AI-Driven Automatic Proctoring System for Secure Online Exams

Aleksei Samarin
ITMO University
St. Petersburg, Russia
avsamarin@itmo.ru

Aleksei Toropov
ITMO University
St. Petersburg, Russia
toropov.ag@hotmail.com

Egor Kotenko
ITMO University
St. Petersburg, Russia
kotenkoed@gmail.com

Alexander Savelev
St. Petersburg Electrotechnical University
St. Petersburg, Russia
algsavelev@gmail.com

Artem Nazarenko
ITMO University
St. Petersburg, Russia
aanazarenko@itmo.ru

Elena Mikhailova
ITMO University
St. Petersburg, Russia
e.mikhailova@itmo.ru

Valentin Malykh
International IT University
Almaty, Republic of Kazakhstan
valentin.malykh@phystech.edu

Alexander Motyko
St. Petersburg Electrotechnical University
St. Petersburg, Russia
aamotyko@etu.ru

Anastasia Mamaeva
ITMO University
St. Petersburg, Russia
asmamaeva.science@gmail.com

***Abstract*—This study presents a microservice-based AI-driven proctoring system for secure and scalable online exam monitoring. The proposed system integrates deep learning and computer vision techniques to analyze multimodal data, including video streams, audio signals, and metadata, to detect dishonest behaviors such as unauthorized assistance, device usage, and unusual gaze patterns. The system architecture ensures seamless integration with online learning platforms, providing a modular and adaptive approach to remote exam supervision. Key components include real-time facial recognition, eye-tracking, head pose estimation, and audio anomaly detection. Explainable AI techniques enhance transparency, allowing educators to interpret decisions and minimize false positives. Experimental evaluation on a controlled dataset demonstrated high detection accuracy and efficiency, validating the system's applicability for automated proctoring. The microservice structure allows for flexible deployment, making it suitable for large-scale educational environments. Future improvements will focus on refining detection models, reducing bias, and addressing ethical considerations related to student privacy. This research contributes to the advancement of AI-powered academic integrity solutions, offering a practical and scalable alternative to traditional proctoring methods.**

***Index Terms*—online proctor, eye tracking, student authentication, academic integrity, behavior detection, regression of student conscientiousness**

I. Introduction

The COVID-19 pandemic significantly impacted education by forcing institutions worldwide to transition from traditional classrooms to distance learning formats rapidly [1], [2]. This accelerated the adoption and development of online education, leading more students to explore online certification programs, including massive open online courses (MOOCs) [3], [4]. Universities expanded their online course offerings, enabling broader resource accessibility and creating opportunities for skill enhancement [5]. Consequently, organizing online exams has become an essential element of contemporary education.

Increased reliance on online learning environments highlights the need for robust mechanisms to evaluate student performance remotely. While online assessments offer flexibility, they introduce significant security challenges, particularly regarding effective monitoring. Traditional exam supervision by teachers proves insufficient or resource-intensive in remote settings, resulting in increased academic dishonesty, such as plagiarism and unauthorized assistance. Research by King et al. [7] indicates a cheating rate of up to 73% in online exams, with Warble [8] reporting that 58% of surveyed students admit to cheating. The emergence of sophisticated cheating methods, including screen sharing, virtual machines, and AI assistance tools, exacerbates the difficulty of ensuring exam integrity.

To address these issues, artificial intelligence (AI) technologies are increasingly employed in developing automated proctoring systems. These systems utilize facial recognition, gaze tracking, keystroke dynamics, and speech analysis to detect suspicious behaviors during exams. AI-driven solutions offer real-time monitoring, immediate alerts, and continuous improvements in detection accuracy through machine learning.

Despite their benefits, AI-based proctoring systems raise ethical and technical concerns, such as privacy violations, algorithmic bias, and accessibility limitations. Students and educators express apprehensions about intrusiveness and potential false accusations caused by technical malfunctions or environmental factors. Additionally, AI proctoring effectiveness varies by testing context, requiring further research to improve robustness and fairness.

This study explores the development and implementation of an AI-driven automatic proctoring system to secure online examinations. By analyzing existing methodologies, addressing limitations, and proposing improvements, the research aims to advance fairness, security, and efficiency in remote assessments. The findings seek to guide educational institutions, policymakers, and technology developers on integrating AI into online examination systems ethically and effectively.

