**Project One**

Justin Starr

Department of STEM

CS 370 – Current/Emerging Trends in CS

Dr. Obafemi Balogun

June 2, 2024

**Project One**

Recently, a European Union (EU) regulator brought to the company’s attention that it may be violating some aspects of the General Data Protection Regulation (GDPR). Specifically, they are concerned that our business model may not conform to some or all of the following principles defined by the law: transparency, purpose limitation, data minimization, accuracy, storage limitation, confidentiality, and accountability. As a company that utilizes neural networks in our personalization algorithms, which play a major role in our business model, it is essential that we address the regulator’s concerns and begin moving forward with bringing the company into GDPR compliance. The following explains the basics of neural networks and how they work, how neural networks are used to create personalization, how portions of the GDPR affect personalization, how the GDPR is affecting the company’s practices, and proposes adaptations to the company’s practices to act in compliance with the GDPR.

Artificial neural networks are essentially a class of machine learning models, which has been inspired by studies of the central nervous system, or biological neural networks (Gulli, A. & Pal, S., 2017). Each network is comprised of many interconnected neurons (nodes), which are organized in layers, and exchange messages when certain conditions are met (Gulli, A. & Pal, S., 2017). These nodes or neurons are capable of processing and learning from data, which enables tasks, such as decision-making and pattern recognition in machine learning (GeeksforGeeks, 2024). Neural networks comprise an input layer, one or more hidden layers, and an output layer.

At the input layer, each feature in this layer is a node that receives input data. This data is then passed into the hidden layer(s) where each neuron processes the inputs, multiplies them by weights, and sums them, where they are then passed through an activation function, allowing the network to recognize patterns (GeeksforGeeks, 2024). This process is continued until the output layer is reached, where a binary result (prediction) is made on the data. It is an iterative process, and through a method called backpropagation, as the network refines its weights, the network becomes better at making predictions, and classifying objects (GeeksforGeeks, 2024).

Neural networks are utilized to aid in the personalization of the user experience by collecting and analyzing large amounts of customers’ data. This data is then used to learn what the specific preferences and needs are of the individual customers, therefore, providing a more tailored and user-centric experience (Zia, 2023). While neural network personalization offers numerous advantages, such as better customer experiences, increased revenue, reduced customer turnover, and data-driven insights, it is essential to understand that there are ethical concerns that arise from utilizing this technology (Zia, 2023). Collecting and analyzing data from customers raises concerns about data security and privacy, and companies must ensure that they adhere to data privacy regulations and secure consumer data from misuse (Zia, 2023). Transparency and trust are concerns that must be addressed due to the risk of data breaches. Customers are becoming more cautious about how their data is being used because of data breaches, and it is the responsibility of these companies to be completely transparent about how they are collecting, handling, and using this information (Zia, 2023). In addition, when considering personalization, it is essential to understand that neural networks often have a “black box” classification system, which can lead to hidden biases. Complex models, such as Deep Neural Networks, which contain up to millions of parameters or weights are considered “black box” because the behavior of the model is incomprehensible, particularly after the model has been trained (Perteneder, 2021). It also includes AI systems whose internal workings are unclear and often hidden from users and sometimes developers, and they do not explain how they arrive at their conclusions (Glover, 2024). These “black box” classification systems are not only problematic because of their lack of transparency, but also because of biases that can be inherited by the algorithms due to human prejudices that exist within the training data, ultimately leading to unfair and inaccurate decisions (Guidotti et al., 2024).

Portions of the GDPR that affect personalization include transparency, purpose limitation, data minimization, accuracy, storage limitation, confidentiality, and accountability, and all should be incorporated into the business model. Being transparent means that people clearly understand who you are, how and why you use their information, and that this is made clear from the beginning by being upfront, open, and honest. (Information Commissioner's Office, n.d., What is transparency?). It must be told to individuals about the processing of their data in a way that is easily understandable and in a way that is easily accessible. Purpose limitation refers to being clear from the beginning, why the personal data of individuals is being collected, what they plan to do with it, and if they plan to do anything else with the data than originally stated, the new use of the data is lawful, fair, and transparent (Information Commissioner's Office, n.d., What is the purpose limitation principle?). Data minimization means that it must be ensured that the personal data being processed is adequate and can fulfill the stated purpose for collecting the data, it must be relevant and is limited to only what is necessary for the stated purpose (Information Commissioner’s Office, n.d., Principle c: Data minimization?). Accuracy means taking all responsible steps that ensure the personal data that is collected is not inaccurate or misleading, it is kept up-to-date depending on what it is used for, and if it is discovered that any of the data is incorrect or misleading, sensible steps must be taken to correct or erase the information as quickly as possible (Information Commissioner's Office, n.d., Principle d: Accuracy). Storage limitation refers to only keeping data for as long as it is needed, and that it is only held onto for the purpose previously stated (Information Commissioner's Office, n.d., Principle e: Storage limitation). Confidentiality means that appropriate security measures are taken to protect the personal information collected, and accountability requires that responsibility is taken for what is done with the personal data and how well you maintain compliance with other principles (Information Commissioner's Office, n.d., Principle f: Integrity and confidentiality, and Accountability principle). Following these principles and adherence to the GDPR will help to ensure the protection of personal information.

