IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 June 2023, accepted 30 June 2023, date of publication 14 July 2023, date of current version 7 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3295412

SURVEY

The Current State and Challenges of Fairness in Federated Learning

SEAN VUCINICH [1] AND QIANG ZHU [2], (Senior Member, IEEE)
[1]Center for Academic Innovation, University of Michigan, Ann Arbor, MI 48103, USA
[2]Department of Computer and Information Science, University of Michigan-Dearborn, Dearborn, MI 48128, USA

Corresponding author: Sean Vucinich (vucinich@umich.edu)

ABSTRACT The proliferation of artificial intelligence systems and their reliance on massive datasets have led to a renewed demand on privacy of data. Both the large data processing need and its associated data privacy demand have led to the development of techniques such as Federated Learning, a distributed machine learning technique with privacy preservation built-in. Within Federated Learning, as with other machine learning based techniques, the concern and challenges of ensuring that the decisions being made are fair and equitable to all users is paramount. This paper presents an up-to-date review of the motivations, concepts, characteristics, challenges, and techniques/methods related to fairness in Federated Learning reported in the literature. It also highlights open challenges and future research directions in evaluating and enforcing fairness in Federated Learning systems.

INDEX TERMS Federated learning, fairness of data, individual fairness, group fairness, fairness of system, algorithmic fairness, fairness measure, fairness evaluation.

I. INTRODUCTION

As the number of decisions made by the automated systems that people interact with daily is ever-growing, an increasing amount of attention is being paid to how the intelligent algorithms in such systems make decisions. These automated systems are demanded for a variety of reasons, from higher productivity compared to a human doing the same task, to their ability to process incomprehensible amounts of data. Unfortunately, just like the humans they often replace, these automated systems are not entirely objective and free from bias when making their decisions. Often, the decisions made by such systems are inherently biased due to existing biases with the data from which the decision models are learned and the system environments/settings with which these models are built.

There are several causes that may lead to unfairness of Machine Learning (ML) algorithms, which include biases from datasets, biases from missing data, biases from algorithmic goals, and biases from proxy variables for protected data attributes. Much work on ML algorithmic fairness has been reported in the literature [1], [2]. Recently, researchers have been attracted to fairness issues in the context of Federated Learning (FL).

Federated Learning, as an emerging Machine Learning approach, aims to train a model across multiple participating clients while retaining independence and privacy of the clients participating in the model training by keeping training data separate from the coordinating server. As an inherently distributed technique, Federated Learning addresses critical data issues (e.g., data access, security, and privacy) while also accounting for non-IID (independent and identically distributed) data due to the nature of the client participation. While these features can help alleviate some of the concerns related to the rise of automated decision systems, this technique still must address issues of bias and fairness in its predictions.

This paper provides an up-to-date review of the current state and challenges on the topic of fairness in Federated Learning systems. Existing surveys on Federated Learning have examined general Federated Learning system architecture, applications, and implementations. Different from these surveys, this paper focuses on reviewing the relevant motivations, concepts, characteristics, challenges, techniques, and

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Yan.

future work of enforcing fairness in FL systems, based on the studies reported in the literature. Fairness in FL has become an increasingly important topic that has attracted much attention from researchers and practitioners recently. This survey advances the understanding of fairness in the context of FL and identifies open challenges and potential future work, which may serve as a guide for the future research. To our knowledge, no similar work has been reported in the literature.

The main contributions of this paper are listed as follows:

- The relevant background, concepts, causes, and challenges of fairness in the FL context are summarized.
- Update-to-date techniques/methods reported in the literature to address challenges of fairness for Federated Learning systems are reviewed.
- Observations about the current state, characteristics and trends of the relevant research are discussed.
- Open challenges and future work for further research opportunities are highlighted.

The rest of this paper is organized as follows. Section II provides an overview of the background techniques that are related to fairness in Federated Learning. Section III presents the core challenges of fairness in Federated Learning. Section IV reviews existing works on fairness in Federated Learning from the literature. Section V discusses observations about current research characteristics and trends as well as open challenges and future research directions. Section VI provides concluding remarks.

II. BACKGROUND AND RELATED TECHNIQUES

In this section, we highlight the background techniques including machine learning, deep learning, algorithmic fairness in ML, and federated learning that are related to the topic of fairness in Federated Learning.

A. MACHINE LEARNING AND DEEP LEARNING

Originally termed by Arthur Samuel in the 1950s [3], Machine Learning is one of the main branches within the field of artificial intelligence [4]. In the past twenty years, Machine Learning has seen an explosion of interest, fueled by advances in computation and research into the backing concepts that have led to significant advances in the field and wide adoption of the technology in commercial use. The general recent trends of modern Machine Learning can be summarized as follows [5]:

- Rule-based systems: This is a category that includes decision trees, tables, and logic programming that share the commonality of utilizing hand-crafted rules and are intuitive to understand.
- Bayesian statistics: This is a field that utilizes Bayesian probability to make inferences, representing a prolific category that has seen a lot of work in recent years due to the emergence of probabilistic programming languages and models.
- Kernel-based algorithms: This includes algorithms that rely on the concept of neighborhood and adherence to a definition of similarity. Examples of such algorithms include k-Nearest Neighbor (KNN) and Support Vector Machines (SVM). These algorithms tend to suffer from poor scaling as the dataset increases, which has led to a resurgence of the next trend.
- Deep Neural Networks (DNN): This technique is an evolution of general neural networks that rely on many layers of neurons (the deep portion of the name) that are stacked in a hierarchical structure for data processing. With typically millions or billions of parameters, these networks are difficult to interpret, but have shown incredible performance in many areas including computer vision, language models, and speech recognition.

Deep Learning builds upon generalized machine learning by introducing representation-based learning in the form of the Deep Neural Network (DNN) [6]. This form of learning utilizes multiple tiers of simple, non-linear representations that transform the input data level by level until a complex function can be learned [7]. Deep Learning has often been compared to FL due to similarities with Deep Learning techniques (primarily DNNs) and how these networks have been utilized for purposes that are now being taken over by FL systems. One notable drawback to the DNN architecture is its complexity, whereas the dataset size increases so does the complexity of the DNN, and proportionally, so does the computation demand of the network. The performance requirements of highly accurate DNNs necessitates the use of a high-performance compute cluster, with the cost and concurrency requirements that are associated with them. To make efficient use of these systems, the introduction of parallelism to code and the distribution of execution across clusters build into forms of distributed learning.

B. ALGORITHMIC FAIRNESS WITHIN ML

As machine learning algorithms and the automated decisions they make become more embedded in people's lives, the concerns about whether these algorithms are fair have led to an ever increasing interest in algorithmic fairness. This increased interest has led to a breadth of literature encompassing defining algorithmic fairness, evaluating fairness of algorithms, and methods to improve fairness. Maintaining fairness of an algorithm is a balancing act of trade-offs, as [1] demonstrates, to achieve higher measures of fairness, accuracy inherently is compromised. As noted in a survey on fairness within machine learning [2], there are four main categories of causes to algorithmic unfairness that can be identified from existing literature:

- Biases included in the datasets: including data from biased measurements or human decisions, from erroneous or biased reports, among other reasons. By their nature, machine learning algorithms replicate these in-built biases.

