Steps for security and privacy protection in NLP-based marking systems

ABSTRACT

This paper provides an overview of the methods and techniques used to ensure the security and privacy protection of Natural Language Processing (NLP) based test scoring systems. NLPs improve the accuracy and efficiency of correction systems. However, these systems process sensitive data such as student responses, which raises security and privacy concerns. We examine the components of such a system and then propose measures such as access controls, homomorphic encryption, firewalls and blockchain mixed together to secure the system. Next, we safeguard privacy through methods such as differential privacy protection, anonymization and pseudonymization of data. In addition, we insist on the integration of a browser monitoring module to detect any cheating during composition. In this article we partly present a system called "GestStudent New Generation" in which we integrate most of the security concepts to secure the whole system and guarantee privacy protection. Finally, we conclude by stressing the importance of continuous evaluation of these security and privacy measures to ensure the trust and reliability of NLP-based examination marking systems.

Keywords: NLP, automatic exam marking, security, data protection, privacy, cryptography, differential privacy, blockchain, GestStudent New Generation.

1. INTRODUCTION

Automated exam marking systems based on NLP technology have developed considerably in recent years.
From 2012, some researchers began by questioning the integrity and confidentiality of NLP-based systems specialy the repression of dissidents [1], compromising privacy/anonymity [2], or profiling [3].
Indeed, these systems raise questions of security and student privacy, although they offer advantages such as speed, accuracy, and consistency. Even if our paper is not the first attempt to comprehensively draw a path to a secure NLP-based system, we ensure reliability, integrity, authentication, confidentiality and availability focusing on key measures that can be implemented.
We offer a comprehensive approach that includes the use of various security measures. The first security issue to be resolved is integrity [4]. Many papers written by some researchers such as Ukwen, David Okore, and Murat Karabatak [5,6] help us to gather enough information on the state of cybersecurity and also a roadmap for the future. Data integrity can be compromised from the training, inference, and final deployment phase to transmission over the network.To deal with this, we use the encryption technique in our article, which hides the real information behind a string generated by the

encryption algorithm. This information can be stored securely and transmitted without risk of compromise in the event of intentional theft. Secondly, data confidentiality must be ensured by setting up an access control system to prevent intruders from gaining access to information in the system. In addition, we explore the importance of privacy-protecting techniques in improving the confidentiality of exam data, such as anonymization and pseudonymization [7]. As this type of system is often exposed to brute force attack in order to make the services unavailable, we suggest to set up decentralized architecture [8] and dispatch the services in multiple nodes. Finally, we give in the form of strategies the factors to focus on such as cheating detection to ensure the reliability of the results provided by the automatic correction system.

2. RELATED WORKS

One of the main challenges of exam marking systems is ensuring the confidentiality of student data. This involves securing the system and preserving personal data. The use of cryptographic techniques is a way to ensure the security and privacy of NLP-based exam scoring systems. This is explained by the set of mathematical tools and methods provided by them. A framework that exposes the most popular secure multiparty computation primitives via common abstractions has been proposed by Knott et al. [9]. This practice reduces the risk of data breaches. In 2018, Zebua and Taronisokhi use the spritz algorithm which is one of the cryptographic algorithms to encode the computerized test text database record to make it harder for attackers to know the original text of the test. review [10].

Using homomorphic encryption is another strategy that allows calculations to be performed directly on encrypted data without having to decrypt it first. In the first review on privacy preservation in NLP-based systems, Mahendran and al. found that encrypting data at rest and in transit will be helpful to prevent unauthorized access [11]. Indeed NLP-based marking systems process sensitive data which could be leaked to unauthorized individuals intentionally or unintentionally. Machine learning applications such as test scoring systems can therefore benefit from the privacy and security offered by homomorphic encryption.
For example, Shiho et al. proposed privacy-preserving financial data analysis systems capable of detecting fraud via homomorphic encryption [12].

Furthermore, blockchain technology has shown promise in ensuring the security and privacy of NLP-based exam correction systems. Blockchain provides a decentralized, unhackable ledger that can be used to store and verify exam results. Tsai et al. proposed a blockchain-based approach that uses smart contracts to automate the grading process and ensure fairness and transparency of the review process [13].

