



HAL
open science

Survey of Bias Mitigation in Federated Learning

Lynda Ferraguig, Yasmine Djebrouni, Sara Bouchenak, Vania Marangozova

► **To cite this version:**

Lynda Ferraguig, Yasmine Djebrouni, Sara Bouchenak, Vania Marangozova. Survey of Bias Mitigation in Federated Learning. Conférence francophone d'informatique en Parallélisme, Architecture et Système, Jul 2021, Lyon (virtuel), France. hal-03343288

HAL Id: hal-03343288

<https://hal.science/hal-03343288v1>

Submitted on 15 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Survey of Bias Mitigation in Federated Learning

Lynda Ferraguig^{1,3}, Yasmine Djebrouni², Sara Bouchenak³, and Vania Marangozova²

¹ESI (Ecole nationale Supérieure d'Informatique) – Algiers, Algeria

²University of Grenoble Alps – LIG, France

³INSA Lyon – LIRIS, France

Abstract

Federated Learning (FL) interestingly allows a set of participants to collectively resolve a machine learning problem in a decentralized and privacy preserving manner. However, data distribution and heterogeneity, that are inherent to FL, may induce and exacerbate the problem of bias, with its prejudicial consequences such as racial or sexist segregation, illegal actions, or reduced revenues. In this paper, we describe the problem of bias in Federated Learning, and provide a comparative review of existing approaches of FL bias mitigation, before discussing open challenges and interesting research directions.

Keywords : Bias, Federated Learning, Machine Learning

1. Introduction

Machine learning (ML) is applied in many areas to extract knowledge from data and guide the decision making process, such in search engines [18], recommendation systems [5] and disease diagnosis [17]. With the rapid growth of data, ML algorithms evolved from centralized to distributed solutions. And to address data privacy issues, Federated Learning (FL) has emerged to allow a set of participants to collectively resolve a machine learning problem without sharing their data. However, FL exacerbates the problem of bias [21, 3]. Bias is a phenomenon that occurs when ML models produce unfair decisions due to the use of incomplete, faulty or prejudicial datasets and models. Bias may have serious consequences such as sexist segregation, illegal actions, or reduced revenues [6, 7, 30]. Federated Learning exacerbates the problem of bias [21, 3], because of the decentralized nature of FL, where data distribution and size are particularly heterogeneous. Furthermore, data privacy constraints in FL do not allow the use of classical ML bias mitigation techniques [32, 10].

In this paper, we study the open challenge of bias mitigation in Federated Learning, and provide a comparative survey of research works in the area. More precisely, we recall the necessary background on Federated Learning, before stating the problem of bias in FL in §2. In §3, we define the criteria that we use to compare bias mitigation approaches in FL, and we review these approaches in §4. Finally, §5 summarizes of the open challenges in FL bias mitigation, and identifies interesting research directions.

To the best of our knowledge this is the first survey that provides a comparative survey of bias mitigation techniques used in Federated Learning approaches.

2. Background and Motivation

2.1. Federated Learning

Today, many parts of the world require technology companies to handle user data carefully, in accordance with user privacy laws. The European Union's General Data Protection Regulation (GDPR) [2] is a prominent example. Federated Learning (FL) [20] is an emerging decentralized approach to design machine learning models that preserve data privacy. In FL systems, multiple clients (*e.g.*, mobile devices, organizations, etc.) collaborate in the training process, while a server orchestrates the learning. Client's raw data is stored locally and never transferred, instead updates containing the clients local models are transferred to the master where they are aggregated to achieve the learning objective and build a global model[21]. Figure 1 de-

