**Project One White Paper**

**Gabrielle Maitland**

**CS 370: Current and Emerging Trends in Computer Science**

**Dr. Ramsey Kraya**

**July 28th, 2024**

As the lead engineer for our social media company, it is my duty to help identify discrepancies in our technological practices and to make recommendations on how we can improve. Because a regulator has brought to our attention possible violations of international laws, we will explore the foundation of our algorithmic personalization models, why they are used, and how we can comply with laws that we may be in violation of.

Neural networks are defined as a series of connected nodes that can pass a signal between one another. These nodes are called neurons, and they were inspired by the neurons observed in a biological brain. Neural networks works not only by passing this information to one node to another, but also by learning from the input and output it experiences (Narayanan, 2019). In neural networking, layers are used to define components that control input, output, and processing. The input layer describes the component that receives information from an external source. That data can be in the form of text, images, and other forms of information. That information is then passed onto the hidden layers of a neural network. Neural networks typically have one ore more hidden layers, which control the processing of data through advanced functions and mathematics. The output layer finally accumulates the processed data from the hidden layer(s) and produces the results for the user (Panagiotis, 2024). These three overarching layers work together as a whole to transform the original data fed to the model into something that the model can draw conclusions from, resulting in the growth of neural networking as we know it today.

Neural networks are used to aid in personalization of the user experience through tracking a user’s activity via clicks, links followed, content interacted with, and more. This data is then used to “recommend” similar content that the network believes the user will find interesting via advanced algorithms (Amazon, n.d). While these things may sound complicated on the surface, they work together to provide a user with an experience that is unique to them, igniting new interests, or revitalizing previous interests.

However, as with any practice, there may be ethical concerns. Some of these concerns include human bias affecting the way these networks operate and therefore the sort of content they contribute with personalization. Human bias is defined as an individual’s tendency to prefer, prioritize, or choose one option over another. In machine learning, there are 3 prominent types of bias: interaction bias, latent bias, and selection bias. Interaction bias can be seen when similar-minded users interact with a network, causing the network to have an easier time interacting with the bias that each user has in common, rather than an option that is common but unexperienced. Latent biases permeate due to historical data from eras in which many others did not have access to the same privileges as other, or from stereotypes that continue to exist in our society. Selection bias occurs when the data selected to train a model is not inclusive of everyone, or the data chosen is not accurately randomized due to their own interests (Google, 2017).

There are always hidden biases to take into account, particularly with a black box classification system. A black box classification system describes models in which the data, variables, and functions being used to produce predictions are being combined in a manner unknown to users (Rudin & Radin, 2019). This is due to the idea that since these systems have the ability to potential identify patterns that users might overlook, black box systems have a higher rate of accuracy than interpretable white box systems (Yasar & Wigmore, 2023). However, this can cause ethical concerns since users are completely blind to how these transformative functions are being carried out. This does not leave the user any room to observe on a detailed level how human bias may have crept into the model.

Citizens of the European Union (EU) are lucky to have such protective measures placed over their data. However, organizations like our own may face difficulty in creating the best experience for our customers due to these limitations. Certain portions of the General Data Protection Regulation (GDPR) affects personalization, such as data minimization, accuracy, purpose limitation, and storage limitation. Data minimization, which is the act of keeping the data that is collected limited to the purpose for why it was collected. Data minimization could affect personalization by forbearing the organization from potentially rolling out new features as soon as planned, as the Terms of Service and Privacy Terms would need to be updated beforehand. Accuracy is in regards to keeping the data collected as up to date without delay. Accuracy laws can cause additional strain on the data collection team, as they will have to continuously check for discrepancies in data and remedy them in a timely manner. Purpose limitation describes collecting data for explicitly laid out purposes, not allowing organizations to process the data in a way that is incompatible with the original purpose. Purpose limitation can be tricky for our organization if we want to try to rollout new features, as these features may not be considered previously outline. Storage limitations stop data from being held longer than originally promised, only being able to be stored for a longer period of time if used for statistics or or research purposes (Art. 5 GDPR – Principles Relating to Processing of Personal Data - General Data Protection Regulation (GDPR), 2021). Storage limitations can hold the organization back from holding on to data that may be beneficial for further personalization, causing us to have to re-collect information to improve the experience for the user. This can create distrust from the user if we constantly have to ask for more permission to source data or update our terms of service.

