Online Exam Monitoring System Based on Federated Learning for Privacy-Preserving AI Proctoring*

Moustapha DER[1,†], Zacharia DAMOUE[1,*,†], and Boudal NIANG[1,†], James K. Tamgno[1,*,†], Ahmed Dooguy KORA[1,*,†], and Samba NDIAYE[2,†]

[1] *Ecole Supérieure Multinationale des Télécommunications (ESMT), Dakar, Sénégal*
[2] *Université Cheikh Anta Diop (UCAD), Dakar, Sénégal*

Abstract

The rapid growth of online education has increased the need for secure and reliable remote examination monitoring systems. However, most existing AI-based proctoring solutions rely on centralized architectures that collect and process sensitive student data, raising important concerns about privacy, data security, and regulatory compliance. To address these challenges, this paper proposes FedProctor, a privacy-preserving online examination monitoring framework based on Federated Learning. The proposed architecture enables student devices to collaboratively train a global model while keeping raw behavioral and biometric data on local devices. Furthermore, FedProctor combines multimodal behavioral analysis with Federated Averaging, Differential Privacy, and Secure Multi-Party Computation to improve both monitoring performance and privacy protection. The framework was evaluated in a simulated federated environment involving 50 virtual clients and a dataset of 5,500 normal and suspicious examination activities. The experimental results demonstrate that FedProctor achieves a cheating detection accuracy of 92.4%, which is close to the 94.1% obtained by a centralized approach, while reducing communication overhead by 86.3%. These findings indicate that Federated Learning can effectively balance detection performance and privacy preservation in online examination monitoring. Finally, the study highlights the ethical aspects of AI-based proctoring, including privacy, informed consent, accessibility, and fairness, while identifying real-world deployment and large-scale evaluation as important directions for future work,

Keywords

Federated Learning, Privacy-Preserving AI, Online Proctoring, Differential Privacy, Ethical AI.

1. Introduction

The rapid expansion of online and distance learning has transformed the way educational institutions conduct assessments. Universities, professional training centers, and certification organizations increasingly rely on online examinations to evaluate learners regardless of geographical constraints. While digital assessment offers flexibility and accessibility, it also raises significant concerns regarding academic integrity. Ensuring that examinations remain fair, secure, and resistant to dishonest behaviors has become a major challenge for educational institutions worldwide. To address this issue, many institutions have adopted Artificial Intelligence (AI)-based remote proctoring systems. These systems employ computer vision, machine learning, and behavioral analytics techniques to monitor students during examinations and detect potentially suspicious activities.

CITA 2026 - Land, Smart Cities, and Sustainable Development, 25-26 June 2026, Cotonou, Benin
* Corresponding author.
† *These authors contributed equally.*
moustapha.der@esmt.sn (M. DER); zacharia.damoue@esmt.sn (Z DAMOUE); boudal.niang@esmt.sn (B. NIANG); james.kouawa@esmt.sn (J. K. TAMGNO); ahmed.kora@esmt.sn (A. KORA); samba.ndiaye@ucad.edu.sn (S. NDIAYE)
0009-0007-7489-6363(M. DER); 0000-0002-8970-0948 (Z. DAMOUE); 0000-0002-9458-0063 (B. NIANG); 0000-0002-7636-2054 (A.D. KORA); 0000-0003-0559-9908 (S. NDIAYE)

We propose a federated architecture for online examination monitoring that enables collaborative model training while keeping raw behavioral and biometric data on student devices. We design a multimodal behavioral analysis framework combining facial expression recognition, gaze tracking, head pose estimation, suspicious activity detection, and temporal behavior modeling. We integrate privacy-enhancing mechanisms, including Differential Privacy and Secure Multi-Party Computation, to reduce the risk of information leakage during model aggregation.

We evaluate the proposed framework in a simulated federated environment using behavioral examination datasets and compare its performance with centralized and local-only learning approaches. We discuss the ethical implications, opportunities, and limitations of privacy-preserving AI-based proctoring systems in educational environments. It is important to note that the present study constitutes an initial evaluation of the proposed framework in a controlled simulation environment. Real-world deployments, large-scale validation studies, fairness assessments, and advanced non-IID optimization strategies remain important directions for future research.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on AI-based online proctoring and Federated Learning. Section 3 presents the proposed FedProctor architecture and methodology. Section 4 describes the experimental setup and evaluation results. Section 5 discusses the ethical implications and limitations of the proposed framework. Finally, Section 6 concludes the paper and outlines future research directions.

2. Related Work

2.1 AI-Based Online Examination Proctoring Systems

The rapid adoption of online education has led to the development of numerous Artificial Intelligence (AI)-based examination monitoring systems. These solutions aim to preserve academic integrity by automatically detecting suspicious behaviors during remote examinations. Recent approaches combine computer vision, machine learning, and behavioral analytics techniques to monitor students in real time. Several studies have explored multimodal proctoring frameworks integrating facial recognition, gaze tracking, head pose estimation, object detection, keyboard activity analysis, and audio monitoring. For example, Naveen et al. [3] proposed a multimodal examination monitoring framework capable of combining multiple behavioral signals to improve cheating detection accuracy.

Similarly, Patil et al. [8] developed an AI-driven proctoring system integrating facial analysis and behavioral monitoring to enhance examination supervision. Other studies have focused on real-time behavioral assessment using deep learning techniques and computer vision models to identify anomalous examination activities [6], [7]. Although these systems demonstrate promising detection capabilities, most rely on centralized architectures where video streams, biometric information, and behavioral data are continuously transmitted to remote servers for processing and storage. Such approaches raise important concerns regarding privacy protection, data security, scalability, and regulatory compliance.