We propose a conceptual framework for a proctoring service built upon a custom-designed transformer-based neural architecture developed specifically for this purpose. The introduced architecture processes multimodal input data related to online assessments and examination administration. The system is capable of performing gaze direction estimation, face and person detection (including multi-person scenarios), and monitoring for participant absence from the designated examination zone. Consequently, it enables the identification of behavioral patterns that indicate violations of examination protocols or academic dishonesty.

II. Related Work

Numerous studies have proposed diverse methods for reliable online proctoring, each targeting academic misconduct detection. This section evaluates existing works, highlighting their strengths, limitations, and comparative features.

A. Methods and Approaches in Online Proctoring

Li et al. [12] introduced a proctoring framework combining two webcams, gaze tracking, and an EEG sensor to detect cheating by external materials. Despite robust hardware capabilities, its narrow focus limits broader applicability.

Wahid et al. [13] developed a web-based exam system emphasizing cheating prevention but heavily reliant on human supervision. This dependence makes it costly and impractical for large-scale use.

Rosen and Carr [14] presented a semi-automated system using a desktop robot with 360-degree camera capabilities. Although effective within its field of view, it fails to detect cheating occurring outside the monitored area.

O'Reilly and Creagh [15] proposed an AI-based Proctoring System (AIPS) fully automating student monitoring. Despite advanced detection capabilities, it generates false positives due to difficulty interpreting contextual nuances, affecting accuracy.

B. Industry Solutions and Their Features

ProctorU, developed by Milone et al. (2017) [16], is a live proctoring system using a webcam and microphone for real-time student monitoring. However, its reliance on human proctors raises security concerns due to potential human error and scalability limitations for large-scale use.

Kryterion, introduced by Prathish et al. (2016) [17], also employs webcam and microphone monitoring. Despite wide adoption, Kryterion faces criticism for security vulnerabilities and limited functionality, as it depends heavily on human proctors and lacks advanced AI detection capabilities.

XProctor, proposed by Slusky (2020) [18], integrates facial recognition, video streaming, and LMS integration, providing a versatile automated monitoring solution. Its effectiveness, however, relies heavily on high-quality hardware and stable internet connections, posing challenges for institutions with limited resources.

TeSLa, developed by Draaijer et al. (2017) [19] and funded by the European Commission, uses biometric methods such as facial recognition, voice analysis, and keystroke dynamics. Although highly secure, its specialized hardware requirements present cost and logistical challenges, limiting accessibility and scalability.

ProctorExam offers dual-view proctoring using screen sharing and smartphone cameras for comprehensive monitoring. Despite effective surveillance, its extensive data collection has raised significant privacy concerns [20], [21], complicating institutional data protection efforts.

Safe Exam Browser securely locks down the testing environment, restricting access to unauthorized resources [22]–[24]. However, it lacks behavioral monitoring features, reducing its effectiveness in detecting activities such as external assistance or secondary device usage.

The reviewed solutions illustrate the evolving field of online proctoring technologies. Existing systems predominantly rely on human oversight or have limited capability in detecting diverse cheating behaviors. AI-driven automated systems show promise but require enhancements in accuracy and the reduction of false positives.

However, existing solutions typically feature narrow integration interfaces, restricting their applicability, or lack transparency due to internal benchmarking limitations. Our proposed approach emphasizes infrastructure flexibility, and this article presents our pipeline's testing results on proprietary data. This research is focused on hybrid models combining gaze tracking, facial recognition, keystroke dynamics, and audio analysis for more reliable proctoring.

III. Problem Statement

There is a pressing need for effective methods to detect dishonest behavior during exams. Most existing solutions are commercial, closed-source, and limited in functionality, relying on simplistic rules that fail to capture complex cheating behaviors and often raise privacy concerns [9], [10].

The expansion of online education requires adopting open technologies to encourage innovation, transparency, and improved monitoring quality [11]. Open-source solutions allow customization and collaboration. In Russia, online proctoring remains underexplored, with limited professional understanding of its potential and implementation.

