2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

# Responsible and Effective Federated Learning in Financial Services: A Comprehensive Survey

Yueyue Shi, Hengjie Song and Jun Xu

*Abstract*— **The financial sector is increasingly leveraging Artificial Intelligence (AI) to deliver intelligent, automated, and personalized services. However, it encounters significant data privacy challenges due to the dispersion of financial data across various entities. Federated Learning (FL) offers a potential solution by facilitating AI model training at the source of data, albeit with certain challenges. Irresponsible utilization of FL can compromise stakeholder interests, and the prevalent heterogeneity in data spaces in numerous financial FL scenarios can impede FL's performance. These complications necessitate the development of a Responsible and Effective Federated Learning (RE-FL) system in finance. In this paper, we explore the interdisciplinary field of RE-FL in finance and guide readers to understand this area thoroughly. We present a taxonomy of RE-FL approaches that address the concerns of stakeholders in FL-based financial services and identify six major dimensions: accountability, controllability, fairness, privacy, security, and effectiveness. We also propose potential directions for future research. To our understanding, this is the first literature review conducted on RE-FL in the financial sector.**

## I. INTRODUCTION

The advent of fintech has yielded ample quality data for financial corporations, enhancing decision-making processes in services like fraud prevention, risk control, and marketing via AI integration, thus invigorating the real economy [1]. However, the sensitivity of certain financial data, held by different institutions, prompts privacy concerns [2]. The intensifying implementation of global privacy-protection regulations may pose challenges to the integration of sensitive data into AI-driven financial services.

Federated Learning (FL) is a promising solution for the development dilemma faced by AI-based financial services [3]. Existing FL techniques address privacy-preserving AI-based financial services in two settings: Horizontal FL (HFL) and Vertical FL (VFL) [4]. HFL-based financial service systems train global Machine Learning (ML) models collaboratively on data in the common feature space when the dimension of these shared features is greater than the dimension of the unique features. Conversely, VFL-based financial service systems train models on overlapping samples with IDs shared by each organizations when the dimension of the unique features is larger than that of the common features.

Yueyue Shi is a PHD candidate of School of Software Engineering, South China University of Technology, 510006 Guangzhou, China msshiyueyue@mail.scut.edu.cn

Hengjie Song is a professor of School of Software Engineering, South China University of Technology, 510006 Guangzhou, China sehjsong@scut.edu.cn

Jun Xu is the executive director of machine learning engineering of DCDA, Standard Chartered Bank, 018981, Singapore xujun@ieee.org

Despite the promising future, FL-based financial service systems encounter various predicaments. One challenge faced by FL-based financial service systems is the need for financial models with high performance. However, these systems are afflicted by the data space heterogeneity [5] within distributed financial datasets. Additionally, there is a growing concern that irresponsible utilization of FL may give rise to counter-effects and trust issues [6]. These issues include compromised stakeholders' trust due to non-accountability, inequitable treatment of different stakeholders, privacy concerns arising from high-level privacy leaks, security threats due to the distributed nature of the data, and uncontrollable decisions and systems. These vulnerabilities significantly hamper the development and deployment of FL-based financial service systems, potentially leading to serious economic and societal problems. Thus, over the years, Responsible and Effective FL (RE-FL) research has focused on various perspectives such as definition, methodology, and assessment of accountability, controllability, fairness, privacy, security, as well as the effectiveness of FL models.

This paper contributes to the existing AI and fintech literature by offering a comprehensive perspective on RE-FL and highlighting its crucial aspects that are often overlooked in existing surveys for building responsible and effective FL-based financial service systems. Specifically, this paper:

(1) provides a detailed analysis of the FL-based financial service systems, with focus on the diverse stakeholders involved, system architectures and learning processes, inherent nature of information asymmetry, basic assumptions, and the significance responsible and effective AI dimensions;

(2) proposes a multi-layered taxonomy of RE-FL, categorizing previous research based on the major techniques that improve FL-based financial service systems. At lower levels, we summarize approaches supporting accountability, controllability, explainability, fairness, privacy, security, and effectiveness. Our taxonomy is the first of its kind and provides a novel perspective on prior research in this domain;

(3) outlines future research directions towards constructing responsible and effective FL-based financial service systems by analyzing inadequacies of current literature and presenting potential ways forward for each direction.

## II. AN OVERVIEW OF FL-BASED FINANCIAL SERVICE SYSTEMS

This section provides background on FL-based financial service systems, including stakeholder introductions, system architectures and learning processes, analysis of information

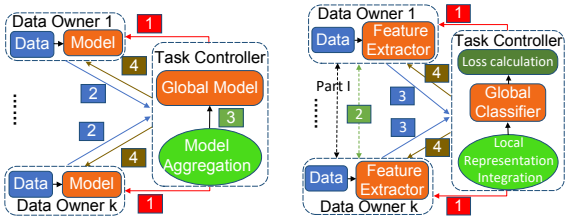
979-8-3503-0123-6/23/$31.00 ©2023 IEEE

[6] asymmetry and basic assumptions, introducing the applications and platforms, and summarizing building requirements for stakeholder responsibility and effectiveness.