Furthermore, centralized monitoring infrastructures increase the consequences of potential data breaches because large volumes of sensitive information are stored in a single location.

In addition to privacy concerns, previous studies have highlighted several challenges associated with AI-based proctoring systems, including algorithmic bias, false-positive detections, lack of transparency, and limited explainability of automated decisions [13], [14]. These limitations have motivated researchers to investigate privacy-preserving alternatives capable of reducing dependence on centralized data collection.

2.2 Federated Learning for Privacy-Preserving Artificial Intelligence

Federated Learning (FL) has emerged as an important paradigm for collaborative machine learning in privacy-sensitive environments. Unlike traditional centralized learning approaches, FL enables multiple clients to train a shared model while keeping raw data locally on their devices. Only model updates are exchanged with a central aggregation server, significantly reducing the exposure of sensitive information.

The pioneering Federated Averaging (FedAvg) algorithm established the foundation for decentralized model training by aggregating locally trained models into a global model. Since then, numerous studies have proposed extensions to address challenges such as communication efficiency, privacy protection, and heterogeneous data distributions. In educational contexts, Latif and Zhai [9] demonstrated the potential of Federated Learning for privacy-preserving assessment systems. Similarly, Liu et al. [15] explored the integration of Federated Learning into intelligent online education environments, highlighting its ability to improve privacy while maintaining acceptable predictive performance. Beyond education, Federated Learning has been successfully applied in healthcare analytics [12], multimedia systems [18], intelligent networks, and edge computing environments [16], [17].

Despite these advances, Federated Learning faces several well-known challenges. Client data is often non-independent and non-identically distributed (non-IID), which can negatively affect model convergence and performance. To address this issue, several optimization algorithms have been proposed, including FedProx, which introduces a proximal regularization term to improve stability under heterogeneous data distributions, and SCAFFOLD, which employs control variates to reduce client drift during training. These approaches have demonstrated improved robustness compared with standard FedAvg in many federated learning scenarios.

2.3 Privacy and Security Mechanisms in Federated Learning

Although Federated Learning reduces the need to share raw data, model updates may still leak sensitive information through inference or reconstruction attacks. Consequently, additional privacy-preserving mechanisms are often incorporated into federated systems. Differential Privacy (DP) is one of the most widely adopted techniques for mitigating information leakage. By injecting carefully calibrated noise into model updates, DP provides mathematically quantifiable privacy guarantees while preserving model utility. Similarly, Secure Multi-Party Computation (SMPC) enables multiple participants to collaboratively compute aggregated results without revealing individual contributions.

Several studies have demonstrated the effectiveness of combining Federated Learning with Differential Privacy and secure aggregation techniques to strengthen privacy protection in distributed learning environments [10], [12]. However, these approaches often involve trade-offs between privacy, communication costs, and model performance. Understanding and balancing these trade-offs remains an active area of research.

2.4 Research Gap and Motivation

The literature review reveals three important gaps. First, most existing AI-based examination monitoring systems continue to rely on centralized architectures that require extensive collection and storage of sensitive student data. While effective from a monitoring perspective, these approaches remain vulnerable to privacy risks and regulatory concerns. Second, although Federated Learning has been extensively studied in domains such as healthcare, IoT, multimedia systems, and intelligent networks, relatively few studies have investigated its application to online examination monitoring.

Existing educational applications primarily focus on learning analytics or automated assessment rather than multimodal behavioral proctoring [13].
Third, current research rarely combines multimodal behavioral analysis, Federated Learning, Differential Privacy, and secure aggregation mechanisms within a unified framework designed specifically for remote examination environments [10]. Moreover, several practical challenges including non-IID data distributions, false-positive detections, fairness evaluation, and large-scale deployment remain insufficiently explored.
To address these limitations, this work proposes FedProctor, a privacy-preserving online examination monitoring framework that combines on-device multimodal behavioral analysis with Federated Learning, Differential Privacy, and Secure Multi-Party Computation [14]. The objective is not to replace existing proctoring systems immediately, but rather to investigate whether privacy-aware collaborative learning can provide a viable foundation for future online examination monitoring solutions.

3. Proposed Methodology

3.1 Overview of the Proposed Framework

This section presents FedProctor, a privacy-preserving online examination monitoring framework based on Federated Learning (FL). The objective of the proposed system is to enable automated behavioral analysis during online examinations while minimizing the collection and transmission of sensitive student information [2]. Unlike conventional AI-based proctoring systems that rely on centralized storage of video streams and biometric data, FedProctor performs behavioral analysis directly on student devices. Raw examination data remain local, while only encrypted model updates are transmitted to the aggregation server [15].
The framework combines four key components:

- On-device multimodal behavioral analysis;
- Federated model training;
- Differential Privacy (DP);
- Secure Multi-Party Computation (SMPC).

The overall architecture is illustrated in Figure 1.

3.2 Federated System Architecture

FedProctor follows a client-server federated learning architecture.

