Stephen Cardone

Professor Balogun

CS-370

January 31, 2021

Artificial Intelligence with GDPR Compliance.

In our social networking company, our business model relies heavily on personalization. We have found that an affective way to support this personalization is by using neural networks to do a large part of the heavy lifting. A EU regulator has brought to our attention that we might be violating some aspects of GDPR. The goal of this paper is to explain what neural networks are, how we use neural networks for personalization, as well as what GDPR is, how GDPR affects the use of neural networks in the context of our company, and what we need to do to be compliant.

Neural networks are an sub category of Artificial Intelligence. The Idea neural networks is to simulat how the biological brain works. In the biological brain, there are neurons that pass signals along to other neurons, and branch to other neurons. The passing of these signals results in the ability for complex thought. A neural network in artificial intelligence will use a similar approach. The basic structure of a neural network is that there are many layers each containing many neurons. The first of these layers is known as the Input Layer. The input layer serves as the beginning of the workflow for the neural network. The input layer will receive all the data that is passed into the neural network. The input layer will then pass the data along to the next layer in the network. Every layer after the input layer that is not the last layer is known as a hidden layer. A hidden layer of a neural network consists of many neurons. Each of these neurons has its own weight that is refined during network training. The weight indicates how significantly the neurons output will affect the data going forward. Additionally, each of these neurons looks at the data in its own way. There can be few or many hidden layers, but eventually, we will end up in the final layer which is known as the Output Layer. The output layer of the neural network is what actually spits out the result of the neural networks assessment. This result will be something that we use to make decisions on how we personalize data for a given user.

Now that we understand the basics of how a Neural Network works, lets discuss how we use neural networks to create personalization, and what concerns we have considered. The power of the neural network is that we do not need to explicitly tell it what to look at. We have found that this works exceptionally well when we feed a significant amount of user data into the network. When this large amount of data is used, we find that the neural network does an excellent job at predicting what a given user would like to see on their page. However, we need to consider the potential for biases. Since we do not explicitly tell the system what to consider and the networks behave more like a black box, it could be making decisions based on categories that would be considered inappropriate, such as race or gender. We need to look at the networks carefully to prevent bias such as this.

The GDPR has significant affects on how we are using user data to achieve our personalization success. There are many principals outlined in GDPR that affect us directly. Firstly, there is the idea of Transparency. Transparency is the idea that we need to make it clear how we are using user data. In the past we have not done a good job of this. To achieve this, we will need to further understand hour our neural networks are working so we can clearly relate to our users how the data is being used. Also, there is there is the principal of Purpose Limitation. This is the idea that data collected can only be used for the purpose we outlined and cannot be reused or used for other purposes. This is specifically important to us. Since we use user data to personalize their own experience, as well as go into our training data sets, we are not following this principal. Another principal is that of Data Minimization. This is especially relevant to us. This principal states that we can only gather data for which we have an intended purpose. We cannot gather as much data as possible. This is particularly relevant because we are always searching for ways to improve our personalization process. One way we do this is by taking in as much user data as possible and looking for new ways to improve personalization. With the principal of Data Minimization, we need to reconsider how we do this. The final principal we will discuss is the principal of Storage Limitation. Storage Limitation is also vitally important to our platform. Since our use of neural networks relies heavily on the ability to train with user data, we are constantly holding on to user data as data sets to be used for model training. With the principal of Storage Limitation, we will need to be very thought out and precise with how we manage training data.

When it comes to legal issues for how we are using data and neural networks for personalization, we will run into many legal issues if we don’t change our ways. At this time, we are collecting use data without any warning. In order to comply with GDPR we will need to outline how each of the data we acquire will be used and we will need to make this information publicly available. I would also advise that we make the user accept the terms of this data collection before allowing them to use our products. Additionally, we will need to change the way that we handle data once it is collected. At this time we are storing data indefinitely with the intention to use it for training our neural networks. At this time we could run into legal issues over this. In order to prevent legal issues, we have to possible options. Firstly, we could stop keeping user data for training purposes. We would then need to find a new way of training our data, whether that is from having some users opt in to allowing us to keep their data for an advantage in the service. Alternatively, we could use a live training method where the neural networks are not trained on stored data, but are trained off live data that we will not need to store. The possibility of not collecting user data at all does not fit with our business model. We rely on user data in order to achieve our edge in the market. Without user data, we will no longer have a viable business plan, so maintaining user data within the GDPR constraints will be necessary.

In order to comply with GDPR, I recommend we take the following actions. Firstly, we need to allocate a team to deeply study our neural networks to identify how we are really using data. This will allow us to outline how we are using data in accordance with the GDPR Transparency principal. Next, we will need to limit the storage of user data. We currently store data indfeinatly, however this goes against the GDPR guidelines. Therefore, we need to allocate a team to re-assess the neural network training data algorithm. We need to identify a way to train our networks without storing user data indefinatly. And finally, we need to clearly explain how we are using user data and have users agree to our GDPR compliant terms before they use our services. In this manor we will be able to maintain our position in the market, while being legally compliant with GDPR so that we and the users can see our services remain up and functional while being GDPR compliant.

Works Cited

Dorschel, A. (2020, May 08). Data Privacy in Machine Learning: A technical deep-dive.

Retrieved February 01, 2021, from <https://medium.com/luminovo/data-privacy-in>

machine-learning-a-technical-deep-dive-f7f0365b1d60

Guide to the UK General Data Protection Regulation (UK GDPR). (n.d.). Retrieved February 01,

2021, from <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the->

general-data-protection-regulation-gdpr/

How to develop Artificial Intelligence that is GDPR-friendly. (2019, June 27). Retrieved

February 01, 2021, from <https://techgdpr.com/blog/develop-artificial-intelligence-ai-gdpr->

friendly/