National College of Ireland

Data Governance and Ethics (H9DGE)
MSc/PGD in Data Analytics
MSCDADJAN20A_O
Continuous Assessment 1

Group: [4]

Members:

[Oluwatobi Ekundayo – x19173105 – Cohort A]
[Shivani Badola – x19177127 – Cohort A]
[Prasanth Mariayagapan Antonyraj – x19198744– Cohort A]

Lecturer: Dr. Vanessa Ayala-Rivera

DATA GOVERNANCE AND ETHICS PROJECT

Covered in this paper are four (4) sections, which are solutions to project questions for the Data Governance and Ethics (H9DGE) module.

I. Background

CleverSoft

CleverSoft is a company headquartered in Dublin which provides software development services. Some of these services include custom application development, IT security, software deployment, software migration, data backup and recovery, system integration, API integration and app maintenance. This study describes the general responsibility for data security for both small and large organizations. We make sure that people can trust the organization to use their data fairly and responsibly. It is about treating people equally and freely and maintaining a balance with the broader interests of society. As Data Protection Officers (DPO) the report below provides the legal requirements related to data protection in Ireland and the European Union.

A. Overview of the legal requirements in data protection for CleverSoft

In this case, GDPR is one of the crucial constraints which needs to be reviewed and followed to get legal consent in Data Protection. As per General Data Protection Regulations (GDPR) CleverSoft is required to provide advice and guidance, encourage good practice, including the rights of people, the management of requests for private information, consent and data security impact assessments. There are two fundamental objectives of GDPR.

- Transparency
- Providing the public their information that is being used

At the same time as that of the **GDPR**, the EU has adopted the provisions of Data Protection Directive 95/46/EC (Regulation, 2018), which protects individuals with regard to the processing and free movement of personal data. It should be clear to individuals that personal data relating to them is obtained, used, evaluated for or analysed and to what level personal data are or will be processed. The GDPR allows CleverSoft to put in place suitable organizational and technological steps to enforce the principles of data security and to protect human rights. Essentially, this means that the company needs to incorporate data security into the processing activities and business practices, from the design phase to the lifecycle. Earlier it was called ‘privacy by design’ and it has had always been part of data protection law. In all you do, data security by design is about recognizing data protection and privacy concerns upfront. It will help you ensure that you comply with the GDPR's fundamental principles and specifications, and forms part of the emphasis on transparency.

On the other hand, in Ireland there is a new legislation for data protection known as **Data Protection Act 2018** which was signed on 24th May 2018 (Regulation, 2018). This legislation would further expand a company’s data privacy, which is lacking in the GDPR. Therefore, as per Irish regulations, the Member State has more flexibility in new data security.

Data protection self-assessment
https://ico.org.uk/for-organisations/data-protection-self-assessment/

Consequently, all the companies be located in Dublin, Ireland, and process the EU personal data must follow the GDPR and Data Protection Act 2018 as legislation for data protection. The organisation is expected to ensure processing of systems and services are available, confidential and resilient. A process to methodical testing and assessment to evaluate how effective the organisational and technical measures taken should be implemented.

B. Roles and Responsibilities for Data Controller and Data Processor

In order to ensure compliance with EU data protection law and the GDPR, the roles and responsibilities of the data controller and data processors are highlighted below:

- **Role - Data Controller**

  The role of the *Data Controller* is to define, determine and decide what information will be processed, which is significant in personal data protection. The data controller is usually an organisation (public authority, a legal agency or legal person etc.).

- **Responsibilities**

1. ***Design processing systems*****:** Data controller must design their processes such that only data absolutely needed for its objective. Also, the controller must limit the access to personal data to only those required for processing. The concept of privacy by design is one that must be adopted by the data controller to implement measures that meets requirement of the regulation.
2. ***Keep data secure*****:** The data controller is expected to ensure the implementation of modern security measures most suitable for the risk relative to the activities been carried out. The security measures must be reviewed by the data controller to validate that it complies with the code of conduct.
3. ***Keep data records*****:** The data controller is expected to also keep a record of processing activities especially those involving sensitive information.
4. ***Carry out data protection impact assessments*****:** In the case of high-risk processing tasks, the data controller must carry out at first a Data Protection Impact Assessment (DPIA). The process operations involved in the assessment is prescribed by the Data Protection Commission.
5. ***Report data breaches*****:** It is the duty of the data controller to inform the Data Protection Commission in the case of any personal data breach without any delay. It is expected the data controller must notify the commission as well as the data subject within 72 hours of becoming aware of the data breach.
6. ***The use of processors that meet the requirements of the legislation*****:** It is expected that the data controller selects only data processors that guarantees their systems meets the requirements that is compliant with the regulation, for any processing activities that will be carried out. The data controller must state the scope of any processing activity required as well as the obligations of the data processor in a binding written agreement (contract).