This paper introduces an AI-based approach utilizing computer vision to monitor academic integrity in online exams. Our automated algorithm detects cheating behaviors, unauthorized materials, and external assistance. We quantify student honesty by analyzing video, audio, head poses, mouth movements, eye gaze, unauthorized persons, and electronic device usage (Fig. 1). The core module is designed to automatically detect segments within each user recording (i.e., video stream) that contain visual patterns indicative of suspicious or non-compliant behavior. These patterns may include prolonged gaze diversion, temporary absence from the examination area,

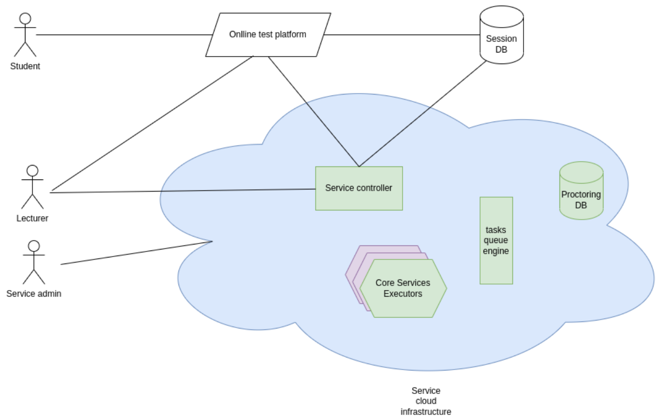

Fig. 1. Design of the AI-driven proctoring system architecture.

or the appearance of additional individuals in the frame, among other anomalies.

Leveraging advanced deep learning models and multimodal data, our system identifies subtle signs of dishonesty, like inconsistent eye patterns or unusual movements. Explainable AI techniques ensure transparent and interpretable results, enhancing educator trust.

Our solution aims to significantly reduce academic dishonesty, minimize false positives, and provide scalable, cost-effective exam monitoring. This research advances AI-driven educational technology and supports future developments in online proctoring.

IV. Proposed solution

This work presents a novel architecture of a neural core based on a modified transformer structure with multimodal input (video and audio), along with a detailed description of the corresponding processing pipeline. A microservice-based system architecture has also been developed, enabling flexible integration with various educational platforms and facilitating scalable deployment. Rather than generalizing existing solutions, we introduce a new approach specifically designed for adaptability and customization in diverse educational scenarios.

The basis of the ecosystem of microservices serves as a core based on artificial intelligence. This system, in essence, solves several regression problems to determine the degree of conscientiousness performance of assignments by students based on several video recordings of passing the exam, as well as according to additional meta-information accompanying this series (information about the task being performed and an audio track of the submission process).

Our model is built upon a stack of 6 succeeding Transformer layers, receiving modality-specific vectorized inputs produced by dedicated pretrained encoders. Gaze direction estimation was handled using the MobileGaze [27] architecture, while facial expression activity and keypoint analysis were performed using the [28] and GFE [29] architectures, respectively. Additionally, a separate vectorization block was implemented to process the outputs of face detection modules.

Each modality was encoded using a pre-trained encoder, followed by a series of Transformer layers trained from scratch on our dataset. To integrate multiple modalities and submodalities, we employed cross-attention blocks at the input stage. In this design, while the modality-specific encoders were pretrained, the core Transformer module with cross-attention was trained from scratch, allowing the model to learn task-specific intermodal dependencies.

A. The Pre-processing Module

The design and development of the solution were grounded in the foundational principles of the LFIEM [25] and UniFi [26] architectures and methods, which were selected as the core frameworks. These architectures provided a robust and scalable basis, ensuring flexibility, interoperability, and efficiency in system integration. By leveraging their structural advantages, the proposed solution benefits from enhanced adaptability and seamless deployment within diverse technological environments.

In addition to the central model, the core of the system also includes a data preprocessing module based on fixed transformations of a local and global nature.

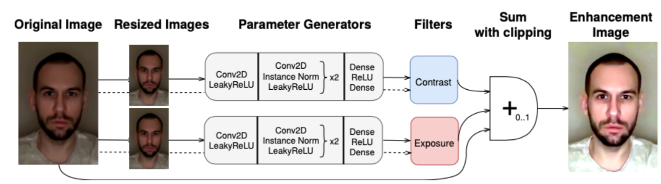

Fig. 2. Preprocessing model structure.

Fig. 2 demonstrates the general neural network architecture of the preprocessing module, which is built on the use of a stack of explicitly defined filters with trainable parameters. It should be noted that this scheme was previously developed in our projects for other purposes, but for this task, it was additionally adapted: only those color correction filters (the so-called saturation, sharpness, and contrast) were used, the contribution of which is most significant for the quality of the algorithms associated with determining the direction of gaze. For this, we conducted several experiments, which allowed us to identify the operations most applicable to our domain.

