Privacy Enhanced, Accuracy Preserved, Federated Learning AI Model to Identify Harmful or Dangerous Acts Involving Minors in Videos Contents

1. Introduction.

Child safety has become a major issue in the internet as the physical & mental state of minors are endangered nowadays with the largely spread online digital platforms. In most of the regions and countries, someone who is under 18 years which is considered as the legal age of maturity, is defined as a minor [17]. Cyber space is a great place for the children or the kids in terms of their education and entertainment. Having said that, now there is a huge platform for the people to create and publish their content in the internet while its being used unethically in certain cases for monetary purposes. Child safety in the internet may have to face with diversified challenges in such situations.

Videos containing dangerous activities such as dares, pranks or dangerous stunts involving children are now also widely spread in YouTube and internet in order to gain more traction of traffic to the channels which may lead to serious injuries and harmful situations. In the other side, this kind of visual content encourages other children who would watch it to try out these activities at home.

Since, current content moderation in this research context is manual with no significant automation. This research proposal focuses on developing a Privacy enhanced, accuracy preserved AI model to automatically identify harmful or dangerous acts involving minors in the videos to ensure the child safety in the cyber space. Proposed model is trained in a federated learning environment preserving privacy while handling the bandwidth limitations, collecting video contents with range of variations and naturally inferring labels for training by giving the space to integrate with existing content reporting processes. This research, also propose developing efficient client and server optimization technique utilizing range of client and server optimizations in different combinations to handle the tradeoff between privacy and model quality. Novel update aggregation method by composing different techniques to preserve the security and privacy in update aggregation is also presented.

2. Background/Literature

Current content moderation in YouTube to remove the videos with harmful acts involving minors is manual and not automated yet [18]. Situation is same with most of the other online video streaming platforms and websites. There are privacy preserving issues, bandwidth limitations in training the machine learning models with video contents in a centralized server. Achieve the expected accuracy is also challengeable because large amount of video data with range of variations is needed. Labeling the video contents naturally in a distributed environment is also important to consider [16].

In 2021 Xiong Li et al developed a methodology to protect privacy in input and inference pipelines using a modified version of the Okamoto-Uchiyama homomorphic encryption mechanism. In this research, machine learning model predictions became verifiable against a schema [1]. Even though the federated learning comes with a mechanism to provide some sort of privacy and security to the existing machine learning model training, by sharing the models updates in the means of gradient information without sharing the direct raw data, sharing model updates in this means also can reveal sensitive information. Existing privacy preserving algorithms are challenged in the federated learning settings. Privacy preserving should also in lined with computational and communicational effectiveness without compromising the accuracy a lot. Federated learning privacy can be considered as global and local privacy. Global privacy is concerned about models' updates produced at each round of training are only shared with the server and all other third parties are restricted of access while the local privacy is concerned about restricting the access even to the servers. This specific research is built on the SMC (Secure Multiparty Computation) and differential privacy like cryptographic protocols with a secure aggregation method to protect the model updates in each iteration. Secure aggregation guarantees the privacy while retaining the original accuracy of the model. However, the proposed method comes with significant communication cost. Other related work has applied global differential privacy considering the communication and model accuracy with carefully selected, hyperparameters. Intermediate solution in between the global and local privacy is presented in the context of meta learning with private differential algorithms alongside combined with the model compression techniques to support privacy preserving and communication at the same time [2].

Shahab Asoodeh et al proposed a theoretical approach to identify the federated learning differential privacy parameters. In this approach only last update from the stochastic decent gradient algorithm is shared. Each training iteration is interpreted using the Markov kernel. Impact of the kernel is quantified by the contraction coefficient of the $E\gamma$ -divergence which ultimately describes the differential privacy. Dobrushin's ergodicity coefficient is generalized in the background to come up with this approach. Convergence rate of the stochastic gradient algorithm is examined to analyze the proposed private federated learning approach [3]. In 2021 Venkatraman Balasubramanian et al introduced a federated learning algorithm called FedCo to predict the user demand for a specific content. Theorical evaluation of user demand behavior is conducted for content management in Edge Data Centers and Mobile Data Centers. [4]. “Privacy-Enhanced Federated Learning against Poisoning Adversaries” research has taken an effort to detect the poisoning behaviors under cipher text in federated learning. PEFL - privacy-enhanced Federated Learning framework can also provide the security against the label flipping and backdoor attacks [5]. Badih Ghazi et al developed and tested a federated learning algorithm to preserve the adversarial robustness with differential privacy simultaneously in the context of learning half spaces for a large number of parameters [6] [7]. Some machine learning models store the training data, inadvertently and implicitly therefore sensitive information can be exposed through a careful model analysis. Nicolas Papernot et al, proposed a combined approach of training disjoint data sets in a Blackbox called PATE (Private Aggregation of Teacher Ensembles). Here models are not published. A particular student model uses these models as teacher models. Teacher is selected through a voting system to learn how to predict. Individual teacher, data or the model parameters

