Received 30 November 2024, accepted 12 January 2025, date of publication 16 January 2025, date of current version 28 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3530474

RESEARCH ARTICLE

Automated Generation of Multiple-Choice Questions for Computer Science Education Using Conditional Generative Adversarial Networks

MUHAMMAD SHOAIB[1,2], GHASSAN HUSNAIN[1], NASIR SAYED[1,3], YAZEED YASIN GHADI[4], MASOUD ALAJMI[5], AND AYMAN QAHMASH[6]

[1]Department of Computer Science, CECOS University of IT and Emerging Sciences, Peshawar 25000, Pakistan
[2]Department of Computer Science and Information Technology, Sarhad University of Science and IT, Peshawar 25000, Pakistan
[3]Department of Computer Science, Islamia College Peshawar, Peshawar 25100, Pakistan
[4]Department of Computer Science and Software Engineering, Al Ain University, Al Ain, United Arab Emirates
[5]Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[6]Department of Informatics and Computer Systems, King Khalid University, Abha 62521, Saudi Arabia

Corresponding author: Ayman Qahmash (a.qahmash@kku.edu.sa)

This work was supported by the Deanship of Research and Graduate Studies at King Khalid University through Large Research Project under Grant RGP2/455/45.

ABSTRACT This work presents a novel perspective towards generating automated multiple-choice questions (MCQs)-a task fundamentally different due to the highly dynamic nature of computer science education, which spans several sub-domains. Taking advantage of Conditional Generative Adversarial Networks (cGANs), our model provides a versatile approach to addressing the need for diversity and context in relevant MCQ generation across proficiency levels, topic areas. Resulting MCQs inspire implementations within a variety of educational environments - from classrooms, to online courses, and finally exams - equipping teachers with an instrument that could be easily adapted based on the specific needs o students. The model is trained on a carefully constructed dataset that includes material from more than 20 subareas in computer science, consisting of materials such as textbooks, online encyclopedias and Q&A websites. Through rigorous evaluation using comprehensive performance metrics, including Question Relevance Score (QRS), Diversity Index (DI), and Difficulty Alignment Accuracy (DAA), we demonstrate the efficacy and robustness of our framework in generating high-quality MCQs. Moreover, we address ethical considerations inherent in AI-driven educational assessment, ensuring fairness, transparency, and accountability in the MCQ generation process. The cGAN architecture facilitates the generation of contextually relevant MCQs across various proficiency levels and subject domains, enhancing the educational assessment process. The comprehensive dataset developed for this study encompasses diverse computer science topics curated from authoritative textbooks, online resources, question banks, and instructor-generated content. Additionally, a user-friendly QT application has been developed, enabling seamless integration of the cGAN model into educational environments. Through rigorous evaluation and ethical considerations, this framework demonstrates its efficacy, ensuring fairness, transparency, and accountability in MCQ generation. This interdisciplinary work represents a significant advancement in computer science education, providing educators with a powerful tool to enhance student engagement and learning outcomes.

INDEX TERMS Automated MCQ generation, conditional generative adversarial networks (cGANs), computer science education, dataset curation, educational assessment.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

I. INTRODUCTION

In the dynamic landscape of educational assessment, the continuous evolution of technology has precipitated groundbreaking methodologies, each aimed at refining and elevating traditional evaluation practices [1]. One particularly innovative approach that stands at the forefront of this transformative wave is the application of Conditional CGANS for the purpose of generating MCQs within the realm of computer science education [2]. This cutting-edge technique represents a paradigm shift, introducing a personalized and adaptive dimension to the evaluation process, strategically personalized to assess students across diverse proficiency levels [3]. In delving into the intricacies of cGAN-based MCQ generation, this approach not only signals a departure from conventional assessment systems but also presents a compelling solution to the inherent limitations and challenges that have long persisted in educational evaluation methodologies [4]. Traditionally, educational assessments have relied on standardized testing and manual question creation, posing inherent challenges such as time constraints for educators, limited personalization, and a one-size-fits-all approach [5]. However, with the advent of transformative methodologies driven by technological advancements, the educational landscape is witnessing a paradigmatic shift. This shift is characterized by a relentless pursuit of methodologies that transcend the limitations of traditional assessment, embracing innovation to better cater to the diverse needs and proficiency levels of the student body [6].