Using differential privacy protection is another essential step in preventing data breaches. This technique ensures that no information about a specific student can be gathered from a query result by adding noise to it. Even if this rigorous technique seems risky because it can exacerbate existing biases in the data and have disparate impacts on data by introducing substantial fairness issues even when slightly imbalanced datasets are used [14], many systems keep using it. In 2022, Abahussein et al. used a different amount of noise on two separate parameters to demonstrate the effectiveness of this method [15]. Similarly, RUAN et al. proposed a differentially private stochastic gradient descent protocol and two optimization methods to avoid attacks that solely depend on model access [16]. They thus protect sensitive data throughout model training and inference.

Tomoaki et al. proposed a k-anonymization algorithm to anonymize the data set by analyzing the correlation between attributes and generates an optimal hierarchy based on this correlation. This method thus guarantees the protection of privacy [7].

The above-mentioned research efforts highlight the importance of ensuring security and privacy in NLP-based exam marking systems.

3. SECURITY COMPONENTS AND METHODOLOGIES OF NLP-BASED EXAM MARKING SYSTEMS

3.1. The components of a NLP-based exam marking system

These systems generally consist of three main components, namely (as illustrated in Figure 1):

- Trained model: this is a NLP trained on data triples (reference response, learner response and the similarity score between these responses)
- User Interface: Allows teachers and students to interact with the grading system.
- Data storage system: to store exam results in a secure space.

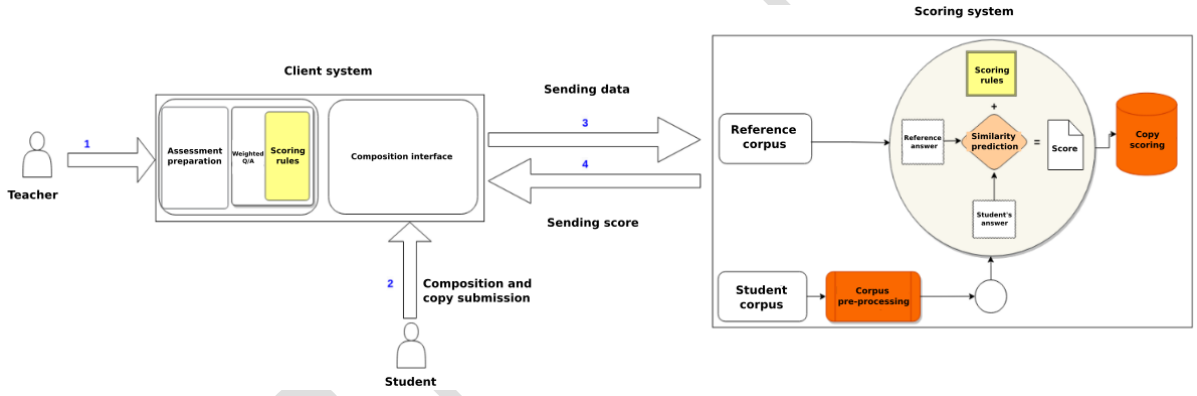

Figure 1: Architecture of the GestStudent New Generation software coupled with a copy grading system

In Figure 1, the architecture of the GestStudent New Generation software is composed of a client system through which the teacher and students can interact with the grading system. The similarity prediction module works with a pre-trained model on student assessment data. Once the similarity prediction between the student's corpus and the reference corpus pre-provided by the teacher has been made, the student's note is stored as well as the corpus is saved for future use.

Each of these components poses potential risks to student security and privacy, particularly regarding the protection of exam data and unauthorized access to results.

3.2. Methods and techniques for securing an NLP-based exam correction system

Different security methods can be implemented in NLP-based exam correction systems, such as the use of access controls, cryptography, blockchain, etc.