- Biases caused by missing data: including missing data entries, data values collected with a sample or selection bias, or from poorly run data collections. This category of bias results in non-representative datasets that can differ greatly from the target populations.
- Biases from algorithmic goals: both due to unreliable algorithms and from algorithms that result in the majority group being favored over minorities due to minimization of prediction errors.
- Biases resulting from the use of proxy variables in the place of protected data attributes. These protected data attributes are typically the variables that distinguish between privileged and unprivileged groups and are not considered permissible for predictions (protected variables are those typically protected from discrimination by law, such as race, age, religious affiliation, gender, etc.). Proxy variables can be used to infer protected attributes that are available to the algorithm, therefore if a dataset includes these proxies, the algorithm may make biased decisions from these inferences.

Lastly, to clarify, throughout this paper, the terms bias, unfairness, and discrimination are used interchangeably to mean similar things. This practice is common amongst fairness literature, as can be seen in the following paper which aims to establish standardized definitions of algorithmic fairness within Machine Learning [8].The concepts of fairness in this paper are in line with those established in the aforementioned paper, however, are expanded to the context of Federated Learning.

C. FEDERATED LEARNING

Originally introduced by researchers at Google in 2016 [9], [10], Federated Learning has shown great promise and has seen immense research progress as the technology becomes more mature and its potential applications are explored. The crux of FL is the concept of training a model without the need for data to be centrally stored or transferred to a centralized location, which necessitates a highly parallel and collaborative system.

As research has continued further in the domain, FL has been applied to a range of different applications, from the Internet of Things to medical applications. Despite the increasing popularity of the technology, its technical components are only now becoming more broadly understood, and resolving the specific challenges in implementing an FL system is now becoming the focus of research. Further driving the adoption of FL systems is the in-built data privacy provided by these systems, as client data is not shared with the central server and training occurs at the client devices rather than adopting a main compute cluster as is the case in traditional machine learning.

Building on the base idea of centralized machine learning systems, Federated Learning differentiates itself through four main criteria:

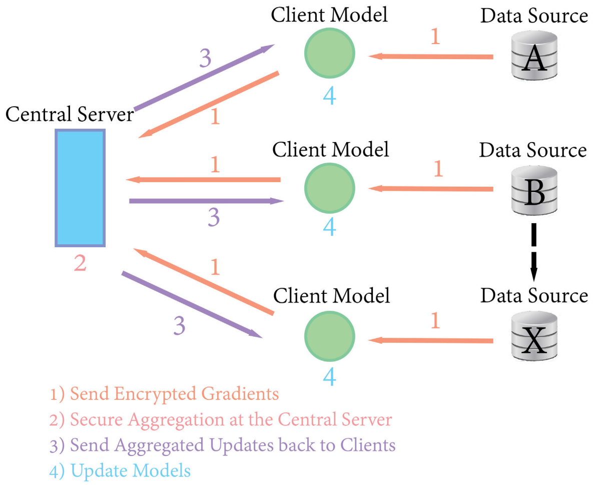
Client Model
Data Source
A
Central Server
Client Model
Data Source
B
Client Model
Data Source
X
1
2
3
4
1) Send Encrypted Gradients
2) Secure Aggregation at the Central Server
3) Send Aggregated Updates back to Clients
4) Update Models

FIGURE 1. Example architecture of a horizontal federated learning system.

- Datasets are not shared; namely, clients retain exclusive access to their training data and share model updates to be aggregated on a central server iteratively. This unbalanced data access leads to the challenge of data heterogeneity,
- Training is distributed and collaborative, taking place typically across many edge devices (e.g., mobile phones, autonomous vehicles, and Internet of Things (IoT) devices). This inherent massive parallelism is both a benefit and a challenge. Private FL systems can typically involve 2-100 clients whereas public (cross-device) FL can reach millions or even billions of clients, compared to distributed machine learning which is typically distributed across several high-performance server clusters.
- Local model training is decentralized, taking place on the clients where the data the clients train their local models on is considered independent and identically distributed (IID). This leads to a challenge at the global model level due to the non-independent identical distribution (non-IID) of client data (from the system perspective), as FL systems must reconcile the model updates of clients as an individual client's data is not representative of the global dataset.
- Privacy preservation is inbuilt as sensitive user data is not shared. Since the only data communicated between the clients and central server are the aggregated model updates, user privacy is retained.

Federated Learning systems come in a variety of architectures, with many improving upon deficiencies of earlier models, being more focused upon specific applications, or expanding the capabilities of the more generalized architectures. As defined in [11], there are two main types of FL architectures, horizontal one and vertical one, which differ in how they are structured, how they deal with client data, and how they learn the model.

Horizontal FL, which has also been referred to in the literature as sample-based FL, focuses on realizing secured

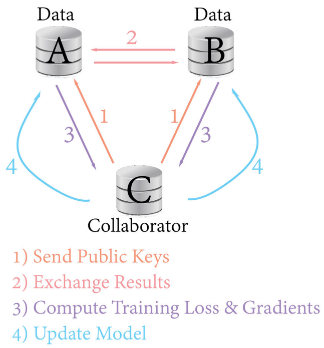

FIGURE 2. Example architecture of a vertical federated learning system.

FL in scenarios where client datasets share the same feature space but have different samples. Features are the same for the local models, while client data differs from client to client. This architecture assumes that each client is honest, with security being inbuilt against the central server. The training process of Horizontal FL is shown in Figure 1, where the basic structure consists of a number of clients with the same data structure computing local model updates and a central server aggregating the locally-computed updates.

Vertical FL, also referenced as feature-based FL, focuses on realizing FL for the scenarios in which features are different among clients while the client datasets share the same sample identifier space. The training process of Vertical FL is shown in Figure 2, wherein the clients perform training to collect, group, and exchange the features of the model utilizing a calculated training loss for building the model. Vertical FL also assumes that clients are participating honestly, but due to the calculation and exchange of features, a compromised client can only expose that client's data. This aspect of in-built privacy helps reinforce the privacy constraints inherent to FL.

D. LIMITATIONS OF PREVIOUS SURVEYS

Most recent surveys on the progress of FL have focused on technical considerations of FL including framework architectures and applications [11], enabling software/hardware tools and platforms for FL [12], system challenges in FL (such as communication costs, security and privacy, and resource allocation) [13].

In the survey [11], the authors present different system challenges of FL with accompanied solutions. They also discuss the three main architectures: Horizontal FL, Vertical FL, and Federated Transfer Learning. Each architecture is discussed from a technical standpoint and presented with current applications. At the end, the authors show that Federated Learning is an efficient technique to build data networks among multiple organizations while preserving user privacy.

The work of [12] presents differences between FL and traditional and centralized ML. The survey covers a range of enabling technologies that are used to learn a model in an FL setting and provides a summary of relevant protocols, platforms and real-world applications of FL. The authors also discuss existing key challenges in recent literature and best practices in the design of FL-based models.

The article [13] presents a comprehensive survey that shows how Deep Learning on Edge Computing is currently used in a variety of applications due to the massive computing power present on edge devices (such as smartphones) nowadays. The authors introduce concepts of FL by discussing the intersection between Deep Learning and Edge Computing as well as discussing the challenges of the implementation process of running Deep Learning on edge devices with suggested solutions. Finally, their work highlights four existing FL applications within the scope of Edge Computing.