At the forefront of this technological frontier is the application of Conditional CGANS, an advanced class of machine learning models renowned for their capacity to generate realistic and contextually relevant content [7]. In the context of educational assessment, particularly within the domain of computer science, cGANs emerge as a transformative force, providing an avenue for the automated generation of MCQs with a level of personalization and adaptability that was hitherto unprecedented [8]. The core essence of employing cGANs for MCQ generation lies in the capacity to tailor assessments to the unique learning trajectories and expertise levels of individual students. By incorporating specific conditions or parameters during the generation process, cGANs facilitate the creation of MCQs that are finely attuned to the proficiency spectrum, ranging from beginner to intermediate and expert levels [9]. MCQs have entrenched themselves as a fundamental and widely embraced component of educational assessments, owing to their unparalleled efficiency in appraising a student's comprehension across a diverse spectrum of subjects [10]. The prototypical format of MCQs is a stem followed by multiple answer choices, providing an equal and unbiased manner to evaluate students. It goes beyond speciality areas and crosses all the branches making MCQs a versatile tool in pedagogical repertoire with universal applicability [11]. The stem is designed to evoke a particular type of response from the examinee, establishing an intellectual climate for further testing [12]. The stem is accompanied by a set of answer options (among which the correct answer and distracters are placed strategically). Its components interact intricately to create one whole, which encompasses the basis of MCQ and ultimately contributes to making them a valid method for assessment [13]. MCQs based systems are considered very objective for binary grading but not only they give a one-dimensional sight of students performance, which is debatable point. This objectivity, in part due to the structured form of MCQs - each question being designed in essentially equal format and scored objectively on responses. [14].

With a continuously shifting educational landscape and an increasing need for more diverse, timely, expert-level questions educators now find themselves faced with a set of complexities that require novel solutions [15]. Even with these various structural changes, it is clear that a fresh approach to MCQ generation was long overdue and adapting limited traditional methodologies only provided the foundation for next-generation alternatives which can continuously adapt based on individual learning styles [16]. Traditional MCQ generation carries with it one of its biggest problems - in the fact that it is resource-heavy. Ensuring that not only the quality of questions are diverse, up to date and expert-level proficient requires significant time and effort from educators as well [17]. Building a comprehensive repository of questions manually, versatile enough to span the range and breadth in any given subject area can be arduous task that could potentially take resources away from other important facets of teaching/learning [18]. As educational institutions work to create high quality curricula and assessments, the resource constraints on traditional MCQ creation methods become more obvious [19].

Traditional question banks, while serving as repositories of knowledge, often fall short in providing adaptive assessments that cater to the unique needs and expertise levels of individual learners [20]. Teachers look for more automation abilities that technology can provide to successfully add insights and improve assessments beyond the manual question making [21]. From this perspective, seeking fresh methods of MCQ generation is in sync with the grand theme for modern education using technology as an enabler to provide a dynamic engagement and focus on student-centric learning. The uptake of modern technological practices, such as Conditional CGANS is a promising potential solution to challenges encountered with outdated MCQ generation methodologies and can pave the way for an evolutionary step into future-proofing assessment environments that are both best-suited and inclusive [21].

Various artificial intelligence and natural language processing tools have been improved significantly for the automating MCQs generation; in this respect it is essential to choose a good model. Though many different tools -that are both AI-based and NLP based - have very promising features, Generative Adversarial Networks (GANs) seem to show the most potential for producing content which is realistic sounding and contextually aware. GANs use a special kind of adversarial training, in which the generator and discriminator take turns trying to outdo each other - promoting creation content creativity/improvisation. Among the GAN variants, Conditional GANs (cGANs) emerge as an ideal choice for MCQ generation in educational settings. The conditional nature of cGANs enables the incorporation of specific parameters or conditions during the generation process. In the context of educational assessments, this means tailoring questions to match the proficiency levels of students, distinguishing between beginners, intermediates, and experts. This adaptability ensures that assessments align closely with individual learning trajectories [22].

The adoption of a cGAN-based MCQ generation system offers several compelling advantages: The cGAN model allows for the creation of MCQs that are tailored to the expertise level of each student. This personalized approach fosters a more engaging and effective learning experience.

By categorizing questions into beginner, intermediate, and expert levels, the system provides adaptive assessments that evolve with the learner's progress, ensuring an appropriate level of challenge [23]. By varying levels of difficulties of questions to cater the customization for students it also promotes greater student engagement, striking a good balance between challenging them to aim high but being there along the way [24].