cannot be directly accessed [8]. K. A. Bonawitz et al, proposed a method to train machine learning models using federated learning techniques along with distributed, gradient and secure aggregation. Efficiency and robustness requirements are also taken in to their consideration [9]. “A Novel Server-side Aggregation Strategy for Federated Learning in Non-IID situations” Research has observed, federated averaging extremely declines in Non-Independent and Identical situations and proposed a novel approach of accuracy-based averaging for server-side aggregation [10].

Research Gap

YouTube & most of the online streaming platforms depend on their community member’s reports to find out the inappropriate video content. Tracking such content is not automated yet and can be reported or flagged as such, manually [18]. Current content moderation in this research context is manual with no significant automation.

These videos, containing the child imagery are very sensitive data and it' s high time to avoid the associated privacy, security and processing risks when dealing with datasets [19]. Transferring this data to a remote server is associated with privacy risks. User information or child information can be vulnerable to middle man attacks. Also, compared to the model, video contents are large in size and uploading such data to a model in a remote server is not practical with bandwidth limitations. In order to correctly identify the videos with harmful acts involving minors, models have to be trained with large variety of data. Finding such variety in a centralized location is very hard when training the model. There can be videos with different variations shared across the globe among different users. Having a mechanism to train the model, using a globally dispersed data set with significant variations, directly impacts on the model quality. Existing manual reporting process in the current video streaming platforms can be integrated with the proposed federated learning set up to naturally infer the labels for the supervised learning in this research context to train the model with new harmful video contents.

Introducing federated learning [16], to handle the privacy issues lead to reduce the quality or the accuracy of the machine learning model. Current systems have mostly experimented only on the Stochastic Gradient Decent algorithm as both the client and server optimization. Range of optimization algorithm combinations on clients and server can be applied in this research context. Making wrong inferences on content moderation can lead to removing the current contents with any control or not identifying the harmful contents properly.

There is an associated privacy and security risk in aggregating the client updates to the global model shared in the server. Client updates are vulnerable to attacks, in the middle. Identified special patterns in the videos with harmful contents are exposed to third parties and can be modified with the intention of avoiding recognizing those patterns for content moderation in this research context. And also, the video contents can be optimized to avoid extracting such patterns for learning in the future to be safe from getting removed or reported on the online platforms. There is a space to develop Novel Aggregation Method by composing different techniques to preserve the security and privacy in update aggregation.

3. Objectives and expected results.

Aim of this research is to Develop AI Model to automatically Identify the Harmful or Dangerous Acts Involving Minors in the Video contents, while handling the tradeoff between data privacy and model quality.

Objectives of this research are listed below.

- Develop AI Model to Identify Harmful or Dangerous Acts Involving Minors in the Video contents
- Develop Federated Learning set up to enable privacy in machine learning model training for this research context
- Enable efficient Client and Server optimization technique for this federated research context
- Develop Novel Aggregation Method by composing different techniques to preserve the security and privacy in update aggregation

This research expects to have a deep learning model to automatically Identify Harmful or Dangerous Acts Involving Minors in the video contents while preserving the model quality and privacy of data in the training pipelines in a federated learning environment. Online video streaming platforms will be able to use this model to automatically identify & remove the inappropriate content in the shared videos without having to manually review and remove them while having the space for improving the model by further training as well. Bandwidth limitations in model training will also be addressed.

4. Research methods and materials.

Develop AI Model to Identify Harmful or Dangerous Acts Involving Minors in the Video contents

Machine learning models should be able to recognize the objects in the video frame by frame in order to extract the intelligence from it. Video clips are labelled or tagged through an annotation process, so that the computer vision machine learning models can learn the frames accordingly. Since these are very sensitive data including the child footages, these data must to be subjected to a serialization or encryption before they are stored.