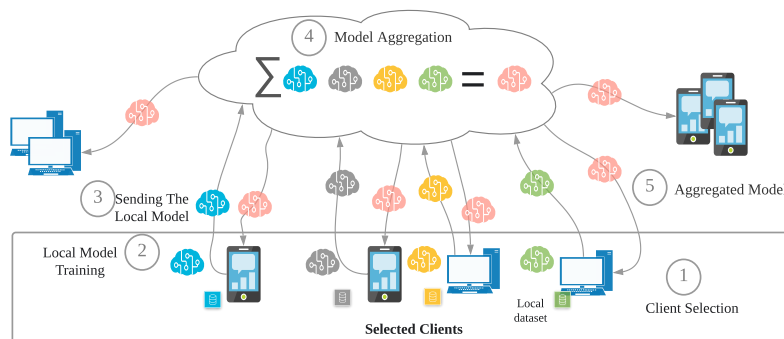


Figure 1: Overview of Federated Learning

picts the architecture of Federated Learning systems. This architecture represents the typical centralized FL system where a master server orchestrates the learning task. Other architectures have been proposed for Federated Learning systems to optimize communication costs or to overcome the single failure point problem of the centralized approach. One example is the peer to peer FL architecture [28]. Federated Learning relies on an iterative process composed of several rounds. Each round includes the following steps as shown in Figure 1. (1) First, the server selects a subset of clients and sends them the actual global model. These clients are selected either randomly or using specific client selection heuristics. For example, in [11] authors propose to select clients based on their training performance. In [35] authors propose the *CS-FedAvg* protocol which chooses clients with lower degree of non-IID data to avoid the accuracy degradation. (2) In the local model training step, each participant trains and optimizes the global model on its local data. (3) Once a client completes training its local model, it sends the model to the server. (4) Then, in the model aggregation step, the server aggregates the received updates from clients using an aggregation method, for instance, the FedAvg algorithm [19] which calculates a weighted average according to the size of each client's data. The aggregated model will be sent to the selected clients of the following round. These steps are repeated until a stopping criterion is reached (*e.g.*, a maximum number of rounds is reached, or the model accuracy is greater than a defined threshold).

2.2. Causes of Bias in Classical Machine Learning

Machine Learning plays a key role in decision making, therefore, it is crucial to ensure that the decisions made by a ML model do not reflect any discriminatory behaviour towards certain groups or populations. One of the factors that can lead to discriminatory decisions in ML is bias in the training data [9]. As defined in [25], bias is the inclination or prejudice of a decision made by a ML system which is for or against one person or group, in a way considered to be

unfair. In many cases, bias results from imbalanced datasets. Data imbalance can refer to an imbalance of classes in the output of the annotated dataset used to train the ML model. For example, for a binary classification task on a dataset containing images of people's skins[14], the objective is to determine from these patients images whether a skin cancer is present or not. If the output classes in the training data are imbalanced (*e.g.*, more sick people), the final model will be biased toward the over-represented class. Data imbalance can also refer to an over-representation/ under-representation of a certain category of persons or groups. For example we can have more samples of males than females in our data even if the output classes are balanced. In[1] , Amazon's recruiting tool was preferring male candidates over female candidates because the latter were under-represented in the training dataset. Both types of data imbalance result in data that lacks diversity which leads to a model biased toward the over-represented group. Moreover, even when the training data is balanced, the final model can exaggerate stereotypes that are present in the training data. In [30] , the authors give the example of a dataset with images of men and women cooking, where children are unlabeled but co-occur with the cooking action, the model could associate the presence of children with cooking. Since children co-occur with women more often than men across all images, a model could label women as cooking more often than we expect from a balanced distribution.

2.3. The Problem of Bias in Federated Learning

Federated Learning is a decentralized ML approach where all clients collaborate to elaborate a global ML model by training models locally on their own data. However, each client's local training process is equivalent to a centralized training, so FL inherits the above-mentioned causes of bias from traditional ML [3] which we refer to as "classical causes". Worse, bias in FL is exacerbated through the model updates shared at each round, since each party will introduce its own bias to the global model. In addition to these classical causes of bias, there are other causes closely related to the specific characteristics of Federated Learning. For example, data heterogeneity which is very present in FL setting and completely absent in a traditional ML setting, favours the construction of biased models [21]. In FL systems, as each client generates its own dataset, the data of a client may not be representative of the global distribution and thus introduce bias to the global model. Moreover, when training a FL model, clients may also leave the training which is known as client dropout [31] due to certain constraints (*e.g.*, low battery, poor connectivity, etc.). This dynamism implies a constant change in the composition of the data, which affects how the global model learns. The client selection process in FL can also introduce bias. If the selection protocol is based on certain attributes such as the device's hardware resources [29], then clients with more robust resources will be more represented. Client participation here is even correlated with socio-economic status[21]. In order to build the overall model, the server follows a strategy dictated by an aggregation algorithm also known as a fusion algorithm. These algorithms vary in the way they aggregate the received updates. For example, some are designed to perform a weighted average based on the dataset size of the party [19] which means that parties with larger datasets influence the overall model more than parties with smaller datasets. Such aggregation strategy implies the effect of over/under representation of certain groups in the overall training data.

