iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

A FEDERATED LEARNING PERSPECTIVE IN AI-DRIVEN ANALYSIS

Dr. Nidhi Bansal
Manav Rachna International Institute of Research and Studies
(Deemed to be University), Faridabad, India
nidhibansal.set@mriu.edu.in

Anil Sharma
Department of Computer Science, College of Science
Anil.sharma@nbu.edu.sa

ABSTRACT
Artificial intelligence (AI) systems are the major components that the core financial activities are relying on. These are mainly consumer lending and credit scoring, fraud detection, and algorithmic trading. To a great extent, these models are said to be able to perform decisions rapidly and to increase the efficiency of the processes, however, they also have the ability to inherit and in consequence to increase the biases, from the training data they are fed. Moreover, the healthcare sector has been investigating the concept of federated learning, a privacy-preserving paradigm in which organizations can train a model together without sharing raw data. Financial institutions can use Federated approaches as a powerful tool to see through data-related bias while they are abiding by strict privacy rules.
This paper locates bias sources in AI-driven finance, studies credit scoring, fraud detection, and algorithmic trading scenarios, weighs the effects on equity and consumer trust, and considers the strategies including federated learning that could be used for bias removal. A comparison chart of fairness metrics such as statistical parity, equal opportunity, and equalized odds is introduced. Figures indicate the stages of federated learning and the compromises involved in the selection of fairness metrics. The article finishes with pointing the next steps in research that would facilitate responsible AI in finance and, additionally, the significance of ethical governance and the healthcare sector's cross-sector lessons.

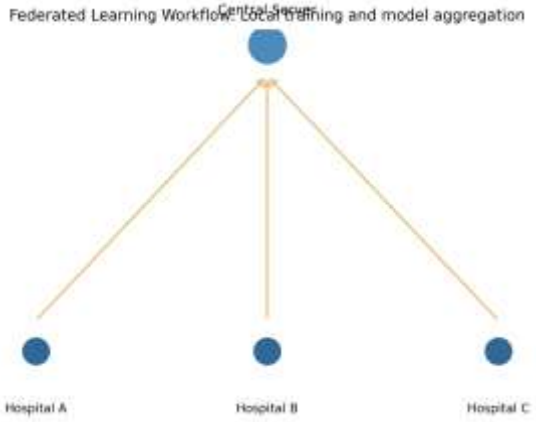
Keywords:
Artificial intelligence, Finance

INTRODUCTION

Financial services increasingly depend on machine-learning models to make decisions that were once the sole domain of human underwriters and analysts. Automated credit scoring systems determine loan eligibility, fraud detectors monitor billions of transactions in real time, and algorithmic trading platforms execute orders at millisecond scales. Yet the training data fuelling these systems reflect historical patterns of discrimination and structural inequities, leading to decisions that can perpetuate or even exacerbate disparities. For example, algorithms may assign lower credit scores to historically marginalised communities or deny loans based on proxies for race, gender or socioeconomic status. Emerging regulations in the United States and Europe demand that financial institutions demonstrate fairness, explainability and privacy compliance in their AI pipelines.
At the same time, the healthcare industry has developed *federated learning (FL)*—a model wherein each hospital or clinic trains the model locally and only shares model updates (e.g., gradients or weights) with a central server. Federated learning, which improves privacy by "bringing the model to the data" rather than collecting sensitive patient data, complies with HIPAA and GDPRFL was initially utilized by Google in 2016 to enhance predictive typing on mobile devices, and since then, researchers have applied it for sepsis prediction across hospitals and medical image diagnosis.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The financial sector could implement federated concepts as training data are tightly controlled and spread across banks, credit bureaus, and payment processors. Data centralization exposes organizations to privacy attacks, regulatory fines, and increased complexity of operations. A federated system may enable several institutions to jointly train fraud-detection or credit-scoring models while raw data remain on-site and the risk of bias amplification is lowered. Figure 1 demonstrates the main federated learning process: clients (e.g., banks or hospitals) locally train a model and only share parameter updates with a central server. The server merges these updates and sends the updated model back to the clients. Going through this cycle again and again results in a global model without actually accessing any institution's data.