7. ***Compliance with codes of conduct and certification*:** Codes of conducts and practices prepared by associations on how GDPR and other Data protection laws should be applied specifically must be adhered to by the data controller.
8. ***Transferring data outside the EU*:** If data is to be transferred to an unapproved international organisation, sector or country, the data controller must provide the appropriate precaution, security and ensure that any data subject can exercise their rights, (Citizensinformation.ie, 2018).

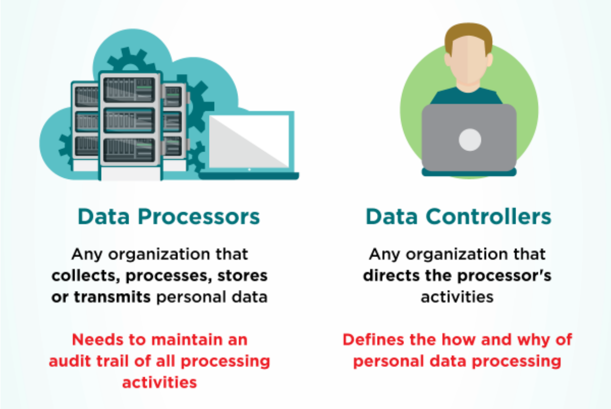
Data Processors
Any organization that collects, processes, stores or transmits personal data
Needs to maintain an audit trail of all processing activities
Data Controllers
Any organization that directs the processor's activities
Defines the how and why of personal data processing

Figure 1: Data Processors vs Data Controllers (Edwards, 2019)

Figure 1 above summaries the difference between the data controller and data processor for quick understanding.

- **Role - Data Processor**

The role of the *Data Processor* is to carry out personal data processing which include collecting, storing, distributing or destroying, as instructed by the data controller. The data processor ensures that records of the company's data processing activities are kept and that the information in those records are available upon request for inspection. The processor is that legal person or agency appointed by the data controller to do the actual data processing.

- **Responsibilities**

1. ***Keep data records*:** A record of every processing activity and those involved in the processing of sensitive information must be kept, as they can be requested at any time for inspection by the Data Protection Commission.
2. ***Keep data secure*:** The data processor must evaluate and implement security measures that complies with the code of conduct to ensure personal data are secured. Some of these measures may include data encryption or anonymisation etc.
3. ***Report data breaches*:** In the case of any personal data breach, the data processor must notify the data controller immediately without any delay. This gives the data controller the notify the commission as well as the data subject just in case the data breach results in the rights and freedom for an impacted data subject, (Citizensinformation.ie, 2018).
4. ***Compliance with codes of conduct and certification*:** An approved code of conduct can be used as an element to measure data processor's compliance with the obligations of

the controller. It is therefore important for a data processor seeking to process information on behalf of a data controller to adhere to such code of conduct.

5. ***Appointment of sub-processors*:** Only with the permission of the data controller can the data processors appoint sub-processors. The processor must ensure the sub-processor personal data are processed in accordance with instructions by the data controller or that which complies with the regulation or law, (Gabel and Hickman, 2019).

C. Factors to consider for data transferring agreement to United States

The agreement on the transmission of data contains all the information relating to the transfer and use of the data and also protects people's intentions. It is mainly between two parties who want to exchange some kind of data between themselves, where the parties could be two or more organizations or nations.

The main objective of the data transfer agreement is to lay down the conditions in which the data will be used after the data has been transferred. The party (CleverSoft) transmitting the data may lay down certain provisions concerning the purpose for the transmission of data and the manner in which it is to be used. This agreement is usually established in situations where exploitation of data may lead to a problematic state of affairs. Such contracts shall be governed by national laws. In the European Law for example, when personal data is transmitted outside the European Economic Area (EEA), a number of requirements need to be fulfilled.

