Applied Data Science - Case Study 1

Data Protection

Q1. Excluding accountability, what are the data privacy principles of the GDPR? You should provide a brief one or two sentence explanation for each, not just a heading. [7 marks]

The GDPR (General Data Protection Regulation) has several key principles aiming to set boundaries in how society may use ones data. These laws are based on several key principles. Below we will discuss these principles and outline what each represents -

- Lawfulness, fairness and transparency

This principle states, according to European Parliament (2018, Recital 39), that in order to collect the data in the first place we must have legitimate grounds. This is summed up from the official documentation which states that we much have an 'appropriate lawful basis' to process and collect this data. The use of the word 'processing' is key as it encompasses the whole timeline for the use of someones data, from collection to deletion. Fairness, is how the processing of personal data may affect the person involved. Those who are processing this data must protect the vital interests of the individual and must not have any adverse impact on the user themselves. Transparency is the way one has to inform right at the start about how the data is going to be used, in an open and honest way.

- Purpose Limitation

Purpose limitation is about the purposes for processing the data, included is the need for the purposes of this processing to be clear from the outset. Furthermore, a record of these purposes must be kept as part of documentation obligations and there is also a need to include this when informing individuals about the privacy of their data. Lastly, if there is a need to use the personal data for a new purpose it needs to be decided whether it is compatible with the original purpose or not. If not then there needs to be some new form of consent. This may involve drawing up a new 'contact' to inform the user informing them of the reasons why one may need their personal data. There is allowance of processing this data for what is known as a 'compatible' purpose without the consent of the individual. These purposes include, archiving purposes in the public interest, scientific/historic research purposes or statistical purposes (European Parliament, 2016, Article 85).

- Data Minimisation

Data minimisations is about making sure we use the personal data as little as necessary. According to European Parliament (2016, Article 5(1)c) we must make sure the personal data collected is “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed”. Hence we must make sure that the personal data we are processing is sufficient to properly fulfil the stated purpose, is relevant, i.e. that it has a clear link to the set purposes and most importantly is limited to what is necessary. Intuitively this implies that ones may not hold the data for more than what is needed to fulfil that purpose.

- Accuracy

The accuracy principle states that we should take all reasonable steps to ensure the personal data is correct and not misleading. The documentation goes further than this by saying the data can’t be misleading ‘to any matter of fact’ (European Parliament, 2016, Article 5(1)d). This suggests that anything other than 100% accuracy will not be tolerated. There is also a need to keep personal data updated but it is acknowledged that this does depend on what the data is being used for. There will always be cases where personal data might be incorrect hence if this is the case, then there must be steps in place to correct or erase this data as soon as possible. Lastly, the company must make sure not to retain old and outdated contacts/data and ensure the erasure of inaccurate personal data without delay.

- Storage Limitation

Storage limitation ensures that no personal data is kept for longer than is needed for the stated purposes. One must be able to justify the length of time that they hold personal data, and periodically the data should be reviewed where you look to erase or anonymise it when it is no longer needed. It is also stated that personal data can be kept for longer if it being kept for public interest archiving, scientific, statistical or historical purposes (European Parliament, 2016, Article 5(1)e). There also needs to be some type of retention policy which stipulates what information you hold, types of records, what it is used for and how long it is intended to be kept. The storage limitation policy needs to be reviewed and changed frequently.

- Integrity and confidentiality (security)

The integrity and confidentiality principle is based on the 'security' principle which is to process personal data securely by means of 'appropriate technical and organisational measures' (European Parliament, 2016, Article 5(1)f). By following this principle, it requires one to consider things like risk analysis, organisational policies and some measures including physical and technical. It also states that throughout the process utilisation of pseudonymisation and encryption is needed. Pseudonymisation is where one cannot tell an individual by the data that you have acquired. Appropriate measures must be in place that enable access and availability of personal data in the event of an incident and also have the processes in place to test the effectiveness of the measures you have in place.

Q2. Indicate two actions the company will need to take in relation to the implementation of the new features described above, because of the GDPR Accountability principle. [3 marks]

- **First action** : Carry out data protection impact assessment
- The new features ACME Reviews ltd want to add to their service include using the personal data for new purposes that they haven't outlined to their users. These purposes are the use of personal data, to generate movies it thinks an individual may like based on other site users and also use data to generate an avatar if one is not supplied by the user. According to the GDPR accountability (European Parliament, 2016, Article 3) a company must maintain assessment and evaluation procedures. Therefore, in order to assess the impact that these new features will have on the existing procedures ACME Reviews ltd have in place, a good starting point would be to carry out data protection impact assessment. This assessment will do the following things; describe the scope, context and purposes of the processing, assess necessity, proportional and compliance measures, identify risk to individuals and any additional measures to mitigate those risks. Applying this will hopefully highlight what action is required to introduce these new features while still being fully compliant with the GDPR principles.