### A. Stakeholders

Based on the analysis of the conceptual framework of AI in financial services, stakeholders in FL-based financial service systems are categorized into four types: (1) task controllers, (2) data owners, (3) application objects and (4) regulators [7]. Task controllers are responsible for building reliable and functional financial models. Data owners help train the models for a reward, and the trained models are then used to provide intelligent financial services to application objects. Regulators oversee the use of FL in the financial services to ensure adherence to laws, regulations, and ethical standards.

### B. System architecture and learning process




(a) The architecture and learning process of a HFL-based financial service system. 1: selection; 2: local model uploading; 3: model aggregation; 4: global model distribution.

(b) The architecture and learning process of a VFL-based financial service system. Part 1: encrypted ID alignment; Part 2: encrypted model training. 1: sending public keys; 2: exchanging intermediate results; 3: computing gradients and loss; 4: updating model.

Fig. 1. An illustration the architecture and learning process of HFL and VFL-based financial service systems.

The HFL-based financial service systems can integrate private distributed data with the same features from large-scale and micro corporations and institutions to train models for intelligent financial services. As shown in Figure 1(a), the HFL-based financial service system updates a global model periodically, with model parameters transmitted to the task controller for fusion or averaging. The updated global model is then disseminated to data owners for local model updating.

The VFL-based financial service systems extend HFL-based systems, involving data partition with sample overlap. For instance, when a bank wants to develop a credit rating model, it seeks help from organizations in other industries to obtain complementary information. A VFL model is constructed by both organizations, and each keeps a partial model. As shown in Figure 1(b), the VFL-based system first uses encryption-based sample ID alignment techniques to validate common sample IDs without revealing private IDs. During model training, data owners and task controllers collaboratively train the VFL model in four stages: (1) sending public keys; (2) exchanging intermediate results; (3) computing gradients and loss; (4) updating models.

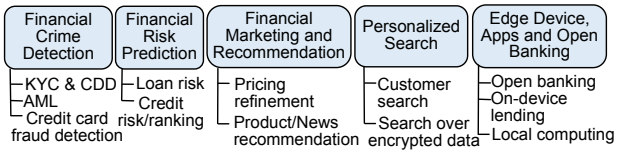



Fig. 2. Application cases of FL-based financial service systems.

### C. Information asymmetry

The distributed design of financial service systems using FL leads to considerable information asymmetry among stakeholders. The asymmetric information consists of: (1) model information from the task controller side, including details about client selection, global update processes, and contribution assessment; (2) model information from the data owner side, encompassing data collection, pre-processing, and local model training; (3) system and data details from the data owner side, including computational and communication capabilities and costs, data quantity and quality, and expenses related to data collection, communication, and computation. Unless shared mutually within the boundaries of the privacy agreement, the asymmetric information within each party remain obscure to other stakeholders.

### D. Common assumptions

Under the strict regulations for financial services, two common assumptions are made in FL-based financial service systems. (1) Honest participants (i.e., task controllers and data owners): They follow the FL training protocol, including truthful uploading local model information or performing aggregation. (2) Non-collusion participants: They do not collude with other participants to manipulate the training process or leak private data.

### E. Application cases and platforms

In recent years, numerous domestic businesses have partnered with financial entities, healthcare firms, and local governments to conduct research and implement FL strategies for financial applications, as illustrated in Figure 2. For example, to capture patterns across financial institutions more effectively, the financial sector has leveraged FL techniques and graph learning approaches to collaboratively learn effective crime detection models [8] [9]. The implementation of FL has also focused on credit risk prediction [10], marketing/recommendation [11], personalized search [12], on-device services and open banking [13], with enhancement attributed to dimension expansion and data feature concatenation. Please refer to Figure 2 for more detailed application cases. Prominent FL platforms used for experimental financial applications include Tencent's PowerFL platform[1], WeBank's open-source FL framework called FATE[2], and Baidu Paddle's PaddleFL platform[3].

[1]https://github.com/Angel-ML/angel
[2]https://github.com/FederatedAI/FATE
[3]https://github.com/PaddlePaddle/PaddleFL

### F. Desirable Responsible and Effective AI Dimensions

According to recent ethics guidelines [6], this paper identifies five key dimensions covered by existing FL literature: (1) accountability; (2) Controllability; (3) Fairness; (4) Privacy; (5) Security. The effectiveness of FL-based financial service systems should also be a focus.

## III. The Taxonomy of RE-FL

This section introduces a taxonomy of RE-FL (see Figure 3). Following the proposed taxonomy, this section examines previous studies and their limitations.

### A. Accountable FL

In the context of information asymmetry, three types of information are required to be audited in order to achieve accountability in FL-based financial service systems: (1) aggregation results, (2) local model information, and (3) local system information. As delineated in Figure 3, based on their technological underpinnings, the accountable FL strategies advanced in HFL systems bifurcate into two categories: accountable HFL devoid of blockchain technology and accountable HFL integrated with blockchain technology.

