GDPR Compliance:

A Framework for MediaSocial

By Benjamin Houldridge

Introduction

MediaSocial uses neural networks to accomplish many tasks including content recommendations and advertising optimization. Neural networks are a type of computer model that consists of hidden, input, and output layers. For our content recommendation model, we format a user’s history and a new content option into data that can be given to the input layer. This first layer of the model is connected through weights to hidden layers. Each neuron multiplies data in the previous layer by the weights and then sums the incoming products. If we were to look at this data in real time it would likely not be intelligible. This process continues with many hidden layers until we come to the output layer. This layer gives us an understandable output. For the case of our content recommendation model, we get out a single number between 0 and 1 that scores how much we should recommend the content input. A 0 would be a piece of content the model does not recommend and a 1 would be full recommendation. When we do this with several pieces of content, we can rank them by the model’s outputs and use this ranking to serve content to the user.

In order for this model to be useful we needed to set weights between the neurons that provide correct outputs. We can do this by starting with random weights and a set of input data that has been labeled with what the correct output should be. We then train the model by running the input data through the model and scoring how close the output was to being correct. We then systematically go back through the weights and slightly change them so that if we ran the same data through the output would be closer to the correct answer. After doing this with enough data the weights will conform to providing accurate outputs. Thought the model will never be perfectly accurate it can come close enough to be extremely valuable to our company.

Personalization

While the model we have built has been trained on millions of individual users the output can be extremely good at selecting content that will be of interest to a single user. This is because the users history is fed into the model and this is all that is needed to make recommendations that will sustain the attention of the user. If we have a hypothetical user that likes and watches cooking videos to the end every night between 10pm and 11pm they may enjoy falling asleep to calm cooking videos. The model doesn’t need to know this to pick up on the pattern and suggest other calming videos like an aesthetic travel montage. This happens because the training data is full of similar users and their preferences are likely to have significant overlap. Much of our success comes from the phenomenon that people don’t know what they want to see until the model suggests it to them. This can however cause concerns as well. We got a complaint from a user that was allergic to shellfish that they continue to get cooking videos about shellfish because the model cannot be told that salmon recipes are significantly different from those about oysters to this particular user. The user doesn’t have control over the model and because it’s a “black box” system we don’t know why it makes the recommendations is does. This means an inherent lack of control in what is seen by the users. This lack of perfect control and the fact that the model does have an error rate means that we have had instances of inappropriate recommendations due to misclassification. Content involving smoking illegal substances has been misidentified as videos about tobacco and improperly pushed users. Because of this we have instituted flagging features allowing users to identify us with mistakes and have them corrected. We also intentionally biased our training data to reduce the misclassification of restricted content. This leads to benign content being classified as sensitive more often, but that is a tradeoff we are willing to make.

**GDPR**

Transparency:

Article 5.2 a data controller “must be able to demonstrate that personal data are processed in a transparent manner in relation to the data subject.” (“IAPP”) We are required to communicate with our users what data we are collecting and for what purposes. They are to be notified on any breach or loss of their data. This all needs to be communicated in a way that is free, and understandable to the users. (“IAPP”)

Purpose Limitation:

Article 5.1 “collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes” We are required to only use data for those purposes outlined in section 5.1 and this data can’t be used or saved for a later undisclosed use. (GDPR)

Data Minimization:

Article 5.1 “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed” We are required to keep the data we collect and store to a minimum. If there is something extraneous we are required to get rid of that data. (ICO)

Accuracy:

Article 5.1 “accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay” We are required to keep data accurate and provide a process to check for, identify, and correct inaccuracies. (ICO)

Storage Limitation:

Article 5.1 “Personal data shall be: kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed”. We are required to understand the data we hold and delete user information when no longer needed including the deletion of user information upon request. (ICO)

Practices

In our practices we have been attempted to maintain not only in compliance with the General Data Protection Regulation but also it’s spirit. We have been transparent with our purpose to provide the most personalized content to users and help advertisers reach their potential customers. We only collect and store data for the purposes that we outline for our users. Though we collect large amounts of data like user history, we have a process by which all new data streams are considered for necessity. We are a fast pace and set limitations on how long we store data. We strive for accurate information not only for regulation but the success of our models and user experience.

Adaptations and Compliance

We have tried to keep up with regulations but can definitely update some policies to maintain compliance. Our user base has grown from it’s originally domestic demographic to all over the world. Out transparency has been text only and in English. With a global userbase we need to expand how we notify our users about what information we collect in there own language and take into account disabilities such as the sight impairment. Some of the data collected is not eventually used by the training models. Our engineers use feature selection methods that sometimes will dismiss data like clickthrough and when this happens we need a process in place to delete this unneeded data.

Sources:

GDPR. “General Data Protection Regulation (GDPR).” *GDPR*, 2025, gdpr-info.eu/.

“IAPP.” *Iapp.org*, 2025, iapp.org/news/a/transparency-and-the-gdpr-practical-guidance-and-interpretive-assistance-from-the-article-29-working-party/. Accessed 4 Aug. 2025.

‌ ICO. “A Guide to the Data Protection Principles.” *Information Commissioner’s Office*, 19 May 2023, ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/.

‌