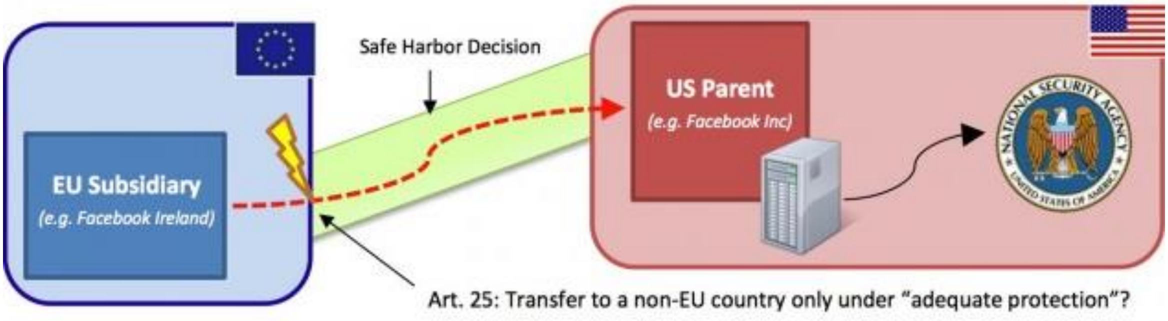

Figure 2 : Sample data transfer by Facebook from EU to USA (Farivar, 2013)

The Data Sharing Agreement must incorporate the requirements of GDPR (General Data Protection Regulation). These arrangements would contain restrictions on data protection contracts. In general, Data can be transferred among different nations for several purposes including commercial use or for personal authentication, identification etc. Some of the use cases and its corresponding protection requirements are discussed below.

EU-US Privacy Shield

This clause protects the constitutional rights of any EU citizen whose private data is transferred to the United States for business purposes. For example, as shown in Figure 2, If Facebook is transferring any data outside EU, then this can only be done if the data transfer medium follows safe and secure data transfer protocols. Without adequate protection, this cannot be carried out. This facilitates the free transfer of information to businesses that are accredited under the Privacy Shield in the US. The provision includes:

EU-US Privacy Shield
https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/eu-us-data-transfers_en

- Powerful data security obligations on businesses that obtain EU personal data
- Protects access to data by the US Government
- Efficient rights and recourse for citizens
- An annual joint review by the EU and the US to track the proper execution of the agreement.

EU-US Umbrella Agreement

An agreement proposed in December 2016 called EU-US Data Protection Umbrella Agreement introduced the high privacy safeguards for transatlantic law enforcement corporation. It provides a detailed collection of data security regulations applicable to all transatlantic exchanges between law enforcement agencies in the field of criminal law. It therefore fulfils the twofold goal of collaborating with our U.S. partners to fight serious crime and extremism while at the same time advancing Europeans' level of security in accordance with their constitutional rights and the rules of EU data protection.

The aspects the company (CleverSoft) needs to consider for **Data Transferring Agreement** (DTA) are:

- Information of the arrangement between the lead consumer company and the lead provider company
- Background details of the organizations involved.
- Guiding principles
- Obligation of both parties
- Costs and payment arrangements.
- Warranties
- Legal title and data transferred and benefit sharing
- Permits, licenses and approvals
- Termination, applicable law and severability

Additional points to ponder include:

- If a transfer agreement is performed independently from the main service agreement, it is important to carefully analyse the relationship with the main agreement. If in reality, clauses that would typically be found in a separate transfer agreement are integrated into the main agreement, it would be appropriate to recognize the broader provisions of the main agreement, (Lyons and Mackay-Temesy, 2020).
- Find the parties referred to in the main agreement and the parties to which the processing is carried out. Consider group businesses, in particular.
- After the termination of the key agreement, is there a need to process personal data or maintain records of personal data? Is that dealt with adequately? Carefully consider the broader provisions on termination of jobs and survival clauses in the agreement as well.
- Consider the role of responsibility opposite the data processing relationship in the main agreement.
- Are any additional onward transfer and subsequent responsibilities discussed in addition to coping with any established direct or indirect transfer of personal data beyond the EEA?

Contents of the Data Transfer Agreement (2020)
https://www.agreements.org/data-transfer-agreement.html/

- Consider the agreement's general sub-contracting provisions Do they comply with the agreed upon sub-processing position?
- Are the interests of the data subjects covered in addition to any third-party rights with respect to community companies?

II. Ethical Issues and Challenges with "Fake News"

Authenticity of facts has been a long-standing concern impacting corporations and culture, both in print and digital media. The scope and effect of content shared on social networks is so instant and intensified that skewed, misleading or incorrect information has an immense potential to create real life impacts to millions of users within minutes. Recently, a variety of societal issues and alternatives to alleviating the issue have been expressed. Alongside, Ethical communication is constantly the cornerstone of global problems being faced today. Fake news is one of the extremely concerning and direct means by which false perceptions have been propagated through social media (Stroud, 2019). In this study they described a pragmatic approach which illustrates the efforts to conceptualize and eliminate fake news. The pragmatist approach to fake news underlines the contradictory ideals and findings at stake.

