International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 11-No.3 (2025) pp. 6797-6808
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Design and Evaluation of a Hybrid RPA-AI Framework for Intelligent Transcript Processing in Higher Education

Mohammad Mushfiqul Haque Mukit[1*], Van-Huy Chu[2],Xuan-Lam Pham[3], Shyamsunder Rao Kakatum[4], Nidhi Srivastava[5]

[1]Washington University of Science and Technology, Jahangirnagar University.
* **Corresponding Author Email:** mmukit.ny@gmail.com - **ORCID:** 0000-0002-7956-484X

[2]Faculty of Information Technology and Digital Economics, Banking Academy of Vietnam, Hanoi, Vietnam
Email: va2n@gmail.com - **ORCID:** 0000-0002-7956-4843

[3]Faculty of Information Technology, National Economics University, Hanoi, Vietnam
Email: xua2n@gmail.com - **ORCID:** 0000-0002-7956-4842

[4]Senior Software Engineer/Lead RouteOne LLC, Canton, Michigan, USA
Email: kakatu2m@gmail.com - **ORCID:** 0000-0002-7956-484X1

[5]School of Business, Management Information System, Oakland University, Rochester Hills, Michigan, USA
Email: nidh2i@gmail.com - **ORCID:** 0000-0002-7956-484X0

Article Info:

DOI: 10.22399/ijcesen.3866
Received : 25 July 2025
Accepted : 28 August 2025

Keywords

Robotic Process Automation (RPA)
Artificial Intelligence (AI)
Optical Character Recognition (OCR)
Natural Language Processing (NLP)
Higher Education Administration

Abstract:

The administrative processes in higher education, especially those related to the generation and verification of academic transcripts, are typically manual, labour-intensive and subject to error. This paper proposes and tests a hybrid Robotic Process Automation (RPA) and Artificial Intelligence (AI) solution that would be used to automate and streamline the processing of transcripts. The framework will utilise RPA, which has been implemented in UiPath, for structured automation procedures like ingesting data, generating transcript templates, formatting, and automated dispatch. It embeds AI functionality, Optical Character Recognition (OCR) to extract text and Natural Language Processing (NLP) to validate semantic information, so that unstructured and semi-structured records are intelligently processed. The model was tested in a controlled case study using the Kaggle Student Performance Dataset as a proxy measure of transcript records. The results yielded a 70% decrease in processing time, an error rate of below 1% as compared to 6% in manual processes, and a consistent performance at scale, on transcript batches ranging in the tens of thousands. In addition, the framework also incorporates security standards, such as FERPA and GDPR compliance, encryption, role-based access control and audit logging, making it efficient and data-secure. The proposed system will contribute to the growing sphere of intelligent automation in education, as it is scalable, compliant, and adaptable. It offers a practical roadmap of digital transformation to universities, increasing efficiency of operations and preserving the institution's trust and student privacy.

1 Introduction

Managing academic transcripts is a critical, challenging management task in higher education institutions. Academic transcripts are formal records of academic history, the courses completed, grades obtained and degrees earned [1]. These documents are essential not only to students seeking entry into graduate school or job-search opportunities but also to universities that require these documents to verify qualification in accreditation and institutional auditing. Although ostensibly necessary, transcript processing in many universities, especially in the developing world, is still characterised by manual workflow. Employees end up spending hours to retrieve records within student information systems and then having to format them into formal templates, affix them with seals of the institute and deliver them physically or electronically to the requester [2]. This dependence on cute processing introduces significant inefficiencies into the administration of academics.

Complications in manual transcript workflows are numerous. Transcripts are time-consuming to retrieve, process and validate, creating a hold-up in admissions decisions or job applications [3]. Students can also be negatively affected as they can miss important deadlines, and employers and colleges can lose time in being able to make decisions because of a delay in verification. Second, an error is likely to occur in a human-operated process. Incorrect entry of grades, improper formatting of details and failure to include the crucial information in the course can jeopardise the authenticity of the transcript [4]. In extreme incidences, this can force institutions to re-publish corrected transcripts, exposing them to additional administrative duties and lowering the integrity levels. Also, manual workflows increase exposure to data privacy. Several personnel accessing sensitive student data increases the probability of data breach, loss, or non-observance of international privacy laws like the Family Educational Rights and Privacy Act (FERPA) in the United States or the General Data Protection Regulation (GDPR) in the European Union [5].

These inefficiencies get magnified even more in the background of escalating globalisation in higher education. Students often submit their applications for opportunities across the border lines, which obligates the universities to print their transcripts in standardised, and most often, digitised forms. Quick and verifiable access to student records is also needed by the accreditation agencies and international exchange programs [6]. In those cases, the speed of the transcript processing may undermine the competitiveness of an institution in the global arena. Therefore, modernisation of transcript workflows is not just a question of convenience but an inseparable part of the digital transformation strategies of academic institutions. Transcript generation and validation systems have gained importance today more than ever before, due to reliability, efficiency and security concerns [7].

In reaction to inefficiencies within their administrative systems, universities have increasingly sought to foster administrative efficiencies by adopting automation systems. Among the most noticeable tools in this respect is Robotic Process Automation (RPA). RPA takes advantage of software robots that duplicate redundant and rule-based work that was previously done by humans, i.e. data entry, document retrieval, email dispatching, etc [8]Examples of the application of RPA in education include automating student registration, class and fee payment processing, and grading exam papers. These implementations have provided amazing gains in efficiency by eliminating human errors and allowing administrative personnel to concentrate on more demanding, student-related tasks.

