Data Quality-Aware Federated Learning for Fake Review Detection

1st Jiantao Xu
Dept. of Computer Science
University of Aizu
Aizuwakamatsu, Japan
d8252108@u-aizu.ac.jp

2nd Chen Zhang
Dept. of Computer Science
University of Aizu
Aizuwakamatsu, Japan
d8252109@u-aizu.ac.jp

3rd Liu Jin
Dept. of Computer Science
University of Aizu
Aizuwakamatsu, Japan
d8242103@u-aizu.ac.jp

* Chunhua Su
Dept. of Computer Science
University of Aizu
Aizuwakamatsu, Japan
chsu@u-aizu.ac.jp

2025 7th International Conference on Software Engineering and Computer Science (CSECS) | 979-8-3315-2221-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CSECS64665.2025.11009627

***Abstract*—Fake reviews are a significant challenge for online consumer and social media platforms, as they mislead consumers and disrupt fair market competition. Traditional centralized detection methods face numerous limitations in handling massive data volumes, heterogeneous distributions, and privacy compliance. To address these issues, this paper proposes a data quality-aware federated learning framework for efficiently detecting fake reviews while preserving the privacy of data holders. Specifically, we introduce a data quality evaluation module during the federated learning training process. This module quantifies the quality of each client's data based on multiple metrics, including annotation accuracy, textual completeness, and user behavior. Higher-quality data clients are assigned greater weights during global model aggregation, thereby mitigating the interference of low-quality or noisy data. Experimental results demonstrate that compared to traditional federated averaging and other baseline methods, our approach significantly improves accuracy, precision, and recall. Additionally, it exhibits greater robustness and faster convergence under various non-independent and identically distributed (non-IID) and multi-noise scenarios. Furthermore, differential privacy ensures the security of participants' raw data and quality evaluation information, enhancing the feasibility of real-world deployment.**

***Index Terms*—Federated Learning, Data Quality, Fake Review Detection, Privacy Preservation, Distributed Machine Learning**

I. Introduction

In modern society, online platforms offer diverse channels for content sharing and feedback, ranging from product evaluations on e-commerce websites to post "likes" on social media. User comments on goods or services have become a critical factor influencing consumer decisions [1], [2]. However, along with the widespread dissemination of online reviews comes a growing concern: the proliferation of fake reviews [3], [4]. Fake reviews may be automatically generated by merchants or third parties, or written by "review farms" with the intent of misleading users, thereby impacting fair market competition and undermining the quality of consumer decision-making [5], [6]. For consumers, fake reviews not only lead to misguided choices but also erode trust in the platform [7]; for merchants, malicious attacks or deceptive promotions damage healthy industry development [8]. Therefore, accurate and timely detection of fake reviews has become an important topic for both academia and industry.

This work was partially supported by JSPS Grant-in-Aid for Scientific Research (C) 23K11103. Chunhua Su is the corresponding author.

In previous studies, fake review detection typically relies on centralized data mining and machine learning methods, pooling data from various sources into a single center for analysis [3], [9]. Although this approach facilitates unified data management and model training, it encounters multiple challenges. First, privacy protection is increasingly emphasized, especially when sensitive information such as user personal data, geographic location, or consumption preferences is involved. Centralized storage poses higher security risks and compliance pressures. Second, review data are often naturally distributed across various industry platforms and regional servers, making them difficult to aggregate by simple centralized collection. In addition, gathering large-scale data necessitates substantial bandwidth and storage resources and may face regulatory or cross-platform collaboration obstacles.

Federated Learning (FL) offers a novel approach to address these issues. FL is a distributed machine learning framework that enables multiple data holders to collaboratively train a global model without sharing raw data [10], [11]. By delegating the training process to local clients (e.g., phones, servers, or edge devices) and uploading only model parameters to the server for aggregation, FL helps mitigate data privacy and data silo issues to some extent [12]. Data in decentralised federated learning is not concentrated in a single central node, and in the event of a data quality problem, it is limited to the local client, rather than triggering a large-scale data quality and security threat as in centralised storage. Nonetheless, several critical challenges remain when applying FL to fake review detection. One key issue is data quality. In a distributed environment, different clients may possess data of varying quality, or even intentionally provide noise or fake data. Distinguishing data quality during training is crucial to improving model performance.