Potential legal concerns may arise from our company’s use of neural networks as a classifier to personalize the user experience. Some of these concerns include the Right to be Forgotten or defying the GDPR. The Right to be Forgotten describes the consumer’s right to request for their data to be deleted and for the holder to delete the data within a month. This would hinder the company from having the ability to use the data for marketing, or inviting a user back to the application. Legally, if the organization does not abide by these rights or the GDPR, we may be fined up to 10 million Euros, or 2% of our global turnover (Fines / Penalties - General Data Protection Regulation (GDPR), 2021). This would be a loss for any company, so we would need to follow the regulations to avoid the ire of the government. One option would be to no longer collect data, but that is not an option. Not collecting data will never be a possibility for this company’s business model in the foreseeable future, as we pride ourselves on having created the best social media experience that our end users have ever had. All of that is thanks to our users for trusting us with using their data to cultivate an environment unique to them. Without data collection, there would be no way to help our users connect to niche influencers, or discover new songs, or add a new destination to their bucket list.

Some of the current best practices for preserving privacy in AI/ML is by only collecting the data necessary to build our intelligence models. This limits having large amounts of data for no reason, and gives us the ability to discard the data when our model have effectively learned from it. Another best practice is to implement transparency in how the data is being used. This gives users a sense of control by having insight on how it affects the model, and how the model makes decisions based on their data (4 Ways to Preserve Privacy in Artificial Intelligence, 2021).

In order to comply with the GDPR, I would recommend that we base our Privacy practices off of their regulations if we intend to sell our product in the EU. This will save us the headache of having to revamp our policy in the future due to delinquency. Even if we are an American-based organization, we are still beholden to their laws if we have a European user base. Since these regulations also help citizens feel safer when agreeing to Terms of Use, it would benefit us to be as transparent as possible. I would also recommend that we appoint a Data Privacy Officer who specializes in compliance with data regulation anywhere we do business. Our reasons for collecting data should be airtight with a strong legal backing in coordination with our legal and compliance teams (CookieYes, 2024).

**References**

*4 Ways to preserve privacy in artificial Intelligence*. (2021, October 28). https://www.boozallen.com/s/solution/four-ways-to-preserve-privacy-in-ai.html

Amazon [AWS]. (n.d.). *What is a Neural Network? - Artificial Neural Network Explained - AWS*. Amazon Web Services, Inc. https://aws.amazon.com/what-is/neural-network/

*Art. 5 GDPR – Principles relating to processing of personal data - General Data Protection Regulation (GDPR)*. (2021, October 22). General Data Protection Regulation (GDPR). https://gdpr-info.eu/art-5-gdpr/

CookieYes. (2024, July 16). *GDPR in the US: A checklist for compliance*. https://www.cookieyes.com/blog/gdpr-in-the-us-a-checklist-for-compliance/

*Fines / penalties - General Data Protection Regulation (GDPR)*. (2021, October 22). General Data Protection Regulation (GDPR). https://gdpr-info.eu/issues/fines-penalties/

Google. (2017, August 25). *3 types of bias in AI | Machine learning* [Video]. YouTube. https://www.youtube.com/watch?v=59bMh59JQDo

Narayanan, R. (2021, December 28). Understanding Key terms in AI - DataDrivenInvestor. *Medium*. https://medium.datadriveninvestor.com/understanding-key-terms-in-ai-415baa8b37a1

Panagiotis, A. (2024, March). *Hidden Layers in a Neural Network*. Baeldung. https://www.baeldung.com/cs/hidden-layers-neural-network

Rudin, C., & Radin, J. (2019). Why are we using black box models in AI when we don’t need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Yasar, K., & Wigmore, I. (2023, March 17). *black box AI*. WhatIs. https://www.techtarget.com/whatis/definition/black-box-AI