3. Comparison Criteria of Bias Mitigation in Federated Learning

Here, we define criteria that characterize bias mitigation approaches in Federated Learning. We will then use these criteria to compare existing approaches of FL bias mitigation in §4.

Type of Data Imbalance. Data class imbalance occurs when there is unfair representation

of (i.e., under-represented) classes in the dataset. In Federated Learning, we distinguish two types of data imbalance. (i) In *Global Imbalance*, the union of the FL clients's datasets results in imbalanced data classes. (ii) In case of *Local Imbalance*, even if the FL clients hold locally imbalanced classes, the global data distribution overall the FL system is balanced [13].

Sensitive feature imbalance refers to the case where the distribution of a sensitive attribute (or a set of sensitive attributes) is not equal among groups that constitute the sensitive feature. For instance, consider the case where the sensitive feature is "gender" where we have two groups "male" and "female". If we have more samples in the first group(male) this leads to what we call *Sensitive feature imbalance*.

Federated Learning Scale. Two main FL scales are usually considered, depending on the nature of FL participants and their number. (i) *Cross-silo* FL involves a small number of relatively reliable FL participants, usually not exceeding 100, which have relatively large local datasets [21]. This is, for example, the case of several banks that collaboratively train fraud detection models. (ii) In *cross-device* FL architecture, there is a large number of participants, with a relatively small amount of data and lower computational resources [21]. This is the case, for instance, of mobile and edge device applications.

Bias Mitigation Techniques. There are several techniques to mitigate bias in Federated Learning which are inherited from classical machine learning methods. Overall, their goal is to ensure that the main factor determining the outcome of a FL model is not a sensitive attribute (e.g., gender or race), since sensitive attributes are the main reasons for bias. The types of techniques for FL bias mitigation differ depending on where they apply in a machine learning pipeline [6], as follows:

- **Pre-processing based bias mitigation techniques** modify training data in order to create a less biased dataset [22], for instance, by sampling or reweighing training samples. Such techniques are generic and can be applied regardless of the underlying ML models.
- **In-processing based bias mitigation techniques** alter the optimization problem related to the underlying classifier, by adding either a discrimination-aware regularizer to the target function or bias mitigation constraints to the optimization formulation [33]. However, this approach is limited to specific ML models and training algorithms. For instance in [16], the proposed technique only works in the case of logistic regression models.
- **Post-processing based bias mitigation techniques** consider the learned model as a black-box, without the ability to modify the training data or the learning algorithm. These techniques are used to help trained classifiers make fairer predictions for a provided dataset.

We believe that the aforementioned techniques can be divided into: **Preventive Techniques** which aim to prevent the construction of biased models without testing a priori if there is bias in training data (*In-processing methods*), and **Reactive Techniques** that aim to correct bias if detected (*Pre-processing and post-processing methods*). In table 1 we divided reactive techniques into "Bias Detection" and "Bias Correction" to indicate if the mentioned paper proposes a clear bias detection mechanism before its correction or it directly assumes there is bias.

Additional Information Sharing. Some FL bias mitigation techniques require additional information. For instance, some techniques assume that FL clients share with the server statistical information regarding the local data distribution of the clients.

4. Existing Works on Bias Mitigation in Federated Learning

Recently, several works have been conducted to address the problem of bias mitigation in Federated Learning. In the following, we provide a comparative survey of these works, based on

the comparison criteria introduced in §3. The different existing approaches of FL bias mitigation are summarized in Table 1.

Astrea. Duan et. al., propose a framework called Astrea to mitigate model bias caused by global class imbalance, they mainly propose two strategies [13]. First, data augmentation for samples belonging to minority classes. Secondly, Astrea proposes to create some mediators which are virtual components created by the server to reschedule the training of clients, their principle purpose is to make the distribution of the collection of data close to the uniform so that a new partial equilibrium is achieved. Once the mediators hold a set of clients with a global data distribution close to the uniform, the training can start. At each communication round, each mediator sends the model to the subordinate clients. Each client trains the model and returns the updated model to the corresponding mediator. Then, the mediator receives the updated model and sends it to the next waiting training client. Once all clients of a mediator has completed the training, the mediator sends the updates of models to the FL server that aggregates them and elaborates the global model. The framework is based on the assumption that each client must share information about the distribution of its local data.