Among the features of this module, it is necessary to note the reluctance to generate uncontrolled local and global artifacts, as in the case of neural network approaches, which, based on a purely generative end-to-end pipeline, are capable of generating unpredictable and inconsistent changes in the shapes of objects, which in turn can lead to the failure or incorrect operation of the general discriminatory system.

B. The Core Regressor Structure

Let us move on to the work done to create the core of the system. The core of the system is based on neural network models with multiple inputs and outputs based on transformers. In this case, the semantic spaces are formed naturally according to the autoencoder training scheme (Fig. 3). Where the input to the first transformer of the chain is a pre-processed reduced image, vectorized and scaled gaze direction, a cast of key points, and face detection information. This information is merged into a single vector space using cross-attention blocks.

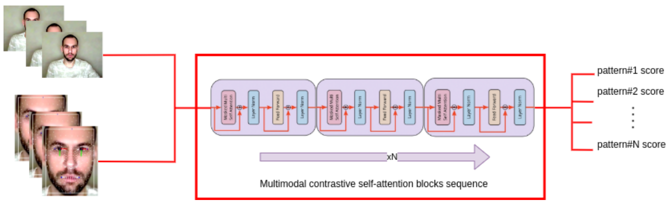

Fig. 3. Structure of the central regressor model.

It should be noted that it is the type of models that constructively belong to this family that are currently state-of-the-art approaches to various discriminatory tasks on video sequences.

C. The Modality Fusion Mechanism

The fusion of multimodal information from different channels occurs through the use of special non-local blocks [32], which is de facto one of the most common approaches to the fusion of modalities in transformer architectures. The scheme of the solution is demonstrated in Fig. 4. Information to the input of these blocks is fed from the outputs of pre-trained encoders, vector representations are pre-scaled. Such blocks are used only at the inputs of the first transformer. As stated above, gaze direction estimation was handled using the MobileGaze [27] architecture, while facial expression activity and keypoint analysis were performed using the [28] and GFE [29] architectures, respectively. Additionally, a separate vectorization block was implemented to process the outputs of face detection modules [30].

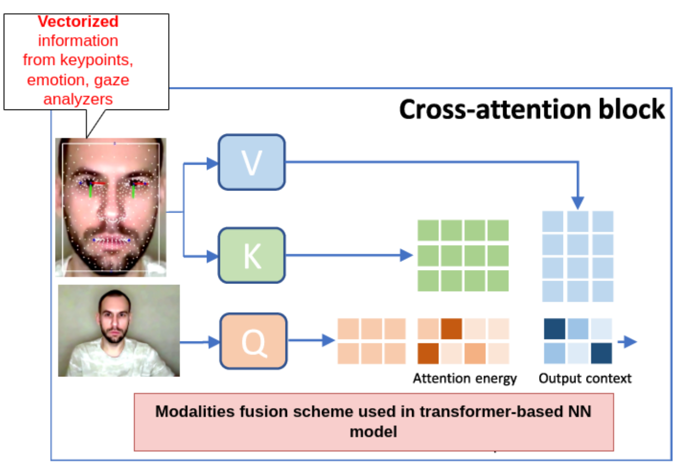

Fig. 4. Modality fusion mechanism.

V. Experimental Setup

A. Dataset Description

In this study, we utilized a custom, manually annotated dataset comprising video recordings of ITMO University students participating in various online assessments. All recordings were used with informed consent from the participants. The dataset included diverse video segments exhibiting patterns of undesirable behavior. Approximately 20,000 sessions contained instances of such behavior, and an additional 20,000 sessions were free of violations.

Based on the annotation results, we applied temporal segmentation to extract video fragments of 5 and 10 seconds in length. The duration of these fragments was established empirically. These fragments contained both acceptable and non-compliant behavioral patterns. Notable categories of undesirable behavior included: prolonged gaze aversion, facial expressions and postures indicative of consulting external sources, temporary absence from the designated examination area, and the presence of unauthorized individuals or objects in the frame. For each pattern, we collected approximately 3,000 samples.

To increase the dataset's diversity and robustness, we applied artificial image degradation techniques, such as intensity reduction, blurring, and contrast suppression. As a result, each behavior class was augmented to a total of 7,000 fragments. These datasets were subsequently split into training, testing, and validation subsets in a 5:1:1 ratio.

