erase any uploaded image/avatar as it qualifies under personal data. (Trussell, 2019)

Ans 4. For providing individualized recommendations, ACME Review Inc will have to incorporate the anonymization of personal data before processing by recommender systems to protect user privacy. Some of the techniques the company can implement to prevent the usage of personal data are cryptography, data perturbation or multi-party computation. (Herce-Zelaya, Tejeda-Lorente, Bernabe-Moreno and Porcel, 2018)

Ans 5. Although deletion of personal data to anonymize user data may seem to be enough to protect user privacy, a new study shows that the data can often be reverse engineered using machine learning to re-identify individuals, despite the anonymization techniques (Rocher, Hendrickx and de Montjoye, 2019). As the uniqueness of each data point(user) increases with each additional personal characteristic (e.g.name, sex, date of birth) captured, hence any arbitrary dataset can be deanonymized. Some new approaches such as differential privacy, which deliberately fuzzes every individual data point to average out across the dataset are more capable to satisfy GDPR data deidentification needs.

Ans 6. The company may make use of a generative model to create the auto-generated avatar images from personal data, there have no practices been introduced to ensure the fair processing of personal data to generate avatars. Hence it could result in algorithmic bias, which could be systemic and unfair discrimination against certain groups in favor of others. Additionally, the processing of data for generating avatars has not been explained. Thus, this practice may result in a violation of GDPR principles pertaining to fairness and transparency. (Salminen, Jung, Chowdhury and Jansen, 2020).

Ans 7. ACME Review Inc. could incorporate an auto-generator tool for avatars, such as Multiavatar. This would not require any personal data, and it could generate over 12 billion cryptographically unique, multicultural avatars. Therefore, the generated avatar under the new mechanism would not qualify under personal data and it would be excluded from GDPR Act. (Lefaix, 2020)

**References**

1.GDPR, 2018. Guide to the General Data Protection Regulation. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation [Accessed 14 Jan 2021].

2. Andreou, A., Venkatadri, G., Goga, O., Gummadi, K., Loiseau, P. and Mislove, A., 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. Proceedings 2018 Network and Distributed System Security Symposium.

3a. Panteli, M., 2019. Recommendation systems compliant with legal and editorial policies. Proceedings of the 13th ACM Conference on Recommender Systems.

3b. Trussell, R., 2019. Copyright and GDPR for photographers - Intellectual Property Office blog. [online] Ipo.blog.gov.uk. Available at: https://ipo.blog.gov.uk/2019/06/11/copyright-and-gdpr-for-photographers/ [Accessed 14 January 2022].

4. Herce-Zelaya, J., Tejeda-Lorente, A., Bernabe-Moreno, J. and Porcel, C., 2018. Adapting Recommender Systems to the New Data Privacy Regulations. 17th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT 2018), Granada (Spain), 303, pp.373-385.

5. Rocher, L., Hendrickx, J. and de Montjoye, Y., 2019. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1).

6. Salminen, J., Jung, S., Chowdhury, S. and Jansen, B., 2020. Analyzing Demographic Bias in Artificially Generated Facial Pictures. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.

7. Lefaix, É., 2020. Multiavatar Generator: 12 milliards d'avatars multiculturels uniques !. [online] Siècle Digital. Available at: https://siecledigital.fr/2020/12/01/multiavatar-generator-12-milliards-avatars-multiculturels-uniques/ [Accessed 14 January 2022].