PRIVACY ASPECTS OF
FEDERATED LEARNING

Legal Review by XayNet & CMS

XayNet

C/M/S/
Law.Tax

A. Brief introduction to Federated Learning

„Machine Learning" describes a process in which an algorithm uses data to train a model (**„Client Model"**). This model may then be used by XayNets **„Clients"** to make decisions or predictions which are not programmed in the algorithm itself, but which are solely based on the model. A Machine Learning model is typically only a vector of parameters of a mathematical function which provides a result y for given value x.

Example: *A very simple model might look as follows:* $(x1, x2) = (1.45, 5.34)$*. This simple model could be used in the algorithm* $y = 1.45x1 + 5.34x2$*, which „predicts" the result* y *for a given* $x = (x1, x2)$*.*

In practice, Machine Learning models consist of hundreds of thousands of parameters.

In **„Federated Learning"**, a company running a parameter server is collecting the trained Client Models from various clients, aggregates such Client Models to one aggregated, **„Global Model"** and sends such Global Model back to the clients. The clients then use this Global Model to train it with their own Client Data, and the process begins again.

The below data flow graphic distinguishes between the different roles: Client is collecting Client Data (1) and is training with such Client Data, either on its own (2a) or with XayNet as its processor (2b). XayNet is aggregating all Client Models to one Global Model (3) and Client is using the Local Modal for its own prediction purposes (4).

Because the quality of the Client Model is decisive for the quality of the prediction and decisions made by Client on the basis of the Client Model, every Client has an interest in aggregating the Global Model with its Client Model to get better results. Since the Global Model is dependent on Clients providing their Client Models, and, because of this, since every Client is committing to the Global Model, the entire procedure is called Federated Learning (FL).

The scope of this paper is to outline how Client Data is anonymized during FL.

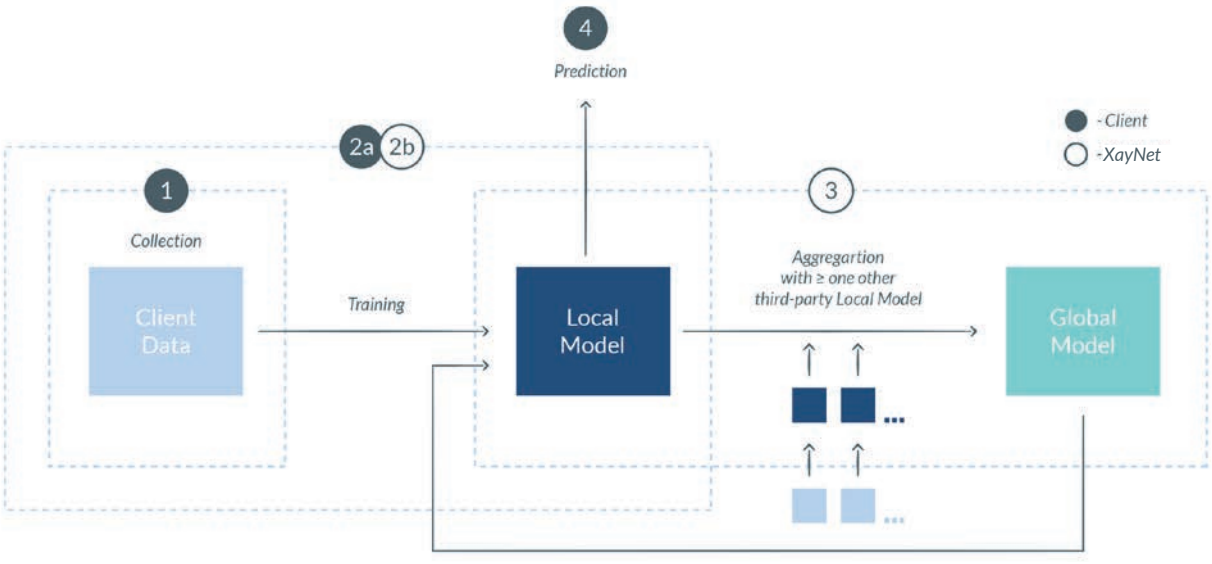

Figure 1: Data Flow in Federated Learning scenarios

B. Requirements for privacy and the protection of data

The law knows many forms of protection of data and information. Whereas the GDPR protects personal data of individuals (I.), there are also provisions on the protection of business trade secrets (II.) or confidential information in general (III.). Companies must consider all of those when using machine learning on potentially protected data.