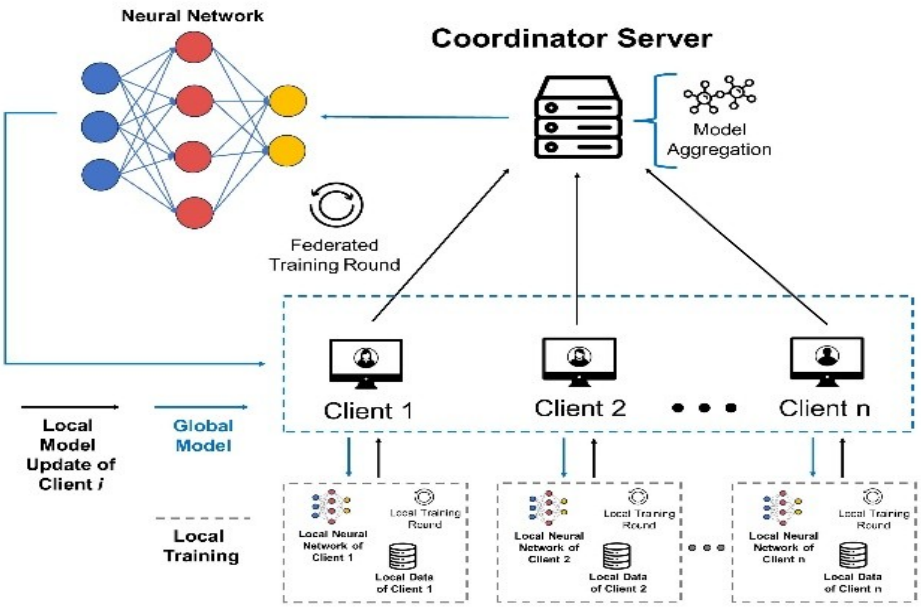

Figure 1: system architecture privacy preserving

Client Layer

Each student device acts as a federated client.

The client is responsible for:

- video acquisition;
- behavioral feature extraction;
- local model training;
- privacy-preserving model update generation.

Raw data never leaves the device.

Server Layer

The aggregation server performs:

- client selection;
- secure model aggregation;
- global model update distribution.

The server does not access individual behavioral records, biometric information, or examination recordings.

$$K=\{1, 2, \dots , N\} \quad (1)$$

(1) be the set of participating clients. At communication round (t), a subset

$$S_t \sqsubseteq K \quad (2)$$

(2) is selected for training. Only the selected clients participate in model updates.

3.3 On-Device Multimodal Behavioral Analysis

FedProctor integrates multiple behavioral modalities to improve the reliability of suspicious activity detection.

3.3.1 Facial Expression Recognition

This module analyzes facial micro-expressions associated with attention changes, confusion, stress, or abnormal behaviors. A lightweight CNN architecture is deployed using TensorFlow Lite or ONNX Runtime [14].

3.3.2 Gaze Tracking and Head Pose Estimation

Eye movements and head orientation are continuously monitored. Features include:
gaze direction;
gaze fixation duration;
head rotation angles;
frequency of off-screen attention.

3.3.3 Suspicious Activity Detection

The system monitors:

- multiple-face appearance;
- unauthorized object presence;
- abnormal keyboard activity;
- unusual mouse interaction patterns.

3.3.4 Temporal Behavioral Modeling

Behavior evolves during an examination session. To capture temporal dependencies, feature sequences are processed using lightweight Long Short-Term Memory (LSTM) networks.
Given a sequence:

$$X = (x_1, x_2, \ldots, x_t) \tag{3}$$

the LSTM generates a hidden representation:

$$h_t = LSTM(x_t, h_{t-1}) \tag{4}$$

which is used for final behavioral classification.

3.4 Federated Learning Framework

FedProctor relies on a Federated Learning (FL) framework to collaboratively train a global cheating detection model without transferring raw student data to a central server. The objective is to leverage behavioral information collected across multiple examination sessions while preserving privacy and minimizing data exposure. Assume a federated system composed of (N) participating clients, where each client (k) owns a local dataset (D_k) containing behavioral observations collected during online examinations [15].

$$n_k = |D_k| \tag{5}$$

(5) denote the number of local samples and

$$n = \sum_{k=1}^{N} n_k \tag{6}$$

(6) the total number of samples across all clients. The global optimization problem is defined as:

$$min_w F(w) \tag{7}$$

$$F(w)=\sum_{k=1}^{N}\frac{n_k}{n}F_k(w) \tag{8}$$

And $F_k(w)$ represents the local loss function associated with client (k).

To solve this distributed optimization problem, FedProctor adopts the Federated Averaging (FedAvg) algorithm as the baseline aggregation strategy. During each communication round, a subset of participating clients receives the current global model, performs local training on its private dataset, and returns updated model parameters to the aggregation server. The server then computes the weighted average of local updates:

$$w^{t+1}=\sum_{k\in S_t}\frac{n_k}{\sum_{j\in S_t}n_j}w_k^{t+1} \tag{9}$$

where (S_t) denotes the subset of selected clients at communication round (t). A major challenge in educational federated environments is the presence of non-independent and non-identically distributed (non-IID) data. Student behavioral patterns can vary significantly due to cultural background, examination conditions, device characteristics, learning habits, and individual cognitive differences. Consequently, local datasets often follow different probability distributions:

$$P_k(X,Y)\neq P_j(X,Y),\quad \forall k\neq j \tag{10}$$

where ($P_k(X,Y)$) denotes the joint distribution of features and labels at client (k). Such heterogeneity may cause client drift, leading local models to converge toward different optimization directions and potentially degrading the performance of the aggregated global model. This issue is widely recognized as one of the main limitations of standard Federated Averaging. To improve robustness in heterogeneous educational environments, FedProctor incorporates several practical mechanisms.