Ordered sequence of frames make a video content. Temporal information can be found in the sequence of frames while spatial information can be found in each frame. In this research, CNN-RNN hybrid architecture will be used to model both the aspects. Convolutional Neural Network (CNN) is supposed to use for spatial information processing while a Recurrent Neural Network (RNN) with GRU layers is supposed to use for temporal information processing. Inception ConvNet extracting features followed by a single layer LSTM RNN machine learning model is proposed for this research. CNN is a type of multilayer perceptron. This kind of deep learning architecture allows the models to understand the complex features in the video contents that are necessary to classify the contents. Basic CNN model architecture with convolutional,

pooling and fully connected (FC) layers which will be used for this research is depicted below. To determine the feature maps, convolutional layer will have set of kernels. To reduce the spatial size in terms of number of parameters and computations pooling layer performs down sampling. Activation functions increase the non-linearity in the feature maps. Fully connected layers take the classification decision based on the information given from the initial layers. LSTM (Long Short-Term Memory) is an extension to RNNs which uses memory blocks with 3 gates namely input, output and forget gates. Cell state is used to remember long term dependencies contrast to the traditional RNNs. LSTM will act as the classifier. In this study, a combination of CNN and extension of the RNN (LSTM) will be used to automatically detect the dangerous acts in the video contents in the internet.

Number of hyperparameters will be selected as a starting point. Fine tuning the hyper parameters results in a better model which is specific to this research. Number of layers in the model, number of units per layer, dropout rate, learning rate will be some of them in the interest. Balancing the number of layers will help to control the complexity of the model which will lead to avoid model underfitting if the number is higher. Number of units should be handled in a way with the choice of contracting or extracting the features minimizing the loss of information.0.2-0.5 is the recommended dropout range. Learning rate should also be balanced. Higher rates are not recommended while starting from lower rate and finding the balance will be the best approach

Organize the data set with relevant classes first before placing them in the simulated federated training environment. Iterate through each and every video in the class and extract the frames to add them in to the list of features of a particular class. Prepare all the videos in the data set as features and labels in the form of NumPy arrays hence the machine learning models can only deal with numerical values. Sequence of video frames can be extracted and save in a 3D tensor. Depending on the length of the video, number of frames per video is different from video to video. Padding has to be applied to the videos incase if the content is supposed to stack in the batches. Frames will be padded or trimmed based on the average frame rate decided on the data analysis stage. Data set will need range of variations in objects and actions across frames in order to generalize the model to classify any other unseen action. Final Model which is fine-tuned will be served with docker using the GPU. Prediction pipeline with same preprocessing steps will be implemented for the real time inference.

Processing and training video data consumes a lot of time and computational power. Model is optimized to use GPU's if available in client devices. This can be speed up by utilizing the GPUs or TPUs. But inefficient input pipelines can act as a I/O bottleneck in this kind of scenarios. Data should be delivered efficiently to the next training/preprocessing step before the current iteration has been finished. Input pipelines are preferred to be more flexible and efficient. Video file should be opened if it not opened, to fetch the data and use them in the training process. If the input pipeline is not optimized, and synchronous, model has to stay idle while the data is being fetched. On the other side, while the model is being trained, input pipeline has to stay idle. *Prefetching* will be applied accordingly to read the data for the n+1th step while the training nth step is being done. This transformation uses an internal buffer and a background thread to prefetch the data before it is requested. Batch size defined for a single training iteration impacts the no of prefetching

elements. Raw video data might need additional computation to deserialize or decrypt them once they are loaded in to the memory. This overhead can be worse, if the prefetching is not performed effectively. *Data loading can be parallelized*, to mitigate the overheads in data extraction and preprocessing. Data preprocessing can be parallelized across multiple GPU cores. Since the video frame data is dependent on each other, both the data loading and preprocessing parallelization, should be carefully done across the classes with the complete batch of data belongs to a particular class. Vectorizing the loading and preprocessing functions by having them operate over a batch of data will work smartly on this kind of scenarios. Cost of these parallelization depend on the other tasks happening on the compute units in the meantime. Operations like opening the data files and reading the content from them are very costly if they happen at each epoch. So, the data set can be cached either in the local storage or in memory. In every transformation, reducing the memory foot print should be taken seriously.

Develop Federated Learning set up for the machine learning model to train the machine learning models locally in a distributed manner to identify the videos with harmful acts involving minors.