Detail on cGAN's application for MCQ generation has quite a few benefits but some challenges need to be noticed: The produced questions have good quality and are true. A rigorous review process which has inputs from educators and domain experts for reviewing the assessments is important to preserve their reliability [26]. Ethical considerations, specifically relating to bias in questions created remain top-of-mind. The need for fair and unbiased assessments is of utmost importance to maintain the sanctity and reliability of education. Educational content is an ever-evolving beast. Feedback, iteration and updating mechanisms must be established in the cGAN model to maintain assessments timely according to new educational standards [25].

Contribution Reflecting Novelty in Work:

- The proposed cGAN-based MCQ generation framework represents a novel contribution to the field of computer science education assessment. Unlike traditional approaches, which often rely on manual question authoring or rule-based generation techniques, our method harnesses the power of deep learning and adversarial training to autonomously generate contextually relevant and diverse MCQs. This innovative approach offers several key contributions:
- Our framework leverages Conditional CGANS, a cutting-edge deep learning architecture, to tailor question generation according to specified difficulty levels and fields of study. By incorporating conditional inputs, our model can produce personalized and adaptive MCQs that cater to the unique learning needs and proficiency levels of individual students.
- We curated a comprehensive dataset comprising a diverse range of computer science topics sourced from authoritative textbooks, online resources, question banks, and instructor-generated content. This extensive dataset ensures the model's exposure to varied linguistic patterns, thematic contexts, and difficulty levels, enhancing the richness and diversity of generated MCQs.
- We devised a robust evaluation framework encompassing multiple performance metrics, including QRS, Diversity Index (DI), Difficulty Alignment Accuracy (DAA), and Semantic Equivalence Score (SES). This holistic approach enables comprehensive assessment of the generated MCQs' quality, relevance, diversity, and alignment with predefined difficulty levels, ensuring the reliability and validity of the evaluation results.
- The integration of a Conditional Generative Adversarial Network (cGAN) into a user-friendly Graphical User Interface (GUI) marks a significant contribution to the field of computer science education. This novel approach streamlines the process of generating contextually relevant and diverse MCQs tailored to specific difficulty levels and subject domains. By leveraging the capabilities of cGANs, the application offers a dynamic and adaptable solution for educators, students, and professionals seeking to create high-quality assessment materials.

II. LITERATURE REVIEW

This research critically examines the impact of ChatGPT in higher education [26], employing thin ethnography to elucidate its perspectives on challenges and opportunities. While personalized learning and enhanced accessibility emerge as potential benefits, concerns loom over issues such as plagiarism and the potential weakening of critical thinking skills. A balanced and cautious approach is crucial to harness the advantages of ChatGPT while addressing and mitigating associated risks in higher education.

This review paper undertakes a meticulous analysis of 207 research papers [27], employing the PRISMA framework alongside content and bibliometric analyses, to delve into the promising potential of Generative Artificial Intelligence (GAI) in education. Study has been carried out, and it identifies the key takeaways like where GAI can be applied to working towards medical & engineering education. The paper closes by arguing for providing directions on future research in the GAI-assisted curriculum design and what benefits will have long term effects of this GAI-aided learning. This expansive review will serve as an invaluable blueprint assisting researchers, educators and most importantly policy makers in the informed integration of GAI into the educational paradigm.

This paper undertakes an extensive exploration of the potential of Artificial Intelligence (AI), specifically focusing on Natural Language Processing (NLP) and Large Language Models (LLMs) such as GPT-4 and BARD, to usher in a paradigm shift in both educational and research domains [28]. At the same time however, it looks into transformational use cases in education like educational assistance and feedback, library evaluation & grades and teaching strategies right up to formulating our career paths or helping students with mental health support as well. The article earnestly discusses the challenges involved in this disruptive journey, which comes with a strong emphasis about ethical considerations as well as algorithmic biases.

This study investigates the underexplored realm of employing generative artificial intelligence (AI) [29], specifically ChatGPT-3.5 and 4, in primary school education. In a project involving 110 pupils aged 8-14 from two schools, the research illustrated that this approach is capable of producing bespoke learning resources which exhibit potential for personalized education. The findings highlight that efforts are warranted to ensure generative AI is integrated in an impactful, inclusive and sustainable manner within the primary school education landscape. Whilst the journey to world-class generative AI continues, this paper is a testament of how it will inevitably play out by demonstrating one step closer towards an optimized learning experience in using sophisticated models which learn with and beyond humans - for human intelligence.