Some ethical issues which can pose different ethical challenges are described below.

- **Issues of the good: Consequences and harms**
  One of the main ethical challenge raised by fake news involves its impact or effect on democratic process that can directly impact social and communal as well as individual happiness. For example, the media around the globe was deeply alarmed during the election of Donald Trump in 2016. A basic pillar of responsible journalism continues to be threatened by the free dissemination of malicious lies. For months, Facebook was accused by Donald Trump's election opponents for allowing fake and hoax news reports to circulate openly via their news feeds. This created a firestorm in media circle in the aftermath of the US presidential election.
- **Issues of the right: Respecting autonomy**
  In the area of property rights, the exchange of information on behalf of people who own the information may not be ethically correct. This right of ownership concept must be recognized as to who is allowed to produce and transfer the information between the other networks.
- **Issues of the virtuous: Judgments and reactions**
  This is about the people's psychological responses within communal contexts. This is an understanding that our conditioned reactions to its specifics and the reactions we expect other members of the group to have in that situation will be a crucial part of every moral situation. Such answers are the details about the morally questionable case, an aspect that is brought to light by the controversies over fake news.
- **Information Overload: Misleading people's emotion**
  The glut of knowledge has given rise to extreme competition for people's attention. Herbert A. Simon, a Nobel Prize-winning economist and psychologist said, "What

information consumes is quite obvious: it consumes the attention of its recipients." One of the first effects of the so-called attention economy is the lack of high-quality information. This result was illustrated by the *OSoMe* team with a series of basic simulations. It portrayed users of social media, such as Andy, called agents, as nodes in an online network of acquaintances. At each phase of the simulation, an agent can either create a meme or post a meme that he or she sees in a news feed. To simulate minimal focus, agents are allowed to access only a handful of articles at the top of their news feeds.

To make it things worse, search engines and social media sites offer tailored reviews based on the large amount of data they provide about previous customer interests. They give priority to information in our feeds that we are more likely to agree with – no matter how fringe and protect us from information that could change our minds. This makes polarization goals simple for us. Nir Grinberg and his collaborators at North-eastern University have recently found that conservatives in the U.S. are more receptive to propaganda.

- **Rise of the bots: Resharing fake news using bots**
  The accuracy of knowledge is further compromised by social bots, who can manipulate all of our perceptual differences. Bots are easy to construct. Social networking sites have the so-called application programming interfaces that makes it very trivial for a single actor to set up and manage thousands of bots. Amplifying a tweet, even with just a few early upvotes on social networking sites like Reddit, can have a major effect on the post's subsequent visibility.

  Few manipulators play on both sides of the divide with different fake news pages and bots, driving political polarization or ad monetization. For instance, a network of inauthentic Twitter accounts that were all orchestrated by the same person was recently discovered at *OSoMe*. Some purported to be pro-Trump backers of the Make America Great Again movement, while others appeared to be Trump's "resisters"; they all called for political donations. These operations amplify the material that preys on affirmation biases and accelerates the formation of unauthentic news.

Challenges fake news can create

Challenges arrive when "fake news" circulates the media space, and some people starts to believe it as well as to seek for it. A huge concern here is the fact that because there is a confirmation bias by users and sharing with other friends, social media leads to the dissemination of false and misleading news like "wildfire". According to an article by Silverman (2016), it was found that 38% of the posts that are being shared on Facebook by three right-wing politics sites included false and misleading information. However, the internet technology offers various ways for people to access data and they generate their own filters for information that they want or do not want.

In the paper by Rubin et al. (2015), three different types of fake news were discussed. each in comparison to legitimate serious reporting, are listed below, indicating there have been at least three different sub-tasks in detecting fake news, namely: fabrication, hoaxing and detecting satire.

Information overload helps fake news spread and social media knows it
https://www.scientificamerican.com/article/information-overload/

- **Social Fabrication**: With the advent of the global social media stage, the embellishment landscape has escalated where each of us can create a personal brand. Hundreds and hundreds of times, the same lie can be told in the social media domain. The contingent liabilities that are hidden in a manufacturing bed will result in a loss of control and a company's ultimate downfall.
- **Large-Scale Hoaxes**: Another form of intentional deception or falsification in the news or social media is hoaxing. Attempts to deceive audiences masquerade as news, and conventional news outlets may pick up and mistakenly confirm them.
- **News Satire**: There are two slightly contrasting forms of news satire. To comment on real-world news events, one type uses satirical satire and sketch humour, whereas the other portrays entirely fictionalized news reports. We separate serious news from funny ones. They will no longer be predisposed to take the data at face value if readers are conscious of the humorous intent. In order to alert users, particularly in decontextualized news aggregators or platforms and technology should recognize humour and show originating sources prominently.

