AI/ML-Based Auto Assessment and Evaluation Platform for EdTech Systems

Rithik Raj Vaishya[1], Shreeya Agrawal[2], Diya Asija[3], Rakshak Gupta[4]
[1]*Data Scientist and Corporate Trainer Maharashtra, India*
[2,3,4]*Student Auro University, Gujrat, India*

***Abstract*—With rapidly increasing use of digital education and training space, the need for quick, scalable and objective assessment platforms has grown. Traditional evaluation methodologies are largely reliant on human-intervention, which leads to lagging feedback, inconsistency and operational cost. Due to adoption of digital based education tools and virtual learning ecosystems has highlighted the shortcomings of manual assessment systems. Manual evaluation methods are time consuming, difficult to scale, and often affected by subjectivity and inconsistency. Auto assessment and evaluation platforms have overcome the limitations and shown to be successful using the potential of AI & ML. This paper will provide a comprehensive study in the field of auto assessment and evaluation platform based on AI. The designs motivation, system architecture, operational workflow, underlying technologies, enabling technologies, applications and performance impact will be discussed. Conceptual and analytical research methodology is used through extensive review of existing literature, systems and industry practices and comparative system analysis to understand their effectiveness and limitations. The report will comment upon the benefits, limitations and future scope research directions to be undertaken towards improvement of accuracy, fairness and flexibility of such automatically evaluating systems, illustrating their increasing application in contemporary education and corporate learning environments training ecosystems.**

***Index Terms*—Auto Assessment, Automated Evaluation, Artificial Intelligence in Education, Learning Analytics, Online Assessment Systems, Machine Learning, EdTech**

I. INTRODUCTION

The rapid growth of EdTech platforms has transformed the modes of delivery of learning content. Today, Learning Management Systems (LMS), Massive Open Online Courses (MOOC), and corporate training platforms enable millions of learners from different geographical and academic backgrounds to have flexible and personalized learning journeys and acquire up-to-date skills for an ever-changing job market.

While content delivery mechanisms evolved markedly in EdTech systems, assessment and evaluation practices did not match up at the same pace. Since assessment is a crucial step in any learning process that determines the learner's status and skills, it is important that evaluation is performed efficiently and reliably. In large digital learning environments, conventional manual approaches cannot keep pace with submission volumes or operate at scale.

As a solution, AI/ML, based auto assessment and evaluation platforms analyze learner responses, content metadata and interaction behaviors to infer the assessment outcomes. These platforms enable large-scale evaluation, automated feedback or continuous instruction planning. This paper discusses the design of such platforms in EdTech systems from a learning effectiveness perspective.

1.1 Background

Recent technological developments as well as enhanced-broadband penetration and adoption of online education models have led to an exponential rise in EdTech platform activity in the recent years. Websites offering online degrees, training videos, certificates and online tests and games function on learning management systems and cloud infrastructures to operate at scale.

One of the main characteristics of EdTech platforms is their scale. The number of learners enrolled in a MOOC can easily reach dozens of thousands. Correspondingly, number of online pages accessed or exercises submitted simultaneously is massive. Human evaluators are incapable of conducting immediate feedback and grading, which erodes the value of assessment.

Besides, learners accessing EdTech courses come from different geographic, academic or socio-economic origins and behave differently. Fixed assessment methods are unsuitable to assess their learning outputs fairly and objectively in diverse contexts. As a consequence, EdTech solutions require scalable, adaptive and learner, aware evaluation mechanisms. The advantage of AI/ML systems is in providing an alternative in this area due to the systems' wide perspectives.

1.2 Problem Statement

As EdTech shifted toward digital, the assessment process remained slow, expensive and relatively ineffective. The main issues of conventional assessment methods in EdTech contexts are the following:

Scalability: The human resource for assessment cannot keep up with the massive amount of research data entering the system in a time-effective manner that can support. Reliability: The external evaluators have difficulty to ensuring inter, and intra, rater consistency. Timeliness: Learners do not receive immediate feedback on their questions, which impacts their motivation and erodes the utility of the assessment. Data: Traditional evaluation processes do not take advantage of the available metadata and learner analytics, which would allow for efficient and informative evaluation decisions.

Therefore, an AI/ML, powered auto assessment system that can be deployed under EdTech settings faces multifaceted challenges that require innovative solutions.

1.3 Auto Assessment in EdTech Platforms

Auto assessment in EdTech platforms is defined as the automated assessment of learner results using AI, Machine Learning and data analytics techniques. While conventional academic systems only deliver mostly summative assessment tools, EdTech evaluation is more focused on formative assessment for continuous learning support.