In traditional centralized approaches, data quality is generally ensured by manual filtering, statistical feature-based cleaning, or anomaly detection [13], [14]. In FL, however, the numerous participants and highly heterogeneous data sources make it difficult to enforce a uniform and rigorous cleaning process across all data [15], [16]. If some clients' data contain a large number of mislabeled instances or originate from "fake review generators," their contributions can negatively affect model updates during global aggregation, or even derail model convergence. Therefore, this paper proposes a Data Quality-Aware Federated Learning framework, aiming to integrate data quality evaluation and weighting strategies throughout the model training and aggregation process. This approach endeavors to leverage high-quality data for effective fake review detection while preserving privacy.

The main contributions of this paper include:

1) Proposing a data quality-aware federated learning framework: to address disparities in data quality in a distributed environment, we establish data quality metrics and handle model updates and aggregation differently based on varying quality levels.

2) Constructing a training process that balances privacy protection and data quality: while assessing data quality, we design privacy-protection measures to ensure that neither raw data nor quality indicators of any client are inappropriately exposed.

3) Conducting experimental validation and analysis: employing multiple real or simulated fake review datasets, we compare the proposed method with various baselines and demonstrate that it significantly improves fake review detection while maintaining privacy.

4) Discussing scalability and generalizability: the proposed framework is not limited to fake review detection but can also serve as a reference for text classification and anomaly detection tasks in other domains.

II. Related Work

Fake review detection has long been a significant research branch in Natural Language Processing (NLP) and Social Computing [3], [9]. Early studies mainly focused on feature-engineering-based supervised learning approaches [17]. For instance, some works identify unusual text by analyzing word frequency or sentiment polarity in reviews, or combine user behavior features (e.g., review frequency, account creation date, IP address distribution) for supplementary judgment [18], [19]. However, these methods typically rely on manually selected features, which encounter bottlenecks in scalability and accuracy as the volume and complexity of reviews increase.

The emergence of deep learning has brought new opportunities to fake review detection. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures with attention mechanisms have been applied in text classification, yielding significant progress [20], [21]. Employing pre-trained language models to encode review texts into vector representations, then fine-tuning or customizing layers for classification, has also proven effective [22], [23]. However, such deep models require extensive and diverse training data to capture the implicit deceptive patterns within reviews. For many platforms, the data are scattered across different servers or subsystems and often contain substantial amounts of privacy-related information, making it risky and non-compliant to centralize all data for training.

Federated Learning (FL) [11], [24] was introduced to harness multi-party data under privacy constraints. Companies like Google first tested FL on mobile devices for keyboard input suggestions—allowing each device to train locally and upload updates to a central server for aggregation—drastically reducing the need to collect raw data. Subsequently, FL gained traction in healthcare, finance, IoT, and other fields, leveraging distributed data from diverse participants for collaborative model training [25]. In text analytics, prior work has applied FL to sentiment analysis, text classification, keyword extraction, etc. [26]. Directly adopting FL for fake review detection, however, introduces two major challenges:

Data heterogeneity (Non-IID problem). Different clients may hold data with markedly distinct distributions. Some clients may have a large volume of high-quality labeled data, while others have fewer samples or poorer quality, causing biases in model updates and degrading global performance.

Malicious participants and low-quality data. The open nature of FL makes it difficult to completely exclude malicious clients. If a client intentionally contributes incorrectly labeled data or has very low data quality, it can severely disrupt global model convergence and performance.