First, behavioral features extracted from facial expressions, gaze movements, head pose estimation, and activity monitoring are standardized through a common preprocessing pipeline before local training. This reduces variability introduced by different devices and examination settings. Second, a partial client participation strategy is adopted. Rather than involving all clients in every communication round, only a fraction of clients participates in model updates.

$$|S_t|=CxN \tag{11}$$

where (C) represents the participation rate. In the experimental evaluation, (C = 0.2), meaning that 20% of available clients participate in each round. This strategy reduces communication overhead while maintaining model convergence [11]. Third, privacy-preserving mechanisms are integrated directly into the federated training process. Before transmission, local model updates are protected through Differential Privacy and Secure Multi-Party Computation techniques, reducing the risk of information leakage from model parameters [16].

It should be noted that the present work does not claim to completely solve the non-IID challenge. The objective is to evaluate the feasibility of privacy-preserving multimodal proctoring under heterogeneous simulated conditions. More advanced federated optimization algorithms specifically designed for non-IID environments, such as FedProx and SCAFFOLD, are not incorporated into the current implementation and will be investigated in future work as part of a broader comparative evaluation framework. The overall federated learning workflow consists of the following steps:

- Selection of participating clients;
- Distribution of the current global model;
- Local multimodal behavioral analysis and feature extraction;
- Local model training;
- Privacy-preserving update generation;
- Secure aggregation of model updates;
- Global model update and redistribution.

This framework enables collaborative learning across distributed examination environments while preserving student privacy and limiting the need for centralized storage of sensitive behavioral and biometric information [17].

3.5 Differential Privacy Mechanism

Although raw data remain local, model updates may still reveal sensitive information. To mitigate this risk, Gaussian Differential Privacy is applied before transmission. Each client computes:

$$\triangle w_k \tag{12}$$

(12) and perturbs the update as:

$$\triangle w_k = \triangle w_k + N(0, o^2) \tag{13}$$

(13) where:

- (o) is the noise scale;
- ($N(0, o^2)$) is Gaussian noise.

The privacy-preserving update is then transmitted to the server. In the experimental evaluation: o = 0.01

Future work will include formal estimation of the corresponding privacy budget ($\mathcal{E}, o$).

3.6 Secure Multi-Party Computation (SMPC)

To prevent the server from observing individual client updates, FedProctor incorporates a secure aggregation mechanism based on SMPC principles. Each client encrypts its local update before transmission. The server only receives encrypted contributions:

$$Enc(\Delta w_k) \tag{14}$$

and computes:

$$\sum_{k=1}^{N} Enc(\Delta w_t) \tag{15}$$

(15) without accessing individual values. Only the aggregated model can be reconstructed. This mechanism strengthens resistance against gradient inspection and reconstruction attacks [10].

3.7 Federated Training Workflow

The complete training process consists of the following steps:

a. Client selection by the aggregation server.
b. Distribution of the current global model.

c. Local behavioral analysis and feature extraction.
d. Local model training.
e. Differential privacy perturbation.
f. Secure aggregation of encrypted updates.
g. Global model update.
h. Redistribution of the updated model.

Algorithm 1: FedProctor Training Procedure

Input:

- Global model (w_0)
- Number of rounds (T)

For each round (t=1, … ,T):

a. Select client subset (S_t)
b. Broadcast global model (w_T)
c. Each client trains locally
d. Apply Differential Privacy
e. Encrypt updates
f. Aggregate updates securely
g. Update global model

Return:

Final global model (w_T)

3.8 Scalability Considerations

The proposed architecture is designed to support large-scale educational environments. Several characteristics contribute to scalability:

- decentralized data storage;
- partial client participation;
- lightweight edge inference;
- reduced communication requirements.

Nevertheless, large-scale deployments involving several thousand concurrent students have not yet been evaluated experimentally. Such validation constitutes an important direction for future work [18].

4. Experimental Results

The objective of the experimental evaluation is to assess the effectiveness of FedProctor from four complementary perspectives: detection performance, robustness to heterogeneous data distributions, privacy preservation, and communication efficiency. In addition, we analyze the impact of the different privacy-preserving mechanisms integrated into the proposed framework. It is important to note that the present evaluation was conducted in a simulated federated learning environment [8]. Therefore, the reported results should be interpreted as an initial validation of the proposed architecture rather than evidence of large-scale real-world deployment.

4.1 Experimental Setup

Experiments were conducted using a simulated federated learning environment composed of 50 virtual clients representing student devices. The simulation was designed to reproduce heterogeneous educational settings involving different behavioral patterns and device capabilities [8]. The global model was trained using the Federated Averaging (FedAvg) algorithm. During each communication

round, 20% of available clients were randomly selected for participation. Each selected client performed three local training epochs before transmitting model updates to the aggregation server [5].

Table 1

Federated Learning Configuration

Parameter	Value
Number of clients	50
Client participation rate	20%
Local epochs	3
Aggregation algorithm	FedAvg

Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score. Additional metrics included communication overhead reduction, false positive rate, and privacy leakage resistance.

4.2 Dataset Description

The evaluation relied on a combination of public behavioral datasets and a custom augmented dataset designed to represent online examination scenarios.

4.2.1 OEP Dataset

The Online Exam Proctoring (OEP) dataset contains behavioral events associated with examination monitoring, including gaze movements, facial expressions, and attention-related indicators.

4.2.2 7WiseUp Dataset

The 7WiseUp dataset provides behavioral and interaction traces collected from online learning activities and was used to enrich temporal behavioral modeling.