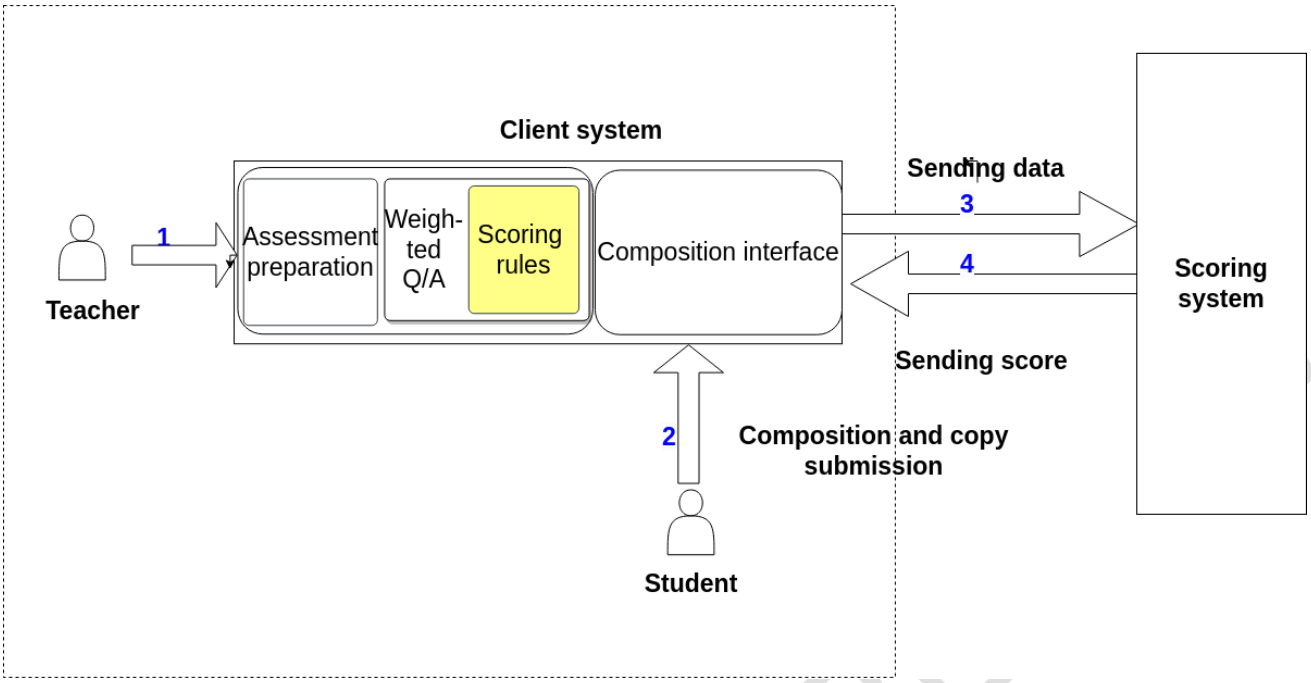

Figure 2: Client system sub part of the Architecture of the GestStudent New Generation software coupled with a copy grading system

- Access Control: This technique involves controlling access to sensitive data, such as student information and exam answers, to prevent unauthorized access. You can see in figure 2 that teachers and students both can access the system. So we need a way to ensure that a student is not going to view exam information before the composition date. Access control can be used in NLP-based exam marking systems to restrict access to sensitive data to authorized personnel only. A common access control measure is the use of role-based access control (RBAC) [17], which assigns permissions based on the user's role within the entity. This model consists of the following components:
    a. Roles: Roles are used to seperate users with similar permissions together in a group. For instance, the “admin” can be used to group users who are allowed to make administrative operations in the system. Similarly, we can have the role “student” for grouping all students and allowing them specific permissions.
    b. Permissions: Permissions are used to specify the operations that users are allowed to perform on NLP services. For example, the permission "edit" would allow a user or a group of users to edit an object in the system.
    c. objects: Objects are the resources on which we are going to set permissions. And the user will have the right to perform actions on objects on which he has permission depending on his role.

The RBAC formula is represented by:
$P(r, o) = \{p1, p2, \ldots, pn\}$

where P is the authorization function, r is the role, o is the object (i.e. the resource being accessed), and p1 through pn are the permissions assigned to this role for this object.

- Data encryption: Exam and student data stored in the system database can be encrypted to protect against unauthorized access. Encryption can be achieved using standard encryption algorithms, such as AES and RSA [11,18]. This technique can also be used in NLP-based exam marking systems to protect student data privacy. Even further, we can proceed to homomorphic encryption. This technique allows calculations to be performed on encrypted data (exam answers) without revealing the underlying data [19]. The effectiveness of encryption can be evaluated using the following formula:

  $$E = \frac{Nt}{Nc} \times 100\%$$

  where E is the encryption strength, Nt is the number of unauthorized access attempts prevented by the encryption, and Nc is the total number of access attempts.