Federated learning workflow showing clients training models locally, sending updates to a central server and receiving an aggregated model in return

Federated learning is a solution to privacy concerns, however, it does not by default solve the problem of bias. The reason is that different data across institutions may make the combined model work on some subgroups better than others. In addition, biases that are introduced during initial training can get bigger over time as the model is retrained with feedback loops. Therefore, the finance industry should use fairness metrics, bias-mitigation techniques and governance frameworks along with federated architectures to achieve fair outcomes.

SOURCES AND TYPES OF BIAS IN AI-DRIVEN FINANCE

- **Selection, measurement and representation bias**

When AI systems produce biased results it is most of the time because their training data contain biased patterns. Selection bias comes about when the data sample does not accurately represent the population. As an example, in credit scoring, training data might be rich in borrowers from certain demographic groups while there might be scarce data for those with no credit history. This imbalance makes the models generalize poorly towards minority applicants, may lead to discriminatory lending practices. Measurement bias derives from the situation when the variables used for model training inaccurately measure the expected constructs. One of the extreme examples comes from the medical field: pulse oximeters systematically overestimate blood oxygen levels in individuals with darker skin tones, which leads to the delay in diagnosis. Those kinds of measurement problems can also happen in finance when aliased variables such as ZIP code or job correlate with race or income thus embedding structural inequities in model inputs. Representation bias is the bias that comes from the underrepresentation or misrepresentation of certain groups in the training data. If communities that have been historically disenfranchised are excluded from financial datasets because they do not have bank accounts or formal credit histories, then models will underpredict their creditworthiness and thus will exclude them further.

In federated settings, these biases can manifest locally. Each participating institution may have data reflecting its specific clientele and business practices. When the central server aggregates model updates, differences in data distributions across clients can yield a global model that performs unevenly across groups. Without careful weighting

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

and fairness constraints, federated learning can reproduce the dominance of well-resourced institutions while marginalising smaller or underrepresented entities.

Implicit and algorithmic bias

Besides biases in the data, developer bias appears when the assumptions of human developers unintentionally influence feature selection, model architecture or label definitions. As an illustration, fraud analysts may decide that transactions from some regions are suspicious based on their prior beliefs rather than objective criteria. These biases get encoded in the supervised labels which are used to train ML models. Furthermore, the objective functions used in training often concentrate on overall accuracy or profit maximisation, thus, overlooking the issue of equity. **Algorithmic bias** is the case when the optimisation process gradually and systematically puts certain groups at a disadvantage. For example, a model may minimise false positives overall but inadvertently increase false negatives for minority borrowers, thus, resulting in fewer loans.

Evaluation and feedback-loop bias

Evaluation bias arises when model evaluation uses metrics that do not represent the experiences of all subgroups. For instance, evaluating a credit-scoring model only on accuracy may hide systematic errors that affect certain demographics. Consequently, the fairness literature offers metrics such as **statistical parity** (ensuring equal selection rates across groups), **equal opportunity** (equal true-positive rates), and **equalized odds** (equal true-positive and false-positive rates). These metrics are often mutually exclusive; the *impossibility theorem* demonstrates that it is impossible to fulfill all fairness criteria at the same time.

_____

The trade-off between statistical parity, equal opportunity and equalized odds showing that no single algorithm can simultaneously optimize all three fairness metrics

Feedback-loop bias occurs when model outputs affect future training data. In credit scoring, applicants denied loans because of a model's predictions never become part of the dataset as repaid or defaulted borrowers, thus, biasing future models. The *FairSense* framework from Carnegie Mellon illustrates how small biases in a model can create vicious cycles: a slightly discriminatory classifier will disproportionately deny loans to minorities; retraining on the resulting dataset will amplify the bias. FairSense recommends monitoring fairness over time and changing decision thresholds to offset feedback loops.

CASE STUDIES: CREDIT SCORING, FRAUD DETECTION AND TRADING BIAS

- **Credit scoring**

Credit scoring models are designed to estimate the probability that an applicant will repay a loan. The traditional methods are based on variables such as credit history, income, and debt-to-income ratios. Machine-learning models use a broader set of features including location, social-media activity, and transaction histories. These new sources of data may also be a source of socioeconomic proxies for sensitive attributes like race, gender, and marital status. The PLOS Digital Health article on bias in healthcare states that if models rely on correlated variables without considering structural inequities, they can perpetuate historical discrimination. The same reasoning goes in finance: a model can penalise applicants from a mainly minority area because it factors property values and employment that have been suppressed by years of redlining and discrimination.