B. Training Details

Our experiments' training procedures, initialization strategies, and optimization parameters are described below.

The model training process included the preliminary initialization of neural network parameters using the He and Xavier initialization strategies [31]. Optimization was performed using the proposed algorithm with empirically selected initial learning parameters. The loss function was defined as a linear combination of categorical crossentropy (for classification tasks) and an MSE-based component (for regression tasks related to the estimation of student conscientiousness).

To enhance the generalization capability of the models, regularization techniques such as L_1 and L_2 weight regularization were applied. The models were trained for approximately 50 epochs depending on the configuration, with early stopping based on the F_1-score (for classification quality determination) evaluated on the validation set.

C. Experimental Results

We trained our system both for binary classification tasks (i.e., detecting the presence of any undesirable behavior pattern) and for multi-class classification, distinguishing between specific types of such patterns. The evaluation was conducted using standard metrics, including precision, recall, and F_1-score:

$$Precision = \frac{TP}{TP + FP},$$
$$Recall = \frac{TP}{TP + FN},$$
$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

TABLE I
TOP QUALITY METRICS OF OUR CLASSIFIER WITH VARIOUS PREPROCESSING CONFIGURATIONS USED ABOVE 3-LAYER TRANSFORMER-BASED REGRESSOR

Preprocessing configuration	classification metrics		
	precision	recall	F1
contrast + exposure	0.813	0.831	0.822
contrast + saturation	0.834	0.820	0.827
saturation + exposure	0.845	0.832	0.839
blur + saturation + exposure	0.861	0.848	0.854
blur + saturation + exposure	0.872	0.860	0.866
blur + saturation + exposure	0.883	0.870	0.877
blur + contrast + saturation	0.895	0.880	0.888
sharp + contrast + saturation	0.902	0.889	0.895
sharp + contrast + exposure	0.913	0.898	0.905
sharp + saturation + exposure	0.926	0.912	0.919
contrast + saturation + exposure	0.937	0.925	0.931

TABLE II
INVESTIGATION OF THE INFLUENCE OF THE NUMBER OF TRANSFORMER LAYERS ON CLASSIFICATION QUALITY METRICS FOR THE MULTI-CLASS BEHAVIOR CLASSIFICATION TASK

Transformer layers number	classification metrics		
	precision	recall	F1
1-layer stack	0.915	0.901	0.908
2-layer stack	0.922	0.910	0.916
3-layer stack	0.925	0.918	0.921
4-layer stack	0.930	0.925	0.927
5-layer stack	0.933	0.929	0.931
6-layer stack	0.937	0.925	0.931
7-layer stack	0.935	0.922	0.928
8-layer stack	0.928	0.925	0.926
9-layer stack	0.940	0.925	0.932
10-layer stack	0.935	0.929	0.932

For the classification setting, the system achieved an average precision, recall, and F1-score. Detailed results of assessing the quality of the classification of 5 patterns (looking away, leaving frame, presence of strangers, incorrect angle, consultation with strangers) of undesirable behavior using different system configurations (different modifications of preprocessing and configuration of layers quantityin our transformer-based regressor) for each task and behavior category are presented in Table 1 and Table 2, respectively.

We also propose a comparative analysis of our model's multi-class behavior classification task performance against several state-of-the-art solutions. Table 3 presents the metrics side-by-side to facilitate a direct comparison. It shows the assessment of the effectiveness of our approach in the context of existing methods.

Empirically, from the experiments conducted, the optimal configuration of the preprocessing module and the stack depth of the transformer-based regressor (6 transformer layers and contrast-saturation-exposure filters based pre-processing module) was determined

VI. CONCLUSION

This study has proposed an AI-driven automatic proctoring system that leverages deep learning and computer vision technologies to monitor student behavior during online examinations.

TABLE III
OMPARISON OF OUR BEST CONFIGURATION WITH SIMILAR SOLUTIONS

Model	Classification metrics		
	precision	recall	F1
VTN [35]	0.65	0.60	0.621
DCGN [34]	0.68	0.62	0.645
HERMES [33]	0.72	0.67	0.694
Our solution	0.937	0.925	0.931

Our research led to the development of a microservice-based architecture designed for seamless integration with online education platforms. The modularity of this architecture ensures adaptability and scalability, allowing institutions to incorporate the system into their existing infrastructures with minimal disruption. The proposed pipeline, consisting of multimodal analysis techniques including video, audio, and metadata processing, demonstrated high accuracy in detecting dishonest behaviors while maintaining a balance between security and user privacy.