*1) Non-blockchain based accountable HFL:* Non-blockchain based accountable HFL approaches employ interactive proof protocol, reputation mechanism and influence function to validate the asymmetric information in FL-based financial service systems. (a) Interactive proof protocols audit aggregation results in HFL systems, as demonstrated in recent studies [14]. The task controller returns the aggregated model in encrypted form or with computed proof for data owners to verify. (b) Reputation schemes assist auditing the correctness of local model information [15]. Reputable data owners participate more honestly and task controllers can design specific contracts with different data owners based on the levels of data quality and corresponding payoffs [16]. If a data owner fails to fulfill contractual obligations, task controllers can adjust reputation and withhold payment. (c) Within HFL systems, influence functions has appeared as a means of auditing a data owner's data quality for task controllers [17].

*2) Blockchain based accountable HFL:* The blockchain-based HFL systems utilize a select group of external organizations or internal data owners as workers, who are responsible for validating the asymmetric information and aggregating local model updates in accordance with the blockchain consensus protocol. (a) At the onset of the HFL process, each data owner creates a block to document the local data information which is then recorded on the blockchain [18]. In the event of any dispute, the task controllers are able to authenticate the integrity of local data by comparing it with the corresponding block stored on the blockchain. (b) After the local training stage is completed, the plaintext or cipertext of local updated model of each data owner is transmitted to randomly chosen workers for the identification and selection of any misbehaving data owners [19]. (c) Following the local updated model audition, a trusted committee is formed consisting of randomly and dynamically selected reputable workers, who proceed with aggregating the local models by majority voting [20]. The trust committee relies on the verifying contract to derive and record verifiable aggregation outcomes in the blockchain. Nevertheless, this approach proves effective only when the workers are honest and free from collusion.

*In summary*, current research on accountable FL revolves around enabling verification through essential information and examining specific activities and roles in HFL settings. Further research is needed on the auditability of financial service systems based on VFL.

### B. Controllable FL

Controllability in FL-based financial systems encounters challenges due to information asymmetry. First, inaccurate predictions may happen, and new FL schemes are needed to incorporate feedback from the application objects. Second, data owners may leave the FL collaboration and not want task controllers to retain knowledge from their data. Expressing deletion intentions manually is challenging in the FL model construction through interactive communication between data owners and task controllers. Third, the FL process incurs communication and computation costs for the participating parties with limited resources and potentially volatile conditions, such as mobile edge devices. One intuitive approach is to customize configurations based on computation and communication budget, including participation frequency and update vector compression. However, ensuring model performance is challenging when allowing a high degree of freedom in tuning these configurations. In this context, researchers have proposed federated feedback learning, federated unlearning and cost-controllable FL as illustrated in Figure 3.

Federated feedback learning has been proposed as an effective method for leveraging both positive and negative user feedback in HFL systems [21]. Specifically, positive feedback samples, which represent correct model predictions, are incorporated into the training data, while negative feedback is utilized as complementary labels for training HFL models. However, this approach is only applicable when the application objects are data owners.

Class-level and client-level federated unlearning approaches have been proposed to delete learned knowledge from the global model in HFL-based financial service systems. (a) Class-level federated unlearning approaches first evaluate term frequency and inverse document frequency to obtain the most discriminative channels of the target class. Then the class-level federated unlearning approaches prune the relevant channel of target class to unlearn the class knowledge of the global model [22]. (b) Client-level federated unlearning aims to scrub data owner's influence on global model in HFL systems. The unlearned model can be reconstruct using historical model updates by the task controller [23]. However, reconstruction based client-level federated unlearning is not controlled by the target data owner. The authors of [24] enhance controllability by formulating unlearning as parameter optimization task