Using EdTech, specific deep learning or NLP models, an auto assessment system can make reliable assessment decisions for a broad range of responses, such as multiple choice, short answer, long essay or programming code. Due to their integration into LMS or other EdTech platforms, these systems analyze more than the final learner responses but also later behaviors, such as number of retries, time intervals and resource references.

The key differentiator between traditional academia assessment systems and EdTech needs is their flexibility: EdTech assessment systems must dynamically change their offered difficulty levels, automatically tailor the assessment feedback to be used and adapt to a broad learner profile spectrum. Auto assessment platforms fulfill these requirements by harnessing learner behavior data and intelligent algorithms

II. ED-TECH PLATFORM DATA AND LEARNING EVIRONMENT

There are lots of Data types available in a EdTech learning environment, by which we can construct an AI based auto assessment system. Degree of insight into EdTech environment and data flow streams is crucial to figure out right assessment architecture for a particular case.

2.1. Different types of Edtech data used

Since knowledge-based assessment agents use multiple types of data streams created in EdTech worlds using AI. Learner profile data allows customization of the assessment to learner ‘s age, prior knowledge and language; course and content meta, data form the basis of the complexity and pedagogical mapping of assessed items and learner progress.

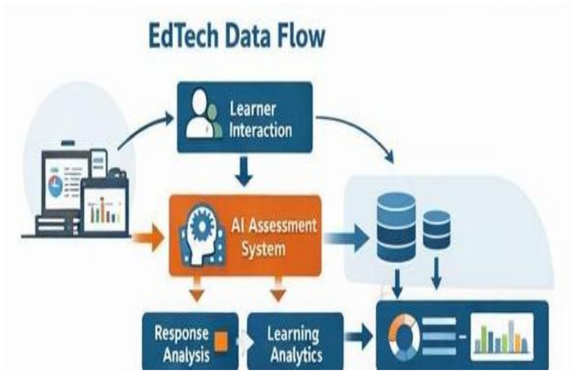
Assessment response data type is the answer submitted by the learner against a question in an assessment task, whereas interaction data correspond to the browsing behavior, level of participation and time spent on, task on, digital platform. In EdTech auto assessment systems, both types of data are used to provide authentic evaluation results.

2.2. Data flow in an EdTech Platform

After receiving the learner request from the front-end interface, the EdTech platform opens the appropriate interaction channel to his front-end and stores the submitted responses and browsing statistics.

Next, a specific assessment module send the learners' answers and interaction metrics to auto assessment engine to be assessed. Then, a Machine Learning engine based on former training data classifies the test

results and returns assessment scores in the output.
Finally, the analytics engine aggregates the results of the assessment with other learning indicators and display the assessment analytics dashboard.



2.3. Data Challenges in EdTech
While there are abundant opportunities to mine knowledge from edtech data, it also hinders the decision, making process due to some difficult issues as follows: Diversity: Edtech data comes from different types of content, learner identities and assessment formats. Noise: Edtech data may have errors, missing values or formatting issues across data, which degrades model effectiveness, Privacy, Heavy regulation in some geographies, e.g. GDPR, prevents sharing or aggregation of the data.
These challenges need to be tackled in order for assessment decisions to be considered valid and reliable.

III. LITERATURE REVIEW

Research of digital assessment has progressed over several years from merely storing and converting data to converting content format, linguistic assessment or ‘intelligent’ analysis.
Initial online assessment models were restricted to converting paper-based tests into electronically administered MCQ questions, with minimal automation and cognitive power. When NLP and deep, learning evolved further, most of the research was on automatic evaluation of textual response and essays. Transformer, based new models, reinforcement learning algorithms were remarkably successful in automating the learning assessment.
Still, many recent works investigate their constraints with respect to understanding, feedback or adaptability. This paper presents the synthesis of the most relevant research to build a functional auto evaluation platform for EdTech context.

3.1 Digitalization & Standardization
Several methods also proposed standards to develop digital assessment as comprehensive assessment system. They provided some guidelines for quick, consistent and correct conversion of Pen and Paper examination to digital.