In this research, a scenario will be simulated to get data from multiple users, and each of the user will label and feed the videos with harmful acts involving minors locally to the machine learning models without sharing the sensitive data. Dockerized environments will be used to simulate training in multiple user machines. Client model updates will be aggregated to update the global model in the server.

Very large amounts of user devices are there in a normal federated learning environment while only a part of them is available for training at a particular time where the device is plugged on, connected to a un metered network or idle. In this simulated research environment, data will be divided among clients randomly and they are available locally. Users will be randomly sampled for the training rounds, in order to select different users or training in each round. As per to the research “Communication-Efficient Learning of Deep Networks from Decentralized Data” published by H. Brendan McMahan et al [13], it takes considerable time to achieve the model convergence with randomly sampled clients in each round and running hundreds of rounds would also be hard.

Data collection plays the most important role in solving supervised machine learning problems. Open-source videos containing dangerous activities such as dares, pranks or dangerous stunts involving children are supposed to collect from the internet. Theses sensitive data that includes the children, will be maintained under stringent and compliant security protocols. Data security will be covered by keeping the firewalls and antivirus software's up to date. All the data stores will also be physically isolated. In order to generalize the machine learning model better, more training samples have to be collected. Classes relevant to harmful and dangerous acts involving minors should be identified beforehand. Each class should maintain a balance in-between the samples to avoid them being overly imbalanced. Variations with all possibilities must be covered in the samples as much as possible, not limiting only to the common variations. This requirement is supposed to fulfilled by setting up a federated learning environment.

Data is usually heterogeneous in federated learning environments. In this research context, different variations of videos including harmful acts involving minors will be labeled and trained in distributed user environments satisfying the requirement of non-i.i.d. nature of federated data in a different way. It's very difficult to find a data set with large range of variations locally to train a model like this. In an environment, where the distributed local training is possible, users dispersed in different geographical locations can report and label the videos to locally train the models. Local training will train the model in different directions with different data variations. By enabling distributed training, locally the model gets exposed to large number of training data. The chance of getting trained on such content is increased. Also, the videos can be labeled or reported incorrectly and can be noisy.

Training video data set with their associated labels will be passed to this process. Based on the current state of the model, predictions will be made and weights are updated through backpropagation algorithm to minimize the error until the model is converged. Metric, Loss Function and the Optimizer are the three main parameters considered for this process. For this study accuracy, sparse_categorical_crossentropy and Adam will be used respectively as a baseline. Learning rate, Epochs, Batch size and Early stopping callbacks will be also set as training hyperparameters. Federated learning mechanism will be implemented in model training to ensure data privacy and security. On device inferencing will also be supported.

Training a machine learning model dealing with video contents is very resource intensive involving very complex tasks like matrix multiplications. Training will be accelerated with GPUs which can process multiple computations simultaneously. Machine leaning models are designed in way, processing can use GPU's if a GPU is available on client devices to speed up the training process. Otherwise, the model will use the existing client CPU's. GPUs memory bandwidths fit in with handling the large amounts of video contents.

After some training iterations, validation accuracy will start decreasing while the training accuracy is still increasing. That means the model can be overfitted to the training data. But our model should perform well on the test or unseen data in order to be a good generalized model. Strickling a balance in number of training epochs will stop the model from learning unwanted patterns from the training data where it will limit the possibilities of generalizing the model. Using complete training data as much as possible will prevent the problem of model overfitting. Full range of variations should be covered for the model to be naturally generalized. When the data set is not that rich enough, weight regularization techniques can be applied. Apart from that drop out functions can be applied in the model design and early stopping callbacks can be applied in the training stage. These techniques constrain the information, model can store. To develop a more generalized model, optimization algorithms should be there in place to focus on most prominent patterns if the model can only memorize few patterns. Machine learning model should have a strong and powerful architecture and should not be overly trained or regularized to avoid model underfitting. In such cases model may not learn the necessary patterns from the training data.

Training the machine learning model on video contents take a very long time. So, starting the training from the scratch when there is a failure would be very painful. Incase if the training process is interrupted due to a system failure or any other reason, training should be able to start

from the place where it was left off without having to retrain. Weighted model checkpoint callbacks will be set to save the model's best weights at each epoch if there is an improvement of the model accuracy. Complete model also can be saved along with the model architecture, weight values, training configurations and optimizer state in SavedModel format or HDF5 format. This research model will use SavedModel format to *serialize* the model to ensure model security. Machine learning models saved in these formats can be restored or loaded to reuse of the model.