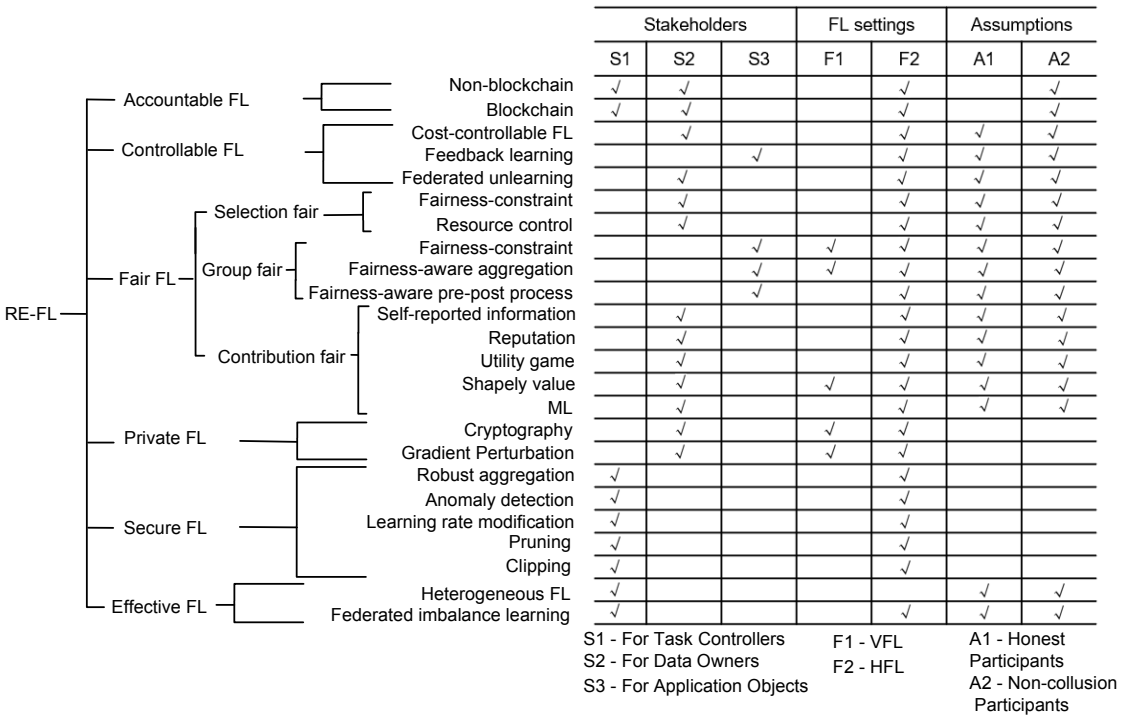



Fig. 3. The proposed taxonomy of RE-FL approaches. In the stakeholders column, √ denotes the corresponding stakeholders related to the responsible and effective AI dimensions. In the FL settings column, √ denotes the FL settings are covered by existing RE-FL literature. In the assumptions column, √ denotes the common assumptions in Section II-D are held by the RE-FL approaches.

subject to constraints on the data owner side, tackled through projected gradient ascent algorithms. The target data owner may further increase the loss of marked data for verification during unlearning via fine-tuning or backdoor injection [25].

Cost-controllable FL has integrated asynchronous FL techniques, model estimation methods and compression ratio adaption approaches to achieve the optimal performance in the case of customized local training budgets. First, when participating parties determine their participation frequency based on their individual budgets, it can result in varying frequencies of participation. Parties with low participation frequency are more likely to upload stale local updates, which can negatively impact the performance of the global model through direct aggregation. In the asynchronous FL approaches, weights are assigned to the uploaded updates based on their level of staleness, and weighted aggregation is performed [26]. The model estimation methods predict the local updates of clients who are not present based on their historical updates in each round of aggregation [27]. Another approach to control communication costs is by integrating FL with model compression methods. Nevertheless, the process of model compression significantly hampers model convergence, subsequently escalating local computational costs. To address this issue, the model compression ratio adaption approaches construct a model compression ratio control problem using the model convergence bound, where the goal is to achieve cost-efficient federated learning [28]. Consequently, each party can train and communicate the local model using personalized compression ratio that align with their budgets, thereby expediting the training process without compromising the performance of the global model.

*In summary*, the existing controllable FL methodologies modify the HFL processes to allow proactive engagement of relevant stakeholders with HFL-based financial service systems. However, these approaches rely on the assumption of honest data owners or application objects, posing a challenge in completing controllable FL with secure FL approaches.

### C. Fair FL

The information asymmetry of local data and local models presents challenges in client selection, model optimization, and reward allocation in HFL-based financial systems. First, task controllers prioritize data owners based on responsiveness or contribution, leading to poor model performance on never-selected data owners' data [29]. Second, biased training data and insufficient consideration of characteristics can cause unfairness for certain application object groups [30]. Traditional fairness AI approaches require sensitive features and training data, making them unsuitable for HFL-based financial service systems. Finally, payoffs for data owners should align with their contributions, but the inaccessibility of training data can result in further inequality [31]. Based on the aforementioned challenges, current fair FL approaches can be categorized into selection-fair, group-fair, and contribution-fair approaches, as depicted in Figure 3.

*1) Selection-fair FL:* There are two main selection-fair FL approaches aimed at achieving fairness within HFL-

based financial service systems during the selection process, consisting of the selection strategy design and the resource control perspectives. First, in order to ensure that each data owner is given a fair chance of being selected, various fairness constraints are applied to the selection function [32]. This is achieved through the introduction of a constant fairness parameter, followed by the application of fairness-constraint optimization methods to calculate the optimized selection probability of each data owner. Second, to enable data owners with low capacities to participate in HFL, the HFL-based systems are adapted based on the data owners' capabilities. Specifically, asynchronous aggregation strategies are proposed to allow for global model updating whenever local updates arrive at the server [33], whilst model compression techniques are used to adapt local model size to the diverse resource capacities [34].

*2) Group-fair FL:* There are three approaches for achieving group fairness in FL-based financial service systems: (1) fairness-constrained objective functions, (2) fairness-aware aggregation strategies, and (3) fairness-aware pre-post process.

In HFL systems, these approaches are employed as follows: (a) Group fairness constraints, such as the difference of equal opportunities (DEO) constraints [30], are applied to limit the degree of unfairness across data owners while optimizing the performance of the global model. (b) Fairness-aware aggregation strategies select the local model that improves the fairness of the global model for fairness-weighted aggregation. Various selection approaches have been proposed, such as FairBest, FairAvg, FairAccAvg [35], and FAIR-FATE [36]. The latter allows for fairness evaluation using a validation dataset, thus adhering to privacy constraints. (c) Fairness-aware pre-post processes is independent of the learning process in HFL systems [37]. PrivFairFL-Pre uses Multiple Party Computation (MPC) techniques to gather aggregated statistical information on label distribution and sensitive attribute values. It assigns weight to each training sample based on the reciprocal number of samples corresponding to its sensitive eigenvalues and labels to tackle potential bias. PrivFairFL-Post is introduced after the learning phase to mitigate bias in predicted outcomes by identifying optimal classification thresholds for each group of sensitive attributes.