- Network Security: Firewalls and other network security tools can be configured to protect against network attacks. Users' personal information may be protected during transmission over the Internet by encrypting data in transit using the TLS protocol.
- Intrusion detection and prevention: This technique involves detecting and preventing unauthorized access to sensitive data. NLP-based exam marking systems can use intrusion detection and prevention techniques to detect and prevent unauthorized access to student data. Audit logs can be used to track user activities and detect suspicious behavior.
- Secure Storage: This technique involves storing sensitive data, such as student information and exam answers, in secure locations to prevent unauthorized access. Cloud-based storage services like Amazon S3 can be very useful in addressing a number of vulnerabilities.
- Blockchain: Implementing a blockchain-based system to ensure data transparency, immutability, and accountability. It can also help prevent fraud and ensure the integrity of exam results [13, 20].
- Decentralization: Decentralization allows data and processing power to be distributed among multiple nodes or machines to avoid a single point of failure.

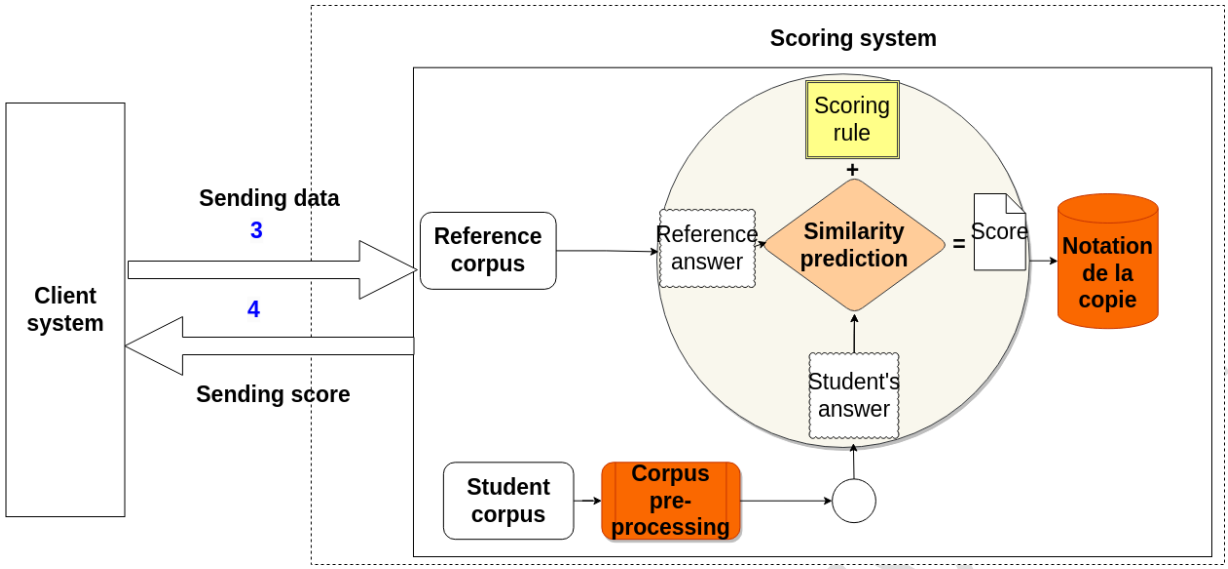

Figure 3: Scoring system sub part of the Architecture of the GestStudent New Generation software coupled with a copy grading system

As shown in figure 3, the system should do many operations for each student composing for the exam. The system can fail as a result of brute force attack which consists of sending a large amount of requests to the server to make it down.

One strategy would be to use blockchain technology. This could provide a safe and transparent way to store test results and other sensitive data. Additionally, by using smart contracts and other blockchain-based tools, it is possible to automate certain aspects of the exam marking process, further enhancing security and privacy [21].

4. Preserving Privacy in Exam Marking Systems

Regarding privacy, exam correction systems may contain sensitive information about students, so it is important to ensure that their privacy is protected. Students have the right to privacy of their personal data, such as their names, addresses and exam results.

To guarantee the preservation of privacy, several techniques can be used.

- Pseudonymization of student data: This technique consists of replacing personally identifiable information with pseudonyms (unique identifiers) in order to protect the confidentiality of student data.
- Differential Privacy Protection: This technique involves adding noise to the data to prevent identification of individual student data. It therefore helps to protect the privacy of student data by adding random noise to exam responses [22].
- Data Anonymization: To maintain student privacy, exam data may be anonymized, removing all personally identifiable information. This ensures that each record in a dataset can be distinguished from at least k-1 other records [23]

5. Assessing Security and Privacy Preservation in NLP-Based Exam Marking Systems

One should constantly evaluate one's system to ensure that it remains non-vulnerable to possible attacks and meets the security requirements for such systems.