4.2.3 Custom Augmented Dataset

To complement publicly available datasets, a custom dataset containing 5,500 behavioral records was constructed. The dataset includes two categories:

- Normal examination behavior;
- Suspicious examination behavior.

Data augmentation techniques such as temporal segmentation, sequence variation, and behavioral event recombination were applied to increase diversity and reduce class imbalance. Although this dataset improves experimental coverage, it was generated under controlled conditions and may not fully represent the diversity of real-world examination environments. Consequently, external validity remains a limitation of the present study [1].

4.3 Detection Performance Evaluation

FedProctor was compared against four baseline approaches:

a. Local-only training;
b. Centralized learning;
c. Standard FedAvg;
d. FedProx.

Table 2

Detection Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Local Only	87.6	88.1	84.5	86.2
Centralized	94.1	94.8	92.3	93.5
FedAvg	90.8	91.5	88.9	90.2
FedProx	91.4	92.1	89.5	90.8
FedProctor	92.4	93.2	90.5	91.8

The proposed framework achieved competitive performance compared with centralized learning, with an accuracy difference of only 1.7 percentage points.

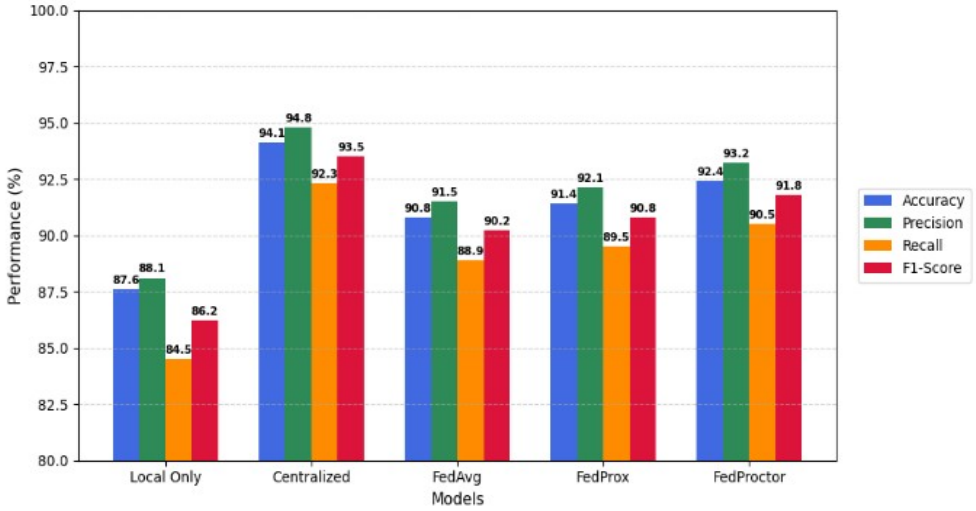

Figure 2: Model Performance Comparison

These results suggest that privacy-preserving federated architectures can maintain effective cheating detection capabilities while substantially reducing data exposure.

4.4 Non-IID Data Analysis

Because student behavior naturally varies across educational environments, evaluating performance under heterogeneous data distributions is particularly important. Three distribution scenarios were considered:

- IID distribution;
- Mild non-IID distribution;
- Severe non-IID distribution.

Table 3

Performance Under Different Data Distributions

Distribution	Accuracy (%)
IID	92.4
Mild non-IID	91.2
Severe non-IID	89.7

Results indicate a gradual decrease in performance as data heterogeneity increases. This observation is consistent with previous Federated Learning studies and confirms that non-IID data remains a challenging problem for distributed training. Nevertheless, the performance degradation remained moderate, suggesting that the proposed architecture can tolerate a reasonable degree of behavioral diversity.

4.5 Ablation Study

To evaluate the contribution of each component, an ablation study was conducted.

Table 4

Ablation Study

Configuration	Accuracy (%)
FedAvg	90.8
FedAvg + Client Selection	91.5
FedAvg + Differential Privacy	91.1
FedAvg + SMPC	90.9
Full FedProctor	92.4

The results show that no individual component alone explains the observed performance.

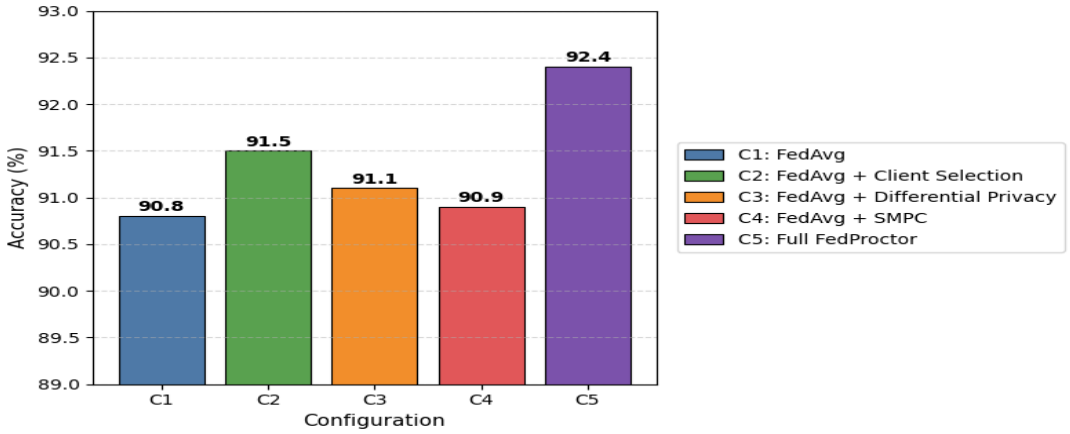

Figure 3: Model Configuration Comparison

Rather, the combination of client selection, privacy-preserving mechanisms, and multimodal behavioral analysis contributes to the overall effectiveness of the framework.