In VFL systems, the VFL problem is formulated as a non-convex constrained problem that incorporates DEO as the constrained term [38]. The task controller can further aggregate the local representation into an unbiased global representation with the aid of adversarial learning techniques [39].

*3) Contribution-fair FL:* Contribution-fair FL focuses on proportionate reward allocation to data owners according to their contributions to the FL model. Available contribution evaluation methods fall into five categories: self-reported information-based, reputation-based, utility game-based, Shapely value-based, and ML-based evaluation methods.

In HFL systems, all of the above five methods can be used for contribution assessment. Self-reported information-based evaluation methods rely on data owners reporting their contribution attributes, including data quality, quantity, computational and communication capabilities, and calculation costs, which may not be a reasonable assumption in FL settings [31]. Alternatively, reputation-based evaluation methodologies rely on calculating the reputation score for data owners by leveraging their historical contributions. This score reflects the reliability and contribution of data owners in terms of the validation accuracy of their local models or similarity scores with the global model [40]. Utility game-based contribution evaluation defines a data owner's contribution in HFL systems as the individual utility output, which includes marginal gain and marginal loss. However, the output for marginal gain and loss depends on the order in which each data owner joins the FL training. Shapely Value is utilized in HFL systems to evaluate the marginal contribution of a data owner, taking into account the problem of joining order by averaging the sum of marginal contributions over all subsets of the data owners [41]. However, applying Shapley Value in HFL systems can cause significant communication and computation overhead. To address the issue of additional communication and computation overhead in contribution evaluation, a more efficient ML-based approach, called FedCCEA, is proposed [42] for HFL systems. Using historical records of the sampled data size and round-wise model accuracy, this method constructs an Accuracy Approximation Model (AAM) that robustly and efficiently approximates the data owners' contribution.

In VFL systems, Shapley Value is used to determine the importance of features [43]. However, directly assessing every prediction using Shapley Value may expose sensitive information about certain features. Thus, it is recommended to compute Shapley Value on sets of features instead of individual features [43]. This solution, however, may still be computationally demanding due to the exponential increase in demand with the training data size.

*In summary*, research on fair FL techniques is in its early stages. These technologies necessitates secure and accountable FL techniques for data authenticity. Moreover, privacy concerns are often ignored, elevating privacy risks.

### *D. Private FL*

In FL-based financial service systems, local data does not require centralization. However, the information asymmetry leads to potential private information leakage, as a malicious task controller or data owner may modify the update information or perform additional learning process for privacy attacks. For example, Generative Adversarial Network (GAN)-based attacks in a HFL-based financial service system with two data owners can generate local data class representatives [44]. Moreover, membership inference attacks can determine if a specific sample was used to train the model [45]. Local data properties can also be inferred by other data owners with auxiliary data [46]. Furthermore, in HFL-based financial service systems with small local training

batches, both original data and labels can be disclosed, posing significant risks to privacy [47].

In VFL-based financial service systems, data owners and task controllers exchange intermediate representations and gradients, not data and labels to protect privacy. However, there are also risks of privacy leakage. First, a malicious data owner can fine-tune the bottom model with an additional layer for label inference using a small amount of auxiliary data [48]. Second, a data owner can recover inputs and labels from batch gradients with additional optimization steps when the batch size is small [47].

*1) Private HFL:* In HFL-based financial service systems, two defense strategies against privacy attacks have been identified in existing research - cryptography and perturbation. First, Cryptography-based strategies for HFL can be grouped into two classes: homomorphic encryption [49] and secure multi-party computation [50]. Homomorphic encryption allows direct aggregation on ciphertext of local updates in HFL-based financial service systems, imitating an additive manipulation of plaintext. Secure multi-party computation is also utilized to secure aggregation on encrypted local updates in HFL, but each data owner must encrypt their local updates using a secure sum protocol to achieve secure aggregation. Despite their effectiveness, HFL cryptography-based protocols lack accountability, preventing updates from being audited and creating opportunities for security attacks from malicious data owners. Second, Perturbation-based defense strategies involve introducing perturbations to the local updates before aggregation, ensuring local data privacy. The perturbation techniques commonly used include combining differential privacy techniques [51], local updates mask [52], and local updates sparsification [53]. However, a recent research has revealed that even with such techniques, a Bayes optimal adversary can still accurately reconstruct the original training samples from the perturbed gradient [54].

*2) Private VFL:* Various gradient-perturbation techniques have been suggested to handle gradient leakage in VFL-based financial service systems. For instance, Jin et al. [47] proposed a method where the data owner generates counterfeit gradients using normal distribution and selects the gradient that closely resembles the authentic one. However, this could lead to loss in main task accuracy. To enhance main task accuracy and mitigate privacy leakage, a Confusional Autoencoder (CAE) has been proposed [55]. It transfers the actual label to a soft one with high probability for each alternative class, increasing the complexity of label leakage attacks and decoupling label inference from the main task.

*In summary*, techniques like gradient-encryption and gradient-perturbation are proposed for privacy in FL-based financial systems. However, a conflict arises between accountability, security, effectiveness, and privacy.