III. Data-Informed Duties in AI Development

Many great things have been accomplished by artificial intelligence (AI)-based technology, such as facial recognition, medical diagnosis, and self-driving vehicles, which also support economic growth, social advancement, as well as enhancement of social well-being and safety.

As AI progresses, how to solve the legal and moral problems associated with AI is one important question. While around 2006 the idea of "machine ethics" was proposed, AI ethics is still in the infancy phase. AI Ethics studies AI-related ethical values, laws, guidelines, policies and regulations. AI ethics is part of technologically advanced ethics that focuses on robots and other artificially smart agents. It can be divided into machine ethics and roboethics (robot ethics). In this study "Data-Informed Duties in AI Development", Pasquale (2019) illustrates the standards followed for data collection, analysis, use and management can inform and complement professional judges. Such a rule would not only provide industry with guidelines to help prevent preventable incidents.

In response to legal conflicts resulting from the deployment of AI, it will also assist a judiciary that is increasingly called upon to establish common law. Where, Tort law will only be among several laws during deployment of AI. A company deploying AI can fail in several ways one particular type is inaccurate and inappropriate data in training sets for machine learning. If such faulty data collection, analysis, and use is repeated or intentional, businesses using faulty data will be forced to compensate those affected by the data use and may be subject to legal costs.

1. Problems Caused by Inaccurate and Inappropriate Data

While many innovations are taking place by AI applications, they are also introducing new risks. The mere fact about AI is that a technology better in general does not mean that it is ideal for all the cases. For example, in deep fake image recognition application there is well-documented failure of AI system as they need to recognise the fake and real images. Similar questions are posed by researchers in ethnic disparities in health care in United States. They already know that skin anomaly-detecting software can fail to function for African Americans while it does for the white patients. Thus, this paper highlights the problem caused by data in AI.

- **Inaccurate Data**
  The task of correcting inaccurate data is nowadays considered as secondary or menial. But such errors could be catastrophic at a certain degree of prevalence. Measurement biases caused by errors in measurement and data collection resulting from defective equipment or software, or human error, must be considered by researchers.
- **Inappropriate Data**
  Once the AI agent is using biased data, the bias will become an ongoing problem. For instance, software used to predict future criminals showed bias toward a certain race bias comes from the data on training involving human biases (Bossmann, 2016). Therefore, it is important to figure out how to program and train AI agents without human biases.

2. COMPLEMENTARY TORT AND REGULATORY REGIMES

In this part, the author described the emerging doctrine and regulations that suggests data driven duties for data science developers. This kind of duties serve two purposes. First to confirm that the machine learning training data sufficiently represents the domain it regulates or impacts and, second to recognize and correct anomalies until they cause great damage. To encourage just and humane developments in AI, it will be essential to establish and sustain these duties.

3. REGULATORY STANDARDS FOR DATA USE AND REPORTING

Data enforcement applies to any law that a corporation must obey, in order to ensure that the confidential digital assets it possesses are secured from destruction, fraud and misuse. Usually, personally identifying information and financial details. There are two terms *Data Compliance* and *Data Security* that are often bundled together. However, they have the same goals to minimise the risk a business is exposed to. While compliance only guarantees that one meets the minimum requirements legally required, conversely data security includes all the processes, procedures and technology that determine how an organisation should look after confidential data and defend against violations.

IV. ACM Principles for Algorithmic Transparency and Accountability

The paper “Statement on Algorithmic Transparency and Accountability” discusses the pervasive use of algorithms and the need to setup standards for identified problems related with the design and technical aspects of algorithms. This is necessary to prevent bias from its inception as well as output errors for some algorithms and analytics, (Council, A.U.P.P., 2017).

Algorithms contribute as inputs to decisions made by humans and are used for making automated decisions. Therefore, institutions or organizations should adhere to the set of standards or principles in-line with ACM Code of Ethics by designing and developing algorithms or systems with these principles in mind. By doing so, potential harm can be minimized. To deploy any of these complex algorithmic-based decision-making systems, system designers are therefore required to build the principles of awareness, access and redress, accountability, explanation, data provenance, auditability and validation and testing into their system.