To address these issues, recent works have begun examining the role of "data quality" in FL. Some studies propose weighted aggregation based on reliability or contribution, measuring the quality of each client's updates from past iterations to determine their weight, thus mitigating the adverse impact of noisy or malicious clients [27], [28]. Other works adopt secure multi-party computation (SMPC) or differential privacy (DP) techniques to safeguard privacy-sensitive information during data sharing and model updates [29]. However, in the context of fake review detection, data quality involves multiple layers: textual authenticity, possible traces of automated generation, behavioral characteristics of reviewers, and labeling accuracy. These factors require more refined and quantitative evaluation.

Furthermore, effectively spotting deceptive text in a federated environment requires integrating advanced NLP models with the flexibility to handle distributed, heterogeneous data [30]. Compared to traditional, feature-engineering-based centralized approaches, emerging methods highlight:

- **Semantic understanding:** High-dimensional semantic features extracted by pre-trained models such as BERT enable better detection of complex deceptive text [31].
- **Cross-domain integration:** Review patterns can differ greatly across platforms, regions, or user groups, necessitating a federated training method that can accommodate multi-domain differences [32].
- **Data quality filtering:** While preserving privacy, training procedures can embed evaluation and screening mechanisms to minimize the impact of invalid or mislabeled data [33].

Overall, existing works have achieved substantial progress in fake review detection and in privacy/security solutions for FL. However, there remains ample space for improvement in systematically incorporating "data quality" into the FL framework and applying it to fake review detection. Starting from this perspective, our study proposes a Data Quality-Aware Federated Learning method, aiming to advance detection accuracy and robustness without disclosing local data.

III. Methodology

This section details the proposed Data Quality-Aware Federated Learning framework and its application in fake review detection. We describe the overall architecture, data quality assessment, the federated learning training process, and security/privacy protections, along with the corresponding algorithm flow and pseudo-code.

A. Overall Framework

As shown in Figure 1, we establish a federated learning setting with multiple clients and a central server. Suppose there are K participants, each holding a portion of review data with potential differences in source, labeling method, noise level, or even language. To train a global model without sharing raw data, we employ a parameter-based FL strategy: each client trains locally on its own data and periodically sends its model parameters or gradients to the server for aggregation. The aggregated global model is then distributed back to each client for further iterative training.

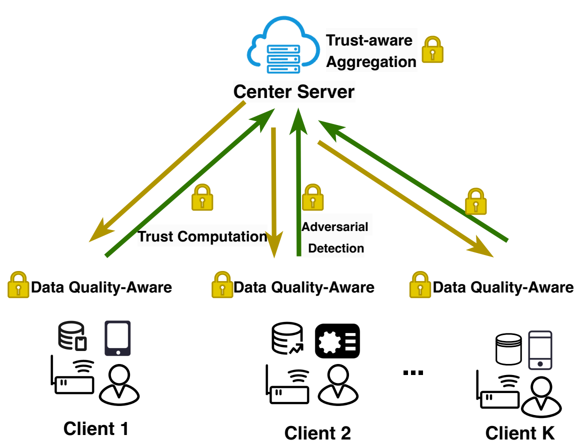

Fig. 1. Illustration of the proposed system architecture.

Unlike conventional FL, we incorporate a "data quality-aware" module into this process. Each client is equipped with a local data quality evaluation sub-module that assesses:

- **Label accuracy:** Estimating annotation reliability with a small validation set or automated consistency checks.
- **Text completeness:** Detecting anomalies like extremely short texts, garbled characters, or repetitive content.
- **Behavioral traits:** If user behavioral data are available (e.g., timestamps, reviewer distribution, review frequency), identifying suspicious patterns of spamming or intentional manipulations.

The resulting "data quality score" is then used to weight each client's contribution during model aggregation on the server. Clients with higher-quality data receive greater weight, thereby reducing the detrimental impact of low-quality data on the global model.