The evaluation of our model pipeline on a dedicated subsample confirmed that its performance metrics are sufficient for real-world deployment in automated proctoring scenarios. The incorporation of explainable AI techniques enhances transparency and trust in the system's decisions, providing educators with interpretable insights into detected anomalies. By automating the proctoring process, our solution reduces the burden on human supervisors while increasing the credibility of online assessments.

Future work will focus on refining detection algorithms to minimize false positives further, improving the adaptability of the system to different educational environments, and incorporating the analysis of additional metadata to enhance the transparency of the examination process.

ACKNOWLEDGMENT

The research was carried out with the financial support of the ITMO University Research Projects in AI Initiative (project No. 640113).

REFERENCES

[1] M. Cote, F. Jean, A. B. Albu, and D. Capson, "Video summarization for remote invigilation of online exams," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, Mar. 2016, pp. 1–9. IEEE.

[2] A. A. Turani, J. H. Alkhateeb, and A. A. Alsewari, "Students online exam proctoring: A case study using 360-degree security cameras," in 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Al-Salt, Jordan, Dec. 2020, pp. 1–5. IEEE.

[3] I. Ahmad, F. AlQurashi, E. Abozinadah, and R. Mehmood, "A novel deep learning-based online proctoring system using face recognition, eye blinking, and object detection techniques," International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 1–8, 2021.

[4] V. Khanna, S. Brodiya, and D. Chaudhary, "Artificial intelligence-based automated exam proctoring system," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 12, pp. 558–560, 2021.

[5] R. Mehmood, F. Alam, N. N. Albogami, I. Katib, A. Albeshri, and S. M. Altowaijri, "UTiLearn: A Personalised Ubiquitous Teaching and Learning System for Smart Societies," IEEE Access, vol. 5, pp. 2615–2635, 2017.

[6] G. Moukhliss, H. Belhadaoui, and R. Hilali, "A new model of automatic and continuous online exam monitoring," in International Conference on Systems of Collaboration: Big Data, Internet of Things & Security (SysCoBIoTS), Casablanca, Morocco, Dec. 12–13, 2019, doi: 10.1109/SysCoBIoTS48768.2019.9028027.

[7] D. L. King and C. J. Case, "E-cheating: Incidence and trends among college students," Issues in Information Systems, vol. 15, no. 1, pp. 20–27, 2014.

[8] D. Varble, "Reducing cheating opportunities in online tests," Atlantic Marketing Journal, vol. 3, no. 3, p. 9, 2014.

[9] Y. Atoum, L. Chen, A. X. Liu, S. D. Hsu, and X. Liu, "Automated online exam proctoring," IEEE Transactions on Multimedia, vol. 19, no. 7, pp. 1609–1624, 2017.

[10] V. Kasinathan, C. E. Yan, A. Mustapha, V. A. Hameed, T. H. Ching, and V. Thiruchelvam, "ProctorEx: An automated online exam proctoring system," Mathematical Statistician and Engineering Applications, vol. 71, no. 3s2, pp. 876–889, 2022.

[11] I. Ahmad, F. AlQurashi, E. Abozinadah, and R. Mehmood, "A novel deep learning-based online proctoring system using face recognition, eye blinking, and object detection techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 10, pp. 1–8, 2021.

[12] X. Li, K.-m. Chang, Y. Yuan, and A. Hauptmann, "Massive open online proctor: Protecting the credibility of MOOCs certificates," in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, Canada, Mar. 2015, pp. 1129–1137.

[13] A. Wahid, Y. Sengoku, and M. Mambo, "Toward constructing a secure online examination system," in Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, Bali, Indonesia, Jan. 2015, pp. 1–8.

[14] W. A. Rosen and M. E. Carr, "An autonomous articulating desktop robot for proctoring remote online examinations," in 2013 IEEE Frontiers in Education Conference (FIE), Oklahoma City, OK, USA, Oct. 2013, pp. 1935–1939.

[15] G. O'Reilly and J. Creagh, "A categorization of online proctoring," in Global Learn, Association for the Advancement of Computing in Education (AACE), Limerick, Ireland, Apr. 2016, pp. 542–552.