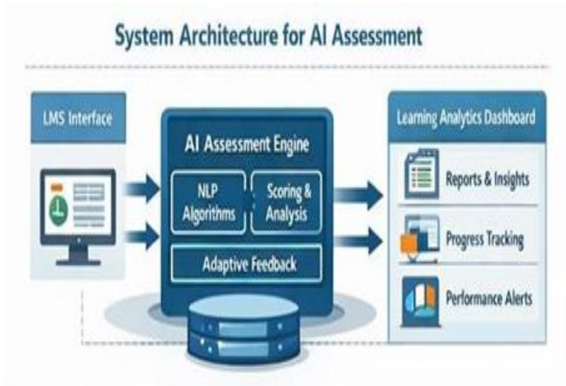
3.2 Content Understanding & Conventional Evaluation
Although the different formats have evolved, the understanding of the content and objectivity has remained unsolved in the traditional evaluation methods. Research in the category sought to classify right/wrong answer and cluster learner responses to suggest objective and reliable grading algorithms.

3.3 Text and Discourse Evaluation
As the language and discourse modeling appeared; many systems tried to evaluate the content by using the language, feature based, and discourse relations but in some systems deep learning algorithms are used to evaluate the language intelligently.

3.4 Auto Assessment Systems
Finally, last developments of the automatic assessment architecture where revealed based on learner, centric, knowledge, driven and adaptive expectations.
In this paper, we use and contribute to the existing research to build an Edtech, auto assessment, based platform.

IV. ARCHITECTURE FOR ED-TECH AUTO ASSESSMENT

An auto assessment module in AI built architecture must be in relation with EdTech system ‘s requirements, infrastructure and expectations.
4.1 General Architecture
The generic EdTech environment, prepared to have the assessment front-end, an LMS (learning management system), an auto assessment agent and a few learner analytics components, how they work together to produce self, evolving evaluation decisions toward the learners.



The EdTech auto assessment agent is a set of modular algorithmic components, composed of the following modules: transfer layer, evaluation engine and the learner profile analytics engine. These modules operate concurrently to the other components of the EdTech environment to provide dynamic evaluation results.

4.2 Layer of integration of EdTech platform
The transfer layer is constituted by the APIs for software development and the real-time data exchange pipeline between the auto assessment and the host LMS, providing system extensibility and scalability in order to meet the EdTech platform high requirements.

V. METHODOLOGY

The methodology for the proposed AI /ML based auto-assessment and evaluation platform in EdTech setting is designed to yield accuracy, scalability, adaptivity in assessments. This methodology leverages structured data precursors, smart model choice, transparent assessment logic to assess heterogeneous responses of learners.
5.1 EdTech Data Preprocessing
Data preprocessing is an important part of a robust auto-assessment system especially in the EdTech setting where robust data are generated. The challenge here is to perform data engineering steps that provide smoothed models while managing large volume of noisy, heterogeneous data. Data Preprocessing pipeline is tailored to deal with EdTech data for building robust learning assessment system.

5.2 Learner Response Data Cleaning
Learner response data can contain noisy, redundant, inconsistent and unwanted information. Issues such as partial submissions, multiple submissions, synonyms, typos, irrelevant responses or extraneous inputs can negatively impact the model performance if not managed effectively. The initial steps of data cleaning in the pipeline analyze the learner submissions for consistency, missing data, outliers. Partially submitted responses are cleaned, pipeline is built to normalize text and standardize responses to correct for several potential inconsistencies. Processing of textual learner responses involves tokenization, stop words removal, lemmatization, normalization for common variations of keywords and words, correction of typos for common errors. Objective responses can be validated against expected formats in order to filter outliers and anomalies. These cleaning procedures to reduce unnecessary variance in learner responses along with completeness checks, create more effective evaluation systems by reducing noise in training datasets.

5.3 Session based Behavior Analysis
Session based behavioral analysis utilizes the fine-grained course data available in most EdTech platforms to summarize learner data for assessment purposes at the session level. Session partitions responses into individual sessions helping to analyze behavioral context, learning behavior, concept retention. The extracted metrics such as time required to complete tasks, number of attempts, time between responses, sequence of navigation are then formatted into behavioral features. Session analysis has also been found helpful in assessing learner fatigue, guessing and skill mastery over multiple sessions, providing additional context for assessment modules to detect knowledge gaps and adaptive assessments.

5.4 What types of AI/ML Models used
The auto-assessment system that is designed as a modular framework uses a tier of various AI models based on the determined task type and desired outcome.

5.4.1 NLP models for Text based subjective answers
The evaluation of subjective content responses will utilize pre, trained Language models using NLP techniques that are true to semantic contextual content rather than mere word matching approaches. These NLP models will match the semantic relevance of learner responses with correct responses and reference answers to account for synonyms, paraphrasing, and different expression styles. With contextualized word

embeddings, sentence transformers, and BERT like language models, semantic equivalency and the depth of learner understanding can be effectively measured.