B. Data Quality Assessment

To conduct data quality assessment under privacy constraints, we can adopt the following approaches:

- **Local evaluation and encrypted upload:** Each client calculates its data quality metrics locally and uploads the resulting quality score in an encrypted form. The server uses the score for weighted aggregation but cannot decrypt its actual value, thus avoiding potential privacy leakage.
- **Adversarial checks:** A portion of commonly shared test data can be distributed to each client, requiring them to make predictions on these samples. The server can compare the results to estimate training quality and refine the data quality assessment.

We define each client's data quality score $Q_k \in [0, 1]$ as:

$$Q_k = \alpha \cdot Q_{label} + \beta \cdot Q_{text} + \gamma \cdot Q_{behavior} \tag{1}$$

where Q_{label} measures labeling accuracy, Q_{text} captures text quality (completeness, readability), $Q_{behavior}$ reflects suspicious behavioral patterns in reverse, and α, β, γ are adjustable weight parameters.

C. Federated Learning with Quality-Aware Training

Within the federated learning process, we extend the classic FedAvg algorithm. Let $\mathbf{w}$ denote the global model parameters. Each training round includes:

1) **Model broadcast:** The server sends the current global model parameters $\mathbf{w}^{(t)}$ to all clients.
2) **Local training:** Each client k uses its local data D_k to train the model for several mini-batch gradient descent steps, obtaining updated parameters $\mathbf{w}_k^{(t+1)}$.
3) **Upload updates:** The client sends $\mathbf{w}_k^{(t+1)}$ and the quality score Q_k (encrypted if necessary) to the server.
4) **Weighted aggregation:** The server updates

$$\mathbf{w}^{(t+1)} = \frac{\sum_{k=1}^{K} Q_k \cdot n_k \cdot \mathbf{w}_k^{(t+1)}}{\sum_{k=1}^{K} Q_k \cdot n_k} \tag{2}$$

where $n_k = |D_k|$ is the data size of client k. In traditional FedAvg, aggregation is typically based only on n_k, but here we integrate Q_k to emphasize contributions from clients with higher data quality.

Clients may recalculate Q_k before each round if data or evaluation strategies change. If privacy-preserving methods are applied, the server may only access homomorphically encrypted $\mathbf{w}_k^{(t+1)}$ and must perform aggregation in the encrypted domain.

D. Security and Privacy Protection

To prevent unnecessary disclosure of sensitive data to other clients or the server, privacy-enhancing techniques—such as differential privacy (DP)—are introduced:

- **Differential Privacy:** During local training, clients can inject noise into gradients to protect data privacy, or similarly add perturbations when uploading quality scores to mitigate inference attacks.

E. Algorithm Flow

Algorithm 1 presents the core pseudo-code for Data Quality-Aware Federated Learning.

Algorithm 1 Data Quality-Aware Federated Learning for Fake Review Detection

Initialize: Server initializes global parameters $\mathbf{w}^{(0)}$ randomly or from pre-training, $t = 0$
while not converged **do**
  Server to Clients: Broadcast current global model $\mathbf{w}^{(t)}$
  for all client k **do**
    Evaluate local data quality Q_k
    Train locally on D_k to obtain $\mathbf{w}_k^{(t+1)}$
    Upload $\mathbf{w}_k^{(t+1)}$ (encrypted) and Q_k (encrypted) to server
  end for
  Server Aggregation:
  $\mathbf{w}^{(t+1)} = \frac{\sum_{k=1}^{K} Q_k \cdot n_k \cdot \mathbf{w}_k^{(t+1)}}{\sum_{k=1}^{K} Q_k \cdot n_k}$
  $t := t + 1$
end while
Return Final global model $\mathbf{w}^{(\text{final})}$

In this way, we can effectively evaluate and leverage data of different quality levels for fake review detection, while safeguarding privacy and security.

IV. Experimental Results and Analysis

We now present experimental evaluations of our approach, including setups, comparative methods, and result analyses, highlighting the role of data quality in improving federated learning.