Regulators have cautioned lenders against using unexplainable (black-box) algorithms to make decisions. The Payments Association conveys that by 2025 about 71 % of financial firms will have AI as a chief tool for risk assessment and compliance. Nevertheless, respondents to the survey indicate that while fraud prevention via AI is a source of comfort for 63 % of them, only 31 % of them trust AI to decide on insurance claims. These contradictions point to the necessity for openness in credit models and for indicating the reasons when denial or approval of applicants is done by automated systems.

Credits through federated learning may be scored better in this way: Banks and credit bureaus collaboratively work on models without the need to pool customer data. Local models on loan histories are trained in each institution; updates are aggregated by the central server. This is a way to attain dataset diversity and thus, it lessens the chance of selection and representation bias. However, fairness constraints must be there to make sure that the updates from smaller institutions are not overshadowed by those of the larger ones.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- **Fraud detection**

Financial fraud is a major contributor to the global economic losses of the last several years in the range of hundreds of billions of dollars. The advent of machine learning has greatly improved the detection of fraudulent activities through its ability to analyze large volumes of transaction data and detect unusual patterns in such data. The Payments Association points out that AI can detect subtle behavior patterns and react very quickly thus, both customer safety and cost reduction of business operations are achieved. Nevertheless, such models require top-notch training data; in case the data have mislabeled or incomplete records, false positives will be generated by the system, and these will upset customers and will result in loss of their trust. Overly aggressive models may block legitimate transactions disproportionately from certain demographics, effectively excluding them from financial services.

Fraud datasets are usually collected from different banks and payment processors to reveal fraud that crosses institutions. Federated learning provides a way to create very effective fraud detection systems without compromising the privacy and compliance of the participants. However, the accuracy of labeling (e.g., deciding which transactions are truly fraudulent) and the rarity of fraud incidents make the training process difficult. If an algorithm is given only a few examples of fraud committed by underrepresented groups, it may not generalize well and hence biased false-negative rates will be produced. Fairness metric equalized odds can be employed to check if the detection rates are the same across demographic groups

- **Algorithmic trading**

Algorithms and high-frequency trading (HFT) (trading) systems make up a large portion of trading volume—50–55 % in some markets. These systems employ AI models to anticipate changes in asset prices and to carry out orders automatically. The problem of bias emerges when models are trained on historical price data reflecting market behaviors that are influenced by government regulations, institutional strategies, and geopolitical events. The Mixflow blog cautions that training on biased data will result in discriminatory outcomes and in market manipulation. For example, the models may be systematic in trading against certain sectors that are related to minority-owned businesses or small enterprises, which deepens the inequality problem.

Regulatory bodies such as the International Organization of Securities Commissions (IOSCO) have provided guidelines to the firms on implementing governance frameworks, conducting fairness audits, and ensuring transparency in AI-powered trading. Federated learning can be the tool that exchanges and brokers use to share insights about market anomalies while keeping the proprietary strategies confidential. The fair trading process requires the surveillance of feedback loops so as to stop these loops from being the cause of algorithmic strategies distorting prices and thus, deepening the biases in the training data.

- **Impact of Bias on Financial Equity and Consumer Trust**

Biases in AI-based financial systems have real implications for equity and trust. Fair credit models have the potential to restrict access to essential needs such as housing, education, and entrepreneurship for those groups that have been marginalized. On the other hand, inaccurate fraud detection systems can incorrectly identify transactions made by certain communities as fraudulent, thus, these communities being excluded from the digital economy. Biases in algorithmic trading can distort the process of capital allocation resulting in the disproportionate harm of small and minority-owned businesses. These wrongdoings not only maintain the wealth gap but also challenge the trustworthiness of the financial system.

Trust from customers is largely based on the belief that financial decisions are conducted in a fair and transparent manner. Surveys show that people have mixed feelings: on the one hand, many people are positive about the use of AI in fraud prevention, on the other hand, they have doubts about its use for high-impact decisions such as insurance payouts or loan approvals. False accusations and rejections that cannot be explained gradually reduce the trust in the system; on the contrary, fair and transparent models have the potential to increase trust and thereby engagement. Thus, financial institutions are under moral and business pressure to proactively tackle the issue of bias.