5.4.2Classification & Regression modeling for Score predictions

For various objectives and scales, classification or regression models will be constructed to use various assessment features like response correctness, session-based metrics, behavioral data, past scoring information to derive scores. These models can support scaling assessment both summative and formative assessment.

5.4.3Learner Group Clustering:

Input data may also be subjected to Clustering algorithms in order to categorize learners into differentiated grouping based on various parameters of performance, engagement and progression. Supervised learning groups provide divisions for grouping learners into high, scoring, mid, scoring, low, scoring cohorts that help to customize assessment items and insights.

5.5 How does the Assessment logic function:

The assessment logic determines how the predictions of models are translated into the scoring or evaluation achievement. The automated assessment platform is designed to maximize the transparency, explainability, scalability and fairness in automated scoring schemes.

5.5.1 Creating Rubrics to standardize assessments:

Rubric driven evaluation techniques allow for consistency and clarity of automated evaluation. Rubrics can provide a set of criteria such as correctness of ideas, conciseness, completeness, clarity, relevance in learning content evaluation. Models can then correspond learner responses to these criteria, in turn enabling explainability in grading that parallels human rubric-based grading.

5.5.2Adaptive Scoring:

In order to maintain fairness and inclusion, the adaptive scoring which considers learners level or knowledge proficiency can be implemented. The evaluation parameters and criteria are adjusted based on the learner level, such that there is emphasis more on foundational understandings for beginner level learners while advanced learners are scored more on higher level ideas and problem-solving issues.

5.6 Summary of Approach:

The methodology focuses on utilizing current state-of-the-art in NLP for non-numerical content understanding, integrating behavioral contextual data for assessment, creating scalable models and task-based design, emphasizing transparent rubric based evaluation and adaptive scoring for personalization to provide effective auto-assessment architecture. This integrated approach endeavors toward the goal of providing an effective AI driven standardized assessment experience for education methodologies.

VI. IMPLEMENTATION DETAILS

The implementation of the course assessment, evaluation, prediction system is modular, scalable, and compatible with existing Ed-Techs. Key technology components are outlined below, as well as the training and evaluation methodology followed in practice.

6.1 Technology components

The system architecture comprises backend services, AI assessment engines, and databases suited for Ed-Techs. The technology options are driven by needs for high availability, low latency and secure handling of massive EdTech data repositories.

6.1.1EdTech Middleware and AI microservices

The driver of the system architecture is an EdTech middleware in the form of a RESTful web service layer, which processes login authentication, session management and communication identifiers via APIs. Ed-Techs offer a baseline LMS, a resilient learning management system with an assessment engine built in; the middleware interacts with the LMS, providing a higher level of inline synthesis systems. The individual assessment activities are implemented modularly in the form of microservices, such that different assessment modules could be independently added, scaled or changed without bringing down the entire system. Container systems and cloud computing platforms were essential for disaster recovery and ease of system maintenance. AI modules handle assessment task and such things as evaluation, scoring, summary generation as well; deployment friendly they could be refreshed with current learning algorithms without affecting the system architecture.

6.2 Data Storage
Learner data, Analytics Learner submission data, interaction tracking data and learner profile data are stored as both relational and NoSQL data sources to address different types of data requirements in an EdTech environment. Learner data include assessment attachment, assessments submissions, interaction logs, learner profile values are stored in their respective sections. Aggregated data collections are designed with a view to subsequent numerical treatments and reporting.

6.3. Training, Deployment Methodology
Models are trained on historic EdTech data while also being released on the live learner assessments. The training/deploying approach is based on use of existing high quality assessment data, and subsequent continuous updates.

6.3.1. Application of EdTech Historical data
The existing EdTech assessment data is analyzed, cleaned, and initially used to train supervised learning models. Expert marking to test Rubric adherence. It quickly proved a reliable and accurate baseline for model diagnostic validation and comparison.

6.3.2 Continual Learning & Model adaptation
Expert solutions and learner submissions are tagged according to marking error, rubrics and rubrics and model outputs are compared and used for ongoing incremental team learning, adaptation to unseen question types, rubrics and changing learner profiles. Summary of implementation approach: The approach balances robustness, ongoing model evolution, and adaptive learning to provide a scalable and flexible system.