- **Second action** : Keep a record of actions taken

- According to the GDPR (European Parliament, 2016, Article 7) we must now have written documentation and an overview of procedures by which personal data are processed. It goes further by stating that processing activities must include significant information about data processing. This is all for the purpose to prove what they do with the data they collect and to have it made available to authorities on request. Because of this, ACME Reviews ltd must maintain an appropriate level of documentation with regards its processing activities. Once these changes have occurred, processing purposes must be documented and logged as this is definitely classed as significant information about data processing, and thus will make the company more compliant.

Q3. Indicate three significant differences between the 2018 and 1998 versions of the Data Protection Act and explain their consequences. [9 marks]

1. Increased Enforcement Powers (Fines) -

There has been a massive emphasis on penalties for failing to protect individuals data from the DPA (1998) to the GDPR (2018). When we focus on the DPA, we see there is a maximum fine of £500,000 for breaking the rules of the data protection act (Walker, 2019). The main condition for this was for personal breaches which may cause harm or financial loss. An example of this according to BBC News (2016), was that the largest fine given under the old laws was a £400,000 to telecoms company TalkTalk, following theft of their customer details highlighting a lack of security of personal data. Under the new rules, there is a more detailed breakdown for how fines will be applied. According to European Parliament (2016, Article 83(5)), there will now be two types of financial jeopardy; fines for personal data breaches and also fines for administrative breaches. These both carry a weight of 2-5% of previous year annual turnover. This has an upper limit of £10,000,000, picking which ever one is higher. An example of this is recent Marriot and British Airways fines of £22,000,000 (Manager, 2020). The consequences this has on how people will follow the new rules is that they will be followed very closely as companies can not afford these right and extortionate fines. It also highlights the importance of privacy and will encourage companies to do the maximum in order to secure this data.

2. Individual Rights (Consent) -

There has been many changes in individual rights from the DPA to the GDPR I. GDPR mainly builds from the basic DPA model, for example, according to the European Parliament (2016), the GDPR states that the consent must be unambiguous and involve affirmative action. This 'unambiguous affirmative action' requires a positive opt-in response, which disregards any pre-ticked boxes or any other method of consent by default. This compares to the DPA where this was not needed and only a negative opt-in response was needed (UK Parliament 1998). Under the DPA the situation might be that a user would opt-in by default and the individual would actively have to change this. The consequences of this would be a clear consent statement presented to the user, an opt-in button or link explicitly explaining why and an explanation of how their personal data is being used. It would put an end to companies putting all-ready ticked permission boxes to the user and encourages the individual to take responsibility for their data. This results in the user being much more informed about how their data is used and now it ultimately means that the emphasis is on individuals owning their data and not the companies involved.

3. Definitions of Personal Data -

According to the European Parliament (2016, Article 2) under the GDPR "the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system". This compares to the DPA's definition of personal data, '"any information relating to an identified or identifiable natural person ('data subject')" (UK Parliament, 1998). It explains this further to define how one can identify a natural person. The consequences of this is that modern technology like the internet of things and location data are now formally included in the definition of personal data. The new definition realises there are new ways to identify a user from their data, such as identification markers, location data, genetic information and more. This broader definition hopes to avoid any confusion in what constitutes personal data, and helps protect ones data.

Q4. Identify a GDPR related issue that the company may have with implementing the plan to provide individualised recommendations and suggest a way these could be addressed to allow this to proceed. [3 marks]

A few questions have to be asked before ACME decide to implement the new features on their current GDPR practises

- Is the need of personal data explained when collected?

- Is the information upfront in an understandable way?
- Are data subjects well informed of the purpose of using their data?

As it stands, I believe ACME will fail in answering these questions with regards to the new GDPR laws. Permission is explicitly asked to use the personal information provided by the user to both to share ratings and comments with other users. This is due to consent and transparency. According to GDPR and the data minimisation principle, that is limiting to the minimum the personal data required for the processing (European Parliament, 2016, Article 2). As a consequence, firstly ACME Reviews ltd need to document their new purposes for the personal data and furthermore, they need to inform the user, specifying the new reason needed to gather this data. This is addressed by updating the 'contract' signed by the user about asking for permissions in using their data for these new purposes. For the user to understand, this needs to be explained in the most obvious way. If each of these questions are answered within the user's contract, then ACME would be acceptable to proceed with implementing this new feature.

Alternatively, ACME could employ the approach of anonymising their data (Tejeda-Lorente et al, 2020). That is the processing of personal data in a manner that makes it impossible to identify individuals from them. One example of this could be turning data into statistics. Anonymised data are no longer considered to constitute personal data and are not subject to data protection regulations. Careful consideration must be in place about the resources needed to identify someone from this anonymised data.

In conclusion the personal data should be the minimum needed for the purpose and must be stored apart from the individual reviews/ratings data generated by the users. These reviews should be anonymised before being stored. An additional step could be to train algorithms used for the recommendation beforehand with pre-loaded user data in a privacy preserving manner.

