Navya Annapareddy and Jade Preston
Machine Learning Systems for Fairness

Problem Description: Currently, there are many contexts where data sharing is difficult or constrained by security and distribution limitations. One common domain where this is a consideration is in Healthcare where data is often governed by data-use-ordinances like HIPAA. On the other hand, larger sample sizes allow models to better generalize on account of the potential for more variability and balancing underrepresented classes.
Federated learning is a type of distributed learning model that allows data to be trained in a decentralized manner. This, in turn, addresses data security, privacy, and vulnerability considerations as data itself is not shared across a given learning network's nodes. Some challenges to federated learning include: node data may not be independent and identically distributed (iid), relatively high levels of communication between network machines is needed, and heterogeneity in the individual nodes with respect to bias and size of data samples.

Considerations: Federated learning has much potential to curb security vulnerabilities normally present in data sharing processes. Data protection would still need to be enforced at the host level but would not require a dedicated TEE to ensemble or share models unless desired. Distributed training methods typically focus on parallelization to obtain less computationally expensive training while federated methods have focused on addressing node heterogeneity. Local systems are considered suitable, but HPC systems might be necessary for high dimensional data. Encryption can be enforced at the model sharing level through secure communication protocols or with data keys (envelope encryption). Finally, secure addition of peers to a network must also be considered.

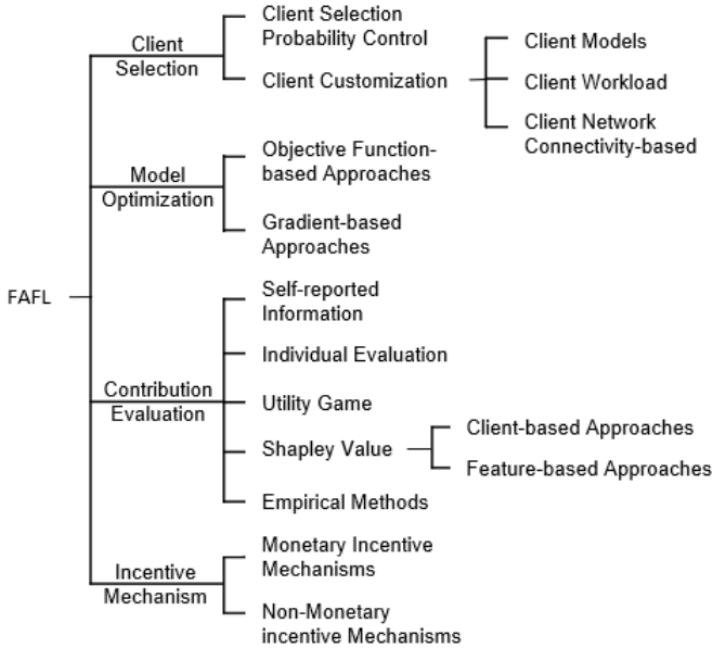

Figure 1: Proposed fairness mechanism taxonomy by Shi et al. [1]

Existing Surveys:
We identified two existing surveys that overview the current state of fairness in federated learning.

Firstly, the paper by Shi et al. takes a taxonomy first approach to classifying fairness approaches. Specifically, the scope of their paper discusses ideal metrics for each classified group of mechanisms but does not benchmark or comprehensively assign existing methods. The authors form a taxonomy of categorized fairness approaches, namely:

1. Accuracy Parity: Mechanisms that consider fairness as a function of fairness
2. Good Intent Fairness: Mechanisms that consider impact and intent metrics
3. Selection Fairness: Mechanisms that deal with client or node selection
4. Contribution Fairness: Mechanisms that equalize contribution of groups
5. Regret Distribution Fairness: Mechanisms that minimize regret metrics within groups
6. Expectation Fairness: Mechanisms that equalize fairness within groups

Secondly, Quy et al. review domain considerations of fairness. [2] Specifically, they examine bias in federated learning datasets in the financial, health, and social domains. They also review considerations of synthetically biased data. In each domain, the authors explored challenges to fairness such as data cardinality, imbalance, and the validity or realisticness of assumptions of independence. They also identify several fairness approaches that were not included in the taxonomy approach taken by Shi et al. such as multi-discrimination, temporal fairness, and distributional fairness.

Gaps in Research:
From these existing survey analyses, we identify several gaps in research. Firstly, there are antiquated categorizations of existing implementations by approach in the taxonomy proposed (according to more recent implementation). Alternative groupings or organization strategies could make more sense in the context of developments in the field.

Both surveys also primarily focused on tabular data with a lack of consideration for image and other high dimensional datasets.

Finally, neither was a comprehensive overview of fairness considerations in the field and only marginally touched on issues like difficulty in aligning metrics.

Proposed Contribution:
In the context of these gaps, we propose a comprehensive survey on the state of fairness in federated learning with several additions on the topics first discussed by these two existing reviews.

We propose a comprehensive survey that includes newer fairness mechanisms and expands the scope of the previous reviews to text, image, and other high dimensional implementations. We aim to differentiate the fairness approach and update the earlier taxonomy or pursue other organization methods based on developments in the field. We also plan to differentiate

considerations for unsupervised and supervised models. Next, we will review security and scaling considerations in the context of fairness, such as the security of metric aggregation and model sending, data robustness, and scalability mechanisms such as peer selection and addition. Finally, we plan on creating a life cycle specific view by building a discrete model process representation for stages of the systems process (ie: pre-training vs. post) that each mechanism is attributed to with the goal of expanding past the assumption of single fairness mechanism systems and instead providing a basis to a build more complex,multi-mechanism, federated learning approaches.