### *E. Secure FL*

The information asymmetry between task controller and data owners in HFL-based financial service systems poses a global model security concern, as local training processes remain unknown to the former. Model security attacks are categorized as backdoor attacks, label-flipping attacks, and byzantine attacks [56].

Current defense strategies aimed at safeguarding HFL systems against model security attacks can be classified into two categories: robust aggregation-based and anomaly detection-based, as illustrated in Figure 3. First, since conventional aggregation strategies like FedAvg are not resistant to outliers, several aggregation strategies based on more robust estimators have been suggested, including Krum [56], which eliminates the updates furthest from its neighborhoods to form the global model; two robust aggregation operators based on the coordinate-wise median and the coordinate-wise trimmed mean [57] and Bulyan [58], which sorts local updates based on geometric distances, filters out harmful local updates, and computes the trimmed median of the remaining updates for aggregation. Second, defense strategies for anomaly detection use ML-based methods to remove harmful updates. AUROR uses k-means to cluster updates and eliminate outlier data beyond a threshold [59], but it is limited to Independent Identical Distribution (IID) FL settings and may disregard clustered information in non-IID data. To address the non-IID issue, anomaly detection techniques and defense strategies using adaptive clipping and noise have been experimented with by [60] to identify harmful local updates in non-IID settings.

*In summary*, secure HFL approaches primarily inspect individual local model information, posing challenges for encrypted HFL-based financial systems. While perturbation techniques can supplement secure FL approaches, further research is needed on the potential trade-off between security, privacy, and effectiveness.

### *F. Effective FL*

Due to information asymmetry, data spaces may vary among different data owners in a FL-based finance service system, resulting in insufficient common features or samples. These issues can lead to unsatisfactory model performance.

As depicted in Figure 3, researchers have proposed three approaches to address the issue of data space heterogeneity: transfer learning-based, data augmentation-based, and knowledge distillation-based. Transfer learning approaches consider one party as the source domain with labeled and the other party as the target domain with only unlabeled or limited labeled samples. Liu et al. [5] proposed a Secure Federated Transfer Learning (SFTL) framework that maps feature spaces into a common subspace and trains local models of the passive party in this subspace. Sharma et al. further improved the efficiency of SFTL using the enhanced secure computation framework SPDZ [61]. However, transfer learning based approaches are only applicable in settings which only involve one source party and one target party. Knowledge distillation-based methods train individual models for each data owner using supervision from the VFL model's privileged information [62]. The VFL model trained on aligned samples acts as a teacher for each data owner's individual model. However, the effectiveness of the VFL

model depends on fully overlapping data and labels, which can be difficult to guarantee. Recent studies have focused on completing non-overlapping samples in the latent feature space. For example, Semi-supervised Federated Cross View Learning leverages semi-supervised learning and cross-view learning to estimate missing features and pseudo-labels of non-overlapping samples in each data owner [63].

*In summary*, FL settings with data space heterogeneity are significantly different from VFL and HFL settings. Novel FL approaches need to be developed to ensure the responsible implementation of heterogeneous FL in financial service systems.

## IV. Open Issues and Future Directions

There are potential problems in FL-based financial service systems that need addressing to enable responsibility and effectiveness. We suggest the following research directions.

### A. Accountable VFL

In VFL systems, data owners may face an information asymmetry problem, resulting in meaningless local outputs due to flawed data collection or free-riders. Furthermore, the data owner with labels cannot verify the trustworthiness of the local outputs, as it lacks access to other parties' data. Thus, auditing information in VFL is necessary to tackle this problem.

### B. Effective Heterogeneous Federated Learning

Despite the existence of various methods proposed by researchers to address the problem of heterogeneous data space by examining non-overlapping samples, they have not taken into account a crucial aspect of the heterogeneous feature space – a significant portion of the distributed features are readily available to the public. In order to enhance the effectiveness of heterogeneous FL, it is desirable for researchers to explore the publicly available datasets in heterogeneous FL.

### C. Controllable Federated Learning with Parameter-Efficient Feedback Learning

Controllable Federated Learning uses positive/negative feedback to improve model performance. However, in real applications, the false feedback may be infrequent. Therefore, it is imperative to conduct further research on how to efficiently leverage the infrequent false feedback to finetune model parameters in a distributed manner.

## V. Conclusion

This paper provides a detailed examination of responsible and effective FL (RE-FL) in financial service systems. The investigation begins with an in-depth review of existing FL-based financial service systems, followed by the introduction of a multi-layer taxonomy based on relevant literature. This taxonomy is designed to evaluate RE-FL approaches, focusing on key criteria such as accountability, controllability, fairness, privacy, security, and effectiveness. Additionally, the paper identifies several promising research directions aimed at enhancing the responsible and effective implementation of FL within financial services. Overall, this survey serves as a beneficial resource for researchers in the fields of computer science and finance, facilitating the further development of FL techniques that incorporate both responsibility and effectiveness.

## References


[1] J. Xu, *The Future and Fintech: ABCDI and Beyond*. World Scientific Press, 2022.

[2] P. Regulation, "General data protection regulation," *Intouch*, vol. 25, 2018.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[4] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 13, no. 3, pp. 1–207, 2019.