A. Experimental Settings

Datasets and Preprocessing: To simulate realistic conditions, we select the publicly available Amazon Review dataset, which contains reviewer IDs, product IDs, review texts, overall ratings, helpfulness votes, and timestamps. We generate corresponding pseudo-labels and perform essential cleaning and labeling, including handling missing values, normalizing text, removing special characters and stopwords, and conducting tokenization and stemming. Additionally, we extract key text and behavioral features. To emphasize differences in data quality, we artificially or randomly introduce noisy data into certain clients (e.g., high rates of misspellings, incorrect labels, or automatically generated reviews), thus simulating malicious or low-quality data sources.

We split the data among 10 or more clients, each with distinct data distributions and quality levels. 2–3 clients contain noticeably noisy data (incorrect labeling rates of 30–50For text representation, we use pretrained word embeddings for encoding, followed by a lightweight classification network for binary fake review classification. We tune local/global epochs, learning rate, etc., via preliminary experiments. The data quality weighting parameters α, β, γ are chosen to balance the impact of various quality indicators.

B. Comparative Methods

We compare our method with:

- **Centralized Training:** Aggregating all data in a single server without privacy constraints (often considered an ideal upper bound).
- **FedAvg:** The standard federated averaging approach without data quality differentiation.
- **FedProx:** Adding a regularization term to the FL objective to mitigate data heterogeneity but not explicitly addressing data quality.

C. Results and Analysis

Quantitative Metrics: Figure 2 shows each method's Accuracy, Precision, Recall, and F1 scores on the test set. Centralized training achieves the highest accuracy but requires collecting all data, losing privacy benefits. FedAvg performs slightly worse, mainly due to data heterogeneity and noisy clients. While FedProx partially alleviates heterogeneity, performance remains unstable under high noise conditions. By introducing quality weighting, our Data Quality-Aware FL significantly dampens the impact of noisy clients, outperforming the alternatives across multiple metrics.

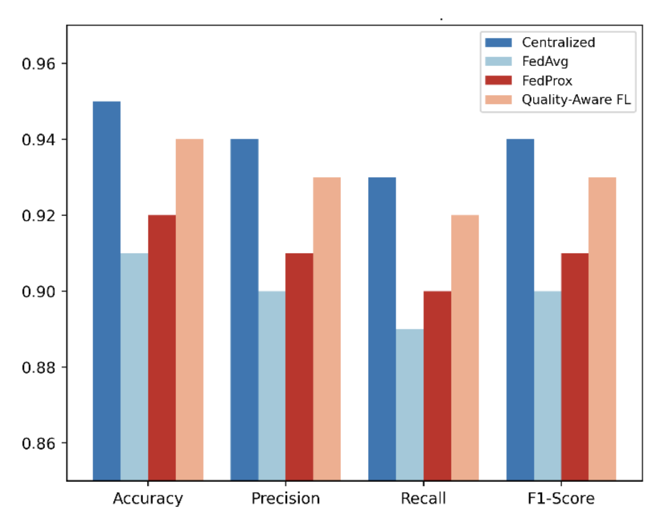

Fig. 2. Performance comparison of different methods.

Convergence Speed and Stability: As illustrated in Figure 3, we track global model convergence over training rounds. The data quality-aware strategy curbs the effect of high-noise clients from the early stages, yielding faster convergence than FL methods without such weighting. Moreover, performance in conventional FL can drastically degrade when the number or severity of high-noise clients grows. The proposed approach, however, maintains better stability via weighted aggregation.