Bias-Free FedGAN. Mugunthan et. al., train a generative adversarial network in a federated fashion [27] with local class imbalance resulting in a biased model towards the majority class[23]. To alleviate this issue, they proposed Bias-Free FedGAN. The main idea is that each client trains a local GAN. Then, clients send their generator and discriminator model parameters to a central server. The server generates a combined meta data using the local generators received from the clients, after that, it trains the averaged model on the meta dataset and sends the final parameters of the generator and discriminator to the clients. The authors assert that Bias-Free FedGAN produces results that are not biased towards a particular class, however privacy aspects have not been taken into account. Training BiasFree FedGAN in a differentially private manner to prevent model inversion attacks is mentioned as a future direction.

FairFL. Zhang et. al. tackle the problem of global class imbalance and the bias related to the presence of sensitive attributes [34]. If, for example, the classification accuracy for classifying the male's job records is higher than those of the female's, then the classifier is considered biased towards males. Essentially, they consider the case where the client's data contains certain private demographic information such as gender, race and age. The goal is to make sure that the global model is unbiased across all demographic groups. To do so, they first define a discrimination index based on F1-score that is chosen as the accuracy measure due to its robustness against the imbalanced dataset. The discrimination index is the difference between the F1 scores of all data samples that belong to a specific demographic group and those that belong to another one. For example, to measure this metric in the case where the demographic attribute is gender, they calculate the F1 score of the "male" instances from which they subtract the F1 score of the "female" instances, the objective is that this difference should be as close to zero as possible indicating that the model treats the demographic groups equally.

Abay et. al.'s approach. Abay et. al., consider the problem of imbalance in some sensitive attributes such as gender and race [4]. This is the case where clients hold datasets with a non-uniform distribution of these sensitive attributes. To alleviate this issue, the authors propose two techniques. The first is a pre-processing method which is re-weighting. Re-weighting [8] rebalances the dataset by affecting weights to samples in the training data. It accesses the entire training dataset and computes weights as the ratio of the expected probability (P_{exp}) to the observed probability (P_{obs}) of the sample's sensitive attribute/label pairing. This technique is used in classical machine learning; however, in the case of FL, there is no access to data. Thus,

authors propose a local-reweighting technique where each client performs reweighting to its own dataset. They show that this technique is effective without compromising the prediction accuracy, even when only a subset of the parties employ it. They also propose a global version of local reweighting with differential privacy under the assumption that each client accepts to share information about its sensitive attributes. The principal limitation of this technique is that it can't be applied to FL systems with dynamic participation, as the global reweighting weights would recalibrate in relation to the number and size of training sets changing over the course of training. Furthermore, authors propose a modified version of prejudice removal technique known in classical ML as an in-processing method which works by adding a fairness-aware regularizer to the loss function. The main idea is that each client applies the Prejudice Remover algorithm [16] to train a less biased local model. The major disadvantage of this technique is selecting a reasonable coefficient for the regularizer, indeed, the increase of this coefficient leads to more bias mitigation, but also a degradation of performance metrics specifically accuracy [16].

FELICIA. Rajotte et. al. propose an approach to alleviate the problem of global class imbalance. The idea is to perform data augmentation to produce a global synthetic dataset yet realistic using private GANs [24].

Approach	FL Scale		Type of Imbalance			Bias Mitigation Technique			Additional Information Required		
	Cross-silo	Cross-device	Class Imbalance		Sensitive Feature Imbalance	Preventive Techniques	Reactive Techniques		Class Distribution	Sensitive Feature	No Additional Information
			Local Imbalance	Global Imbalance			Bias Detection	Bias Correction			
Astrea [13]		✓		✓			✓	✓			✓
Bias-Free FedGAN [23]	✓		✓	✓				✓			✓
FairFL [34]	✓				✓		✓	✓			✓
Abay et. al.'s approach [4]	✓				✓	✓		✓		✓	
FELICIA [26]	✓				✓	✓		✓			✓

Table 1: Existing approaches of bias mitigation in Federated Learning

As a summary of the existing approaches of FL bias mitigation, we notice that:

- These approaches focus mainly on the cross-silo scenario, which is explained by the fact that in the case of the cross-device, bias correction is relatively more difficult, because the data changes dynamically from one round to another.
- None of these approaches proposes a post-processing technique to mitigate bias in FL, even though this type of approach may be adequate to the constraint of Federated Learning, as it acts on the final model available in the case of FL, and considers the ML model and the data as a black-box.