The survey from [14] discusses the development process of FL and how the common challenges are resolved during implementation of such systems. The authors provide an overview of the evolution of FL frameworks over time and how these frameworks were developed in response to different challenges encountered. Furthermore, the authors discuss the future direction of FL applications, while primarily focusing on the domain of industrial engineering applications.

To our knowledge, there is no survey about the current state and challenges of fairness in FL, which will be presented in this paper. The overview, classification and analysis of related concepts, challenges and techniques of fairness in FL given in this paper will facilitate identifying new problems and developing novel solutions by researchers, developers and probationers in the field.

III. CORE CHALLENGES OF FAIRNESS IN FL

The concept of fairness within Federated Learning can be defined broadly in two ways: fairness in the context of data and fairness in the context of the system. Fairness of data can be defined similarly to fairness within the broader domain of Machine Learning, typically defined as individual-based fairness and group-based fairness. Fairness has been typically considered as an optimization problem in the domain of ML, where it acts as a constraint on the performance of a given model in respect to model accuracy [15].

A. FAIRNESS OF DATA - INDIVIDUAL FAIRNESS

Individual fairness is based upon the principle of similarity, that is, within the context of a given task, similar individuals can be assumed to be classified as similar [16]. Measuring this similarity is the core challenge of assessing individual fairness, as similarity of individuals can differ depending on the underlying specific data and applications. Another way to define individual fairness, as noted by [17], is the constraints that bind pairs of individuals rather than the average that binds a group.

One of the difficulties with defining individual fairness is the assumptions that one must make when selecting an approach, as these assumptions can make individual definitions of fairness impractical. The article [17] also explores existing definitions of statistical and individual fairness, their impact on intersectionality, and other pertinent questions on the topic of fairness within ML. Intersectionality in this context refers to how different types of bias can interact for individuals who belong to multiple protected classes. This work can serve as a good starting point for further reading into individual fairness. Much of the focus on improving individual-level fairness in FL looks at the stage of client selection, as ensuring similarity at this stage of the training process can address unfairness before model updates.

B. FAIRNESS OF DATA - GROUP FAIRNESS

Group fairness focuses on the issues that arise from quantifying the sensitive characteristics of populations (e.g., gender, race, and age) and mitigating the bias that can exist inherently from the use of such sensitive data. This notion is connected to statistical parity, which is the property of population sub-groups receiving identical classifications as that of the entire population [16]. Statistical parity aims to standardize the outcomes across both protected and non-protected groups. However, this can have the effect of unfair outcomes from the perspective of the individual.

The goal of group fairness is to ensure that the predicted outcome and the sensitive attributes of the data are independent, and any effect of potential biases are minimized. As a result, statistical parity between the protected and non-protected groups is ensured. The core challenge of group fairness under Federated Learning comes from the inherent use of sensitive data characteristics to attempt to mitigate potential bias, which stands in stark contrast to the core principle of FL that protects client privacy by not allowing access to client data [18].

It should be noted that the decisions and predictions made by these automated systems and algorithms will be biased towards the privileged groups and individuals (that is, there is inherent bias in these systems, and it is a fact that should be accounted for) [19]. This fact inevitably leads to the concern and need to reduce discrimination in models that interpret sensitive data, as miscalibrations caused by bias in these models can cause harm to some groups [20].

The type of data may also influence how the various approaches to addressing fairness challenges should work. For example, as seen in the article [21] in regards to tabular data, the context around the boundaries of features and what data these features relate to is important. Using a “month” feature, for example, any relevant method should understand the valid range of the feature (i.e., 1 - 12) and the valid type of data (i.e., an integer rather than a string or float (e.g., 6.5 is invalid)). This would differ from the approach needed in an image-based dataset, where features may not be explicitly defined as they are in tabular data. Any approach to addressing the challenges of fairness needs to be tailored to the type of data being used, including temporal vs. spatial properties to be discussed in Section IV-D.

C. FAIRNESS OF THE SYSTEM

Fairness of FL systems differs from that of fairness in the context of data, as it looks at the clients within the network and the challenges that arise from the asymmetric nature of the network. Fairness of the system can be defined as balancing the contributions and performance of clients across the network. This balancing is done with the goal that clients are equitably participating in the model learning, while minimizing the differences due to factors such as geographic location, client data distribution, and individual client performance.

Due to the distributed structure of a Federated Learning system, there are multiple challenges that come from the concept of system fairness. As the network is asymmetric and client data is inherently unbalanced, both the quality of the data used and rewards for a given client’s contributions are difficult to address. The limited bandwidth and inconsistent data distribution from network clients, along with the associated expensive communication costs, can result in difficulty of maintaining reasonable and satisfactory performance for the diverse clients on the network. The distributed nature of the FL network is a challenge, as such a network (especially mobile based ones) may have millions of clients across a large geographic area, compared to traditional distributed machine learning models which may only reside on 10s nodes within a data center [22].

It should be noted that there can be additional causes of bias in the system that are independent of the type of data, resulting in or resulting from the local or global models being biased. One example of such additional biases is caused by the choice of hyper-parameters used on the local models that can inadvertently introduce biases in the global model. Since these parameters may affect the learning process of both the local and global models in an FL system, they can have an out-sized influence on the outcomes of the model. Some approaches to mitigating FL model biases have examined methods that prevent this cause of bias entirely by not relying on hyper-parameter tuning of the FL models [23].

Similar to biases with causes independent of the data, the models in an FL system can become biased due to malicious attacks resulting from malicious clients. The most pertinent example of such attacks on FL systems with regards to fairness are the poisoning attacks. These attacks are characterized by malicious clients submitting bad updates with the intent of either preventing the global model from converging or covertly introducing artificial bias to the model [24]. Addressing the negative effects on model bias and overall system fairness these attacks can have without adversely over-detecting poison attempts is an emerging area of focus [25], [26].

There has been some work integrating complementing technologies with FL to better address challenges of privacy and integrity. For example, FL systems are integrated with blockchains, a distributed ledger technology whose records are linked securely using cryptographical hashes [27]. With blockchain records being immutable, blockchain empowered systems can support both integrity and privacy at a high level. To achieve improved scalablity of these systems, a lightweight blockchain approach like [28] could be utilized. Enforcing fairness in such systems is challenging as demonstrated in [29] and [30].

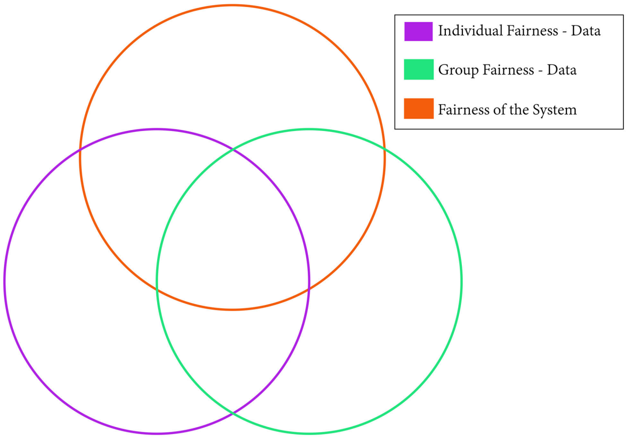

FIGURE 3. Simple diagram of how the categories of fairness may overlap.