VII. RESULTS AND LEARNING ANALYTICS

The success of an AI/ML based auto-assessment and evaluation platform is predicated not only on automation, scalability and throughput, but also on the accuracy and value of its outputs. In EdTech environments, assessment systems must be able to demonstrate end, to, end performance benefits while simultaneously providing learner engagement and instructional insights.
This section highlights the evaluation results and learning analytics outcomes of the auto-assessment system described in this paper, on the basis of key performance metrics and dashboard level insights.

7.1 Performance Metrics
The auto-assessment platform exhibited multiple quantitative and qualitative performance metrics, emphasizing both efficiency and pedagogical outcomes within an EdTech setting.

7.2 Grading Precision and Recall
This metric quantifies how well the auto-assessment outcomes approximate/integrate manual grading by teachers/instructors. In a purely objective assessment environment such as quizzes and multiple-choice tests, the auto, assessor reported no errors of equivalence, as there was no subjective evaluation or semantic interpretation involved. For subjective assessment environments such as sentences/phrase, level input or essay, type types, the assessment leverage NLP models, and adapt semantic equivalence, concept coverage and context fit based on the question and assessment type. When compared to instructor- graded samples, the auto, assesses demonstrated enough agreement that could establish reliability with manual graders, as well as consistency over large numbers of submissions. Such uniformity is of great consequence in EdTech platforms, where instructor variability may be interpreted as systemic unfairness and sub-optimal learner experience.

7.3 Time to Feedback
Time to Feedback measures the speed of the AI/ML auto-assessment system. In traditional assessment ecosystems, turnaround takes between several hours to a few days, which restricts timely feedback and formative assessments. Our system produces auto-assessment results in time, to feedback near instantaneously, consistent with learner inputs. Automated auto-assessment and feedback, enabled assessment also brings significant time savings, and supports increased learning iterations, which is crucial in EdTech systems.

7.4. Learner Proficiency Growth
This metric encapsulates the aggregated impact of the auto-assessment system on learner performance over time, in raw form and in the context of adaptive assessment and personalized difficulty adjustment. This is a linear trend based on the data of performance

across multiple assessments, and is corrected for learner ability, as well as context, domain and question origin. The analytics outcomes demonstrated clear learning impact, with significant performance gains owing to feedback regression and useful coverage, at the speed of automated auto-assessment. Learners can thus quickly learn from mistakes, identify weak points in their knowledge and learn others' approaches. This can be illustrated with learner progression visualizations, which corroborate learning improvement benefits.

7.5. Auto-assessment Learning Analytics Dashboards

With respect to providing AI/ML, powered formative assessment outcome summaries, the platform enables a comprehensive assessment data dashboard, which enhances both transparency and user instructional analytics to enable value creation via data-driven insights.

7.5.1 Learner Analytics Demographs

A learner dashboard can include outputting summary visualizations such as career temporal graphs, failure trend graphs, skill mastery level charts, performance distribution histograms, repetition effect diagrams and individual attempt, to, score comparisons. Such graphical summaries are invaluable for learner sensing, to contextualize the assessment experience, and allow intrinsic motivation building.

7.5.2 Instructor Analytics Pallets

Instructors can access a global learner analytics dashboard that unifies many sources of assessment data from the course or assessment platform. Such dashboards show learner performance distribution maps, indicator of missed concept clusters, learner engagement, performance by skill and domain types, and so on. These analytics can confer to instructors the knowledge about their learner groups, enabling tailored interventions on a large scale, and course design adaptations for future offerings.

VIII. USE CASE SCENARIOS

AI/ML based auto-assessment and evaluation platforms can be deployed in a variety of assessment scenarios in EdTech systems. The following use cases demonstrate varied kinds of assessment in real learning environments, and showcase how automated assessment methods help with assessment scale, speed, accuracy and engagement.

8.1 Automated quiz grading

Automated quiz grading is among the most common use cases for AI, based assessment in EdTech platforms. Online quizzes are often comprised of multiple-choice questions, true/false questions, fill, in, the, blanks and short, answer question formats. Manual grading of quizzes can be computationally infeasible when done at scale in online systems hosting thousands or millions of learners.

Automated quiz grading systems automatically evaluate quiz responses by ranking them with respect to answer keys or model answers. Further, machine, learning classifiers are used to determine grades of responses when they vary with respect to answer key, such as with regard to spelling errors or descriptive liberties. These systems allow for consistent grading standards while dramatically reducing assessment time.