4.6 Privacy and Communication Efficiency Analysis

One of the primary objectives of FedProctor is reducing the exposure of sensitive student information. Compared with centralized architectures that continuously transmit video streams and biometric information, the proposed framework exchanges only model parameters [11].

Table 5

Communication Efficiency

Method	Relative Communication Cost
Centralized	100%
FedProctor	13.7%

This corresponds to an 86.3% reduction in communication overhead. To assess privacy protection, reconstruction attack simulations were performed. Attack success rates remained significantly lower than those observed in centralized settings due to the combined effect of local training, Differential Privacy, and secure aggregation. However, a formal privacy budget analysis was not performed. Consequently, the privacy guarantees provided by the current implementation should be interpreted as empirical rather than formally certified.

4.7 False Positive Analysis

False accusations represent a critical concern in academic integrity systems.

Table 6

Summarizes error-related metrics

Metric	Value (%)
False Positive Rate	4.7
False Negative Rate	5.3

Although the false positive rate remains relatively low, incorrect identification of honest students may still occur. Therefore, FedProctor should be considered a decision-support tool rather than a fully autonomous disciplinary system. Human review remains essential before any academic decision is made.

4.8 Discussion of Results

The experimental results suggest that FedProctor can achieve a favorable balance between privacy preservation and detection performance. Compared with centralized approaches, the framework substantially reduces communication requirements and limits the transmission of sensitive behavioral information while maintaining competitive predictive performance. Nevertheless, several limitations should be acknowledged. First, all experiments were conducted in a simulated federated environment rather than under real examination conditions. Second, the custom dataset may not fully capture the diversity of global student populations [15].

Third, fairness metrics across demographic groups were not evaluated. Finally, more advanced federated optimization algorithms such as SCAFFOLD and adaptive federated methods were not considered in the present study. Future work will focus on large-scale real-world deployments, formal privacy budget estimation, fairness evaluation, and comparative analysis against state-of-the-art federated learning algorithms.

5. Discussion and Ethical Considerations

5.1 Discussion of Findings

The experimental results indicate that FedProctor can achieve competitive cheating detection performance while significantly reducing the transmission of sensitive behavioral and biometric information. Compared with centralized learning, the proposed framework achieved only a moderate decrease in predictive performance while substantially lowering communication requirements and limiting direct exposure of raw examination data. These findings suggest that Federated Learning represents a promising approach for balancing academic integrity requirements with privacy protection objectives. By keeping examination recordings and behavioral observations on student devices, the framework reduces reliance on centralized storage infrastructures, which are often associated with privacy, security, and regulatory concerns [17].

The results also demonstrate that integrating Differential Privacy and Secure Multi-Party Computation into the federated training process can strengthen privacy protection without causing a substantial degradation in model performance. Nevertheless, the trade-off between privacy and utility remains an important challenge that requires further investigation through formal privacy budget analysis and large-scale experimentation. The non-IID experiments further highlight the impact of heterogeneous behavioral distributions on federated model performance [13].

Although the observed degradation remained moderate in the simulated environment, the results confirm that data heterogeneity remains a critical challenge for educational Federated Learning systems. Future work should explore advanced optimization strategies specifically designed to address client drift and distributional divergence. Overall, the current findings provide preliminary evidence supporting the feasibility of privacy-preserving online examination monitoring. However, they should not be interpreted as definitive proof of effectiveness under real-world deployment conditions.

5.2 Privacy and Data Protection Considerations

Privacy protection constitutes one of the primary motivations for the development of FedProctor. Traditional AI-based proctoring systems frequently require continuous transmission and centralized storage of video recordings, audio streams, facial images, and behavioral logs. Such architecture increases the risks associated with unauthorized access, data breaches, excessive surveillance, and secondary use of personal information [14]. FedProctor adopts a privacy-by-design approach in which raw examination data remain locally stored on student devices. Only model parameters are exchanged during collaborative training, while Differential Privacy and Secure Aggregation mechanisms provide additional protection against information leakage.

Despite these advantages, privacy risks cannot be considered eliminated. Recent studies have demonstrated that model updates may still reveal partial information through sophisticated inference or reconstruction attacks. Consequently, privacy-preserving technologies should be viewed as risk-reduction mechanisms rather than absolute guarantees. Furthermore, institutions deploying AI-assisted proctoring systems must ensure compliance with applicable legal and regulatory frameworks governing personal data processing, transparency, and user consent.

5.3 Fairness, Accessibility, and Inclusiveness

Fairness represents one of the most frequently discussed concerns surrounding AI-assisted examination monitoring. In the current study, no demographic fairness evaluation was conducted. Consequently, the experimental results do not provide evidence regarding performance differences across demographic groups, cultural backgrounds, gender categories, age groups, or students with disabilities [10].