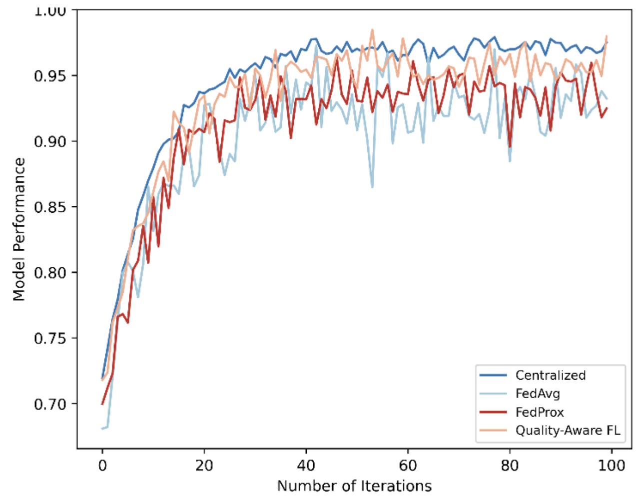

Fig. 3. Training convergence comparison of different methods.

Varying Data Distributions: As shown in Figure 4, under IID distributions (similar data across clients), performance differences are modest. In contrast, when Non-IID conditions prevail (highly imbalanced or divergent data), the data quality-aware FL distinctly outperforms others. This is because higher-quality client updates are emphasized, mitigating disruptions caused by low-quality data.

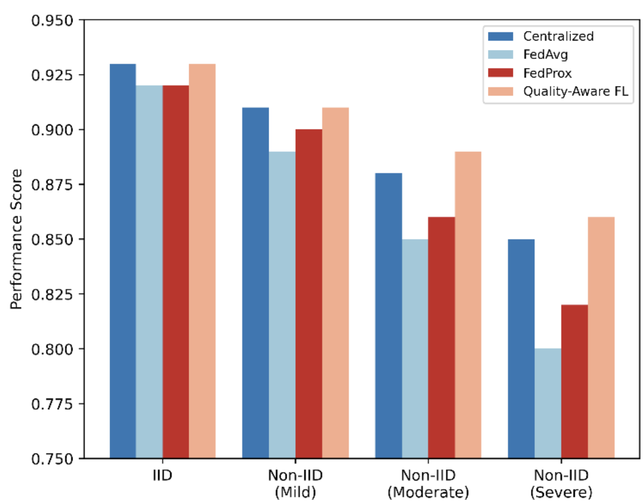

Fig. 4. Comparison of IID vs. Non-IID scenarios.

D. Analysis of Detection Performance on Artificially Injected Fake Reviews

To further validate the capability of our method in detecting artificially injected fake reviews, we introduced 500 artificially generated fake review samples into the experimental dataset. These samples simulate potential review manipulation behaviors, such as fake endorsements or misleading reviews, to assess the model's performance under controlled conditions.

Experimental results indicate that our proposed Data Quality-Aware Federated Learning method successfully identified 472 fake reviews, achieving a detection rate of 94.4%. In contrast, the FedAvg method detected only 392 fake reviews (78.4%), and the FedProx method detected 405 fake reviews (81%). These results demonstrate that the data quality-aware strategy effectively identifies injected noise, significantly outperforming traditional federated learning methods.

E. Discussion

Our experiments underscore the advantages of data quality-aware federated learning for fake review detection:

- **Improved accuracy:** Quality scoring for aggregation effectively counters the influence of noisy or malicious data.
- **Enhanced robustness:** Even under extreme data distribution shifts or malicious clients, our method maintains stable performance.
- **Preserved privacy and scalability:** Without centralizing raw data, FL retains privacy benefits and potential for large-scale deployment.

V. Discussion

Despite the promising results, several limitations and directions remain for future research. Accurately evaluating data quality in real-world settings can be challenging due to limited annotations or highly heterogeneous data. Moreover, malicious clients may engage in collaborative adversarial behaviors, requiring complementary robust training methods. Privacy-preserving techniques often increase computational and communication overhead, necessitating a balance between security and efficiency. Lastly, designing adaptable quality metrics for multilingual or cross-domain fake review detection is another avenue worth exploring.