D. OVERLAP OF FAIRNESS IN FL

The above categories of fairness are not mutually exclusive. The ML based systems often run into multiple fairness challenges since they can overlap. An example of this challenge overlap would be a hypothetical FL system for diagnosing melanoma, a skin cancer, based on photos of an area of skin, with clients participating from locations around the world. In this system, individual fairness would be a concern for the individual clients participating, as an accurate diagnosis is of utmost importance to the client. Group fairness concerns manifest themselves at the system level, where ensuring that different skin type groups are diagnosed fairly and there is no in-built bias in the system would be of paramount importance. Fairness of the system would then apply to how the clients are participating in the system and ensuring that regardless of where a client may be located, their participation is accounted for fairly. This overlap of fairness can be modeled by a simple Venn diagram, as shown in Figure 3. With the base challenges of fairness established and outlined, the next section will organize recent works in the field that aim to address the challenges.

IV. STATE OF EXISTING WORKS ON FL FAIRNESS

While the challenges of fairness related to data are common to the broad ML field beyond just Federated Learning, fairness of the system is quite unique to FL due to the distributed structure that is utilized. Much work has been done since FL emerged, with a lot of interest being focused upon applications and the privacy preserving aspects. More recently, some work has been done to address the challenges of fairness as it becomes clearer how impactful fairness is in the context of an FL system. In this section, we review existing works that aim to address the challenges presented in Section III.

A. ALGORITHMIC APPROACHES TO ADDRESSING FAIRNESS

In this subsection, we summarize the characteristics of recent algorithms that aim to address the challenges of fairness within a FL system.

Asynchronous Federated Learning systems deal with many of the same challenges as synchronous FL systems but must address them in distinct ways due to the difference among approaches. The study [31] explores asynchronous FL frameworks that aim to resolve the issue of straggling clients. This is a problem widely present within synchronous FL systems wherein the system must wait to collect all client models before performing the model aggregation. While waiting for straggling clients, the system may suffer from degraded performance. The paper proposes to overcome this by using asynchronous FL frameworks and by attempting to utilize both client availability and long-term fairness to reduce the training latency caused by client selection.

Due to the different approaches to dealing with client data, horizontal and vertical FL systems must address the challenges of fairness in different ways. In [22], the authors propose an algorithm, FedFa, that aims to achieve better fairness and accuracy within horizontal federated learning systems through the introduction of a double momentum gradient optimization scheme, and an appropriate weight selection algorithm to assist training aggregation with more fair weights. This can be contrasted by the study [32], in which the challenges of fairness within a vertical federated learning system are discussed. The authors propose a framework to approach the challenge of fairness as a constrained optimization problem that can be solved by an asynchronous gradient coordinate-descent ascent algorithm.

The client selection stage is a key point at which fairness criteria can be enforced, as the FL system is able to better accommodate for changes in the distribution of clients at this time. The paper [33] explores the impact of client selection, demonstrating that during the client selection stage of FL training, both performance (in terms of the efficiency of training) and fairness of the final model can be adversely impacted by the choice of clients selected. The authors propose a method to guarantee fairness during client selection, termed RBCS-F, which aimed to solve this challenge through approaching client selection as a Lyapunov optimization problem. A different way to deal with the clients interacting with the FL system is explored in the paper [34]. The authors propose an improved aggregation algorithm, called Center Dropout, which selects a random assortment of the clients participating in the FL system and increases the amount of local learning as to allow for underrepresented clients (the centers) not to be overwhelmed by the learning of bigger clients during aggregation.

Achieving group fairness requires different considerations compared to achieve individual fairness. The authors of [18] explore these considerations and proposes a method to achieve group fairness within a Federated Learning system and resolve the challenge of fairness vs client privacy through the combination of Secure Multiparty Computation (MPC) and Differential Privacy (DP). A different approach to enforcing group fairness is investigated in [35], wherein the authors propose an optimization algorithm, called FedMinMax. This algorithm introduces minimax group fairness within an FL

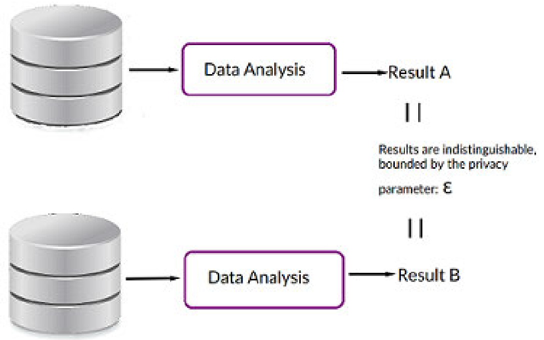
Data Analysis
Result A
Results are indistinguishable, bounded by the privacy parameter: ε
Data Analysis
Result B

FIGURE 4. A simple overview of how differential privacy functions.

context, where the participating clients have limited access to a subset of demographic groups during training. Enforcing group fairness while preserving privacy foundation is another challenge explored in [36]. The authors introduce an algorithm that seeks to enforce group fairness within private federated learning. The algorithm accomplishes this goal through extending a method of differential multipliers to empirical risk minimization with fairness constraints.

Enforcing fairness within specific niches can introduce additional challenges, as explored in the study [37]. In this work, the challenge of fairness is approached in the context of Intelligent Transportation systems, where system heterogeneity is often limited due to the difficulty of addressing fairness. The authors propose an algorithm, called the system heterogeneous fair federated learning (SHFF), to introduce an equipment influence factor that allows the global fairness to be controlled according to the needs of the system.

B. DIFFERENTIAL PRIVACY AND OTHER APPROACHES FROM DIFFERENT DOMAINS

In this subsection, we summarize the characteristics of recent approaches that aim to address the challenges of fairness within a FL system by utilizing techniques originally from other domains.

New approaches that seek to enforce fairness while preserving the privacy in a FL system can incorporate techniques from other domains. Differential Privacy (DP) is a concept in data analysis that aims to ensure that the change of any database entry does not fundamentally change the outcome of the analysis performed [38]. Figure 4 demonstrates an overview of a simple generic DP architecture.

The article [36] demonstrates that the use of DP in an FL context can affect under-represented groups within a model. The authors propose an algorithm, called FPFL, to ensure that group fairness is enforced within a private (privacy-preserving) federated learning system.

Differential Privacy applied at the global model level can be utilized to better control the trajectory of model bias. The study [39] investigates DP in the context of an Internet of Things (IoT) network. The authors propose the idea to control the quality of the global FL model shared with the devices, in each round, based on client contribution and expenditure (participation costs). This is accomplished through DP to curtail global model divergence based on the learning contribution, while the expenditure costs are controlled through adaptive computation and transmission policies for each device, thereby mitigating utility unfairness. The authors also investigate the topic of utility fairness within IoT devices that are participating in a DP-based FL system, identifying unfairness in utility that occurs due to the global model being applied to non-heterogeneous devices.

Investigating the impact of non-IID (independent and identically distributed) data on DP-based FL systems, the article [40] demonstrates the negative impact of this type of data on both model fairness and performance.

Addressing bias in the predictions of a model can help reinforce an ethical framework. The study [41] presents an ethical federated learning model that incorporates differential privacy, federated learning, and fairness metrics to address ethical concerns resulting from prediction bias.