Apart from the risks that bias poses to individuals, it can also put businesses at legal and reputational risks. Regulatory agencies that oversee compliance with anti-discrimination laws, data-protection standards, and fair lending regulations are increasingly scrutinizing AI systems. Consequences may be in the form of monetary penalties, corrective measures, and judicial proceedings. What is more, negative media exposure of biased algorithms can tarnish the brand and thus, lose the trust of the customers.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

MITIGATION FRAMEWORKS AND ETHICAL AI STRATEGIES

- **Fairness metrics and trade-offs**

Mitigating bias begins with defining and gauging fairness. Statistical parity calls for the selection rate (e.g., loan approvals) to be the same for the protected and non-protected groups. Equal opportunity requires equal true-positive rates which means that individuals who are qualified have an equal chance of receiving a positive decision. Equalized odds takes this further by necessitating equal true-positive and false-positive rates for all groups. These metrics frequently oppose each other: fulfilling statistical parity may cause false positives for one group to increase while for another to decrease. Consequently, practitioners have to explicitly decide which concept of fairness correlates with their ethical and legal duties. Figure 2 demonstrates the trade-offs between these metrics. The impossibility theorem indicates that one cannot fulfill all fairness definitions at the same time when base rates are different.

- **Bias-mitigation techniques**

Bias can be regulated at three different points: *pre-processing*, *in-processing* and *post-processing*.

Pre-processing methods change the training data with the aim of lessening bias before model training. The techniques may be reweighting of samples to reflect the population demographics, removal or transformation of sensitive attributes, and augmentation of synthetic data. As an illustration, oversampling of the underrepresented groups or the use of fairness-aware resampling may solve the problem of selection bias.

In-processing methods bind fairness constraints right into the learning algorithm. Examples of these are adversarial debiasing, which continues the model training while also trying to prevent the auxiliary classifier from predicting sensitive attributes, and the other strategy, which adds the regularisation terms to the loss function to penalize unfair outcomes and enforce equal opportunity or equalized odds.

Post-processing techniques change model output after the training. Methods such as thresholding achieve statistical parity or equal opportunity by calibrating decision boundaries separately for each group. **FairSense** is the post-processing example that tracks fairness through time and changes decision thresholds to counter feedback-loop bias

Mitigation methods in federated settings need to be capable of local execution without access to raw data. **Federated debiasing** performs local pre-processing and in-processing at individual clients with the server coordinating global fairness constraints. Additionally, differential privacy may be used along with FL to keep the contribution of each individual private while lessening the measurement bias. The most recent study presents the idea of *fair federated averaging* which adjusts the weights of client updates to achieve equality between groups.

- **Governance and ethics**

Technological fixes need support from governance structures and ethical supervision. Regulators and industry organisations are putting together guidance for responsible AI use. IOSCO advises the use of robust governance, data-quality standards, and transparency for algorithmic trading. The Payments Association points to the need for explainability and human oversight as components of fraud detection. Financial institutions may set up interdisciplinary committees to include data scientists, ethicists, legal experts and community representatives to model evaluation and outcome monitoring. Reports that are disclosed to the public can become a means of accountability and trust building.

COMPARATIVE TABLE: FAIRNESS METRICS IN FINANCE

The table below provides a summary of commonly employed fairness metrics in terms of their definitions, advantages, and limitations. Practitioners evaluating metrics should take into account that these metrics align with their objectives and regulatory requirements and are also aware of the trade-offs that exist between them.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Metric	Definition	Advantages	Limitations
Statistical parity	Ensures that the proportion of positive decisions (e.g., loan approvals) is equal across protected and unprotected groups.	Easy to implement; encourages equal opportunity.	May overlook differences in qualifications and increase error rates for certain groups; can be at odds with merit-based selection.
Equal opportunity	Ensures equal true-positive rates across groups i.e. qualified individuals have an equal chance of receiving a positive decision.	Concentrates on fairness for those who merit a positive outcome; balances accuracy and access.	False positives are not controlled; may result in different selection rates; cannot be reconciled with statistical parity when base rates vary.
Equalized odds	Require equal true-positive and false-positive rates for groups.	Offers a complete fairness notion by balancing both error types; reduces disparate impact.	Difficult to accomplish; overall accuracy may be lowered; threshold tuning needs to be done carefully; statistical parity and equal opportunity may be in conflict.