In the federated averaging algorithm, there will be two optimization functions, namely client optimizer and server optimizer. Client optimizing algorithm will do the local model computations on each distributed client and the server optimizing algorithm will apply the averaged client updates to the server model. Training will start with regular stochastic gradient algorithm with a small learning rate which is not optimized. Learning rate will be experimentally finetuned for this context during the research.

To complete the training round faster, client to server communication will be optimized using a lossy and lossless compression combination. uniform quantization methos proposed by Suresh et al. [15] will be used. quantization_bits and threshold and number of clients in training is supposed to adjust for an effective compression. Threshold will be changed accordingly since the bias values in the output layers of classification models like this are more sensitive to the noise. Number of quantization bits will be experimented by decreasing the value while checking the model accuracy/quality. If model degrades, with reduction increase the value and find a balance. Higher quantization reduces the model quality. Noise introduced through quantization will be considered to be evened out by increasing the clients per training round in the simulated environment.

Federated Learning Architectural steps for content moderation in this research is listed below.

1. Generic shared model is trained in the server with the content moderation training data from clients.
2. Shared model is downloaded and trained on selected number of distributed clients with heterogeneous videos with harmful acts involving minors locally.
3. Updated weights of the locally trained models on the content moderation deep learning neural network are sent to the server by each client preserving the privacy.
4. Client's optimized weight updates are aggregated at the server to improve the shared global model through a Federated Averaging Algorithm.
5. Sending general model to the local clients and sending back the summary of optimized model weights will be repeated.

Since the machine learning models are trained locally with in the clients and only the model weights are transmitted to the server, raw sensitive data is not exposed to any threat.

Enable efficient Client and Server optimization technique for this federated research context

Depending on the desired customization level, research on the optimization techniques in this federated learning environment will be conducted in different ways. This research will experiment on sophisticated optimization techniques considering different learning rates and different optimization algorithms on both the clients and server with several combinations. Previous federated learning researches has mostly used Stochastic Gradient Decent algorithm. This research will deploy several optimization algorithms like RMSProp, Adam, AdaDelta, AdaGrad, AdaMax, Nadam and Ftrl in different combinations along with an experimental evaluation of other associated parameters to find and enable the most efficient optimization technique for this specific content moderation.

Develop Novel Aggregation Method by composing different techniques to preserve the security and privacy in update aggregation

Most of the existing researches have considered differential privacy or single isolated aggregation technique in federated learning for the update aggregation. This research proposes to extend the default mean aggregation by different other aggregation techniques to preserve security and privacy issues in upgrade aggregation. Extension is done by composing. Different aggregation techniques like zeroing, clipping, differential privacy and secure aggregation will be utilized interchangeably to full fill the requirements.

Quantile Matching, Differential Privacy and Secure Aggregation, Quantile Matching techniques together will be composed together to experiment and develop composed aggregation method to improve the privacy in existing federated learning environments in this research context. It's better to adapt a norm bound during the training rather than using a fixed bound. Quantile matching which is proposed by Andrew et al [14]is the recommended way to decide the norm bound while its compatible with differential privacy. Learning rate will be increased to quickly adapt to the correct quantile but with higher variance. Quantile matching will be implemented in a way, noise is added to the differential privacy to preserve the model accuracy in the meantime. This research will address the tradeoff between utility and privacy or the difficulties in training a model with higher accuracy comparing to the training a non-private model. Differential privacy will be implemented by using adaptive clipping and gaussian noise. Differential Privacy Stochastic Gradient Decent algorithm Abadi et al., "Deep Learning with Differential Privacy" [11] is supposed to use in the context of Private Recurrent Models for content moderation. Privacy leakage of sensitive data in training the models can be bounded and quantified using differential privacy. This approach mitigates the risk of exposing sensitive data. Models are not learning anything significant about the users but the patterns exist different client data. Federated averaging algorithm is defined with client and server optimizers along with a model update aggregation method. Basic federated averaging algorithm will be changed in two ways to guarantee the user level differential privacy. Model updates from the clients are supposed to clipped before sending to the server to bound a client's maximum influence, first. Enough noise should be added to the sum of client updates before averaging to reduce the worst-case client influence. To clip the updates, adaptive clipping method [12] will be used. Here fixed clipping norm won't be set explicitly. Adding noise, reduces the model utility. So, the amount of the noise in the average