Proposed Implementations to Include:
Table 1 displays the implementation papers we plan to incorporate into this research plan. The table is a breakdown of the dataset type and the metrics used within each paper. Afterward each of the implementation papers are summarized. The only paper left without a summary is 'Federated Learning on Clinical Benchmark Data: Performance Assessment'; the rights to review this text still need to be obtained [4]. We have provided a table breakdown and summarized each of the papers to show where our project can fill the gaps in literature found (dataset or metrics used).

Implementation Papers	Type	Metric
FedScale: Benchmarking Model and System Performance at Scale	Image, Video, Text Audio	Accuracy, Loss
Federated Learning on Clinical Benchmark Data: Performance Assessment	Signal (ECG), Image	AUROC and F1
Fair Federated Learning via Bounded Group Loss	Text	Custom Loss
Fairness and Accuracy in Federated Learning	Synthetic Image, Text	Accuracy, Loss
Reputation Mechanism, Collaborative Fairness, Adversarial Robustness in FL	Image, Text	Pearson Correlation Coefficient
Ditto: Fair and Robust Federated Learning Through Personalization	Image, Text	Accuracy, variance of robustness
Utility Fairness for the Differentially Private Federated Learning	Image	Accuracy; negative log likelihood loss
Resource Management and Fairness for FL over Wireless Edge Networks	Image	Accuracy, Loss

Table 1: Implementation Papers

FedScale: Benchmarking Model and System Performance of Federated Learning at Scale

Lai et al. introduce Fedscale. [3] Fedscale is novel in that it is an open source benchmarking platform that incorporates 20 federated learning datasets. These datasets encompass a myriad of data types from video to text data. Additionally, the authors employ an evaluation system called Fedscale RunTime. To that end the authors complete multiple experiments using Fedscale Runtime to demonstrate Fedscales benchmarking enablement capabilities.

Fair Federated Learning via Bounded Group Loss

The authors of this paper propose a methodology for federated learning that adheres to group fairness requirements. [5] Hu et al. do this by expanding on the idea of bounded group loss. Bounded group loss is the idea that the average loss across all data points with each group are the same. Additionally, within this paper the authors introduce an optimization formulation guaranteeing adherence to implemented group fairness requirements.

Fairness and Accuracy in Federated Learning

Huang et al. introduces FedFa. [6] FedFa is a federated learning optimization algorithm. This algorithm incorporates a double momentum gradient methodology to assist with speeding up model convergence. Additionally, Fed Fa implements a weighting scheme based on device learning performance. Ultimately, within this paper, federated learning unfairness is dealt with regarding object utilization of clients.

A Reputation Mechanism Is All You Need: Collaborative Fairness and Adversarial Robustness in Federated Learning

Xu et al. introduces a methodology that balances fairness and robustness within Federated Learning. [7] The authors do this by rewarding participants based on their contribution. Each participant is evaluated based on their workload gradient. Quality participants are rewarded but poor performers are removed each round of federated learning.

Ditto: Fair and Robust Federated Learning Through Personalization

This paper asserts that robustness and fairness are competing concepts in many federated learning frameworks. [8] Li et al. discuss the idea that recent papers attempt to ensure fairness while also remaining robust, but none of the papers directly assert that fairness and robustness are in fact competing concepts. Addressing this idea the authors introduce a methodology with an objective that makes federated learning more personalized.

Utility Fairness for the Differentially Private Federated Learning

Because devices are inherently different, this results in utility unfairness. [9] Alvi et al. propose moderating the quality that is sent to the global model. Evaluating the contributions of every device, a utility function will assess the learning gain and cost per federated learning round. The quality sent to the global model will be assessed to mitigate for malicious users.

Resource Management and and Model Personalization for Federated Learning over Wireless Edge Networks

The estimated amount of data by 2025 is projected to be enormous. [10] Balakrishnan et al. state there are three factors contributing to this phenomenon: large number of machines, large number of big models run on these machines and 5G is often used in federated learning. In this paper, these authors proposed incorporating resource management for federated learning.

Ultimately, they generated a federated meta-model by personsonalizing models across clients and Federated Learning benchmarks.

Current Progress:
Currently, we have reviewed implementations of fairness in federated learning and are exploring how to categorize them in the context of the existing proposed taxonomy and fairness approach classes. Specifically, we have focused on papers and fairness mechanisms applied to image, signal, and other non-tabular dataset types as we propose expanding the scope of our survey beyond previous analyses. We also propose contributing a life cycle view of fairness approaches and are in progress of reviewing how best to categorize mechanisms in this context.

Our immediate next steps consist of 1) assigning each implementation for review against existing groupings 2) reviewing gaps in the existing groupings 3) proposing new or augmenting existing groupings to account for these gaps 4) proposing life cycle stages to which mechanisms can be classified under and 5) classifying approaches and gathering insights into how this view can help support future research.

Conclusion:
In conclusion, after reviewing existing fairness implementations and surveys in the field of federated learning, we have identified research gaps and propose an up to date survey furthering analysis on existing fairness approaches and their considerations. Specifically, we aim to do so by contributing a discrete life cycle approach, reviewing SOTA fairness mechanisms, and identifying remaining challenges in the context of aligning fairness approaches.

References

1. https://arxiv.org/pdf/2110.00530.pdf
2. https://arxiv.org/pdf/2111.01872.pdf
3. https://arxiv.org/pdf/2105.11367.pdf
4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652692/
5. https://arxiv.org/pdf/2203.10190
6. https://arxiv.org/pdf/2012.10069
7. https://arxiv.org/pdf/2011.10464
8. http://proceedings.mlr.press/v139/li21h/li21h.pdf
9. https://arxiv.org/pdf/2109.05267
10. https://www.mdpi.com/2224-2708/10/1/17/pdf?version=1616039824