[16] A. S. Milone, A. M. Cortese, R. L. Balestrieri, and A. L. Pittenger, "The impact of proctored online exams on the educational experience," Currents in Pharmacy Teaching and Learning, vol. 9, no. 1, pp. 108–114, 2017.

[17] S. Prathish, K. Bijlani, et al., "An intelligent system for online exam monitoring," in 2016 International Conference on Information Science (ICIS), Chengdu, China, Dec. 2016, pp. 138–143. IEEE.

[18] L. Slusky, "Cybersecurity of online proctoring systems," Journal of International Technology and Information Management, vol. 29, pp. 56–83, 2020.

[19] S. Draaijer, A. Jefferies, and G. Somers, "Online proctoring for remote examination: A state of play in higher education in the EU," in International Conference on Technology Enhanced Assessment, Springer, Amsterdam, Netherlands, Oct. 2017, pp. 96–108.

[20] Testinvite, "Safe Browser for Online Exams," 2023. [Online]. Available: https://www.testinvite.com/help/safe-browser-for-online-exams.html. [Accessed: Oct. 10, 2023].

[21] PSI, "Online Proctoring Privacy Concerns," 2023. [Online]. Available: https://www.psionline.com/privacy-concerns. [Accessed: Oct. 10, 2023].

[22] Safe Exam Browser, "Official Documentation," 2023. [Online]. Available: https://www.safebrowser.org/docs. [Accessed: Oct. 10, 2023].

[23] Talview, "Safe Exam Browser Limitations," 2023. [Online]. Available: https://www.talview.com/glossary/safe-exam-browser. [Accessed: Oct. 10, 2023].

[24] MoodleDocs, "Safe Exam Browser Features," 2023. [Online]. Available: https://docs.moodle.org/en/Safe_Exam_Browser. [Accessed: Oct. 10, 2023].

[25] O. Tatanov and A. Samarin, "LFIEM: Lightweight Filter-based Image Enhancement Model," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 873-878, doi: 10.1109/ICPR48806.2021.9413138.

[26] A. Samarin, A. Nazarenko, A. Toropov, E. Kotenko, A. Dzestelova, E. Mikhailova, V. Malykh, A. Savelev, and A. Motyko, "Universal Filter-Based Lightweight Image Enhancement Model with Unpaired Learning Mode," 2024 36th Conference of Open Innovations Association (FRUCT), Lappeenranta, Finland, 2024, pp. 711-720, doi: 10.23919/FRUCT64283.2024.10749962.

[27] Y. Valikhujaev, "MobileGaze: Pre-trained mobile nets for Gaze-Estimation," Zenodo, 2024, [Online]. Available: 10.5281/zenodo.14257640, doi: https://doi.org/10.5281/zenodo.14257640.

[28] X. Ren, A. Lattas, . Gecer, J. Deng, C. Ma, X. Yang, "Facial Geometric Detail Recovery via Implicit Representation," 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition, 2023.

[29] Y. Zhang, X. Zheng, C. Liang, J. Hu, and W. Deng, "Generalizable facial expression recognition," in Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XIV. Berlin, Heidelberg: Springer-Verlag, 2024, p. 231–248.[Online]. Available: https://doi.org/10.1007/978-3-031-72630-9 14

[30] J. Deng, J. Guo, Z. Yuxiang, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage dense face localisation in the wild," in arxiv, 2019.

[31] S. Kumar, "On weight initialization in deep neural networks," 04 2017.

[32] R. G. Praveen and J. Alam, "Recursive joint cross-modal attention for multimodal fusion in dimensional emotion recognition," in 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024, pp. 4803–4813.

[33] G. J. Faure, J.-F. Yeh, M.-H. Chen, H.-T. Su, S.-H. Lai, and W. H. Hsu, "Hermes: temporal-coherent long-form understanding with episodes and semantics," 2024. [Online]. Available: https://arxiv.org/abs/2408.17443

[34] F. Mao, X. Wu, H. Xue, and R. Zhang, "Hierarchical video frame sequence representation with deep convolutional graph network," in Computer Vision – ECCV 2018 Workshops, L. Leal-Taix´e and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 262–270.

[35] S. Ahmed, M. Khan, H. Qaisar, and G. Sukthankar, "Malicious or benign? towards effective content moderation for children's videos," The International FLAIRS Conference Proceedings, vol. 36, 05 2023.