5. Summary and Open Research Directions

The problem of bias in machine learning has been extensively studied, and a variety of techniques have been proposed [15, 12, 30, 9]. The direct application of classical ML bias mitigation techniques in the case of Federated Learning is not always possible due to the constraints of data privacy. Moreover, FL exacerbates some causes of bias and gives rise to new ones. Finding techniques that can both measure and reduce bias without directly accessing sensitive information is challenging because most current techniques require access to the data to overcome the bias problem. That is why correcting bias in training data is an important research direction in the FL field [21]. Furthermore, most of the existing works on FL bias mitigation focus on the cross-silo case, because it is more difficult to correct the bias in the cross-device case where the evolution of the data is very dynamic [21]. Thus, further research in this area is needed.

Bibliography

1. Amazon scraps secret ai recruiting tool that showed bias against women. – <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>. [Online; accessed 10-04-2021].
2. General Data Protection Regulation. – <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679>. [Online; accessed 10-04-2021].
3. Abay (A.), Chuba (E.), Zhou (Y.), Baracaldo (N.) et Ludwig (H.). – Addressing unique fairness obstacles within federated learning. 2021.
4. Abay (A.), Zhou (Y.), Baracaldo (N.), Rajamoni (S.), Chuba (E.) et Ludwig (H.). – Mitigating bias in federated learning. *arXiv preprint arXiv:2012.02447*, 2020.
5. Aher (S. B.) et Lobo (L.). – Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data. *Knowledge-Based Systems*, vol. 51, 2013, pp. 1–14.
6. Bellamy (R. K.), Dey (K.), Hind (M.), Hoffman (S. C.), Houde (S.), Kannan (K.), Lohia (P.), Martino (J.), Mehta (S.), Mojsilovic (A.) et al. – Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943*, 2018.
7. Bogroff (A.) et Guegan (D.). – Artificial intelligence, data, ethics an holistic approach for risks and regulation. *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series*, no19, 2019.
8. Calders (T.), Kamiran (F.) et Pechenizkiy (M.). – Building classifiers with independency constraints. – In *2009 IEEE International Conference on Data Mining Workshops*, pp. 13–18. IEEE, 2009.
9. Calders (T.) et Zliobaite (I.). – Why unbiased computational processes can lead to discriminative decision procedures. vol. 3, 2013, pp. 43–57.
10. Calmon (F.), Wei (D.), Vinzamuri (B.), Natesan Ramamurthy (K.) et Varshney (K. R.). – Optimized pre-processing for discrimination prevention. – In Guyon (I.), Luxburg (U. V.), Bengio (S.), Wallach (H.), Fergus (R.), Vishwanathan (S.) et Garnett (R.) (édité par), *Advances in Neural Information Processing Systems* volume 30. Curran Associates, Inc., 2017.
11. Chai (Z.), Ali (A.), Zawad (S.), Truex (S.), Anwar (A.), Baracaldo (N.), Zhou (Y.), Ludwig (H.), Yan (F.) et Cheng (Y.). – Tifl: A tier-based federated learning system. – In *Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 125–136, 2020.
12. d'Alessandro (B.), O'Neil (C.) et LaGatta (T.). – Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big data*, vol. 5, n2, 2017, pp. 120–134.
13. Duan (M.), Liu (D.), Chen (X.), Tan (Y.), Ren (J.), Qiao (L.) et Liang (L.). – Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications. – In *2019 IEEE 37th International Conference on Computer Design (ICCD)*, pp. 246–254, Los Alamitos, CA, USA, nov 2019. IEEE Computer Society.
14. Esteva (A.), Kuprel (B.), Novoa (R. A.), Ko (J.), Swetter (S. M.), Blau (H. M.) et Thrun (S.). – Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, vol. 542, n7639, 2017, pp. 115–118.
15. Hardt (M.), Price (E.) et Srebro (N.). – Equality of opportunity in supervised learning. *arXiv preprint arXiv:1610.02413*, 2016.
16. Kamishima (T.), Akaho (S.), Asoh (H.) et Sakuma (J.). – Fairness-aware classifier with prej-