This article explores the impact of ChatGPT, a generative AI tool, on education in China [30]. While recognizing the possibility for personalized learning and digital transformation, it also discusses difficulties such as fears of academic integrity or lack of ability in critical thinking. This study will integrate the ChatGPT model as an educational tool carefully based on a systematic review of Chinese scholars and experts opinions. The new framework, ''Development-Administration-Teaching-Student (DATS)'', is designed to inform future implementations grounded in the insights of developers, administrators and teachers/students. The author discuss ChatGPT as a platform China had already mastered and the necessity for responsible deployment to confront education in China where it is today.

This research investigates the application of ChatGPT, a generative AI, in science education, focusing on its proficiency in answering scientific questions, educational applications for teachers, and its role as a research tool [31]. The results shows that the ChatGPT can be employed to solve questions yet impart challenges on how it would convey information as such a strong claim. Educators have a role to model responsible AI use and stress the importance of critical thinking, in addition to setting clear student expectations.

This article employs critical analysis and paradox theory to explore the multifaceted role of Generative AI in education [32]. The paper argues it as a fovour, and a foe face accessible yet also limiting sometimes without going into full lock down state of use in mainstream acceptance. This specific analysis brings a level of clarity that furthers our overall understanding - which is incredibly valuable for administrators, policy makers and other stakeholders involved in the future of education.

This paper explores ChatGPT's potential in education, emphasizing its role as a conversational agent to enhance teaching and learning [33]. While the study points out limitations of this generative AI model, such as require to know human-like text and training data dependency (existing models better for image-textbert generation), it underscores its potential promise needs in new areas like prompt engineering. It highlights the role of human agents not only to navigate possible pitfalls but also in order for ChatGPT innovation within education is leveraged effectively, with some adjustments made now.

This paper explores the transformative potential of Generative Pre-trained Transformer (GPT) AI in education [34], specifically focusing on its role in enhancing collaborative learning and knowledge building. To further advance this argument - that integrating AI can be aligned with pedagogical ideals, the study presents practical designs based on ''knowledge building'' pedagogy. This paper is written for educators, researchers and all others intending to use Generative AI in teaching practices with a view towards generating insights that will open up discussions furthering best-pactices.

This study addresses the limitations of traditional learning technologies in universities by proposing the integration of intelligent chatbots powered by generative AI [35]. The study highlights how chatbots show promise in helping to boost learner engagement, offer immediate feedback and tailor learning experiences. By leveraging generative AI tools the proposed concept aims to allow online student-to-student interaction and self-paced learning in an automated course creation module for instructors. Designed for university educators, administrators and academics researching AI in education, the research is to help this work better drive effective implementation through a more joined up approach of delivering an enhanced learning experience at scale.

III. METHODOLOGY

A. DATASET DEVELOPMENT

1) DATA COLLECTION

The dataset for this study was meticulously curated from authoritative textbooks integral to the computer science degree program. These textbooks spanned an array of comprehensive subjects, encompassing the foundational and advanced knowledge areas of programming languages, theoretical computer science, networks and communication, databases, artificial intelligence, operating systems, software engineering, web technologies, cybersecurity, computer graphics, human-computer interaction, software architecture, Internet of Things (IoT), big data and data science, software testing and quality assurance, mobile application development, and computer ethics and professionalism. Table 1 illustrates the distribution of samples across various computer science fields, forming a comprehensive dataset for assessing educational proficiency levels.

2) ONLINE EDUCATIONAL RESOURCES

Complementary to traditional textbooks, pertinent content was also sourced from reputable online educational

TABLE 1. Distribution of samples across computer science fields, forming a comprehensive dataset for varied educational proficiency levels.

Field	Number of Samples
Programming Languages	5020
Theoretical Computer Science	3540
Networks and Communication	6330
Databases	4945
Artificial Intelligence (AI)	5155
Operating Systems (OS) and Hardware	4235
Software Engineering	3860
Web Technologies	2825
Cybersecurity	2338
Computer Graphics	1620
Human-Computer Interaction (HCI)	1442
Internet of Things (IoT)	1232
Big Data and Data Science	1148
Mobile Application Development	1422

platforms, ensuring a contemporary and diverse representation of computer science concepts. This involved meticulous consideration of content quality, relevance, and adherence to the specified topics.