Specific legal concerns may arise from the company’s use of neural networks as a classifier to personalize the user experience. Particularly, as it concerns the GDPR, if a company is collecting and processing data of EU or EAA citizens, it must comply with the GDPR (Langeland, 2023). One legal concern is that because the company utilizes neural networks as a classifier to personalize the user experience, it is also known that this is accomplished with the use of “black box” models. This directly conflicts with the GDPR’s emphasis on transparency and the right for individuals to understand or know how their personal information is processed. According to the Information Commissioner’s Office (n.d.), failure to comply with the principles of the GDPR can leave you vulnerable to heavy fines, which could mean a fine of up to 17.5 million euros, or 4% of your annual worldwide turnover. Another legal consideration would be concerning purpose limitation and not knowing every aspect of how the AI model will use personal data as well as how the model will make decisions based on that data. It is possible that previously collected data may be used to train the model for future purposes for which the data was never intended to be used.

Not collecting data from users would significantly impact the user experience. Essentially, this would be taking a step backward, in the wrong direction, concerning improving user experiences and the technology available today that can be utilized to do this. Also, because the company is an industry leader in user experience, our customers expect the software to anticipate individual user needs based on recommended posts, recommendations for friend requests, groups to join based on shared interests, etc. None of this is accomplished without being able to collect data from individual users and learning about them and their interests on a personal or more intimate level.

Current trends (best Practices) in artificial intelligence and machine learning aimed at preserving privacy include having an understanding of your use of AI, practicing transparency and ensuring explainability, incorporating ethical testing, de-identifying personal data, minimizing personal data in AI algorithms, conducting privacy assessments specific to AI functions, and establishing accountability and internal governance (Davis, 2023). Understanding the use of AI means sitting down with internal teams and determining exactly what personal data is being collected in AI applications, knowing how and why they are using it, and what they do with the information once they are done with it. Customers should have a clear understanding (transparency) of how AI platforms use their personal information and whether or not decisions have been made for them or could have affected them (Davis, 2023). Incorporating ethical testing helps to ensure a more reliable product, delivering fair and explainable decisions that are unbiased and not prejudicial. Incorporating methods of anonymization and aggregation enables companies to detach personal data from individual identities (Davis, 2023).

To comply with the GDPR, the above best practices must be followed. The company needs to begin by looking at all the different types of data collected from its users. Afterward, the company should determine whether each type of information is essential for accomplishing the company’s goals and remaining a leader in user experience. Once data is obtained from a user, it should be immediately broken down into essential and non-essential information, and non-essential data should be erased immediately. Essential data should then be anonymized and classified with similar types of data. From here, algorithms can be better designed to recognize significant differences in the data and make useful, unbiased recommendations for users. Establishing a team responsible for monitoring the performance of team members to ensure accountability is taken when systems do not perform as expected will help ensure the highest level of standards including in ethics, in all capacities, not only in the design phase but all phases of the software development lifecycle.

**References**

Davis, S. (2023, February 23). *AI and data privacy: Safeguarding sensitive information*. Logic

20/20. <https://logic2020.com/insight/ai-data-privacy-strategies/>

GeeksforGeeks. (2024, January 3). *What is a neural network?*

[https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/#](https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/)

Glover, E. (2024, May 22). *What Is Black Box AI?* Built In. [https://builtin.com/articles/black-](https://builtin.com/articles/black-" \o "https://builtin.com/articles/black-box-ai)

[box-ai](https://builtin.com/articles/black-" \o "https://builtin.com/articles/black-box-ai)

Guidotti, R., Monreale, A., & Pedreschi, D. (2024). *The AI Black Box Explanation*

*Problem*. Guiding Tech Media. <https://www.kdnuggets.com/2019/03/ai-black-box-explanation-problem.html>

Gulli, A., & Pal, S. (2017). *Deep learning with keras : Get to grips with the basics of*

*keras to implement fast and efficient deep-learning models*. Packt Publishing, Limited.

Information Commissioner’s Office. (n.d.). *UK GDPR guidance and*

*resources*. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/>

Langeland, J. (2023, August 14). *GDPR in the US: Compliance Simplified for*

*Businesses*. Termly. <https://termly.io/resources/articles/gdpr-in-the-us/>

Perteneder, F. (2021, April 29). *Understanding Black-Box ML Models with Explainable*

*AI*. Dynatrace LLC. <https://engineering.dynatrace.com/blog/understanding-black-box-ml-models-with-explainable-ai/>

Zia, T. (2023, July 6). *AI-Powered Personalization: How Machine Learning is Transforming*

*Customer Experience*. Techopedia. <https://www.techopedia.com/ai-powered-personalization-how-machine-learning-is-transforming-customer-experience>