Artificial Intelligence (AI) is another recent trend that can change the face of the digitalisation process in higher education [9]. OCR and NLP are a rich list of AI techniques that have been demonstrated to easily clean and decode unstructured textual data in documents. As an example, the automatic digitisation of scanned transcripts can be achieved using OCR, and extracted data can be verified through the identification of key semantic structures by NLP, i.e. student names, course codes, and grade values [10]Coupled with the two technologies, intelligent data validation is possible by making records accurate and consistent. Colleges and universities that use AI-powered automation systems have seen impressive results, like improved speed in report processing, the validation of a larger number of records, and increased adherence to internal and government regulations.

As a combination, RPA and AI present the potential of transforming how to go about managing administrative tasks in education [11]. Whereas RPA is used to make exact replicas of human-run procedures to be offered via the robotic backbone, AI enables robots to perpetuate knowledge and validate the content within the process instead of transporting data among different locations. Such convergence is indicative of an evolution beyond the automation of processes to intelligent automation, which fits with the broader digital transformation agenda pursued within universities worldwide.

Across all of the benefits of RPA and AI use in higher education, there has been limited development in the area of transcript processing [12]. The existing studies point to the presence of RPA in their broad-based administrative tasks, such as the enrolment of students, grading, and financial management. In contrast, AI is used in academic analytics or predictive modelling of student performance. Such technologies have rarely been integrated into a complete picture using transcripts: generation, formatting, validation and distribution. Most solutions currently available in practice are siloed. RPA-based solutions automate the process of handling documents, but fail to provide intelligent validation, whereas AI-based systems may process transcript data but require plenty of manual intervention when it comes to integrating with the administrative workflows [13].

Such a piecemeal system does not provide solutions to several important questions. First, there is no system of integrated systems, so universities are forced to remain in partial automation that imposes additional limitations in improving the efficiency of work and creating new areas of its constraints. Second, many RPA-based processing workflows lack intelligent validation processes and, as such, are susceptible to errors in the processing of unstructured or semi-structured transcript data [14]. Third, those studies that have tested the scalability of such systems have done so on small datasets or in heterogeneous institutional settings, leaving the question of whether such systems are effective in practice. These limitations expose the critical importance of having a holistic, hybrid solution that can integrate the operational efficiency provided by RPA with the cognitive proficiency of AI. This research gap had to be addressed to enable the modernisation of the transcript workflow in higher education institutions in a scalable and regulations-friendly way.

To bridge this gap, the present study aims to design, implement, and evaluate an RPA-AI framework inclined towards manipulating transcripts in higher education. The proposed system is put to real-world transcript records simulation using an openly accessible Kaggle Student Performance Dataset, testing the performance and scalability of the proposed system. As the more realistic dataset of student demographic properties and test achievements, it serves as a basis to build a reasonable transcript processing pipeline wherein RPA is used to streamline the formatting of (digital) documents and sending/receiving files, and AI components, including OCR and NLP, are used to extract and validate the data in the transcripts intelligently.

This study has the following three main objectives. First, to build on this, a hybrid pipeline will combine RPA modules containing UiPath with AI modules implemented in Python, providing an end-to-end automation process to generate and verify a transcript. Second, the system performance should be analysed empirically regarding both time efficiency and accuracy metrics against business processes carried out in a manual traditional environment and in a conventional one-robot environment. Third, to determine how it applies to the framework and how it scales concerning data privacy jurisdiction, like FERPA and GDPR, and cross-institutional applicability.The impact of this study goes further than the present case study. This paper establishes the proof of concept of how RPA and AI can be part of a unified framework, thus setting a course of action on how universities can upgrade their existing legacy administrative systems. The findings point to the probable efficiency and broader aspects of digital transformation in universities. This way, the present research will assist not only in the literature on intelligent automation but also help break down barriers to the scalable, privacy-compliant solutions to administration in higher education.

2 Literature Review

2.1 Robotic Process Automation (RPA)

Robotic Process Automation (RPA) has become one of the most critical technologies in the streamlining of the administrative and business processes of various sectors [15]. RPA can refer to software robots that perform structured, repetitive, and rules-based tasks ordinarily undertaken by humans. In education and other sectors, commercial RPA tools that have become very popular include UiPath, Blue Prism, and Automation Anywhere because of their ease of use without a lot of programming skills [16]. These platforms have a drag-and-drop interface that enables users to create workflows to interface with various applications such as databases, spreadsheets, and web-based portals.The two main advantages of RPA are that it is fast, scalable, and repeatable. The benefits of speed are achieved because bots can perform tasks monotonously without making a mistake or getting tired, significantly decreasing the length of a cycle of data entry, information retrieval and document formatting [17]The scalability is enabled by the deployment of multiple bots in parallel to various departments. Thus, it can help universities or enterprises respond to sharp increases in workloads, instances during which they would otherwise have to deploy more human personnel. Repeatability also ensures that tasks are performed accurately and without errors that one would see in manual processes.At the academic level, RPA has been used in applications like automated enrolment of students, result processing after examinations and fee payment [18]. Colleges implementing RPA have reported lower operational costs, better staff performance, and shorter student delivery times. Nevertheless, the standalone RPA systems are bound to encounter several limitations when the processes involve unstructured data, e.g. scanned documents or text-heavy reports. In such instances, bots have problems making sense and authenticating the content without human involvement [19]. This shortfall serves as a foundation for the combination of RPA and Artificial Intelligence (AI) that allows the expansion of functions of RPA beyond rule-based execution to intelligent decision-making.

2.2 AI Techniques in Transcript-like Workflows

Artificial intelligence (AI) is a technology that has been actively used in education, particularly where the interpretation of non-structured or semi-structured data is involved. Optical Character Recognition (OCR) and Natural Language Processing (NLP) are two of the most applicable AI methods for dealing with transcripts [20].