Differential Privacy can be incorporated at a local (single client) level or a global (system wide) level. In [42], the authors discuss the fairness and privacy effect of local DP and global DP when applied to federated learning by designing a fair and privacy quantification mechanism. They demonstrate an acceptable trade-off among accuracy, privacy, and model fairness while quantifying the level of fairness based on the constraints of three definitions of fairness, including demographic parity, equal odds, and equality of opportunity. This study also shows that privacy can come at the cost of fairness, as stricter privacy can intensify discrimination.

Similar to how differential privacy has been used, Proportional Fairness (PF) has its roots in telecommunications and communication networks. Based on cooperative game theory, PF aims to model a given problem as a cooperative bargaining game where players can improve their utility through collaboration [43]. The study [44] investigates PF in the context of a Federated Learning system. The authors propose a novel algorithm, called PropFair, to find fair solutions for FL systems by modeling the overall system as a cooperative game, which allows PF to be modeled as Nash bargaining solutions.

C. FRAMEWORKS AND MODELS OF FAIRNESS CONSTRAINTS

As a result of the growing popularity of federated learning and the recognition of the challenges of fairness, there have been many frameworks and models proposed to better address these challenges in FL systems. In this subsection, we summarize such frameworks and models.

As the quality of clients can greatly impact the performance of an FL system, ensuring that quality clients are chosen to participate is of importance. One approach explored in [45] is to adopt a framework that seeks to address collaborative fairness by considering participant (client) reputations, which are based upon participation to the central model.

Another study [46] investigates how heterogeneous clients can impact the final model. The authors introduce a formal definition of fairness in Federated Learning, namely, fairness via agent-awareness (FAA) that takes into account the different contributions of heterogeneous clients. The authors also present a framework, FOCUS, that utilizes client clustering to achieve higher fairness measured under FAA compared to the standard FedAvg algorithm.

The client selection phase, taken as an individual program, can be modeled as an optimization problem. In the study [33], the authors present two key findings. First, they introduce a model with the fairness guaranteed client selection as a Lyapunov optimization problem. Second, they introduce a C2MAB-based method for estimation of the model's exchange time between each client and the server, which is then used to design a fairness guaranteed algorithm, called RBCS-F, for problem-solving.

A framework that aims to enforce both group and individual fairness would need to reconcile the challenges of both types of fairness and introduce improvements to both. The study [47] proposes a framework, called GIFAIR-FL, to accommodate both group and individual (personalized) fairness settings. This framework is based on the addition of a regularization term that is used to penalize the loss of client groups, resulting in the improved diverse and fair solutions.

Enforcing group fairness requires considering sensitive demographic attributes, with the goal of de-biasing (improving group fairness) a model's output so that the distribution of results across demographic groups is equitable. In [48], the authors propose a framework that incorporates a Variational AutoEncoder (VAE) to aid with semi-centralized adversarial training with the goal of improving group fairness. This encoder is paired with a decoder on the central server side, ensuring client privacy as the encoder remains client-side, while providing greater fairness on sensitive attributes.

When working with unknown test data, balancing fairness and accuracy can be exceptionally challenging. The paper [49] demonstrates that, in the context of unknown test data, introducing fairness constraints to the central FL model will not achieve model fairness. The authors, therefore, propose a fairness-aware agnostic FL framework, called AgnosticFair, that uses kernel reweighing functions to achieve high accuracy and fairness when being used on unknown test data.

In the field of dermatology, the accuracy of a model is of paramount importance, as a misdiagnosis of a skin condition could be life changing. Models used in dermatology have also historically been biased against people of color as the datasets used to train these models tend to be predominantly of fair-skinned individuals. The authors in the study [50] examined how the problem of dermatological disease diagnosis is being addressed with existing deep learning and FL based solutions but suffering from imbalanced data that affects the system's performance, causing significant diagnosis disparities. They propose a fairness-aware FL framework for achieving high fairness and accuracy in the context of dermatological disease diagnosis.

Critical energy infrastructure (CEI) systems are vital to the health of every nation's economy and society. The challenges that these systems face are of utmost importance to the ongoing development of a country while also making them prime targets for cyber-attacks and data leakage. In the study [51], the authors design an asynchronous FL framework that incorporates fairness-awareness and time-sensitive task allocation mechanisms for the use in CEI systems, with the goal of enforcing privacy protections and fairness while addressing the challenge of client node scheduling within CEI systems.

In the Internet of Things, devices are often constrained on both computation and connectivity. The paper [52] proposes an analytical fairness-aware FL model that aims to improve performance on resource-constrained IoT FL systems. The fairness-aware model aims to accomplish this goal, along with addressing the challenges that FL systems face with client scheduling and parameter transmission failure, by introducing a reliable statistically re-weighted aggregation (RSRA) scheme to guarantee the fairness of local clients.

Blockchain technology has the potential to be very impactful on data storage and privacy, as a blockchain allows for greater transparency and accountability. At the same time, blockchains have inherent challenges with privacy due to their core nature, that being two-fold; the blockchain is immutable and data cannot be deleted, and all users with access to the blockchain can view the entire blockchain. By combining blockchains with federated learning systems, the benefits of both technologies can be used to complement each other and offset the downsides [27].

In [29], the authors propose an FL framework with the aim of attempting to simultaneously address the challenges of achieving fairness, integrity, and privacy preservation for all clients by utilizing blockchain technology, local differential privacy, and zero-knowledge proofs. A similar study [30] proposes an architecture based on a trustworthy blockchain implementation for FL with the aim to enhance the accountability and fairness of FL systems. This is accomplished through two contributions: a smart contract-based data-model provenance registry to enable accountability and a weighted fair data sampler algorithm to enhance fairness in training data.

Approaches that are complex and robust may potentially adversely categorize and discard rare client updates, causing unfairness. A simple approach with robustness in mind would address these constraints, which is the approach investigated in the study [53]. The authors propose a general framework for personalized federated learning, called Ditto, that has the goal of inherently providing fairness and robustness, with an addition of a scalable solver.

In [54], the authors propose a theoretical framework with the goal of demonstrating that FL can boost model fairness in comparison to non-federated/distributed algorithms. This framework, called FedFB, provides a private fair learning algorithm designed to be trained on decentralized data, outperforming the FedAVG algorithm. The authors demonstrate that federated learning-based systems are able to outperform those based on non-distributed learning algorithms.

TABLE 1. Examples of spatial-temporal data.

Name	Type
Building Price	Spatial
Nearby Points of Interest	Spatial
Streets	Spatial
Temperature	Temporal
Air Quality	Temporal
Precipitation	Temporal
Building Permits	Spatial-Temporal
Traffic Collisions nearby	Spatial-Temporal
Cell Data Coverage	Spatial-Temporal

D. EVALUATION AND METRICS FOR EVALUATING FAIRNESS FOR FL SYSTEMS

Building on the general challenges with fairness comes the simple question of how to evaluate exactly how fairness is accounted for within a FL system. In this subsection, we summarize some recent efforts in introducing metrics for evaluating fairness, and investigations into fairness evaluation in existing FL systems.