3) EXISTING QUESTION BANKS

To enrich the dataset, select question banks related to computer science education were integrated. These sources served as reservoirs of validated questions, ensuring a broad spectrum of topics and difficulty levels.

4) INSTRUCTOR-GENERATED CONTENT

In order to tailor the dataset to the specific needs of the study, supplementary content was developed by domain experts and educators. This content was aligned with the curriculum objectives and provided a bespoke dimension to the dataset. Table 2 provides a snapshot of varied multiple-choice questions, showcasing the diversity of the comprehensive dataset used for computer science education assessment.

5) CUSTOM MCQS BANK

This study leverages a diverse array of online resources, including LeetCode, HackerRank, Interview Cake, Codewars, GeeksforGeeks, CodinGame, Sphere Online Judge, HackerEarth, and Project Euler, to curate a rich dataset for the automated generation of multiple-choice questions (MCQs) in computer science education. By tapping into these platforms, which offer a wealth of coding challenges and problem-solving exercises, we ensure the breadth and depth of topics covered in our MCQ bank. Furthermore, our dataset is augmented by insights from prominent textbooks such as " Grokking artificial intelligence algorithms [36]," "Introduction to Algorithms [37]", "Cracking the code: Co-coding with AI in creative programming education [38]", "MCQs in Computer Science [39]", "Data Structures and Algorithms in Python [40]," and "Computer Networking: A Top-Down Approach [41]." Drawing from these authoritative sources, we enrich our MCQ bank with a wide spectrum of topics, ranging from fundamental algorithms and data structures to advanced networking concepts.

6) CATEGORIZING DATA FOR MODEL DEVELOPMENT

Rigorous criteria, grounded in pedagogical principles, were employed to categorize questions into distinct difficulty levels—beginner, intermediate, and expert. The complexity of concepts, problem-solving requirements, and depth of understanding were integral considerations in this categorization process. A panel of subject matter experts, possessing substantial proficiency in the diverse fields of computer science, was engaged to validate and assign difficulty levels to questions. Their expertise ensured a nuanced and accurate classification reflective of the intended difficulty spectrum. Deliberate efforts were made to achieve a balanced distribution of questions across difficulty levels. This equilibrium ensured the representation of each proficiency tier, contributing to the overall robustness of the dataset. Choices accompanying each question were meticulously annotated with difficulty levels. This granular annotation facilitated the generation of varied and contextually relevant multiple-choice options for each question. The table 3 presents the variety of computer science questions sourced from textbooks, online resources, question banks, and instructors, collectively forming a comprehensive dataset for educational assessment in computer science.

B. MODEL DEVELOPMENT

1) MODEL ARCHITECTURE

The text-based MCQs are tokenized, transforming them into sequences of numerical tokens. This step captures the semantic meaning of words and allows the model to work with numerical representations of the input. To create uniform input sequences, padding is applied to the tokenized sequences. This ensures that all sequences have the same length, facilitating efficient batch processing during model training. Tokenized sequences are embedded into high-dimensional vectors using pre-trained word embeddings or embedding layers. Embeddings capture the semantic relationships between words and enable the model to understand the context of the MCQs. Metadata, including difficulty level and field of study, is appropriately encoded and combined with the tokenized text data. This allows the model to consider additional contextual information during MCQ generation (see Figure 1 for the proposed cGAN for text generation).

This table 4 below illustrates the step-by-step preprocessing of a programming-related Multiple-Choice Question (MCQ). The process includes tokenization to break down the text into individual tokens, padding to ensure uniform sequence length, embedding to convert tokens into high-dimensional vectors, and metadata processing, incorporating difficulty level and field of study. The resulting processed data is essential for inputting into the Conditional Generative Adversarial Network (cGAN) model for MCQ generation.

TABLE 2. Varied multiple-choice questions representing a snapshot of the comprehensive dataset for computer science education assessment.