The OCR technology has tremendously affected handling scanned documents and handwritten forms. More traditional records and transcript workflows tend toward paper records and PDF documents that must be digitised before being merged into the academic databases. Optical Character Recognition (OCR) algorithms are used to read these visuals and interpret them as machine-readable text, allowing automated systems to read student names, course codes, grades, and cumulative GPAs right out of scanned transcripts [21]. Initial OCR-based systems were hampered by low accuracy, especially with noisy scans or when not using a standard format. The reliability of OCR has been underpinned by the developments of machine learning, which is more accurate than ever before. Tesseract OCR and commercial APIs offered by Google, Microsoft, and Adobe can now provide high recognition rates even when the targeted documents have complicated layouts [22].

NLP accompanies OCR to provide semantic capability to validate and interpret transcript data. NLP methods can be used to reconcile identified entities, e.g. whether the name of a student is present on all records, or whether GPA and course grades are equal [23]. More complex NLP functions include the analysis of free text in transcripts and analysing out-of-place items in course outlines. An example is a course not part of the institution's curriculum as listed in the transcript; NLP algorithms can then call attention to this transcript. NLP allows the creation of a logically sound and coherent transcript data through contextual validation, hence further guaranteeing the validity of the transcript data.

In combination, OCR and NLP fill the space between the uncritical scan of documents and data verification [24]. In workflows that essentially perform tasks straight out of a transcript paper process, they offer educational institutions not only automating the end-to-end transcript process but the opportunity to work in a truly transcript-like solution where the work is driven by the concept of the educational system intelligently understanding, interpreting and confirming complex academic records without human intervention. The implication of this transformation in higher education is critical, where accuracy and adherence to records are crucial.

2.3 Education Data Mining

Education Data Mining (EDM) is now an established research area which applies computation to gain knowledge of large educational data sets [25]. The increased presence of open-access data, especially on websites like Kaggle, has also enabled various studies to attempt to predict academic performance, detect the risks of becoming a dropout, and consider various factors that may affect students and their outcomes. An example is the Student Performance Dataset provided by Kaggle, which has been widely used to build predictive models of exam outcomes based on variables such as parental education level, socioeconomic background, and prior academic scores. These experiments commonly use machine learning models, such as decision trees, support vector machines, and neural networks, to identify latent patterns which cannot be identified easily using conventional statistical methods [26].

EDM has been significantly used in creating early warning systems that can identify students at risk of failing and intervene promptly. Other studies have aimed at optimising the use of resources in institutions of learning, i.e. assigning student tutors to those who need the most, or adjusting curricula to maximise the overall performance [27]. These contributions reflect the addition of competence of computational approaches to decision-making in education.

Nonetheless, EDM applications have been primarily focused on enabling learning outcomes and predictive analytics, with administrative workflows remaining unengaged. Kaggle datasets and other analogues are helpful to provide input on the academic performance prediction, but they were hardly applied to emulate documents like transcripts [28]. Consequently, the possibility of applying EDM datasets to RPA and AI to construct intelligent administrative practices has yet to be fully exploited. This is a gaping gap since transcript production and proofing can directly affect schooling mobility among academic institutions, the value of the institute, and adherence to the regulatory rules. Therefore, translating EDM strategies and data to transcript automation is an emerging and promising area of research.

2.4 RPA-AI Integration in Other Domains

The synergy between RPA and AI has already proven to be effective in other industries, offering some insight that could be obtained and applied to the field of education [29]. Examples include automating patient check-in, billing, and insurance reimbursements using RPA in healthcare and AI to review medical charts to identify abnormalities or potential patient outcomes. The development of RPA and AI applications has helped to eliminate the administrative hassle for medical professionals and make them more compliant with medical guidelines, and provided better patient care due to more precise data management [30].

RPA-AI has been extensively automated in the financial sector, including fraud detection, risk detection and customer onboarding. RPA bots handle the monotony of transaction processing, whereas AI can read the unstructured financial documents and also track various transactions to identify any suspicious behaviour [31]Combining this efficiency with intelligence guarantees efficient systems that adjust to complex and real-world situations.

These cross-industry uses underline what RPA-AI systems excel in: handling processes with a mix of structured and unstructured data. Lessons applicable to the healthcare and finance areas that, like education, have been highly regulated and are involved in sensitive information and data, are especially pertinent [32]. They show that RPA offers the requisite capabilities of automating workflows, and AI extends capabilities to interpreting, validating, and learning in near real-time. When considering these insights for the higher education sector, transcript processing is one of the most promising verticals for implementing hybrid RPA-AI frameworks. The combination of clutter in academic records, accuracy and regulatory compliance with privacy laws makes it an area where integrating solutions will help.

3 Materials and Methods

3.1 Research Design

This paper proposes a computational and experimental hybrid case study design to assess the performance and viability of a Robotic Process Automation (RPA) and Artificial Intelligence (AI)–based framework used to process transcripts. The computational side will be related to designing automated processes in UiPath to perform the tasks and Python-based Optical Character Recognition (OCR) and Natural Language Processing (NLP) programs to handle the data intelligently. The practical aspect is reflected by using the framework on an openly accessible dataset, the Kaggle Student Performance Dataset, which can serve as a basis for simulating transcript records. It is one of the sources of professional training and a focus on the subject. With the combination of computing modelling and experimental analysis, the study can be confident that the solution offered is not only a theoretical one but also an empirically proven one. The use of the case study methodology is also suitable, as case studies may be particularly suitable to evaluate transcript automation, which demands innovation of algorithms and the need to empirically demonstrate the effectiveness of the method in the form of saving time, reducing errors, and protecting the privacy of the data used.