Therefore, this work does not claim that Federated Learning automatically eliminates bias or guarantees equitable outcomes. Although decentralized learning may enable training on more diverse data sources, fairness remains an empirical question that must be evaluated through dedicated experiments and appropriate fairness metrics. Similarly, accessibility challenges remain largely unexplored. Students with neurodiverse conditions, atypical behavioral patterns, motor impairments, or limited access to high-quality hardware may interact differently with AI-based monitoring systems. These differences could influence system performance and user experience. Future research should incorporate fairness audits, demographic performance analyses, accessibility assessments, and participatory evaluation methodologies involving diverse student populations.

5.4 False Positives and Human Oversight

One of the most critical ethical concerns associated with automated examination monitoring is the risk of false-positive detections. A false positive occurs when a legitimate student's behavior is incorrectly classified as suspicious or fraudulent. In educational contexts, such errors may have significant academic, psychological, and reputational consequences.

The experimental evaluation reported a false positive rate of 4.7%, indicating that incorrect classifications remain possible despite the overall performance of the proposed framework. Consequently, automated behavioral analysis should not be used as the sole basis for disciplinary decisions [11]. FedProctor is designed as a decision-support system rather than an autonomous decision-making mechanism. Suspicious events identified by the model should be reviewed by qualified human examiners before any academic action is taken. Maintaining meaningful human oversight is essential for ensuring procedural fairness, contestability of decisions, and protection of student rights.

5.5 Limitations of the Present Study

Several limitations must be acknowledged. First, all experiments were conducted in a simulated federated environment involving virtual clients rather than real students participating in live examinations. Consequently, the reported results may not fully reflect operational conditions encountered in practice. Second, the custom behavioral dataset was generated under controlled conditions and may not capture the full diversity of examination environments, behavioral patterns, and cultural contexts observed across educational institutions. Third, although Differential Privacy and Secure Multi-Party Computation were incorporated into the architecture, a formal privacy budget analysis was not performed. Therefore, the privacy guarantees provided by the current implementation remain empirical rather than mathematically quantified [13].

Fourth, fairness, accessibility, and demographic performance variations were not experimentally evaluated. These aspects remain open research questions. Finally, only a limited number of federated optimization strategies were considered. Comparative evaluation against additional approaches such as SCAFFOLD, FedNova, and adaptive federated optimization methods constitutes an important direction for future work.

5.6 Future Research Directions

Several opportunities exist for extending the proposed framework. Future work will focus on conducting real-world pilot deployments involving students and educational institutions under

authentic examination conditions. Such studies will provide a more realistic assessment of usability, scalability, communication efficiency, and detection reliability [1]. Additional research should investigate formal Differential Privacy guarantees, advanced secure aggregation mechanisms, and stronger defenses against inference attacks. Furthermore, fairness-aware evaluation protocols should be integrated to measure demographic performance variations and identify potential sources of algorithmic bias. Accessibility-oriented adaptations should also be explored to ensure that AI-assisted proctoring systems remain inclusive and equitable for diverse learner populations [14].

Finally, comparative studies involving state-of-the-art federated optimization algorithms and larger-scale federated infrastructures will be necessary to assess the long-term viability of privacy-preserving examination monitoring systems. The results presented in this study should therefore be viewed as an initial step toward the development of privacy-aware, transparent, and responsible AI-assisted examination monitoring frameworks rather than as a fully validated operational solution [12].

6. Conclusion and Future Work

This paper presented FedProctor, a privacy-preserving online examination monitoring framework based on Federated Learning. The proposed approach was motivated by the growing need to balance academic integrity requirements with the protection of students' sensitive behavioral and biometric data. Unlike conventional centralized proctoring systems, FedProctor performs multimodal behavioral analysis locally on student devices and relies on collaborative model training without transferring raw examination data to a central server. The proposed framework integrates facial expression recognition, gaze tracking, head pose estimation, suspicious activity detection, and temporal behavioral modeling within a federated learning architecture. To strengthen privacy protection, Differential Privacy and Secure Multi-Party Computation mechanisms were incorporated into the collaborative training process.

Experimental evaluation conducted in a simulated federated environment involving 50 virtual clients demonstrated that the proposed framework could achieve competitive cheating detection performance while significantly reducing communication requirements. FedProctor achieved an accuracy of 92.4%, compared with 94.1% for a centralized baseline, while reducing communication overhead by 86.3%. These results suggest that privacy-preserving federated architectures may provide a viable alternative to traditional centralized proctoring approaches. The study also highlighted the influence of heterogeneous data distributions on federated model performance and demonstrated that acceptable detection accuracy can be maintained under simulated non-IID conditions. Furthermore, the integration of privacy-enhancing mechanisms was shown to preserve model utility while limiting direct exposure of sensitive student information.

Nevertheless, several limitations must be acknowledged. First, all experiments were conducted in a simulated federated environment rather than in real examination settings. Second, the custom behavioral data set was generated under controlled conditions and may not fully represent the diversity of real-world educational contexts. Third, formal privacy budget analysis was not performed, and therefore the privacy guarantees of the current implementation remain empirical rather than mathematically quantified. In addition, fairness, accessibility, and demographic performance variations were not evaluated and should not be inferred from the present results. Consequently, the findings reported in this work should be viewed as preliminary evidence supporting the feasibility of privacy-aware AI-assisted examination monitoring rather than as definitive validation of a production-ready system.

Several directions for future research emerge from this study. First, large-scale pilot deployments involving real students, heterogeneous devices, and authentic examination environments are necessary to assess operational feasibility and user acceptance. Second, future work should incorporate formal

Differential Privacy accounting and stronger protection mechanisms against inference and reconstruction attacks. Third, fairness-aware evaluation protocols should be developed to investigate demographic performance differences and potential sources of algorithmic bias. Finally, comparative analyses involving advanced federated optimization algorithms such as FedProx, SCAFFOLD, FedNova, and adaptive federated learning approaches will be essential to improve robustness under highly heterogeneous data distributions.