Future work will refine and automate data quality assessment, integrate more advanced NLP models for multi-platform and multilingual feature learning, and investigate more efficient secure multi-party computation protocols. By continuously enhancing and expanding the approach, we aim to develop data quality-aware FL into a broadly applicable methodology that addresses the growing need for trustworthy text analytics in various fields.

VI. Conclusion

Focusing on fake review detection in a federated learning environment, this paper tackles the common issue of varying data quality in distributed, multi-source settings. We introduce a Data Quality-Aware Federated Learning framework that incorporates data quality evaluation and weighting throughout the training and aggregation steps, enabling more effective utilization of high-quality data while preserving privacy. Experimental results confirm that the proposed approach outperforms traditional FL methods in both accuracy and robustness, especially under Non-IID and noisy conditions.

Future research can delve deeper into data quality assessment techniques by leveraging advanced NLP and user behavior analytics, and explore higher levels of privacy protection for cross-platform, cross-regional collaboration. In essence, data quality-aware FL shows considerable potential for not only fake review detection but also other text analytics tasks, promising a more secure and fair digital ecosystem.

REFERENCES

[1] A. K. Kushwaha, S. Jadon, P. Kamal, M. L. Saini, and V. M. Shrimal, "Comments and feedback verification system using large language model," in *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE, 2024, pp. 1–6.

[2] R. Bharti, U. Khatri, and O. A. Duralia, "Examining the role of digital marketing in shaping consumer communication and behavior," *Feedback International Journal of Communication*, vol. 1, no. 4, pp. 192–201, 2024.

[3] K. Mane and S. Dongre, "A review of different machine learning techniques for fake review identification," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*. IEEE, 2024, pp. 476–481.

[4] C. R. Veda, N. Apoorva, N. Vishnu, M. S. Velpuru, H. Akshaya, and S. P. Siripuram, "Fake review identification using hybrid fusion of machine learning and natural language processing techniques," in *2024 3rd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*. IEEE, 2024, pp. 1–6.

[5] P. Sikchi, S. Jagtap, and J. Kundale, "Fake review detection using neural network," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2024, pp. 1–6.

[6] B. A. Reddy, G. S. Reddy, K. Lokesh, and M. Belwal, "Detecting fraudulent reviews in e-commerce platforms," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–7.

[7] M. Ennaouri, M. Raoui, A. Zellou, and M. H. El Yazidi, "Fake review detection using extravagant words: A comparative study of k-means, k-mode, and hierarchical classification," in *2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE)*. IEEE, 2024, pp. 1–7.

[8] Y. Zhou, "The silent saboteur: The impact and management of malicious word-of-mouth in the digital age," *Highlights in Business, Economics and Management*, vol. 41, pp. 381–386, 2024.

[9] A. Vashist, A. Keswani, V. Pareek, and T. Jain, "Detecting fake reviews on e-commerce platforms using machine learning," in *2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. IEEE, 2024, pp. 1–6.

[10] A. Korkmaz and P. Rao, "A selective homomorphic encryption approach for faster privacy-preserving federated learning," *arXiv preprint arXiv:2501.12911*, 2025.

[11] M. Bharathi and T. A. S. Srinivas, "Federated learning unveiled: From practical insights to bold predictions."

[12] R. Thakur, S. Patel, N. Singh, A. Barde, and S. Barde, "Tackling the complexities of federated learning," *Generative Artificial Intelligence for Biomedical and Smart Health Informatics*, pp. 343–353, 2025.

[13] V. S. de Feiter, J. M. Strickland, and I. Garcia-Marti, "Advancing data quality assurance with machine learning: A case study on wind vane stalling detection," *Atmosphere*, vol. 16, no. 2, p. 129, 2025.

[14] H. Zhang, X. Jia, and C. Chen, "Deep learning-based real-time data quality assessment and anomaly detection for large-scale distributed data streams," 2025.

[15] H. Xu, J. Li, W. Wu, and H. Ren, "Federated learning with sample-level client drift mitigation," *arXiv preprint arXiv:2501.11360*, 2025.