Our aim is to enable the use of machine learning with protected data, but to ensure that the protection is kept in place.

I. PERSONAL DATA

„Personal Data" is, according to Art. 4 (1) GDPR, any information relating to an identified or identifiable natural person. Data, which is solely related to machines or companies, for example, is not considered Personal Data in terms of the GDPR but may of course nevertheless be subject to confidentiality obligations or qualified as trade secrets.

In a GDPR context, data is anonymous and therefore not Personal Data if modified in such a way that data subjects can no longer be identified. According to recital 26 of the GDPR, to determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. Further, to ascertain whether means are reasonably likely to be used to identify the natural person, all objective factors, such as the costs of and the amount of time required for identification, shall be assessed, taking into consideration the available technology at the time of the processing and technological developments (see recital 26 GDPR).

This means that no absolute anonymity is required for data to be considered anonymous. When assessing the likelihood of so-called re-identification attacks, one should not only consider the effort necessary for completing such re-identification attacks, but also the criticality and sensitivity of the data targeted by the attacks.

Example: *Since XayNet together with its Clients strives to remove sensitive/critical data at the beginning of the process (see below, C.I), one may assume that the likelihood for a third-party to attack such model is much lower, given the amount of time and effort necessary for any viable attack. This means that the threshold for data to be considered anonymous is also lower.*

II. TRADE SECRETS

As regards trade secrets, one may consider Art. 2 (1) lit. c) Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure: Trade secrets must be subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret. This does not require a perfect protection of confidential information (which would not be possible in practice anyway) but the application of reasonable means, taking into consideration the criticality of the information to protect. Those means typically include technical (e.g., encryption), organizational (e.g., access rules) and legal (e.g., contractual penalties) measures. The surprising consequence: If no adequate measures are applied, confidential information does not benefit from the protection of the EU trade secret directive.

III. CONTRACTUAL CONFIDENTIALITY OBLIGATIONS

In confidentiality obligations and non-disclosure agreements, one usually finds a wording similar to „appropriate precautions", „adequate means", „keep confidential" or „duties of a prudent businessperson". Those wordings typically shall not ensure an absolute confiden-

tiality (which is practically impossible, at least as soon as human beings are involved) but use undefined legal terms (*unbestimmte Rechtsbegriffe*) which are subject to an interpretation according to the individual case.

IV. SUMMARY

In essence, the aforementioned legal obligations are united by the obligation to apply „reasonable" means which certainly have to consider the concrete circumstances and the criticality of the information to protect. A "perfect protection" is, of course, typically required. Nevertheless, XayNet's Federated Learning solution strives to reach the latter.

C. Privacy behind FedML

Federated Learning consists of several techniques which aim at ensuring that Client Data is not made accessible to third parties, including other Clients of XayNet. In the following, we describe techniques which are used in Federated Learning scenarios to „anonymize" Client Data in a way that their substance and content is hidden from others, but they can nevertheless form the basis for the training of machine learning models. We are convinced that some of those techniques described in the following are already sufficient so that inferences about individual Client Data are reasonably not possible anymore, and this must apply even more if all techniques are applied cumulatively, which is the case with Federated Learning provided by XayNet.

I. COLLECTION OF CLIENT DATA

As one of the first steps, XayNet together with its Clients strives to limit the amount of Client Data that is processed in a FL algorithm to those which are strictly necessary for the Machine Learning to function. During such a process, much Personal Data is typically removed because data such as names or social security numbers are, due to their randomness, usually not suitable for Machine Learning models - which seek to identify patterns.

Example: *If a model shall be created which allows predictions about the number of staff being ill at a certain point in time during the year, then the names of the employees are not necessary to train such a model.*

II. TRAINING

As regards the remaining data which is used as Client Data, one has to bear in mind that such data is not stored in a Client Model in plain form, but, since it is model data after having been trained, stored by means of a parameter vector (see above, A, for an example). This parameter vector does not directly provide any information about the data which has been the basis for the training.