From a pragmatic perspective, automated quiz grading delivers immediate feedback to learners, allowing them to quickly identify their knowledge gaps. EdTech course instructors also benefit from lighter administrative loads, allowing for greater exploration of self, directed learning techniques and pedagogical improvements. Such use cases lend well to the formative assessment characteristics of EdTech environments by enabling continuous improvement.

8.2 Automated evaluation of assignments in a MOOC environment

MOOCs have high enrollment rates, often in the hundreds, of, thousands. As such, grading assignments in the same MOOC typically requires numerous graders, but even then, the evaluation process is slow and inconsistent. Large cohort sizes in online MOOCs also make student, by, student manual assessment infeasible.

AI, based auto-assessment tools apply Natural Language Processing (NLP) techniques to compare various dimensions of text or other feedback to an answer key in order to evaluate learner submissions. Machine learning classifiers can also help identify desiderata across other objective assignment artifacts. In performing this task, the auto-assessment platform addresses the challenge of grading large cohorts across time.

In the context of a MOOC, auto-assessment platforms enable unbiased, equal treatment of all MOOC learner submissions for assignments. They may also catalyze the development of objective metrics of achievement and measure progress towards defined learning objectives, thus fostering peer-to-peer interactions and motivation as well.

8.3. Customized assessment pathways

Customized assessment pathways are a more sophisticated implementation of an automated auto-assessment platform within an EdTech context. Typically, use cases in this family leverage the learner models created via automated assessments in order to design and deliver adaptive assessments that cater to the particular learner.

Automated assessment, derived learner models estimate learner cognitive readiness, and guide adaptive assessment decisions. For example, focus of future assessments may be shifted dynamically to match individual learner priorities and strengths. This type of personalized formative assessment enables learner engagement in the learning process, while enhancing learning efficacy.

Given the learner models generated by these customized use cases, instructors may also design more customized instruction based on the particular needs of their learners, ascertain which assessment techniques work well for which learners, and increase learner completion and reinsertion rates.

8.4 Summary of how auto-assessment use cases impact EdTech systems

These concrete use cases shed light on how an auto-assessment system grows to encompass various forms of learner evaluation. Speed and consistency available via automation aid systems in providing large-scale, objective, equitable, and customized management of assessment at different levels of the EdTech ecosystem.

IX. ADVANTAGES AND LIMITATIONS

There are many benefits of using an auto assessment platform based on AI/ML in current EdTech systems which are used for solving the inherent challenges of traditional modes of assessment, but it also exhibits similar issues that must be taken into consideration.

9.1 Advantages

The major advantages of the proposed auto assessment platform include:

Scalability of assessment platforms - EdTech systems cater to large number of learners who are geographically distributed. Traditional manual assessment of processes will not be feasible when that number is very high. Automated assessment of systems will enable the educational institutions/organization to evaluate thousands of learners in parallel without proportionate investments in human resources. This is highly applicable for MOOCs, certification programs and corporate training systems.

Accuracy and objectivity in evaluation by removing grader bias or variability. Human grading is subjective and is highly variable depending on grader experience, familiarity with learner‘s responses and grading inconsistencies. Standard rubrics and data driven scoring are used in the proposed platform will standardize the evaluation criteria for all learners. This will make the grading fair and reliable.

Delivery of continuous, personalized feedback and reinforcement for each learner. As highlighted earlier, the immediate feedback delivered to the learners through the proposed system, will include wrong answers, concept notations, points of prior learning, etc. Hence, the system will direct the learners in their learning mode while reducing time that would have been spent on manual evaluation. This framework will support higher level of engagement in learners.

9.2 Limitations

The performance of the auto assessment system is highly dependent on the training data quality and usability. Biased or incomplete training data may result in non, fair evaluation process for different learners especially when they exhibit different linguistic styles of communication, learning styles and assessment ability. Although, the ML based evaluation systems work well with semi, structured or well, structured learning responses, evaluation of creative, open-ended or ambiguous responses may pose as a challenge for the proposed system to perform similar to a human evaluator.

X. ETHICAL, PRIVACY AND REGULATORY ISSUES

The use of AI driven assessment systems in EdTech systems have many ethical and legislative implications especially on learner data privacy, fair judgment, education and data security and transparency. Since assessment results can influence the educational institutes or learner's career choices, it is imperative that the evaluation system behaves ethically and thoughtfully.

Learner Data privacy is an important concern in emerging EdTech platforms that collect sensitive learners' data like learner identifiers, behavioral data, action logs, learning outcome data, etc. The proposed auto assessment platform incorporates data anonymization, bundling and encryption for viewing data system. Confidentiality measures like access barriers and regular audit may be employed in the proposed implementation for data security.