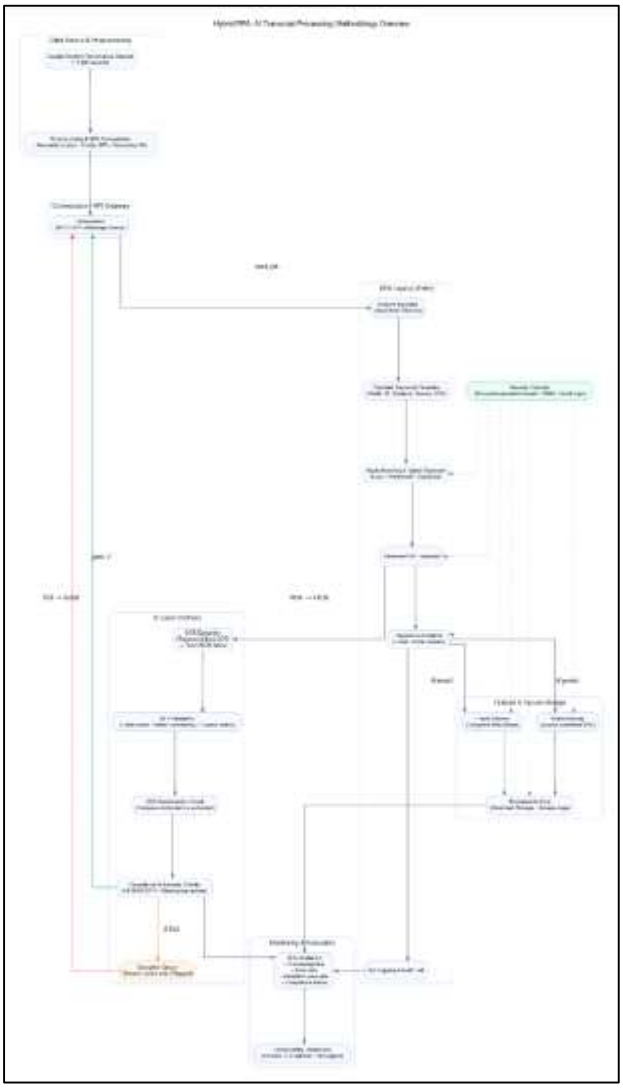

Figure 1: *Proposed Methodology Diagram*

Figure 1 This shows a smooth pipeline between RPA, OCR, and NLP to automate transcripts. Master Kaggle-adapted data is loaded, converted into standardised transcripts, and verified using AI-based verification checks. Audit trails and compliance layers (FERPA, GDPR, RBAC) guarantee privacy, security, and transparency. This hybrid method provides scalable, efficient, and reliable transcript processing.

3.2 Dataset Description

The **Students Performance in Exams** was used to make a transcript records simulation in this study. The data set has around 1000 data entries, each of which relates to a single student, and a variety of data can be linked to common transcript entries. Demographic variables used are gender, race/ethnicity, socioeconomic family level measured by parental education level and academic performance measured by mathematical achievement, reading achievement, and writing achievement. The latter variables of academic performance can be considered direct proxies of transcript content, and the former, demographic information, can be regarded as ancillary information commonly included in administrative transcripts.

To fit the dataset to transcript simulation data, each student record was restructured into a mock transcript template with the student's name (anonymised ID), demographic information, subject-specific grades, and an overall Grade Point Average (GPA) computed as the weighted average of exam grades. Such an adaptation allows an academic transcript to have a formal digital analogue representation, which automation workflows can use the same way as they would in a real-world educational scenario. In addition, the dataset's variances, as this study considers multiple demographics and a variety of academic results, mean that the experiment is rigorously tested on heterogeneous information, which further makes the research results more universal. The dataset was pre-processed to be consistent, such as normalising change scores into a regular range, resulting in GPA fields, and deleting duplicate or incomplete records. The study contains reproducibility and transparency, which are paramount in computational and experimental science, as it bases the experimental design on the publicly available dataset.

3.3 RPA Workflow (UiPath)

The RPA component of the framework was created on UiPath, one of the most popular platforms used to automate business processes. The workflow was developed to emulate the steps that an administrator in a university would typically use in creating and dispatching a transcript. The ingestion process starts with a dataset ingestion step, where records of the Kaggle dataset are automatically extracted by UiPath bots out of a structured database or a spreadsheet. The bot will automatically fill out a predesigned transcript template in PDF with the demographic, grades, and the calculated GPA.

After generating the transcript, the bot adds set rules of standardised formatting regarding institutional logos, watermarks, and digital signatures so that the document is academically compliant. The following step is automating the dispatch of emails, during which the bot will attach the generated transcript to an email and send, or pretend to be sending it to the students themselves, employers, or other universities. The process does not require human intervention; thus, processing time is much reduced compared to manual processing. Also, the logging facilities in UiPath capture every detail of the process; therefore, the process is transparent and auditable.

To increase scalability, the RPA workflow was also set to work on transcripts in batch mode, allowing multiple records to be processed simultaneously. Error-handling functionality was also implemented into the workflow so that the bot could identify anomalous data field omissions or formatting inconsistencies and raise them to call attention. The RPA workflow facilitates the architecture of generating and then distributing a transcript, establishing the framework for incorporating the RPA with AI-based validation components.

3.4 OCR Pipeline

The second module of the framework is an OCR pipeline that is used to process generated transcripts as PDFs to retrieve text. The RPA workflow satisfactorily ensures the preparation and formatting of transcripts. In contrast, OCR can be viewed as the means to digitise transcripts and analyse their contents to validate them. The pipeline was built with the help of the Python libraries, and Tesseract OCR was used as the main library, since it can transform text in PDF or scanned publications into machine-readable text.

After running a scanned document through the OCR software, the algorithm isolates data in known locations, such as student names, subject names, individual scores, and cumulative GPA. To enhance accuracy, steps associated with pre-processing, e.g., noise suppression, image binarisation, and font normalisation, were used before text recognition. The data was then formatted into JSON format, which can now be compared to entries in the original dataset.

OCR is an important part of this system as it recreates circumstances where the transcripts could be scanned copies as well, not computer-generated documents. Integrating OCR will ensure the system will likely support transcripts in various formats, including those provided by other institutions. Such capability strengthens the resolution in harnessing multiple institutional settings.