- udice remover regularizer. – In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 35–50. Springer, 2012.
17. Kourou (K.), Exarchos (T. P.), Exarchos (K. P.), Karamouzis (M. V.) et Fotiadis (D. I.). – Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, vol. 13, 2015, pp. 8–17.
 18. McCallumzy (A.), Nigamy (K.), Renniey (J.) et Seymorey (K.). – Building domain-specific search engines with machine learning techniques. – In *Proceedings of the AAAI Spring Symposium on Intelligent Agents in Cyberspace*. Citeseer, pp. 28–39. Citeseer, 1999.
 19. McMahan (B.), Moore (E.), Ramage (D.), Hampson (S.) et y Arcas (B. A.). – Communication-efficient learning of deep networks from decentralized data. – In *Artificial Intelligence and Statistics*, pp. 1273–1282. PMLR, 2016.
 20. McMahan (B.), Moore (E.), Ramage (D.), Hampson (S.) et y Arcas (B. A.). – Communication-efficient learning of deep networks from decentralized data. – In *Artificial Intelligence and Statistics*, pp. 1273–1282. PMLR, 2017.
 21. McMahan (H. B.) et al. – Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, vol. 14, n1, 2021.
 22. Mehrabi (N.), Morstatter (F.), Saxena (N.), Lerman (K.) et Galstyan (A.). – A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*, 2019.
 23. Mugunthan (V.), Gokul (V.), Kagal (L.) et Dubnov (S.). – Bias-free fedgan, 2021.
 24. Mukherjee (S.), Xu (Y.), Trivedi (A.) et Ferres (J. L.). – privgan: Protecting gans from membership inference attacks at low cost. *arXiv preprint arXiv:2001.00071*, 2019.
 25. Ntoutsis (e. a.). – Bias in data-driven artificial intelligence systems—an introductory survey. *WIREs Data Mining and Knowledge Discovery*, vol. 10, 02 2020.
 26. Rajotte (J.-F.), Mukherjee (S.), Robinson (C.), Ortiz (A.), West (C.), Ferres (J. L.) et Ng (R. T.). – Reducing bias and increasing utility by federated generative modeling of medical images using a centralized adversary. *arXiv preprint arXiv:2101.07235*, 2021.
 27. Rasouli (M.), Sun (T.) et Rajagopal (R.). – Fedgan: Federated generative adversarial networks for distributed data, 2020.
 28. Roy (A. G.), Siddiqui (S.), Pölsterl (S.), Navab (N.) et Wachinger (C.). – Braintorrent: A peer-to-peer environment for decentralized federated learning. *arXiv preprint arXiv:1905.06731*, 2019.
 29. Wang (S.), Tuor (T.), Salonidis (T.), Leung (K. K.), Makaya (C.), He (T.) et Chan (K.). – Adaptive federated learning in resource constrained edge computing systems. *IEEE Journal on Selected Areas in Communications*, vol. 37, n6, 2019, pp. 1205–1221.
 30. Wang (T.), Zhao (J.), Yatskar (M.), Chang (K.-W.) et Ordonez (V.). – Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. – In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5310–5319, 2019.
 31. Wu (W.), He (L.), Lin (W.), Mao (R.), Maple (C.) et Jarvis (S. A.). – Safa: a semi-asynchronous protocol for fast federated learning with low overhead. *IEEE Transactions on Computers*, 2020.
 32. Zemel (R.), Wu (Y.), Swersky (K.), Pitassi (T.) et Dwork (C.). – Learning fair representations. – In Dasgupta (S.) et McAllester (D.) (édité par), *Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research*, volume 28, pp. 325–333, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.
 33. Zhang (B. H.), Lemoine (B.) et Mitchell (M.). – Mitigating unwanted biases with adversarial learning. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, Dec 2018.
 34. Zhang (D. Y.), Kou (Z.) et Wang (D.). – Fairfl: A fair federated learning approach to reducing demographic bias in privacy-sensitive classification models. – In *2020 IEEE International*

Conference on Big Data (Big Data), pp. 1051–1060, 2020.

35. Zhang (W.), Wang (X.), Zhou (P.), Wu (W.) et Zhang (X.). – Client selection for federated learning with non-iid data in mobile edge computing. *IEEE Access*, vol. 9, 2021, pp. 24462–24474.