Question	Choices	Correct Answer
1. What is the purpose of a semicolon in C++?	a. Terminate a program	b. Separate statements
	b. Separate statements	
	c. Declare a variable	
	d. Define a function	
2. In Python, what is the output of print(3 * 4)?	a. 7	b. 12
	b. 12	
	c. 0	
	d. 34	
3. What is a spanning tree in graph theory?	a. A tree with a single node	b. A tree that spans all nodes
	node b. A tree that spans all nodes	
	c. A tree with no leaves	
	d. A tree with two branches	
4. What is the time complexity of a binary search algorithm?	a. O(1)	b. O(log n)
	b. O(log n)	
	c. O(n)	
	d. O(n^2)	
5. What is a characteristic of Cloud Computing?	a. Local storage only	c. Dynamic scaling of resources
	b. On-premise servers	
	c. Dynamic scaling of resources	
	d. Decentralized architecture	
6. In computer networks, what does DHCP stand for?	a. Dynamic Host Configuration Protocol	a. Dynamic Host Configuration Protocol
	b. Distributed Host Control Protocol	
	c. Domain Host Configuration Process	
	d. Digital Host Control Protocol	

The table 4 illustrates the preprocessing steps undertaken for a programming-related Multiple-Choice Question (MCQ).

TABLE 3. Variety of computer science questions sourced from textbooks, online resources, question banks, and instructors, forming a comprehensive dataset.

Field	Textbooks	Online Resources	Question Banks	Instructor-Generated
Programming Languages	2151	1075	1434	360
Theoretical Computer Science	2212	708	354	265
Networks and Communication	1770	708	885	177
Databases	2065	708	472	295
Artificial Intelligence (AI)	1888	849.6	566	236
Operating Systems (OS) and Hardware	1517	809	1011	202
Software Engineering	2065	1180	295	0
Web Technologies	1676	931	652	279
Cybersecurity	1843	737	590	368
Computer Graphics	1888	1062	295	295
Human-Computer Interaction (HCI)	1685	842	842	168
Internet of Things (IoT)	1991.25	1106.25	442.5	0
Big Data and Data Science	1831.034	915.5172	610.3448	183.1034
Mobile Application Development	1770	944	708	118

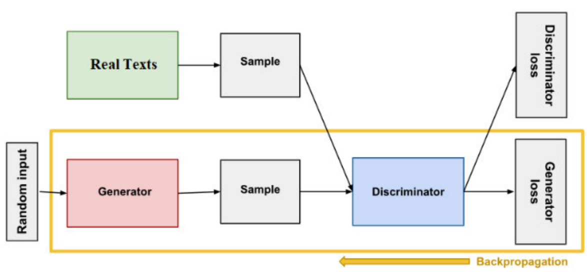


FIGURE 1. Proposed cGAN for text generation.

MCQ Generation:- It completely depends on how you design a cGAN. Model architecture is an important variable in how well the model can learn these patterns, ask appropriate questions that make sense given context of input and adapt depending on user response. The cGAN used in this research is designed especially for the requirements of educational assessment. It is made up of a generator and discriminator which keep trying to out play each other in the

TABLE 4. Preprocessing steps for a programming-related MCQ.

Step	Processed Data
Original Text	"In Python, what does the len() function do?"
Tokenization	['In', 'Python', ',', 'what', 'does', 'the', 'len()', 'function', 'do', '?']
Padding	['In', 'Python', ',', 'what', 'does', 'the', 'len()', 'function', 'do', '?', '0', '0']
Embedding	[0.15, 0.72, 0.88, -0.45, 0.92, -0.81, 0.65, 0.42, 0.78, -0.63, 0.0, 0.0]
Metadata (Difficulty Level, Field of Study)	[2 (Intermediate), 1 (Programming)]

zero sum game designed for our MCQ generation. The generator (a recurrent neural network, RNN) is implemented to contain several layers of Long Short-Term Memory (LSTM) units in order to maintain the natural process through time and enhance connectedness between words. Dropping out is coupled with each LSTM layer to avoid overfitting and improve generalization. The discriminator, a convolutional neural network (CNN), serves as the discerning component, evaluating the authenticity and relevance of generated questions. Convolutional layers, complemented by max-pooling operations, enable the extraction of salient features crucial for discriminating between genuine and synthetic MCQs.

These changes were made to adapt the cGAN for MCQ generation. A conditional input layer receiving metadata, such as the desired difficulty level and subject matter, was added. By tuning the cGAN, explicit proficiency tiers and subject domains can be targeted in the generated questions.

Training the model on balanced data across fields and difficulty levels involves using Adam optimization with changes in the learning rate to aid convergence. During the training process, for each iteration, a generator and discriminator are set up in an adversarial training paradigm, where one generates some images to try to fool the other. The Table 6 contains the hyperparameters used for training cGAN. Moreover, the cGAN architecture, as shown in Figure 2, provides details about how the model was structured and built.