In [55], the authors propose a metric (i.e., the Federated Shapley value) and a method (i.e., the Sharp Federated algorithm) to use Shapley values to determine feature importance for both client and host model features, and balance the model interpretability and data privacy in vertical FL systems. This Shapley value metric can also be used as a method to better understand the value (quality) of client data while also reinforcing the measures of general model performance. Building upon this work, the study [56] evaluates the Federated Shapley value and proposes a new measure, called the Completed Federated Shapley value, to overcome some potential unfairness associated with the original Shapley value.

Personalized FL systems (that is, an FL system wherein each client has a personalized model to address the issue of data distribution) have their own inherent challenges with evaluating fairness. The authors of [57] study the challenges in evaluating personalized FL systems and propose a set of performance and fairness metrics to aid in assessing the effectiveness of a personalized FL system.

Spatial-temporal datasets that focus on urban environments have inherent fairness challenges due to the socio-economic factors inbuilt in those environments, as explored in [58]. Table 1 details common data types used in urban spatial-temporal datasets. Spatial data is categorized as that which does not vary significantly over time, such as the road system, while Temporal data by contrast varies regularly, such as weather. On the other hand, spatial-temporal data vary in both time and space.

An FL system that works upon these datasets therefore must consider and actively adjust client participation for this pollution of the data so that the system does not reinforce these biases. The paper [15] explores fairness within this context, investigating existing metrics and approaches for the measurement and evaluation of fairness within spatial-temporal data based FL systems. The authors also discuss how these metrics may be changed to better address existing challenges within FL.

A study in [59] delves into how the performance of an FL system is related to how similar the local data is distributed amongst clients (that is, the greater the difference in distribution, the lower the accuracy of the model). The authors demonstrate that fairness and accuracy will be negatively impacted, as models that show highly different local data distributions exhibit both higher bias and a significant decrease in fairness compared to the impact to accuracy.

As addressing the impact of fairness for FL systems becomes more widespread, the potential trade-offs between preserving privacy and fairness become more of a concern. The authors of [19] explore these trade-offs by auditing a privacy preserving FL system to evaluate the fairness of its output, using entropy to determine the similarity of the input data and to compare against the output to detect bias.

V. FURTHER DISCUSSION

In the previous sections, we have provided an overview of background techniques, challenges of the various types of fairness in FL, and existing techniques/methods dealing with various aspects of FL fairness. In this section, we will present the selection method of publications in this survey, observations on existing FL fairness studies, as well as open challenges and future work.

A. OBSERVATIONS ON EXISTING FL FAIRNESS STUDIES

As the topic of fairness in Federated Learning is an emerging domain, the number of papers covering the challenges associated with fairness in FL is not very large. The papers included in this study were mainly selected from two sources: the Google Scholar search engine and the DBLP computer science bibliography. Google Scholar has limitations as a search engine (limited results per query, incomplete coverage of scholarly articles, limited coverage of articles in non-English languages, etc). However, recent studies into the platform [60] conclude that it remains the most comprehensive scholarly search engine. To supplement this, the DBLP computer science bibliography that indexes over 6.6 million publications as of 2023 (dblp.org) was also used to source papers. Note that only studies that either were published in major computing journals or were pre-prints of works awaiting publication were included in this survey. In addition, studies cited in a selected paper were also included if they are relevant.

As the Federated Learning field continues to mature, one of the trends that will continue to be seen is the adoption and application of techniques that originate from other fields. This was explored in section IV-B, which examined applications of two concepts, Differential Privacy and Proportional Fairness. Both of these concepts originated from different domains in computer science.

Figure 5 shows the numbers of publications for the general FL, the Differential Privacy concept, and the overlap between the two by year. The data in this figure is from the

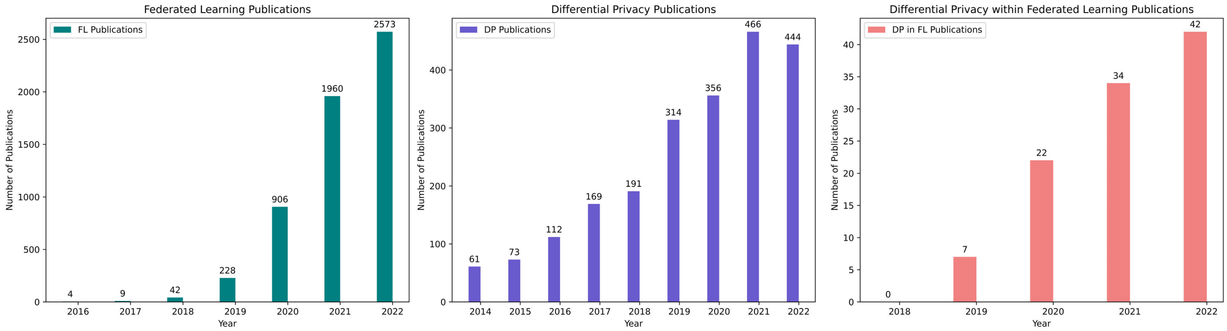

FIGURE 5. A comparison of publications for general FL, DP, and the overlap between the two.

TABLE 2. Literature coverage of fairness.

Related Study	Individual	Group	System Fair-ness	Evaluation of Fairness
Yue-2022	✓	✓		
Zhu-2022, Jiang-2021, Lyu-2020, Chu-2022, Du-2021, Ruckel-2022, Lo-2022, Padala-2021, Li-2021, Yu-2022			✓	
Linardos-2022, Huang-2020, Zhang-2022, Lu-2021, Li-2022	✓		✓	
Liu-2021, Pentyala-2022, Papadaki-2022, Rodriguez-2021, Amiri-2022, Yang-2022, Xu-2022, Zeng-2021, Singh-2020, Singh-2023		✓	✓	
Gu-2022, Alvi-2021	✓	✓	✓	
Wang-2019, Fan-2022, Divi-2021, Yan-2021, Mashhadi-2022, Ozdayi-2021, Salem-2022				✓

*Each related study is labeled by the last name of the first author with the publication year for the corresponding reference.

DBLP computer science bibliography. Looking first at the FL publications, it is quite clear that, after the initial Google publications in 2016 [9], [10], there has been an exponential increase in interest in FL over the years. Similarly, there is a clear trend visible in the DP graph, where there is a year-to-year growth in inclusion of the DP concept in publications. This trend can be explained by the concept's usage in machine learning, where it is a popular inclusion for introducing and addressing privacy constraints. Additionally, this trend can also be seen to a lesser extent in the graph for publications with the overlap between DP and FL, which demonstrates the adoption of DP in FL although the domain of data (the number of years) is small. When techniques and concepts from other domains gain popularity, if they show an applicability in FL, then the crossing interest will show a similar growth.

Table 2 summarizes the related studies included in this survey and their coverage of the different types of fairness (along with the topic of evaluation of fairness). Several trends can be observed by this categorization:

- With the related studies that fall under both group fairness and system fairness, the overarching theme is mitigating bias of under-represented groups through improved algorithmic approaches to the FL training process that lead to greater fairness performance in the global model.
- The majority of the studies that touch upon both individual fairness and system fairness are those that investigate fairness of client selection and propose methods to improve it. As touched upon in section III-A, the client selection stage is a key point for potential improvements to the FL training process, as it occurs before the central model updates, which provides an opportunity to better enforce the similarity notions of individual fairness.
- Studies that only touched upon system fairness were largely focused on improving the FL system itself, with the goal of improving fairness being accompanied by other goals generally around improving overall FL performance.