- Security Audit: An in-depth system security assessment can be performed by an independent security auditor. The audit may include an analysis of the system configuration, an assessment of security vulnerabilities, an analysis of identity and access management and an assessment of compliance with security standards.
- Penetration test: Penetration tests are simulations of computer attacks that aim to assess the security of a system. They therefore make it possible to identify vulnerabilities and weaknesses in the system and can be used to improve system security.
- Architectural Assessment: Assessing the architecture of the exam marking system can help identify potential system vulnerabilities and security issues. This assessment may include an analysis of data processing processes, authentication and authorization mechanisms, and communication protocols.
- Incident Management Assessment: Incident management describes the processes used to detect, report, and manage security incidents or data breaches. An incident management assessment examines whether processes are adequate to minimize risks and respond quickly in the event of an incident.
- Privacy Shield Analysis: Privacy Shield Analysis can help identify potential privacy risks to student data. This analysis may include an assessment of data protection policies, consent management and users' rights to their personal data.

6. Strategies for Implementing Security Measures and Preserving Privacy in NLP-Based Exam Marking Systems

Table 1: Choice of security measures and privacy preservation in NLP-based exam correction systems

Measures	System security	Protection of private life	security level	Cost optimization	Maintenance frequency
Access control	Role-based access control with strong authentication	-	High	Free software	Regular
Data encryption	Data encryption using the AES algorithm	Encryption of personal data	High	Open source libraries	Regular
Decentralization	Decentralized architecture using blockchain	Use of smart contracts	Very high	pre-built or public infrastructure	Occasional
Regular tests and audits	Regular penetration testing and vulnerability assessments	Regular updates and fixes	High	Free software	Regular

Table 1 provides an overview of the different security and privacy measures that can be taken in a NLP-based exam marking system, as well as some cost optimization tips, the level of security afforded and the frequency of recommended maintenance. By analyzing these measures in a structured way, although a decentralized architecture can offer high security, it can be dispensed with if it requires a large investment.
A decentralized architecture can be used to avoid a single point of failure and increase overall system security. But it does require some investment in the servers needed to form the cluster. To implement it, it is necessary to determine the number of nodes with the formula:

$$N = n \times (1 - p)^c$$

where N is the number of nodes required to keep the system operational, n is the initial number of nodes, p is the probability of failure of a node, and c is the number of nodes required for consensus.

Even better, in the GestStudent New Generation software, we have added a browser monitoring module to ensure the integrity and authenticity of the exam results. Although such a measure presents risks in terms of ensuring the protection of students' privacy, we were able to develop concepts around the subject symbolizing actions that could be considered suspicious. It is :

- collapsing the composition window;
- opening new tab different from composition tab in browser;
- opening new window, different from the composition window;
- click on the taskbar on Windows or on the launcher on Linux Ubuntu;
- clicks it in an area of the screen outside the composition window;
- the composition screenshot.

All of these actions are grouped under the concept of “Digital Cheating”.
When the system detects one of its events, the student is automatically disconnected from the composition interface due to cheating and their grade at the end of the exam will be purely and simply 00/20.

Note that Table 1 also highlights the importance of regularly testing and auditing the system in order to identify and correct possible vulnerabilities or weaknesses.

7. CONCLUSION

In artificial intelligence-based exam correction systems, it is essential to ensure security and prevent privacy breaches. This requires careful thought and the implementation of appropriate measures. In this review, we have identified several key measures that can be taken to improve the security and privacy of NLP-based exam marking systems, including the use of encryption, access control, anonymization data and browser monitoring.
By implementing these measures, NLP-based exam marking systems can be made safer and more reliable, while improving the overall quality of education.
However, it is important to continue to explore new approaches such as emotion detection which could be used to identify and prevent cheating attempts. Emotion detection technologies can also make it possible in other places to analyze the behavior of learners in order to know what orientation to give to the course to make it more explicit.

REFERENCES

[1] Zhang, Boliang, et al. "Be appropriate and funny: Automatic entity morph encoding." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014.

[2] Coavoux, Maximin, Shashi Narayan, and Shay B. Cohen. "Privacy-preserving neural representations of text." arXiv preprint arXiv:1808.09408 (2018).

[3] Wang, Jingjing, et al. "Cross-media user profiling with joint textual and social user embedding." Proceedings of the 27th International Conference on Computational Linguistics. 2018.