update at the end of each round will be controlled. The proposed strategy will identify, how much noise, the model can take in with a smaller number of clients per round with an acceptable model utility. Amount of noise can be increased, proportionally to the number of clients. Since this research has a large data set to support many clients per one iteration this is feasible to test. Series of models will be trained with N clients per round, in the first place increasing the noise amount. More specifically, ratio of the noise standard deviation to the clipping norm will be increased. Actual magnitude of the noise is supposed to change from round to round since this research uses adaptive clipping. With same number of clients, and same amount of training iterations more accuracy can be achieved with higher noise. Higher the noise, lesser the model quality is. To achieve the targeted privacy level with less noise and high accuracy model will be trained with more clients per round. Poisson sub sampling will be applied to provide tighter privacy guarantee. Clients per round is stochastic with mean clients per round. Federated averaging process will be built with non-adaptive server optimizer since noise can cause very large momentum accumulation.

Cryptographic protocol called Secure Aggregation, will be used to encrypt the client updates in a way the server has to decrypt their sum. But the experiment is supposed to managed in a way sufficient number of clients reports back in the simulated environment since with insufficient number of clients server model does not learn anything. Secure Aggregation protocol will operate only on integers. But the client model updates are in the form of floating-point numbers. Larger values will be clipped and bound for an integer. Clipping bound will be adaptively determined. Sum of the integers will be mapped back to the floating points. Precision won't be affected much in discretization since adaptive clipping is used. Parameters will be tuned, considering the fact that the secure aggregation is adding the model weight updates after averaging.

Data can be compromised and corrupted in the client environments and it affects the final model accuracy. Clipping and zeroing will be composed together in the aggregation algorithm accordingly to address this issue. Zeroing aggregation technique will be in place to replace the values larger than a defined threshold or relative to the previous round values by zero. Adaptative zeroing will be composed with a quantile estimator in the proposed research environment to secure the model from learning wrong information or patterns. Clipping to bound L2 norm will be used in a way that adapt moderately quickly to a moderately high norm. In this way global server model is secured to improve the robustness to outliers. Techniques discussed so far in this section will be composed together experimentally to Develop a Novel Aggregation Method to preserve the privacy and security in model update aggregation.

Evaluation

In this research even though a federated learning environment is set using a federated data set, centralized test data set is also allocated for further evaluation. Trained weights taken from the federated clients will be applied to a standard model and that model will be evaluated with a centralized data set to test the global server model. Global model accuracy will also be tested with different client and server optimization techniques and with the novel composite aggregation algorithm. Since the federated environment is simulated, this research has the access to the downloaded models in clients for a federated evaluation. So, a local evaluation by averaging up the losses across all data batches will be conducted locally in the clients.

Tools and Technologies

- PyCharm IDE - Community Version
- Data Annotation Tools
- TensorFlow/ Keras
- AI/Machine Learning /Deep Learning
- Federated Learning
- Client and Server optimization
- Update aggregation
- GPU computing
- Dockers
- Python

5. Research ethical issues

No significant ethical issues in this research were identified. All the sensitive data, related to this research will be kept and maintained under stringent and compliant security protocols.

6. Implementation.

To achieve the objectives mentioned in section 3, following tasks will be carried out throughout the research.

- Develop AI Model to Identify Harmful or Dangerous Acts Involving Minors in the Video contents
    - Data Collection
    - Data Annotation
    - Data Analysis
    - Data Set Preparation
    - Data Preprocessing
    - Feature Extraction
    - Implement, Optimize and Analyze the Data Input Pipeline
    - Machine Learning Model Design & Compilation
    - Handling Model Overfitting and Underfitting
    - Accelerated Training Implementation using GPUs
    - Server Model Fine Tuning
    - Server Model Deployment
    - Implement, Optimize and Analyze the Data Inference Pipeline
- Develop Federated Learning set up to enable privacy in machine learning model training for this research context
    - Federated Data Set Preparation
    - Federated Training Implementation
    - Federated Training Pipeline Testing

- Enable efficient Client and Server optimization technique for this federated research context
    - Client optimization
    - Server optimization
    - Update aggregation
    - Lossy compression
- Develop Novel Aggregation Method by composing different techniques to preserve the security and privacy in update aggregation
    - Implement Quantile Matching
    - Implement Zeroing
    - Implement Clipping
    - Implement Differential Privacy
    - Implement Secure Aggregation
    - Composing, Tuning and Analysis of above aggregation techniques