B. OPEN CHALLENGES AND DIRECTION OF FUTURE WORKS

One major open challenge in this field would be the current limitations with defining and measuring fairness within a Federated Learning system. Not all recently proposed approaches include how a given FL system performs in regard to fairness, and typically, the measures that are provided are not uniform or comparable across implementations. As fairness within the broader Machine Learning field has progressed as the field matured, so too must fairness within FL, and perhaps inspiration on how standardized measures were implemented can be taken from this closely related field. Generalized benchmarks and datasets dedicated to evaluating fairness in the distributed manner of FL should be one direction of future efforts.

Another major open challenge is maintaining a good balance among competing constraints of fairness, integrity, accuracy, privacy, scalability, and robustness without losing one or the other. As addressing fairness within an FL system is a balance of trade-offs, achieving fairness can be thought of as a multi-objective task, which weighs the various objectives and contexts around the fairness that is set as a goal. Equally as important is ensuring that maintaining these fairness constraints does not impose a negative effect on the system, falling victim to over-correcting biases. Building and utilizing datasets representative of that are both equitable to under-represented populations and realistic is a great challenge but would allow for more transparency and ease in the evaluation of these systems. Only limited work has been reported in this area. Further research efforts are required in this direction.

Another area of interesting future work is the improvement and optimization of fairness evaluation methods so they can be integrated into real-world use cases. With current methods (such as the Completed Federated Shapley framework [56]), one of the main factors limiting their adoption is the time and cost of computing contribution valuations. This restricts these methods to unrealistic use cases where aspects of the FL system (namely the client pool or models) need to be restricted to where it is not representative of how FL systems are used in real world conditions. Future work in this area can hopefully introduce improvements with these methods so they can better be integrated and used to enforce fairness guarantees in FL systems.

The other related problems that continue to be challenging and require further studies for fairness in FL systems include interpretability of fairness enforcement, fairness mechanisms resistant to different types of privacy attacks, domain/application-specific fairness schemes, theoretical analysis of fairness properties, improved algorithms for fairness optimization, and fairness quantification in complex settings.

Another general trend of improvement in this field is the availability of open code to reproduce results. A significant number of studies covered in this survey did not readily provide code. As a result, this survey is unable to comment on the reproducibility of results for the studies covered.

VI. CONCLUSION

As distributed systems become more embedded in the frameworks of society that we interact with every day, it is imperative that the decisions that they make are as fair and ethical as possible. A Federated Learning system can provide great privacy benefits to the participants of the system, but these benefits need to be ensured alongside high fairness constraints to ensure the goals of the system are accomplished. Although the FL fairness has attracted much attention from researchers recently, the research advance is still in its infancy stage. The reported studies have covered various types of fairness including individual and group fairness of data as well as fairness of the system. Some of them handle multiple types simultaneously, while most of them deal with the system fairness in some way, which is unique for a FL system. A number of algorithmic approaches to addressing fairness in FL have been proposed. Some concepts such as differential privacy from other fields have been leveraged to benefit the relevant study in the FL context. Developing a uniform or comparable fairness measure/evaluation for FL systems as well as balancing FL fairness and other constraints (e.g., accuracy, integrity, privacy, scalability, etc.) continue to be major challenges in research on the fairness of Federated Learning.

REFERENCES

[1] J. Kleinberg, S. Mullainathan, and M. Raghavan, "Inherent trade-offs in the fair determination of risk scores," 2016, *arXiv:1609.05807*.

[2] D. Pessach and E. Shmueli, "A review on fairness in machine learning," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–44, Mar. 2023.

[3] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Develop.*, vol. 3, no. 3, pp. 210–229, Jul. 1959.

[4] S. J. Russell, *Artificial Intelligence a Modern Approach*. London, U.K.: Pearson, 2010.

[5] F. Pereira and S. Borysov, *Machine Learning Fundamentals Mobility Patterns, Big Data and Transport Analytics*. Chichester, U.K.: Elsevier, 2019.

[6] T. Ben-Nun and T. Hoefler, "Demystifying parallel and distributed deep learning: An in-depth concurrency analysis," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–43, Jul. 2020.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] S. Verma and J. Rubin, "Fairness definitions explained," in *Proc. IEEE/ACM Int. Workshop Softw. Fairness (FairWare)*, May 2018, pp. 1–7.

[9] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," 2016, *arXiv:1610.02527*.

[10] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[11] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[12] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: A survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140699–140725, 2020.

[13] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

[14] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Comput. Ind. Eng.*, vol. 149, 2020, Art. no. 106854.

[15] A. Mashhadi, A. Kyllo, and R. M. Parizi, "Fairness in federated learning for spatial–temporal applications," 2022, *arXiv:2201.06598*.

[16] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proc. 3rd Innov. Theor. Comput. Sci. Conf.*, Jan. 2012, pp. 214–226.

[17] A. Chouldechova and A. Roth, "The frontiers of fairness in machine learning," 2018, *arXiv:1810.08810*.

[18] S. Pentyala, N. Neophytou, A. Nascimento, M. De Cock, and G. Farnadi, "PrivFairFL: Privacy-preserving group fairness in federated learning," 2022, *arXiv:2205.11584*.

[19] A. B. Salem, B. Khalfoun, S. B. Mokhtar, and A. Mashhadi, "Quantifying fairness of federated learning LPPM models," in *Proc. 20th Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2022, pp. 569–570.

[20] A. Chouldechova, D. Benavides-Prado, O. Fialko, and R. Vaithianathan, "A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 134–148.

[21] A. Giloni, E. Grolman, Y. Elovici, and A. Shabtai, "FEPC: Fairness estimation using prototypes and critics for tabular data," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 4877–4884.

[22] W. Huang, T. Li, D. Wang, S. Du, J. Zhang, and T. Huang, "Fairness and accuracy in horizontal federated learning," *Inf. Sci.*, vol. 589, pp. 170–185, Apr. 2022.

[23] A. Abay, Y. Zhou, N. Baracaldo, S. Rajamoni, E. Chuba, and H. Ludwig, "Mitigating bias in federated learning," 2020, *arXiv:2012.02447*.

[24] G. Xia, J. Chen, C. Yu, and J. Ma, "Poisoning attacks in federated learning: A survey," *IEEE Access*, vol. 11, pp. 10708–10722, 2023.

[25] A. K. Singh, A. Blanco-Justicia, and J. Domingo-Ferrer, "Fair detection of poisoning attacks in federated learning on non-i.i.d. data," *Data Mining Knowl. Discovery*, pp. 1–26, Jan. 2023.

[26] A. K. Singh, A. Blanco-Justicia, J. Domingo-Ferrer, D. Sánchez, and D. Rebollo-Monedero, "Fair detection of poisoning attacks in federated learning," in *Proc. IEEE 32nd Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2020, pp. 224–229.

[27] F. Yu, H. Lin, X. Wang, A. Yassine, and M. S. Hossain, "Blockchain-empowered secure federated learning system: Architecture and applications," *Comput. Commun.*, vol. 196, pp. 55–65, Dec. 2022.

[28] Z. Tian, M. Li, M. Qiu, Y. Sun, and S. Su, "Block-DEF: A secure digital evidence framework using blockchain," *Inf. Sci.*, vol. 491, pp. 151–165, Jul. 2019.