A decision-making system evaluates and organizes comprehensive information to support problem-solving. An algorithm applied to such systems analyses huge amounts of data to infer information for decision-making. Example of such algorithmic-based decision-making systems include medical diagnosis software, global positioning system (GPS), virtual call centre solution etc

Principles applicable to algorithmic-based decision-making systems

1. ***Accountability***: Decision made by an algorithmic-based decision-making systems should be the responsibility of any organisation using such system. Even when results obtained by the system can cannot be explained in detail, the organisation must be responsible for the decisions made. For instance, the medical diagnosis software is a tool whose decision affects human life directly. Doctors using the software must be responsible for any conclusive decision derived from the system
2. ***Auditability***: Auditability in an algorithmic decision-making system requires “logging and record keeping”, Garfinkel et al. (2017). It becomes essential for regulatory compliance and dispute resolution. Therefore, it is important that data collected, algorithms and models used, and decisions made are constantly recorded for auditing purposes. This becomes useful for any scenario where malicious act or harm is suspected. For instance, in a scenario where virtual call centre solution deducts huge amount of customers credit, customers private voice call gets sent to different customers or data from the solution gets leaked to the public, the system can easily be audited to figured out the root cause.
3. ***Validation and Testing***: A precise and meticulous method should be utilised for the validation of developed models applied in a decision-making system. Techniques such as red-teaming strategies applied in computer security, regression tests and other stress-test concepts should be employed to increase the confidence in the system (Garfinkel et al., 2017). It is essential that the methods applied, and results are well documented. For instance, the global position system (GPS) must go through several model validation to ensure results of the test-cases or use-cases are successful. Application of this solution can potentially misinform a user or cause an accident if it does not go through rigorous validation and testing.

REFERENCES

Bossmann, J., 2016. Top 9 Ethical Issues In Artificial Intelligence. [online] World Economic Forum. Available at: https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/ [Accessed 12 November 2020].

Citizensinformation.ie., 2018. *Obligations of data controllers and processors under the GDPR*. [online] www.citizensinformation.ie. Available at: https://www.citizensinformation.ie/en/government_in_ireland/data_protection/obligations_under_general_protection_regulation.html# [Accessed 9 November 2020].

Council, A.U.P.P., 2017. Statement on algorithmic transparency and accountability. Commun. ACM.

Edwards, J., 2019. *GDPR Overview: Complying with EU Laws for Personal Data*. [online] blog.ipswitch.com. Available at: https://blog.ipswitch.com/gdpr-eu-personal-data. [Accessed 9 November 2020].

Farivar, C., 2013. EU Reevaluating Data Sharing Agreement With US In Wake Of NSA Leaks – Ars Technica. [online] Arstechnica.com. Available at: https://arstechnica.com/tech-policy/2013/07/eu-reevaluating-data-sharing-agreement-with-us-in-wake-of-nsa-leaks/?amp=1 [Accessed 27 November 2020].

Gabel, D. and Hickman, T., 2019. Chapter 11: Obligations of processors – Unlocking the EU General Data Protection Regulation | White & Case LLP. [online] www.whitecase.com. Available at: https://www.whitecase.com/publications/article/chapter-11-obligations-processors-unlocking-eu-general-data-protection [Accessed 9 November 2020].

Garfinkel, S., Matthews, J., Shapiro, S.S. and Smith, J.M., 2017. Toward algorithmic transparency and accountability. Communications of the ACM, 60(9), pp.5-5.

Lyons, L. and Mackay-Temesy, T., 2020. Data Transfer Agreements - Taylor Wessing's Global Data Hub. [online] Globaldatahub.taylorwessing.com. Available at: https://globaldatahub.taylorwessing.com/article/data-transfer-agreements [Accessed 12 November 2020].

Pasquale, F., 2019. Data-Informed Duties in AI Development. Columbia Law Review, 119(7), pp.1917-1940.

Regulation, G.D.P., 2018. General data protection regulation (GDPR). Intersoft Consulting, Accessed in October 24, 1.

Rubin, V.L., Chen, Y. and Conroy, N.K., 2015. Deception detection for news: three types of fakes. Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

Silverman, C., 2016. Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate. [online] Buzzfeednews.com. Available at: https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis [Accessed 22 November 2020].

Stroud, S.R., 2019. Pragmatist Media Ethics and the Challenges of Fake News. Journal of Media Ethics, 34(4), pp.178-192.