Example: *If the Machine Learning algorithm based on the Local Model is, put simply,* $y = 1.45x1 + 5.34x2$*, then there is no direct possibility to calculate from the value of* y *and the learned constants 1.45 and 5.34 part of the Client Data which has been used to learn that model.*

III. AGGREGATION TO LOCAL MODEL AND GLOBAL MODEL

Aggregation is an anonymization technique. The more data is used for aggregation, the more secure it becomes. XayNet's FL algorithms usually use at least three Clients.

Example: *Since the age of three persons with 32, 34 and 38 years is aggregated to 34.67, this aggregate data cannot be used to single-out an individual from the Source Data. It's just impossible.*

There are two aggregations taking place in FL:

- First, during the training aggregation is done based on the Global Modal from the previous iteration to get the Client Model as an outcome. This is an aggregation since new training results are aggregated with the Global Model which existed already before the training took place.

- Second, all Client Models are aggregated to get the Global Model. Although the Client Model consists already of aggregated data, this second aggregation is effectuated in a way that prevents the aggregator (typically a third party or XayNet) from accessing the Client Models, which dramatically increases the security and thus the efforts to make inferences about Client Data. This is achieved in two cumulative steps:

  - 1) The Clients only share, with the aggregator, Client Models which have been masked before sharing with a random mask. This makes it mathematically impossible to decrypt the Client Model from the masked Client Model alone without the masking key.

    Example: *If the masked Client Model is known (here: by the aggregator), there are endless possible combinations of Client Model and mask. In other words: If 5 = a + b, there are endless possible combinations of a and b.*

  - 2) Clients do not share their masks with the aggregator, but only with other Clients, encrypted using the latter's public key. Those other Clients decrypt the masks with their private key and share the sum of all decrypted masks of all Clients with the aggregator. The aggregator subtracts that sum from the sum of masked Client Models received from the Clients in step 1 above. This produces the new Global Model, aggregated from various Client Models, without ever granting the aggregator access to individual Client Models.

This means: In the first step, each Client accesses and uses its own Client Data. In the second step, no Client Data at all is being used, even the Client Model cannot be accessed by the aggregator.

IV. CONTRACTUAL SAFEGUARDS

Apart from the more technical safeguards described above, XayNet has and will put in place various technical, organizational and legal measures to further minimize risks in the extremely (!) unlikely event that a person could make relevant inferences from model data about individuals.

Those measures include,
but are not limited to:

- If the Client Data is critical or sensitive, XayNet will not pass the models to third parties which are not contractually bound, and XayNet will not make the models public, in order to minimize the number of persons having access to the models. To increase the level of security even more, the models could be hosted in a cloud environment instead of providing the model itself, only giving third parties access via an API (black box) and not directly (white box). Based on this, direct access other than via an API is technically impossible, which makes attacks very difficult, not to say impossible.

  Example: *Most attacks on machine learning models require direct access to the models in order to analyze their structure and to make inferences. Attacks are no longer possible or extremely limited if attackers do not have direct access, but only use an API for black-box attacks which only permits simple read requests, but no analyses of box internals.*

- XayNet will not unreasonably provide other Clients or third parties with the information on which Clients have contributed to the models, nor in which way they have done that. Such information would, however, be necessary to facilitate attacks.

  Example: *Even in the extremely unlikely scenario that an attacker could get the information that the number of staff being ill in a company at a certain point in time is 50 %, then the attacker would still not know to which company this single data point applies.*

- XayNet concludes agreements with its Clients which shall ensure the confidentiality of the models and which shall forbid Clients to access the models other than with XayNet's software, which means that, legally, Clients are not allowed to make any attacks on the models.

D. Summary

The Global Models used in XayNet's Federated Learning Platform are, in our view, anonymous in the sense of the law.

PLEASE NOTE:

Our intention with this paper is to inform the public about privacy techniques applied in Federated Learning services provided by Xayn AG. This paper pursues a general approach, does not consider any individual Client case, reflects our personal legal view on FL and is certainly not to be considered as legal advice. Nothing in this document shall be construed as a warranty of any kind. If you have any questions regarding Federated Learning services provided by Xayn AG, please write us an e-mail: info@xayn.com.