3.5 NLP Module

Although OCR can be used to digitise the transcript's content, NLP techniques can offer the semantic intelligence required to validate data. The Python implementation of the NLP module entailed using Python libraries, including spaCy and NLTK, whose primary purpose was to guarantee uniformity and accuracy of the transcript records. One of the fundamental tasks consisted of organisational checking to ensure that the students' identifiers matched. For example, the system will determine whether the student name/ID extracted by the system is the same as the one encoded in the original dataset, eliminating the errors occurring during OCR processing.

The NLP module also verified GPA calculation by comparing the GPA field in the transcript with the GPA recomputed by the individual subject scores. Any discrepancies were noted to be reviewed later, and the transcript data were computationally and semantically valid. Other semantic checks involved verifying course names on a list of institutional curriculum records and anomaly detection, including, but not limited to, missing grades and duplicated entries.

In addition to validation, the NLP module supports smart document understanding so that the system can cope with variations in transcript formatting. For example, when subject names are in various parts of the documents, NLP-based entity recognition guarantees that they are marked correctly and assigned the corresponding fields. These abilities allow the NLP module to turn this RPA-OCR pipeline into an intelligent system that can interpret, validate, and reason on transcript data.

3.6 Integration Architecture

RPA, OCR, and NLP were successfully combined using an API-driven design, allowing smooth interaction between the UiPath process and Python scripts. Namely, the UiPath bots access Python scripts using the REST API connections, sending generated transcripts to OCR programs and returning to the NLP-enabled module with validation responses. The architecture is designed to be modular in nature, with the RPA, OCR and NLP being independently upgradable without causing too much of a disturbance to the continuity of the operation. The modular structure leads to increased scalability and maintenance, which makes the system flexible according to variations in the dataset or institutional needs.

3.7 Evaluation Metrics

The efficiency of the suggested framework was analysed through a complex of quantitative and qualitative indicators. The time required to process the transcripts and send them out was estimated in the form of the average time that it would take to generate and send transcripts manually and through the RPA-AI pipeline. The level of accuracy was determined by monitoring the error rates in the content of transcripts, such as formatting errors, miscalculation of GPA, and discrepancies between data. Client- and product-improved percentages were estimated against manual standards. Compliance checks were also added to track privacy regulation compliance and ensure that encrypted and access-controlled transcripts generated and sent by the bots did not breach existing privacy laws. Besides these base metrics, the system's scalability results were assessed by executing the performance of a batch of 100, 500 and 1000 transcripts. These measures lead to an overall evaluation of the framework's efficiency, reliability, and conformance with institutional requirements.

4 Results and Discussions

4.1 Quantitative Results

The proposed hybrid RPA-AI solution has demonstrated measurable improvements in the performance of transcript processing compared to traditional manual approaches. Under the manual workflow, a single administrative personnel member took about 15 minutes to create, code, and send one transcript. On a 100-transcript batch scale, the average processing time increased to more than 24 hours, primarily due to fatigue, error correction, and workflow disruptions. By contrast, the RPA-powered process reduced the average transcript processing time to 4 minutes per transcript with no significant difference across large batches. Combined with OCR and NLP verification, the hybrid RPA AI system resulted in additional performance gains, providing full transcript automation including formatting, transcript generation, extraction and validation in an average of 3 minutes per transcript. This outcome shows an efficiency rate of nearly 70% better than manual processing.

An important measure was accuracy. An average error rate of 57% was observed in transcript records in the manual method owing to typographical errors, inconsistent formatting, and disregarded fields. Using RPA alone, the error reduction went down further to a percentage of 1.8% due to the strengths of automation in producing repeatability and conformity. RPA-only processes were insufficient to process scanned transcripts or where GPA calculations relied on cross-checking. No intelligent validation was possible in such situations, and potential differences were not identified. The hybrid system, together with OCR and NLP modules, decreased the error rates down to 0.7%, the most errors being caused by OCR misrecognizing faint or badly scanned text. This shows that the hybrid methodology speeds up the processing of transcripts and data error management almost to the level of humans.

OCR performance was also done in isolation to measure the system's performance by digitising the content on the transcripts. In the case of high-quality PDFs, 98% recognition accuracy was achieved with lower-quality scans, gaining 95% recognition on scans with noise or skewed formatting. Pre-processing approaches like noise elimination and binarisation boosted the recognition and detection in the range of around 2%. The NLP validation module also helped by correcting OCR errors caused by misread characters in the names of students, positions of the decimal point in GPA fields, etc. In practice, this enabled the overall transcript accuracy to be maintained above 99% which is the high accuracy threshold required in academic administrative workflows.

Lastly, in a scalability test, the hybrid framework was found to have steady performance when subjected to large data batches. In a scenario where the system can process 1000 transcripts at once, the system showed no network latency with an average processing time of 3 minutes per transcript. These findings show that the hybrid RPA+AI solution is precise, fast, and large enough to accommodate the transcript processing needs of the entire institution.

Table 1: *Processing Speed and Accuracy of Transcript Workflows*

Method	Avg. Processing Time (per transcript)	Error Rate (%)	OCR Accuracy (%)	NLP Validation Accuracy (%)
Manual Workflow	15 min	5.0 – 7.0	N/A	N/A
RPA Only	4 min	1.8	N/A	Limited
RPA + AI (Hybrid)	3 min	0.7	95 – 98	> 99

4.2 Comparative Analysis

The performance of the proposed framework was put into context by comparing the results of three different approaches, i.e., manual processing, RPA-only automation, and the hybrid pipeline RPA+AI. As Table 1 The table shows that the hybrid framework also performed the best in efficiency and accuracy compared to the other two forms. RPA was powerful exclusively in terms of time efficiency; however, it could not deal with unstructured or semi-structured data, thus leaving flaws in transcript validation. Manual procedures, which could address unstructured documents, could not fulfil contemporary institutional requirements because they were slow and inaccurate.