Q5. When a user decides to close their account on the website, the company is required to delete their data. In order to continue to provide the useful ratings and review comments to other users, the company would like to turn this data into anonymous data by disconnecting it from the user personal details (name, city, etc.) with the consent of the user. Is the deleting of the personal data sufficient to achieve this? Explain why it is/is not sufficient. [3 marks]

This response is based on the anonmyisation and pseudonymisation principles. Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific person without the use of additional information. This additional information must be kept carefully and separate from the personal data. The most important point about this, is that pseudonymised data can still be used to identify and single individuals out when combined with their data from other sources. Hence, this is still classed as personal data and is subject to data protection regulations.

Anonymisation refers to processing personal data in a way that makes it impossible to identify an individual from it. This must take into account all feasible ways to convert this data back to a form which one could use to identify the individual concerned. The important thing to note is that anonymised data is no longer considered personal data and therefore will not be subject to the data protection regulations. Do we class creating an avatar in this way as a form of pseudonymisation or anoymisation? I believe the answer to be pseudonymisation. We see that we have disconnected the personal details to the ratings and review data of an individual which is the definition of pseudonymisation. This data is then deleted. However, the review data could still be tracked back and attributed to a certain individual and hence it is not sufficient to process the data in this manner when a user decides to close their account. This is now pseudonymised data and as said before, still counts as personal data, hence when the company is asked to delete all data on the closure of a users account, this approach would not be sufficient. (Tejeda-Lorente et al, 2020)

Q6. Other than a lack of consent, suggest a reason that allowing the system to generate the avatar image in the way described would not be compatible with the GDPR. [2 marks]

According to the legislation, ones processing of personal data must be fair (European Parliament, 2016, Article 5). That is, considering how the processing might affect the individual concerned, only handling the data in a way that the user would expect and not deceiving or misleading the user. I would argue that using the personal data in this way to create a personalised avatar does not meet any of the fairness contents of the GDPR. There has been no consideration about the adverse impacts that sharing such an image can have on the user. One example of this could be potential bullying or cyber-bullying from peers about the generated image. In terms of the handling of the data, the user does not have any clue that their data is being used to generate this image, and this is a huge breach of their first principle of GDPR about being transparent (European Parliament, 2016, Article 2). It could therefore be argued that the use of the data in this way is misleading. They have stated their use of the data, and what they have actually done with the data is something totally different, in using the data for the creation of this avatar. ACME should take all steps to protect users information and to share this avatar based on their personal information could be potentially harmful emotionally and/or to their reputation (Hunton, 2017, p. 6).

Q7. Assuming explicit informed consent was obtained, indicate an alternative approach that could be employed to provide a unique system generated avatar image for each user that would be compatible with the GDPR and explain why this would be compatible. [3 marks]

There are a lot of GDPR issues raised with the current version of creating an avatar based on personal data. I believe one of the first steps is to update our contract to the user in telling them about the processes we have in place to use pseudonymization on their data, and what that data will then go on to be used for. According to the European Parliament (2016), Pseudonymisation is defined as, “the processing of personal data in such a way that the data can not be attributed to a specific data subject without the use of additional information, was long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or an identifiable individual.” (European Parliament, 2016, Article 4(3b)). Using this we can separate the personal data from the other data collected for example, movie reviews and comments. This data can now be used a little more freely under the GDPR. After we have gained the necessary permissions from the user, we can now create an avatar based solely on what films this user likes or the films that similar users like based on this pseudomised data. I believe this is compatible with the GDPR. As it can be noticed, the handling of data in this form is fair, transparent and lawful. I believe we have achieved this, the way we process the data here is totally lawful and transparent, in updating the user on the new purposes for their data. We have safeguarded them by making sure you cannot detect the individual from the data, hence making this approach fair. The other GDPR principles are met, with consent being given in the updated contract and purposes. The right documentation of these processes will also have to be done in order to meet GDPR Accountability principle.

Bibliography

1. European, T. E. P. a. t. C. o. t., 2018. [Online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
2. EU Parliament, 2016. *General Data Protection Regulation.* [Online] Available at: https://gdpr-info.eu
3. Hunton, 2017. Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR. *Centre for information policy leadership,* p. 6.

4. Manager, D. P., 2020. *GDPR fines in 2020.* [Online]

Available at: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/

5. News, BBC, 2016. *Talk Talk fined $400,000 for theft of customer details.* [Online]

Available at: https://www.bbc.co.uk/news/business-37565367

6. Tejeda-Lorente, 2020. Adapting Recommneder Systems to the new data privacy regulations.

7. UK Parliament, 1998. *The Data Proctection Act 1998.* [Online]

Available at: legislation.gov.uk

8. Walker, D., 2019. What is Data Protection act 1998.