[29] T. Rückel, J. Sedlmeir, and P. Hofmann, "Fairness, integrity, and privacy in a scalable blockchain-based federated learning system," *Comput. Netw.*, vol. 202, Jan. 2022, Art. no. 108621.

[30] S. K. Lo, Y. Liu, Q. Lu, C. Wang, X. Xu, H.-Y. Paik, and L. Zhu, "Toward trustworthy AI: Blockchain-based architecture design for accountability and fairness of federated learning systems," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3276–3284, Feb. 2023.

[31] H. Zhu, M. Yang, J. Kuang, H. Qian, and Y. Zhou, "Client selection for asynchronous federated learning with fairness consideration," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2022, pp. 800–805.

[32] C. Liu, Z. Fan, Z. Zhou, Y. Shi, J. Pei, L. Chu, and Y. Zhang, "Achieving model fairness in vertical federated learning," 2021, *arXiv:2109.08344*.

[33] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya, "An efficiency-boosting client selection scheme for federated learning with fairness guarantee," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1552–1564, Jul. 2021.

[34] A. Linardos, K. Kushibar, and K. Lekadir, "Center dropout: A simple method for speed and fairness in federated learning," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, Sep. 2022, pp. 481–493.

[35] A. Papadaki, N. Martinez, M. Bertran, G. Sapiro, and M. Rodrigues, "Minimax demographic group fairness in federated learning," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2022, pp. 142–159.

[36] B. Rodríguez-Gálvez, F. Granqvist, R. van Dalen, and M. Seigel, "Enforcing fairness in private federated learning via the modified method of differential multipliers," 2021, *arXiv:2109.08604*.

[37] Y. Jiang, G. Xu, Z. Fang, S. Song, and B. Li, "Heterogeneous fairness algorithm based on federated learning in intelligent transportation system," *J. Comput. Methods Sci. Eng.*, vol. 21, no. 5, pp. 1365–1373, Nov. 2021.

[38] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation*. Xi'an, China: Springer, Apr. 2008, pp. 1–19.

[39] S. A. Alvi, Y. Hong, and S. Durrani, "Utility fairness for the differentially private federated-learning-based wireless IoT networks," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 19398–19413, Oct. 2022.

[40] S. Amiri, A. Belloum, E. Nalisnick, S. Klous, and L. Gommans, "On the impact of non-IID data on the performance and fairness of differentially private federated learning," in *Proc. 52nd Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. Workshops (DSN-W)*, Jun. 2022, pp. 52–58.

[41] M. Padala, S. Damle, and S. Gujar, "Federated learning meets fairness and differential privacy," in *Neural Information Processing*. Bali, Indonesia: Springer, Dec. 2021, pp. 692–699.

[42] X. Gu, Z. Tianqing, J. Li, T. Zhang, W. Ren, and K.-K.-R. Choo, "Privacy, accuracy, and model fairness trade-offs in federated learning," *Comput. Secur.*, vol. 122, Nov. 2022, Art. no. 102907.

[43] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Apr. 1998.

[44] G. Zhang, S. Malekmohammadi, X. Chen, and Y. Yu, "Proportional fairness in federated learning," 2022, *arXiv:2202.01666*.

[45] L. Lyu, X. Xu, Q. Wang, and H. Yu, "Collaborative fairness in federated learning," in *Federated Learning: Privacy and Incentive*. 2020, pp. 189–204.

[46] W. Chu, C. Xie, B. Wang, L. Li, L. Yin, H. Zhao, and B. Li, "FOCUS: Fairness via agent-awareness for federated learning on heterogeneous data," 2022, *arXiv:2207.10265*.

[47] X. Yue, M. Nouiehed, and R. Al Kontar, "GIFAIR-FL: A framework for group and individual fairness in federated learning," *INFORMS J. Data Sci.*, vol. 2, no. 1, pp. 10–23, Apr. 2023.

[48] Y. Yang and B. Jiang, "Towards group fairness via semi-centralized adversarial training in federated learning," in *Proc. 23rd IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jun. 2022, pp. 482–487.

[49] W. Du, D. Xu, X. Wu, and H. Tong, "Fairness-aware agnostic federated learning," in *Proc. SIAM Int. Conf. Data Mining (SDM)*. Philadelphia, PA, USA: SIAM, 2021, pp. 181–189.

[50] G. Xu, Y. Wu, J. Hu, and Y. Shi, "Achieving fairness in dermatological disease diagnosis through automatic weight adjusting federated learning and personalization," 2022, *arXiv:2208.11187*.

[51] J. Lu, H. Liu, Z. Zhang, J. Wang, S. K. Goudos, and S. Wan, "Toward fairness-aware time-sensitive asynchronous federated learning for critical energy infrastructure," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3462–3472, May 2022.

[52] Z. Li, Y. Zhou, D. Wu, T. Tang, and R. Wang, "Fairness-aware federated learning with unreliable links in resource-constrained Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17359–17371, Sep. 2022.

[53] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and robust federated learning through personalization," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 6357–6368.

[54] Y. Zeng, H. Chen, and K. Lee, "Improving fairness via federated learning," 2021, *arXiv:2110.15545*.

[55] G. Wang, "Interpret federated learning with Shapley values," 2019, *arXiv:1905.04519*.

[56] Z. Fan, H. Fang, Z. Zhou, J. Pei, M. P. Friedlander, C. Liu, and Y. Zhang, "Improving fairness for data valuation in horizontal federated learning," in *Proc. IEEE 38th Int. Conf. Data Eng. (ICDE)*, May 2022, pp. 2440–2453.

[57] S. Divi, Y.-S. Lin, H. Farrukh, and Z. B. Celik, "New metrics to evaluate the performance and fairness of personalized federated learning," 2021, *arXiv:2107.13173*.

[58] A. Yan and B. Howe, "EquiTensors: Learning fair integrations of heterogeneous urban data," in *Proc. Int. Conf. Manage. Data*, Jun. 2021, pp. 2338–2347.

[59] M. S. Ozdayi and M. Kantarcioglu, "The impact of data distribution on fairness and robustness in federated learning," in *Proc. 3rd IEEE Int. Conf. Trust, Privacy Secur. Intell. Syst. Appl. (TPS-ISA)*, Dec. 2021, pp. 191–196.

[60] M. Gusenbauer, "Google scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases," *Scientometrics*, vol. 118, no. 1, pp. 177–214, Jan. 2019.

SEAN VUCINICH received the B.S. degree in computer science from California State University, Monterey Bay. He is currently pursuing the M.S. degree in computer and information science with the University of Michigan-Dearborn. He is also with the Center for Academic Innovation, University of Michigan. His research interests include federated learning, algorithmic fairness, distributed machine learning, and ethics in artificial intelligence systems.

QIANG ZHU (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Waterloo, Canada, in 1995. He is currently a Professor and the Chair of the Department of Computer and Information Science, University of Michigan-Dearborn, USA, where he has been honored as the William E. Stirton Professor, an ACM Distinguished Scientist, and an IBM CAS Faculty Fellow. His current research interests include query optimization for advanced database systems, big data processing and analytics, streaming data processing, spatio-temporal data processing, data mining, federated learning, and data security and privacy.