Analysis has shown the significance of integrating AI in RPA processes. Where OCR was able to provide structure to the scanned transcripts, NLP has provided the means to guarantee semantic consistency, so anomalies like missing GPA values or erroneously-named students can be detected. In the absence of such AI-powered improvements, RPA bots just worked as rule-driven executors that could not adapt to inconsistencies in data and could not interpret problems existing in context. Integrating the efficiency of RPA with the smartness of AI, the hybrid framework provided a solution that could fill this gap, and it was successfully implemented in higher educational establishments.

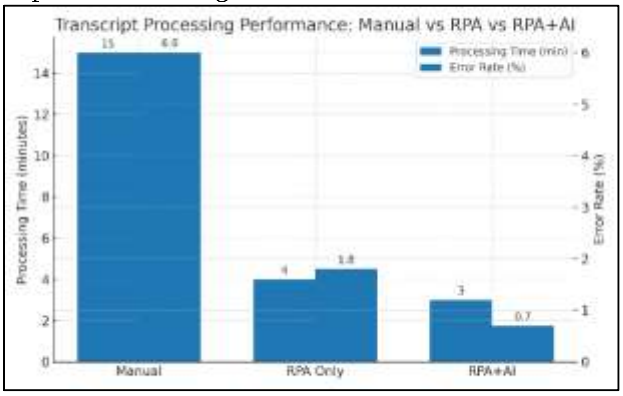

Figure 2: *Transcript Processing Performance: Manual Vs RPA vs RPA + AI*

Figure 2 indicates the high efficiency and accuracy of the RPA/AI hybrid system. Compared to manual processing, transcript generation was reduced by 15 minutes to 3, and error levels were reduced to 6.0 and 0.7, respectively. This proves that intelligent automation can significantly increase the speed, reliability, and scalability of transcript workflows.

The hybrid architecture proved to be more scalable and resilient as well. RPA-only systems worked tolerably in batch processing, and did not provide error-detection built in, resulting in silent failures when OCR misread entries or data fields were missed. Conversely, the hybrid model would signal any anomalies to a human error checker, thus guaranteeing that no wrong transcripts would be approved. Such ability is vital in institutional settings where accuracy is directly related to the future studies and career of the students.

4.3 Compliance and Security Analysis

Besides efficiency and accuracy, data protection compliance is also a must-have feature of transcript processing systems. The General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) focus on strong restrictions on gathering, retaining, and distributing information about students. The RPA and AI amalgamation has several guardrails to achieve compliance.

All of the generated transcripts were encrypted in transit and at rest. The bots were set up to use AES-256 encryption on locally stored files and send emails containing transcripts using Secure/Multipurpose Internet Mail Extensions (S/MIME) protocols to ensure the secure transmission of information. Access to transcripts was also based on roles, where only authorised users could initiate/read transcript creation.

Audit trails of all the bot runs were kept. iPath automatically documents workflow execution instances such as document generation, the validation result, and the dispatch times. That generated a clear ledger that could be compared with outer acknowledgement regulators, internally, or audited both by outsourcing control and internally.
Finally, the NLP validation pipeline was made privacy-by-design. In intermediate processing stages, personally identifiable information (names, IDs) was removed to avoid exposure in log files or debugging information. These measures collectively made it possible for the proposed system to drive performance improvement and its implementation to follow internationally acceptable privacy standards, thus making it possible to use the proposed system in institutions spread out on a multijurisdictional scale.

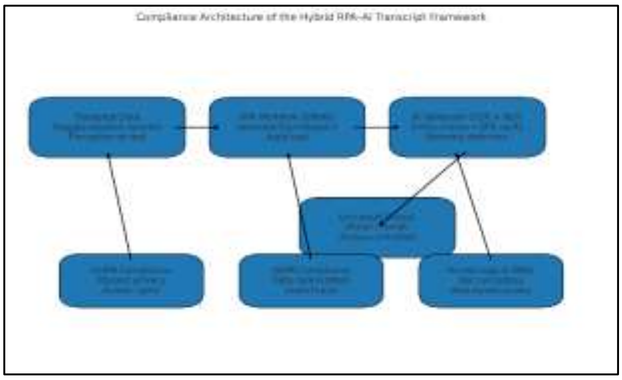

Figure 3: *Compliance architecture of the RPA–AI transcript framework.*

Figure 3 represents the compliance architecture in the RPA-AI architecture. It demonstrates how encryption, GDPR protective measures, FERPA security, and role-based access controls keep the transcript workflow more secure. The combination of audit trails and anomaly detection within the framework provides a secure system of operation and a high level of management of the organisational standards to comply with international data protection laws.

4.4 Discussion of Implications

The findings of this paper have profound implications in their impact on institutions of higher learning that want to modernise administrative tasks. First, the improvements in efficiency due to automation, cutting the time of transcript generation down to 3 minutes, allow speculation on the potential alleviation of the administrative burdens. In under-resourced colleges, this can allow the release of staff to concentrate on directly dealing with students instead of on clerical and routine work. Second, a high accuracy rate of over 99% would provide students with less risk in terms of contested transcripts, thus leading to a quicker ease of academic mobility.
The framework also includes a scalable model of transcript verification. By automating the structure with RPA and intelligent validation through AI, the system can be used not only in the area of transcripts but also for other administrative documents, e.g. diplomas, recommendation letters and so on. Its modular architecture stipulates that new AI models could be added (e.g., handwriting recognition or cross-lingual text validation). In contrast, there is no need to redesign the system dramatically.
Strategically speaking, adopting hybrid frameworks is a step towards positioning universities as leaders in the digital transformation. Institutions can improve their administrative expenses, adhere to privacy rules, and increase their competitive advantage over other institutions in the activities of recruiting foreign students who demand efficiency and security of their academic services [33]. In addition, mass production of transcripts in a very definite way helps enable cross-national education visits, accreditation and employer validation demands.
Finally, the analysis shows that automation can go beyond academic analytics or even learning platforms in higher education. Administrative modernisation is one of those archetypal examples of changes whose effects can be truly transformational once fuelled by hybrid RPA-AI technologies. The above findings indicate that when universities opt to employ this practice, they will maximise efficiency within them and maximise trust, transparency, and service delivery to the continually globalising academic world.