FUTURE DIRECTIONS FOR RESPONSIBLE AI IN FINANCE

The AI-based finance landscape is changing swiftly. Some directions are expected to improve fairness, transparency, and efficiency:

Standardised reporting and auditing: Industry-wide fairness reporting standards, similar to financial audits, would make it possible for all the stakeholders to assess models in a uniform manner. FairSense-type tools that simulate long-term fairness effects could become standard.

Cross-sector collaborations: Federated learning initiatives in healthcare could be a source of inspiration for the finance industry. Collaborative federated networks might help banks, regulators, and consumer advocacy groups to jointly train and evaluate models without any privacy compromises.

Adaptive and explainable models: Progress in interpretable machine learning will enable models to offer explanation of their decision, thus increasing transparency and easing regulatory compliance. To gain consumer trust, companies will have to combine explainability with fairness constraints.

Regulatory innovation: Policymakers are setting up sandboxes where AI systems can be tested under controlled conditions while bias is monitored. Regulators may impose fairness metrics as a requirement or impose the use of federated architectures in certain sectors with high-impact applications.

Ethical data stewardship: Organizations ought to put money into fair data collection methods so that the disadvantaged groups will be included and measuring instruments will be calibrated across different demographics. Continuous monitoring and community involvement can help in identifying unintentional biases as well as guiding the corrective actions.

CONCLUSION

AI-powered finance is at a turning point. Those models that offer great efficiency and innovation are also capable of, if left unmonitored, continuously transferring the past injustice. Selection, measurement, and representation biases in data, along with implicit and algorithmic biases in model design, can lead to discrimination in areas such as credit scoring, fraud detection, and algorithmic trading. Federated learning, a model developed in healthcare with the aim of securing patient privacy, is a way for banks to jointly work on building strong models without the need to centralize sensitive data. However, it is crucial that federated architectures be complemented with fairness metrics, debiasing methods and ethical governance so as to guarantee fairness.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The way forward demands collaboration between different disciplines. Data scientists have to create models that are a good trade-off between accuracy and fairness. Regulators need to devise policies that balance the encouragement of innovation and the protection of consumers. Institutions are to infuse ethical factors into their AI lifecycles. With the use of federated approaches, open fairness metrics and solid mitigation frameworks, the financial sector will be able to use the AI power not for charging further but for widening the circle of opportunities.

REFERENCES

1) Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: challenges and applications. *International journal of machine learning and cybernetics*, *14*(2), 513-535.
2) Mohri, M., Sivek, G., & Suresh, A. T. (2019, May). Agnostic federated learning. In *International conference on machine learning* (pp. 4615-4625). PMLR.
3) B. D. Zeisl and A. Narayanan, "FairSense: preventing feedback loops and bias amplification," *Tech Xplore* (Carnegie Mellon University), 2024.
4) Kabade, S.; Sharma, A.; Kagalkar, A. Securing Pension Systems With AI-Driven Risk Analytics and Cloud-Native Machine Learning Architectures. *IJERET* **2024**, *5* (2), 52-64. https://doi.org/10.63282/3050-922X.IJERET-V5I2P106.
5) Varma, S. C. G. (2024). Ethical implications of AI-driven decision-making in cloud-based services. International Journal For Multidisciplinary Research., DOI: 10.36948/ijfmr.2024.v06i06.31583
6) Gopal Varma, S. C. Artificial Intelligence in Cloud Computing: Building Intelligent, Distributed, and Fault-Tolerant Systems. *IJAIBDCMS* **2022**, *3* (1), 37-45. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V3I1P105.
7) Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: Fundamentals, enabling technologies, and future applications. *Information processing & management*, *59*(6), 103061.
8) Yuan, L., Wang, Z., Sun, L., Yu, P. S., & Brinton, C. G. (2024). Decentralized federated learning: A survey and perspective. *IEEE Internet of Things Journal*, *11*(21), 34617-34638.