7. Bibliography.

[1] X. Li, J. He, P. Vijayakumar, X. Zhang, and V. Chang, "A verifiable privacy-preserving machine learning prediction scheme for edge-enhanced hcpss," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[2] K. M. Rahman, F. Ahmed, N. Akhter, M. Hasan, R. Amin, K. E. Aziz, A. K. Islam, M. S. Mukta, and A. K. Islam, "Challenges, applications and design aspects of Federated Learning: A survey," *IEEE Access*, vol. 9, pp. 124682–124700, 2021.

[3] S. Asoodeh, W.-N. Chen, F. P. Calmon, and A. Ozgur, "Differentially private federated learning: An information-theoretic perspective," *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021.

[4] V. Balasubramanian, M. Aloqaily, and M. Reisslein, "FedCo: A FEDERATED Learning controller for content management in MULTI-PARTY edge systems," *2021 International Conference on Computer Communications and Networks (ICCCN)*, 2021.

[5] X. Liu, H. Li, G. Xu, Z. Chen, X. Huang, and R. Lu, "Privacy-Enhanced federated learning AGAINST Poisoning Adversaries," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2021.

[6] Badih Ghazi and Pasin Manurangsi and Ravi Kumar Ravikumar and Thao Nguyen, "Robust and Private Learning of Halfspaces," *IEEE Transactions on Information Forensics and Security*, pp. 1603-1611, 2021.

[7] M. Abadi, Ú. Erlingsson, I. Goodfellow, H. B. McMahan, N. Papernot , I. Mironov, K. Talwar, and L. Zhang, *On the Protection of Private Information in Machine Learning Systems: Two Recent Approaches. Proceedings of 30th IEEE Computer Security Foundations Symposium (CSF)*, 2017.

[8] Nicolas Papernot and Martín Abadi and Úlfar Erlingsson and Ian Goodfellow and Kunal Talwar, Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data. Proceedings of the International Conference on Learning Representations, 2017.

[9] K. A. Bonawitz and Vladimir Ivanov and Ben Kreuter and Antonio Marcedone and H. Brendan McMahan and Sarvar Patel and Daniel Ramage and Aaron Segal and Karn Seth, Practical Secure Aggregation for Federated Learning on User-Held Data. I NIPS Workshop on Private Multi-Party Machine Learning, 2016.

[10] J. Xiao, C. Du, Z. Duan, and W. Guo, “A novel Server-side Aggregation strategy for FEDERATED learning in NON-IID SITUATIONS,” *2021 20th International Symposium on Parallel and Distributed Computing (ISPDC)*, 2021.

[11] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, “Deep learning with differential privacy,” *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016.

[12] J. Lee and D. Kifer, “Scaling up differentially private deep learning with fast per-example gradient clipping,” *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 1, pp. 128–144, 2020.

[13] C.-W. Ching, H.-S. Huang, C.-A. Yang, Y.-C. Liu, and J.-J. Kuo, “Efficient communication topology via partially differential privacy for decentralized learning,” *2021 International Conference on Computer Communications and Networks (ICCCN)*, 2021.

[14] Om Thakkar and alen Andrew and H. Brendan McMahan, “Differentially Private Learning with Adaptive Clipping,” Computing Research Repository - arXiv, 2021.

[15] M. Schonfeld and M. Werner, “Distributed privacy-preserving mean estimation,” *2014 International Conference on Privacy and Security in Mobile Systems (PRISMS)*, 2014.

[16] “Introduction to federated learning,” *KDnuggets*. [Online]. Available: https://www.kdnuggets.com/2020/08/introduction-federated-learning.html. [Accessed: 15-Sep-2021].

[17] “Child safety policy - youtube help,” *Google*. [Online]. Available: https://support.google.com/youtube/answer/2801999?hl=en. [Accessed: 02-Sep-2021].

[18] “Report inappropriate content - android - youtube help,” *Google*. [Online]. Available: https://support.google.com/youtube/answer/2802027#zippy=. [Accessed: 02-Sep-2021].

[19] "Federated Learning” *TechTalks*, 10-Aug-2021. [Online]. Available: https://bdtechtalks.com/2021/08/09/what-is-federated-learning/. [Accessed: 03-Sep-2021].