[5] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 70–82, 2020.

[6] Y. Zeng, E. Lu, and C. Huangfu, "Linking artificial intelligence principles," *arXiv preprint arXiv:1812.04814*, 2018.

[7] E. Mogaji and N. P. Nguyen, "Managers' understanding of artificial intelligence in relation to marketing financial services: insights from a cross-country study," *International Journal of Bank Marketing*, vol. 40, no. 6, pp. 1272–1298, 2022.

[8] H. Zhang, J. Hong, F. Dong, S. Drew, L. Xue, and J. Zhou, "A privacy-preserving hybrid federated learning framework for financial crime detection," *arXiv preprint arXiv:2302.03654*, 2023.

[9] T. Suzumura, Y. Zhou, R. Kawahara, N. Baracaldo, and H. Ludwig, "Federated learning for collaborative financial crimes detection," in *Federated Learning*. Springer, 2022, pp. 455–466.

[10] D. Kawa, S. Punyani, P. Nayak, A. Karkera, and V. Jyotinagar, "Credit risk assessment from combined bank records using federated learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 4, pp. 1355–1358, 2019.

[11] Z. Li, M. Bilal, X. Xu, J. Jiang, and Y. Cui, "Federated learning based cross-enterprise recommendation with graph neural networks," *IEEE Transactions on Industrial Informatics*, 2022.

[12] J. Yao, Z. Dou, and J.-R. Wen, "Fedps: A privacy protection enhanced personalized search framework," in *Proceedings of the Web Conference 2021*, 2021, pp. 3757–3766.

[13] G. Long, Y. Tan, J. Jiang, and C. Zhang, "Federated learning for open banking," in *Federated Learning: Privacy and Incentive*. Springer, 2020, pp. 240–254.

[14] G. Han, T. Zhang, Y. Zhang, G. Xu, J. Sun, and J. Cao, "Verifiable and privacy preserving federated learning without fully trusted centers," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 3, pp. 1431–1441, 2022.

[15] Z. Liu, Y. Chen, H. Yu, Y. Liu, and L. Cui, "Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 4, pp. 1–21, 2022.

[16] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 72–80, 2020.

[17] A. Li, L. Zhang, J. Wang, J. Tan, F. Han, Y. Qin, N. M. Freris, and X.-Y. Li, "Efficient federated-learning model debugging," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021, pp. 372–383.

[18] W. Zhang, Q. Lu, Q. Yu, Z. Li, Y. Liu, S. K. Lo, S. Chen, X. Xu, and L. Zhu, "Blockchain-based federated learning for device failure detection in industrial iot," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5926–5937, 2020.

[19] P. Ramanan and K. Nakayama, "Baffle: Blockchain based aggregator free federated learning," in *2020 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2020, pp. 72–81.

[20] Z. Peng, J. Xu, X. Chu, S. Gao, Y. Yao, R. Gu, and Y. Tang, "Vfchain: Enabling verifiable and auditable federated learning via blockchain systems," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 173–186, 2022.

[21] R. Sharma, A. Ramakrishna, A. MacLaughlin, A. Rumshisky, J. Majmudar, C. Chung, S. Avestimehr, and R. Gupta, "Federated learning with noisy user feedback," *arXiv preprint arXiv:2205.03092*, 2022.

[22] J. Wang, S. Guo, X. Xie, and H. Qi, "Federated unlearning via class-discriminative pruning," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 622–632.

[23] G. Liu, X. Ma, Y. Yang, C. Wang, and J. Liu, "Federaser: Enabling efficient client-level data removal from federated learning models," in *2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)*. IEEE, 2021, pp. 1–10.

[24] A. Halimi, S. Kadhe, A. Rawat, and N. Baracaldo, "Federated unlearning: How to efficiently erase a client in fl?" *arXiv preprint arXiv:2207.05521*, 2022.

[25] X. Gao, X. Ma, J. Wang, Y. Sun, B. Li, S. Ji, P. Cheng, and J. Chen, "Verifi: Towards verifiable federated unlearning," *arXiv preprint arXiv:2205.12709*, 2022.

[26] D. Stripelis, P. M. Thompson, and J. L. Ambite, "Semi-synchronous federated learning for energy-efficient training and accelerated convergence in cross-silo settings," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 5, pp. 1–29, 2022.

[27] H. Zhang, T. Wu, S. Cheng, and J. Liu, "Cc-fedavg: Computationally customized federated averaging," *IEEE Internet of Things Journal*, 2023.

[28] Z. Jiang, Y. Xu, H. Xu, Z. Wang, J. Liu, Q. Chen, and C. Qiao, "Computation and communication efficient federated learning with adaptive model pruning," *IEEE Transactions on Mobile Computing*, 2023.

[29] L. Li, M. Duan, D. Liu, Y. Zhang, A. Ren, X. Chen, Y. Tan, and C. Wang, "Fedsae: A novel self-adaptive federated learning framework in heterogeneous systems," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–10.