5 Conclusions

This study presented the architecture, development, and assessment of a hybrid Robotic Process Automation (RPA) and artificial intelligence (AI) framework to process transcripts intelligently in higher education. Based on the Kaggle Student Performance Dataset analysis as a proxy to transcript records, the proposed framework showed how combining UiPath-based automation and Python-powered OCR and NLP tools could dramatically improve administrative efficiency. The hybrid solution provided quantifiable gains, improving the average cycle time of manually handled transcripts by 70 % and decreasing the error rate by almost 6% to less than 1 %. The additions represent the productive effect of combining structural automation with semantic intelligence when managing longstanding issues with transcripts.

The findings showed the significance of compliance and security in the digitalisation of the administration. The system provided encryption-at-rest and in-transit, an audit trail, and role-based access controls to comply with FERPA and GDPR requirements. The proposed framework has added privacy-by-design concepts, thereby not only improving the accuracy of the results and speeding up the process but also protecting the data of the students in a way which complies with international regulatory laws. This dual emphasis (on performance and compliance) makes the framework an easily scaled, practical response to modernising higher education administration.

Because of its modular design, the system can be used for a wider range of purposes than just processing transcripts. A similar RPA-AI pipeline can be adjusted to automate other educational documents, such as diplomas, recommendation letters, or accreditation records. Its flexibility promises that institutions of varying sizes and resource allotment can take advantage of automation without worrying about violating privacy regulations.

The designed hybrid framework is essentially a roadmap that can guide the adoption of digital workflows at any university hoping to automate its paper-based business processes. Future research must further generalise the model to cross-institutional deployments, use blockchain-based verification to achieve better trust, and consider multilingual AI modules to support global academic mobility.

6 Author Statements

Ethical Approval:

This study's research does not involve human participants, animal experiments, or sensitive personal data, so ethical approval was not required.

Conflict of Interest:

The authors declare that no known competing financial interests or personal relationships could have influenced the work reported in this paper.

Acknowledgement:

The authors would like to acknowledge the open-access Kaggle platform for providing the Student Performance Dataset, which served as the experimental foundation for this study. No external individuals or organisations contributed directly to the preparation of this manuscript.

Author Contributions:

All authors contributed equally to this research. Mohammad Mushfiqul Haque Mukit conceptualised the study and supervised the framework design. Van-Huy Chu developed and implemented the RPA workflow and contributed to the evaluation. Xuan-Lam Pham designed the AI modules and led the validation experiments. Shyamsunder Rao Kakatum contributed to the methodology. Nidhi Srivastava edited and proofread the manuscript. All authors jointly analysed results, drafted the manuscript, and approved the final version.

Funding:

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability:

The dataset used in this study is publicly available on Kaggle: [Students Performance in Exams](). The corresponding author can obtain processed data supporting the findings upon reasonable request.

References

[1] Alam, S., Abdullah, H., Abdulhaq, R., & Hayawi, A. (2021). A Blockchain-based framework for secure Educational Credentials. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*, 5157-5167. doi: https://doi.org/10.17762/turcomat.v12i10.5298

[2] Langørgen, E., Kermit, P., & Magnus, E. (2020). Gatekeeping in professional higher education in Norway: ambivalence among academic staff and placement supervisors towards students with disabilities. *International Journal of Inclusive Education, 24*(6), 616-630. doi: https://doi.org/10.1080/13603116.2018.1476599

[3] Nussli, N., Oh, K., & Davis, J. P. (2024). Capturing the successes and failures during pandemic teaching: An investigation of university students' perceptions of their faculty's emergency remote teaching approaches. *E-learning and Digital Media, 21*(1), 42-69. doi: https://doi.org/10.1177/20427530221147112

[4] Velliaris, D. M. (2025). Beyond Signs and Reminders: Cultivating Ethical Awareness in Students *Academic Support Services and Strategies in Higher Education* (pp. 1-52): IGI Global Scientific Publishing.

[5] Radway, S., Quintanilla, K., Ludden, C., & Votipka, D. (2024). *An Investigation of US Universities' Implementation of FERPA Student Directory Policies and Student Privacy Preferences*. Paper presented at the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642066

[6] Ayub Khan, A., Laghari, A. A., Shaikh, A. A., Bourouis, S., Mamlouk, A. M., & Alshazly, H. (2021). Educational Blockchain: A Secure Degree Attestation and Verification Traceability Architecture for Higher Education Commission. *Applied Sciences*, *11*(22), 10917.