[4] Hu, Yupeng, et al. "Artificial intelligence security: Threats and countermeasures." ACM Computing Surveys (CSUR) 55.1 (2021): 1-36.

[5] Ukwen, David Okore, and Murat Karabatak. "Review of NLP-based systems in digital forensics and cybersecurity." 2021 9th International symposium on digital forensics and security (ISDFS). IEEE, 2021.

[6] Thejaswini, S., and C. Indupriya. "Big Data Security Issues and Natural Language Processing." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2019.

[7] Tomoaki Mimoto, Anirban Basu and Shinsaku Kiyomoto, "Towards practical k-anonymization: Correlation-based construction of generalization hierarchy", Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016) - Volume 4: SECRYPT , p. 411-418, 2016.

[8] Bharimalla, Pranab Kumar, et al. "A blockchain and NLP based electronic health record system: Indian subcontinent context." Informatica 45.4 (2021).

[9] Knott Brian, Venkataraman Shobha, Hannun Awni, Sengupta Shubho, Ibrahim Mark, and Maaten Laurens van der. 2021. Crypten: Secure multi-party computation meets machine learning. Adv. Neur. Inf. Process. System 34 (2021)

[10] ZEBUA, Taronisokhi. Encoding the record database of computer based test exam based on spritz algorithm. Lontar Komput. J.Ilm. Teknol. Inf, 2018, vol. 9, no. 1, p. 52.

[11] Mahendran, Darshini, Changqing Luo, and Bridget T. Mcinnes. "Privacy-preservation in the context of Natural Language Processing." IEEE Access 9 (2021): 147600-147612.

[12] Shiho Moriai. 2019. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. In ARITH 2019. IEEE, 198

[13] Tsai, CT; Wu, JL; Lin, YT; Yeh, MKC Design and development of a Blockchain-based secure scoring mechanism for online learning. Educ. Technol. Soc. 2022, 25, 105–121

[14] Farrand, Tom, et al. "Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy." Proceedings of the 2020 workshop on privacy-preserving machine learning in practice. 2020.

[15] Abahussein, S., Cheng, Z., Zhu, T., Ye, D., Zhou, W. (2022). Privacy-Preserving in Double Deep-Q-Network with Differential Privacy in Continuous Spaces. In: Long, G., Yu, X., Wang, S. (eds) AI 2021: Advances in Artificial Intelligence. AI 2022. Lecture Notes in Computer Science(), vol 13151. Springer, Cham. https://doi.org/10.1007/978-3-030-97546-3_2

[16] RUAN, Wenqiang, XU, Mingxin, FANG, Wenjing, et al. Private, Efficient, and Accurate: Protecting Models Trained by Multi-party Learning with Differential Privacy. arXiv preprint arXiv:2208.08662, 2022.

[17] Zaidi, Tanzeel, et al. "Fabrication of Flexible Role-Based Access Control based on Blockchain for Internet of Things Use Cases." IEEE Access (2023).

[18] DENIS, R. and MADHUBALA, P. Hybrid data encryption model integrating multi-objective adaptive genetic algorithm for secure medical data communication over cloud-based healthcare systems. Multimedia Tools and Applications, 2021, vol. 80, p. 21165-21202.

[19] KANG, Ha Eun David, KIM, Duhyeong, KIM, Sangwoon, et al. Homomorphic Encryption as a secure PHM outsourcing solution for small and medium manufacturing enterprise. Journal of Manufacturing Systems, 2021, vol. 61, p. 856-865.

[20] KAREEM, Abdulkareem Saber and SHAKIR, Ahmed Chalak. Verification Process of Academic Certificates Using Blockchain Technology. Studies, 2023, vol. 18, No. 1, p. 62-75.

[21] KOBAYASHI, Reiji and KANAI, Atsushi. Blockchain-based Self-scoring Method. In: 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE). IEEE, 2021. p. 1-5.

[22] HA, Trung, DANG, Tran Khanh, DANG, Tran Tri, et al. Differential privacy in deep learning: an overview. In: 2019 International Conference on Advanced Computing and Applications (ACOMP). IEEE, 2019. p. 97-102.

[23] MAJUMDAR, Rwitajit, AKÇAPINAR, Arzu, AKÇAPINAR, Gökhan, et al. LAView: Learning analytics dashboard towards evidence-based education. In: Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge. Society for Learning Analytics Research (SoLAR), 2019.