[30] M. Donini, L. Oneto, S. Ben-David, J. S. Shawe-Taylor, and M. Pontil, "Empirical risk minimization under fairness constraints," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[31] T. H. T. Le, N. H. Tran, Y. K. Tun, M. N. Nguyen, S. R. Pandey, Z. Han, and C. S. Hong, "An incentive mechanism for federated learning in wireless cellular networks: An auction approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4874–4887, 2021.

[32] T. Huang, W. Lin, L. Shen, K. Li, and A. Y. Zomaya, "Stochastic client selection for federated learning with volatile clients," *IEEE Internet of Things Journal*, 2022.

[33] W. Wu, L. He, W. Lin, R. Mao, C. Maple, and S. Jarvis, "Safa: A semi-asynchronous protocol for fast federated learning with low overhead," *IEEE Transactions on Computers*, vol. 70, no. 5, pp. 655–668, 2020.

[34] D. Yao, W. Pan, Y. Wan, H. Jin, and L. Sun, "Fedhm: Efficient federated learning for heterogeneous models via low-rank factorization," *arXiv preprint arXiv:2111.14655*, 2021.

[35] G. Yu, L. Ma, W. Du, W. Du, and Y. Jin, "Towards fairness-aware multi-objective optimization," *arXiv preprint arXiv:2207.12138*, 2022.

[36] T. Salazar, M. Fernandes, H. Araujo, and P. H. Abreu, "Fair-fate: Fair federated learning with momentum," *arXiv preprint arXiv:2209.13678*, 2022.

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

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

[39] T. Qi, F. Wu, C. Wu, L. Lyu, T. Xu, Z. Yang, Y. Huang, and X. Xie, "Fairvfl: A fair vertical federated learning framework with contrastive adversarial learning," *arXiv preprint arXiv:2206.03200*, 2022.

[40] Z. Song, H. Sun, H. H. Yang, X. Wang, Y. Zhang, and T. Q. Quek, "Reputation-based federated learning for secure wireless networks," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1212–1226, 2021.

[41] T. Song, Y. Tong, and S. Wei, "Profit allocation for federated learning," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 2577–2586.

[42] S. K. Shyn, D. Kim, and K. Kim, "Fedccea: A practical approach of client contribution evaluation for federated learning," *arXiv preprint arXiv:2106.02310*, 2021.

[43] G. Wang, C. X. Dang, and Z. Zhou, "Measure contribution of participants in federated learning," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 2597–2604.

[44] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the gan: information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, 2017, pp. 603–618.

[45] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *2019 IEEE symposium on security and privacy (SP)*. IEEE, 2019, pp. 691–706.

[46] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *2019 IEEE symposium on security and privacy (SP)*. IEEE, 2019, pp. 739–753.

[47] X. Jin, P.-Y. Chen, C.-Y. Hsu, C.-M. Yu, and T. Chen, "Cafe: Catastrophic data leakage in vertical federated learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 994–1006, 2021.

[48] C. Fu, X. Zhang, S. Ji, J. Chen, J. Wu, S. Guo, J. Zhou, A. X. Liu, and T. Wang, "Label inference attacks against vertical federated learning," in *31st USENIX Security Symposium (USENIX Security 22), Boston, MA*, 2022.

[49] M. Kim, Y. Song, S. Wang, Y. Xia, X. Jiang *et al.*, "Secure logistic regression based on homomorphic encryption: Design and evaluation," *JMIR medical informatics*, vol. 6, no. 2, p. e8805, 2018.

[50] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[51] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.

[52] D. Scheliga, P. Mäder, and M. Seeland, "Precode-a generic model extension to prevent deep gradient leakage," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1849–1858.

[53] J. Sun, A. Li, B. Wang, H. Yang, H. Li, and Y. Chen, "Soteria: Provable defense against privacy leakage in federated learning from representation perspective," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9311–9319.

[54] M. Balunović, D. I. Dimitrov, R. Staab, and M. Vechev, "Bayesian framework for gradient leakage," *arXiv preprint arXiv:2111.04706*, 2021.

[55] T. Zou, Y. Liu, Y. Kang, W. Liu, Y. He, Z. Yi, Q. Yang, and Y.-Q. Zhang, "Defending batch-level label inference and replacement attacks in vertical federated learning," *IEEE Transactions on Big Data*, 2022.

[56] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[57] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5650–5659.

[58] R. Guerraoui, S. Rouault *et al.*, "The hidden vulnerability of distributed learning in byzantium," in *International Conference on Machine Learning*. PMLR, 2018, pp. 3521–3530.

[59] D. Cao, S. Chang, Z. Lin, G. Liu, and D. Sun, "Understanding distributed poisoning attack in federated learning," in *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2019, pp. 233–239.

[60] S. Andreina, G. A. Marson, H. Möllering, and G. Karame, "Baffle: Backdoor detection via feedback-based federated learning," in *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2021, pp. 852–863.

[61] S. Sharma, C. Xing, Y. Liu, and Y. Kang, "Secure and efficient federated transfer learning," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 2569–2576.

[62] Z. Ren, L. Yang, and K. Chen, "Improving availability of vertical federated learning: Relaxing inference on non-overlapping data," *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022.

[63] Y. Kang, Y. Liu, and X. Liang, "Fedcvt: Semi-supervised vertical federated learning with cross-view training," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 4, pp. 1–16, 2022.