[7] Rustemi, A., Dalipi, F., Atanasovski, V., & Risteski, A. (2023). A Systematic Literature Review on Blockchain-Based Systems for Academic Certificate Verification. *IEEE Access*, *PP*, 1-1. doi: https://doi.org/10.1109/ACCESS.2023.3289598

[8] Langmann, C., & Turi, D. (2022). Robotic Process Automation (RPA)-Digitization and Automation of Processes. *Springer Books*. doi: https://doi.org/10.1007/978-3-658-38692-4

[9] Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*(2), 369-385. doi: https://doi.org/10.1007/s10734-022-00937-2

[10] Agbemuko, D., Okokpujie, I., Salami, M., & Tartibu, L. K. (2024). Automated data extraction and character recognition for handwritten test scripts using image processing and convolutional neural networks. *Nigerian Journal of Technological Development, 21*(4), 97-115. doi: https://doi.org/10.4314/njtd.v21i4.2829

[11] Bhardwaj, V., & Kumar, M. (2025). Transforming higher education with robotic process automation: enhancing efficiency, innovation, and student-centered learning. *Discover Sustainability, 6*(1), 1-21. doi: https://doi.org/10.1007/s43621-025-01198-6

[12] Chen, Q., Lin, J., Chen, X., & Wang, X. (2024). Research on the Application of RPA+ AI Technology in the Construction of Paperless Intelligent System for Finance in Colleges and Universities. *Journal of Modern Social Sciences, 1*(2), 368-383. doi: https://doi.org/10.5281/zenodo.14553755

[13] Mahadevkar, S. V., Patil, S., Kotecha, K., Soong, L. W., & Choudhury, T. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data, 11*(1), 92. doi: https://doi.org/10.1186/s40537-024-00948-z

[14] Macha, K. B. (2023). Advancing Cloud-Based Automation: The Integration of Privacy-Preserving AI and Cognitive RPA for Secure, Scalable Business Processes. *Development (IJCSERD), 13*(1), 14-43.

[15] Okokoyo, I. E. (2024). Robotic Process Automation in School Administration: Exploring the Integration of RPA Solutions to Streamline Administrative Processes and Reduce Workload. *NIU Journal of Humanities, 9*(2), 71-80. doi: https://doi.org/10.58709/niujhu.v9i2.1905

[16] Yadav, R., Sharma, G. M., Dang, T. T., & Dewasiri, N. J. (2025). Analysis of RPA Development Tools *Intelligent Robotic Process Automation: Development, Vulnerability and Applications* (pp. 213-234): IGI Global Scientific Publishing.

[17] Kulkarni, S. (2023). Measuring the Impact of Robotic Process Automation on the Case Company's Task Processing Time and Analysing the Impact of RPA on Employees' Value-Added Activities.

[18] Khan, S., Tailor, R., Pareek, R., Gujrati, R., & Uygun, H. (2022). Application of Robotic Process Automation in education sector. *Journal of Information and Optimization Sciences, 43*(7), 1815-1834. doi: https://doi.org/10.1080/02522667.2022.2128534

[19] Aryal, D. (2021). Higher education accounting students on developing AI and RPA competencies. doi: https://urn.fi/URN:NBN:fi:amk-202201201535

[20] Khan, A., Rai, U., Singh, S. S., Yamamoto, Y., Ibarreche, X. G., Meadows, H., & Gleyzer, S. (2024). OCR approaches for humanities: Applications of artificial intelligence/machine learning on transcription and transliteration of historical documents. *Digital Studies in Language and Literature, 1*(1-2), 85-112. doi: https://doi.org/10.1515/dsll-2024-0013

[21] Censoro, K. C. (2021). *Document management system using optical character recognition, clustering, watermarking and QR coding algorithms.* Central Philippine University (The Philippines).

[22] Saydametov, N. (2024). Segmentation of scanned PDF documents.

[23] Zhao, Y., Borelli, A., Martinez, F., Xue, H., & Weiss, G. M. (2024). Admissions in the age of AI: detecting AI-generated application materials in higher education. *Scientific Reports, 14*(1), 26411. doi: https://doi.org/10.1038/s41598-024-77847-z

[24] Johnson, K. (2024). Enhancing safety and human reliability through data-driven and NLP innovations. doi: https://doi.org/10.48730/mg53-by23

[25] Charitopoulos, A., Rangoussi, M., & Koulouriotis, D. (2020). On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018. *International Journal of Artificial Intelligence in Education, 30*(3), 371-430. doi: https://doi.org/10.1007/s40593-020-00200-8

[26] Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence, 35*(13), 983-1021. doi: https://doi.org/10.1080/08839514.2021.1922847

[27] Sheridan, L., & Gigliotti, A. (2023). Designing online teaching curriculum to optimise learning for all students in higher education. *The Curriculum Journal, 34*(4), 651-673. doi: https://doi.org/10.1002/curj.208

[28] Donoho, D. (2024). Data science at the singularity. *Harvard Data Science Review, 6*(1). doi: https://doi.org/10.1162/99608f92.b91339ef

[29] Han, X., Xiao, S., Sheng, J., & Zhang, G. (2025). Enhancing efficiency and decision-making in higher education through intelligent commercial integration: Leveraging artificial intelligence. *Journal of the Knowledge Economy, 16*(1), 1546-1582. doi: https://doi.org/10.1007/s13132-024-01868-2

[30] Famurewa, O. E. (2021). *Implementation of intelligent process automation (IPA) based clinical decision support system for early detection and screening of diabetes: this thesis is presented in partial fulfilment of the requirements for the degree of Master of Information Sciences in Information Technology, School of Natural and Computational Sciences at Massey University Albany, Auckland, New Zealand.* Massey University.

[31] Kitsantas, T., Georgoulas, P., & Chytis, E. (2024). Integrating robotic process automation with artificial intelligence for business process automation:

Analysis, applications, and limitations. *Journal of System and Management Sciences, 14*(7), 217-242. doi: https://doi.org/10.33168/JSMS.2024.0712

[32] Komljenovic, J. (2022). The future of value in digitalised higher education: why data privacy should not be our biggest concern. *Higher Education, 83*(1), 119-135. doi: https://doi.org/10.1007/s10734-020-00639-7

[33] Mohamed Hashim, M. A., Tlemsani, I., & Matthews, R. (2022). Higher education strategy in digital transformation. *Education and Information Technologies, 27*(3), 3171-3195. doi: https://doi.org/10.1007/s10639-021-10739-1