Equity and lack of bias may creep in when trained on past imperfect data and subsequently exhibit in matured models, non, respect for minority specific data patterns and unfair judgments. The proposed auto-assessment system has stress on collection of diverse training data and follow, up bias audits. The assessment outputs are checked periodically for any inequities. Explainability insights-based AI models will be users' trust in the system behavior.

Justice and attribution are crucial for system acceptance. The proposed system deploys explainability methods to get the rationale behind evaluation decisions. This explanation removes the in, between mechanics from users' perspective and generates confidence in the auto assessment system.

10.1 Compliance with regulation and policy

The development of auto-assessment system should comply with all institutional requirements for assessment, grading and data security. The system is designed to facilitate audit transparency, workflows with clear documentation for the multiple evaluation steps and evaluation results. The assessment system can be tailored to meet the institutional accreditation, data security and assessment legislations like Title IX requirements of the US Department of Education, etc. Responsible inclusion of the proposed auto assessment system will help us use this emerging technology to assist educators in their critical roles.

XI. FUTURE SCOPE

Auto assessment and evaluation system, as described in the proposed framework, is just the beginning of many possible application areas of the EdTech platforms. Many future innovations can build upon this foundation in order to enhance assessment and learning experiences.

One important future scope is to add adaptive assessment and evaluation features to support risk differentiation of learners. Adaptive assessment dynamically changes according to the rate of correct and incorrect responses as well as current level of learners, thus providing the learning support at the right cognitive levels.

Additionally, multilingual evaluation support will be significant. This is increasingly relevant now that mobile based and online EdTech systems are accessed by a global and diverse population of learners who communicate in different languages and dialects. AI models for multi, lingual responsiveness, would hereafter be more expensive and require large amount of multi, lingual training data.

Furthermore, use these ongoing evaluation results to identify skill gap zones for each learners and influence learner's future registrations accordingly. This enables combination of assessment results with their future selections for a holistic and more comprehensive future framework for automated assessment for learning management systems in EdTech.

11.1 Emerging AI technologies and support for the auto assessment framework

Generative AI models may be integrated to the auto assessment framework to automate the delivery of feedback. These models have the potential to generate feedback, explanations, elaborations and examples that will inform learners about the assessment scores and the next steps for their learning endeavors.

Multimodal assessment using multiple modalities like videos, images, audios and simulations may be coupled together under the proposed auto assessment framework for assessment of non, text questions and responses.

XII. SUMMARY AND CONCLUSION

This paper largely focused on design considerations for developing a state-of-the-art AI/ML auto assessment and evaluation platform for EdTech systems. The tool architecture of the AI/ML auto assessment system, workflows, data pipelines, analytic driven assessment processes, evaluation rubrics and feedback mechanisms were highlighted in the described auto assessment platform. The suitability of the auto assessment platform to EdTech systems for scalable assessment environments was reinforced.

This paper outlined the use of AI/ML auto assessment framework will aid in overcoming the scaling challenges of traditional assessment mechanisms in EdTech. This work demonstrated that learner evaluation time, instructor's evaluation effort and evaluation intricacies can be massively reduced with respect to the use of the auto assessment platform in EdTech environments. The challenge of designing warm and open auto assessment process that is aligned with pedagogy also was discussed.

The research also threw light on the importance of responsible auto assessment decision systems which incorporate multiple considerations for pedagogical, technical and legal impacts caused while designing this effective auto assessment framework.

All in all, this research emphasized on how education and assessment industries will leverage the advancements of AI/ML auto assessment technology to make education more affordable, equitable and efficient in coming years.

REFERENCES

[1] R. Clark and R. Mayer, "E-Learning and the Science of Instruction," Wiley, 4th ed., 2016. https://www.wiley.com/en-us/e-Learning+and+the+Science+of+Instruction%3A+Proven+Guidelines%20+for+Consumers+and+Designers+of+Multimedia+Learning%2C+4th%20+Edition-p-9781119239086

[2] G. Siemens, "Learning analytics: The emergence of a discipline," Amer. Behav. Sci., vol. 57, no. 10, pp. 1380-1400, 2013.

[3] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," in The Adaptive Web, P. Brusilovsky et al., Eds. Berlin, Germany: Springer, 2007, pp. 3-53.