Overall, FedProctor contributes to the growing body of research on privacy-preserving educational technologies by demonstrating how Federated Learning can be leveraged to reduce dependence on centralized data collection while maintaining competitive behavioral monitoring performance. The proposed framework represents an initial step toward the development of more privacy-aware, transparent, and responsible AI-assisted online examination monitoring systems.

Declaration on Generative AI

During the preparation of this work, the author used **ChatGPT** to support the literature search, synthesize related work, and assist in drafting the related work section. After using this tool, the author carefully reviewed, verified, and edited all generated content to ensure its accuracy, originality, and scientific validity. The author takes full responsibility for the content of this publication.

References

[1] M. Malhotra and I. Chhabra, "MANIT: a multilayer ANN integrated framework using biometrics and historical features for online examination proctoring," IEEE Access, vol. 13, pp. 11245–11263, 2025, doi: 10.1038/s41598-025-15486-8.

[2] M. Der, A. D. Kora, B. Niang, and S. Ndiaye, "Automated Detection of Suspicious Behavior During Online Exams Using Artificial Intelligence and Computer Vision," Ingénierie des systèmes d'information, Sep. 2025.

[3] A. K. Naveen, S. Verma, and R. Gupta, "AutoOEP: A Multi-modal Framework for Online Exam Proctoring," in Proc. Int. Conf. Artificial Intelligence and Smart Systems, 2025, pp. 1–8.

[4] M. Der, A. D. Kora, and S. Ndiaye, "Two-factor Biometric Authentication System Based on Facial Recognition Using the SVC Model: The Case of Senegal," in Proc. Conf., May 2025.

[5] M. Der, A. D. Kora, and S. Ndiaye, "Study of AI-based Architectures for Remote Examination Monitoring Using Machine Learning," Journal / Conference Proceedings, May 2025.

[6] H. Sharma and V. K. Pant, "AI-Driven Online Exam Proctoring: An Enhanced Machine Learning Approach," Journal of Recent Innovations in Computer Science and Technology, vol. 5, no. 2, pp. 34 –49, 2025, doi: 10.70454/JRICST.2025.20405.

[7] A. G. Kumar, P. Natarajan, and S. Iyer, "AI-Powered Digital Behavior Analysis System (AIDBAS): Real-Time Proctoring Using Computer Vision and Machine Learning for Online Examination Integrity," SSRN Electronic Journal, 2026, doi: 10.2139/ssrn.6735562.

[8] S. Patil, A. More, and P. Deshmukh, "A Multi-Modal AI Proctoring System for Remote Examinations," International Journal of Engineering Research & Technology (IJERT), vol. 15, no. 5, pp. 221–228, 2026.

[9] E. Latif and X. Zhai, "Privacy-Preserved Automated Scoring using Federated Learning for Educational Research," arXiv preprint arXiv:2503.11711, 2025.

[10] C. Jin, Y. Wang, and H. Zhao, "DMAFL: Effective defense against malicious attacker federated learning framework via blockchain and TFHE," Journal of King Saud University Computer and Information Sciences, vol. 37, no. 4, pp. 1021–1037, 2025, doi: 10.1007/s44443-025-00256-3.

[11] J. Moreno, L. Silva, and F. Costa, “Introducing a Quality-Driven Approach for Federated Learning,” Sensors, vol. 25, no. 10, p. 3083, 2025, doi: 10.3390/s25103083.

[12] B. Li, Y. Wu, J. Song, and Y. Wang, “A Survey of Federated Learning in Education: Opportunities, Challenges and Future Directions,” IEEE Access, vol. 11, pp. 45812–45831, 2023.

[13] P. Rodrigues and M. Fernandez, “Ensuring academic integrity through automated online exam proctoring: a decade long systematic review,” Discover Education, vol. 5, no. 1, pp. 1–28, 2026, doi: 10.1007/s44217-026-01224-3.

[14] S. Ibrahim and A. K. Singh, “A comprehensive review of the changing landscape of academic dishonesty in automated proctoring in the era of artificial intelligence,” Discover Education, vol. 5, no. 1, pp. 1–31, 2026, doi: 10.1007/s44217-026-01275-6.

[15] Y. Liu, J. Kang, D. Niyato, and S. Xie, “Privacy-Preserving Federated Learning for Intelligent Online Education Systems,” IEEE Transactions on Learning Technologies, vol. 18, no. 2, pp. 155–169, 2025, doi: 10.1109/TLT.2025.3489123.

[16] T. D. Nguyen, P. Rieger, H. H. Nguyen, and K. Hamlen, “Privacy-Preserving Deep Learning via Differential Privacy and Federated Learning: A Survey,” *IEEE Access*, vol. 10, pp. 111350–111374, 2022.

[17] A. Cotovanu, M. H. Choudhury, and P. Brusilovsky, “Ethical Challenges of AI-Based Online Proctoring Systems: A Systematic Review,” *Computers and Education: Artificial Intelligence*, vol. 4, 2023.

[18] S. Costan, R. Gonzales, and J. O. C. Cruz, “Online Proctoring in Higher Education: Privacy, Fairness, and Student Acceptance,” *Education and Information Technologies*, vol. 29, no. 3, pp. 3115–3134, 2024.