[16] F. Alhafiz and A. Basuhail, "The data heterogeneity issue regarding covid-19 lung imaging in federated learning: An experimental study," *Big Data and Cognitive Computing*, vol. 9, no. 1, p. 11, 2025.

[17] A. Parekar, P. Musale, S. Joshi, C. Attar, R. Shaikh, R. Salve *et al.*, "Detection of fake online review using machine learning," *Shweta and Attar, Chirag and Shaikh, Rihan and Salve, Rajeshwar, Detection of Fake Online Review using Machine Learning (November 15, 2024)*, 2024.

[18] N. Aminuddin, T. Tajuddin, S. Mutalib, S. R. M. Dawam, N. R. Ali, and M. Din, "Deep learning-based classification of toxic online comments using long short-term memory (lstm) for sentiment analysis," *Journal of Information System and Technology Management*, vol. 9, no. 37, pp. 352–368, 2024.

[19] Gagandeep and J. Verma, "Natural language processing for sentiment analysis in social media posts to identify suspicious behaviour," *Abhigyan*, p. 09702385241284879, 2024.

[20] J. T. Jayasinghe and S. Dassanayaka, "Detecting deception: Employing deep neural networks for fraudulent review detection on amazon," 2024.

[21] M. A. Wani, M. ElAffendi, and K. A. Shakil, "Ai-generated spam review detection framework with deep learning algorithms and natural language processing." *Computers (2073-431X)*, vol. 13, no. 10, 2024.

[22] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models." *Future Internet*, vol. 17, no. 1, 2025.

[23] L. Petrillo, F. Martinelli, A. Santone, and F. Mercaldo, "Toward the adoption of explainable pre-trained large language models for classifying human-written and ai-generated sentences," *Electronics*, vol. 13, no. 20, p. 4057, 2024.

[24] K. Lazaros, D. E. Koumadorakis, A. G. Vrahatis, and S. Kotsiantis, "Federated learning: Navigating the landscape of collaborative intelligence," *Electronics*, vol. 13, no. 23, p. 4744, 2024.

[25] W. Hassan and H. Mohamed, "Applications of federated learning in ai, iot, healthcare, finance, banking, and cross-domain learning," in *Artificial Intelligence Using Federated Learning*. CRC Press, 2024, pp. 175–195.

[26] S. Mehta and A. Bhalla, "Enhanced sentiment classification with federated learning cnns: Exploring five sentiment categories," in *2024 3rd International Conference for Advancement in Technology (ICONAT)*. IEEE, 2024, pp. 1–5.

[27] V. Altomare, D. Thakur, A. Guzzo, and F. Piccialli, "Client specific dynamic aggregation for non-iid federated learning," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 7622–7631.

[28] H. Kabbaj, R. El-Azouzi, and A. Kobbane, "Robust federated learning via weighted median aggregation," in *2024 2nd International Conference on Federated Learning Technologies and Applications (FLTA)*. IEEE, 2024, pp. 298–303.

[29] F. Meisingseth and C. Rechberger, "Sok: Computational and distributed differential privacy for mpc," *Cryptology ePrint Archive*, 2024.

[30] L. Tian and N. Jiang, "Research on detection methods for text generated by large language models based on multi-model ensemble," *Applied and Computational Engineering*, vol. 106, pp. 59–67, 2024.

[31] N. Ma and W. Di, "Research on machine reading comprehension integrating longdistance semantic relations and linguistic features," 2024.

[32] Z. Qi, L. Meng, W. He, R. Zhang, Y. Wang, X. Qi, and X. Meng, "Cross-training with multi-view knowledge fusion for heterogenous federated learning," *arXiv preprint arXiv:2405.20046*, 2024.

[33] N. Chakraborty, A. Sharma, J. Dutta, and H. D. Kumar, "Privacy-preserving data quality assessment for time-series iot sensors," in *2024 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*. IEEE, 2024, pp. 51–57.