[4] T. K. Landauer et al., "The reliability, validity, and utility of the automated essay scoring system," in Automated Essay Scoring: A Cross-disciplinary Perspective, M. D. Shermis and J. Burstein, Eds. Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 157-180

[5] V. González-Calatayud, P. Prendes-Espinosa, and R. Roig-Vila, "Artificial Intelligence for Student Assessment: A Systematic Review," Appl. Sci., vol. 11, no. 12, p. 5467, 2021. doi: 10.3390/app11125467.

[6] J. Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers, ACL

[7] N. Reimers and I. Gurevych, Sentence-BERT, EMNLP

[8] Shermis & Burstein, Automated Essay Scoring: A Cross-Disciplinary Perspective, Routledge. https://www.routledge.com/Automated-Essay-Scoring/Shermis-Burstein/p/book/9780415890485

[9] Baker & Inventado, Educational Data Mining and Learning Analytics, Springer, 2014. https://link.springer.com/chapter/10.1007/978-1-4614-3305-7_2

[10] Uto & Okano, Robust Neural Grading for Open-Ended Responses, Computers & Education, 2020. https://www.sciencedirect.com/science/article/pii/S036013152030162 2

[11] Bodily & Verbert, Review of Automated Feedback and Assessment in Learning Analytics, IEEE Transactions on Learning Technologies, 2017. https://ieeexplore.ieee.org/document/7964702

[12] Piech et al., Deep Knowledge Tracing, NeurIPS, 2015. https://papers.nips.cc/paper/5654-deep-knowledge-tracing

[13] Ke, Z., & Ng, V., Automated Essay Scoring: A Survey of the State of the Art, Proceedings of IJCAI, 2019. https://www.ijcai.org/proceedings/2019/0720.pdf

[14] Hattie, J., & Timperley, H., The Power of Feedback, Review of Educational Research, 2007. https://journals.sagepub.com/doi/10.3102/003465430298487

[15] Bodily, R., & Verbert, K., Review of Research on Automated Feedback and Learning Analytics, IEEE Transactions on Learning Technologies, 2017. https://ieeexplore.ieee.org/document/7964702

[16] Holmes, W., Bialik, M., Fadel, C., Artificial Intelligence in Education: Promises and Implications, Pearson, 2019. https://curriculumredesign.org/wp-content/uploads/AIED-Book.pdf

[17] Luckin, R. et al., "Intelligence Unleashed: An argument for AI in education," Pearson, 2016. https://www.pearson.com/content/dam/one-dot-

com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed-Publication.pdf

[18] Burrows, S., Gurevych, I., Stein, B., "The eras and trends of automatic short answer grading," International Journal of Artificial Intelligence in Education, 2015. https://link.springer.com/article/10.1007/s40593-014-0026-8

[19] Williamson, D. M., Xi, X., Breyer, F. J., "A framework for automated scoring," Educational Measurement, 2012. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2012. 00240.x

[20] Shute, V. J., "Focus on formative feedback," Review of Educational Research,2008. https://journals.sagepub.com/doi/10.3102/0034654307313795

[21] Hew, K. F., Cheung, W. S., "Students' and instructors' use of MOOCs," Educational Research Review, 2014. https://www.sciencedirect.com/science/article/pii/S1747938X140000 64

[22] Nicol, D. J., Macfarlane-Dick, D., "Formative assessment and self-regulated learning," Studies in Higher Education, 2006. https://www.tandfonline.com/doi/10.1080/03075070600572090

[23] Baker, R., Hawn, A., "Algorithmic bias in education," International Journal of AI in Education, 2021. https://link.springer.com/article/10.1007/s40593-021-00228-4

[24] Pardo, A., Siemens, G., "Ethical and privacy principles for learning analytics," British Journal of Educational Technology, 2014. https://berajournals.onlinelibrary.wiley.com/doi/10.1111/bjet.12152

[25] GDPR – General Data Protection Regulation, European Union, 2018. https://gdpr-info.eu

[26] Doshi-Velez, F., Kim, B., "Towards a rigorous science of interpretable ML," arXiv, 2017. https://arxiv.org/abs/1702.08608

[27] Crossley, S., McNamara, D., "Multilingual NLP in educational assessment," Language Resources and Evaluation,2019. https://link.springer.com/article/10.1007/s10579-019-09450-9

[28] Kasneci, E. et al., "ChatGPT for education: Opportunities and challenges," Learning and Individual Differences, 2023. https://www.sciencedirect.com/science/article/pii/S1041608023000195

[29] Holmes, W., Tuomi, I., "Generative AI and the future of assessment," OECD Education Working Papers, 2022. https://www.oecd.org/education/generative-ai-and-the-future-of-assessment.htm