2) TRAINING PROCESS

Training the cGAN is a complex and stage-wise process, necessary to equip it with skills to create rich informative MCQs. This part presents the complete methodology covers key aspects such as number of epochs, batch sizes, learning rates and regularization techniques through which these models were developed. And the cGAN is trained for a large number of epochs to understand how MCQs are generated. During training the model continuously performs gradual optimization over several epochs, finding a balance between convergence and avoiding complete lack of generalization (overfitting). This duration is determined empirically to give an excellent but subtle view of the rich variety half-a-million rows in your data. A batch size of 64 is used to

TABLE 5. cGAN architecture layers and parameters for text-based MCQ generation.

Layer Type	Parameters
Input Layer (Embedding)	Input: Text-based MCQs (Tokenized and Padded)
Conditional Input Layer	Input: Metadata (Difficulty Level, Field of Study)
Generator (RNN or LSTM)	LSTM Layer 1: Units=256, Activation='tanh', Return Sequences=True, Dropout=0.2
	LSTM Layer 2: Units=256, Activation='tanh', Return Sequences=True, Dropout=0.2
	Dense Layer 1: Units=128, Activation='tanh'
Conditional Input Layer	Input: Metadata (Difficulty Level, Field of Study)
Discriminator (CNN)	Conv1D Layer 1: Filters=64, Kernel Size=3, Activation='relu'
	MaxPooling1D Layer 1: Pool Size=2
	Conv1D Layer 2: Filters=128, Kernel Size=3, Activation='relu'
	MaxPooling1D Layer 2: Pool Size=2
	GlobalMaxPooling1D Layer
	Dense Layer 1: Units=128, Activation='relu'
	Dropout Layer 1: Rate=0.5
	Dense Layer 2: Units=1, Activation='sigmoid'

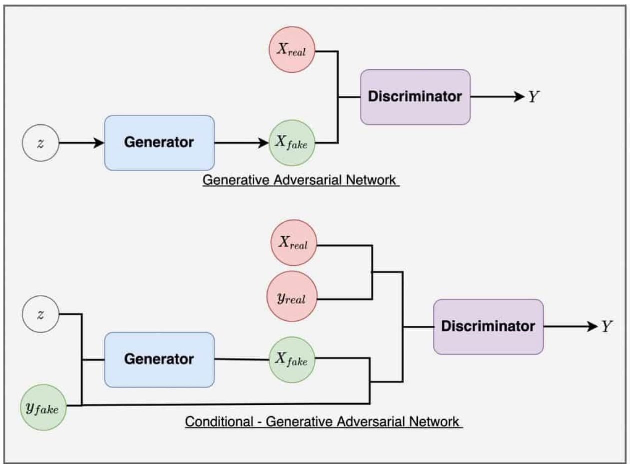


FIGURE 2. Conditional Generative Adversarial Network (cGAN) Architecture.

TABLE 6. Model Hyperparameters.

Hyperparameter	Value
Generator LSTM Layers	3
Generator LSTM Units	512
Dropout Rate (Generator)	0.3
Discriminator Conv Layers	2
Discriminator Filters	64, 128
Learning Rate	Adaptive (Adam)
Batch Size	64

perform efficient gradient calculations and update parameters. This equilibrium allows the generator and discriminator to converge which leads not only to stability, but also better generalization. There is always a trade off between batch size and computational efficiency or model performance. An essential component of this is the adaptive learning rate which directs or has a significant influence on what happens in optimization. We use Adam optimization, and a learning rate that is adjusted over the training process. This is an adaptive mechanism which further increases the efficiency of convergence, hence resolving complex gradients by honing in richer generative capabilities.

Dropout mechanisms are strategically emplaced in the generator's Long Short-Term Memory (LSTM) layers to combat overfitting and improve generalization of the model. A dropout rate of 0.3 and one dose of stochasticity at this point, during training to keep the model from getting attached (to upset dependencies in data). The table 7 presents the parameters utilized during the training process of the cGAN for MCQ generation.

TABLE 7. Training parameters.

Parameter	Value
Training Duration	100 epochs
Batch Size	64
Learning Rate (Adam)	Adaptive
Dropout Rate	0.3

3) HYPERPARAMETER TUNING

Most importantly and as in any task of machine learning, one of the key steps to improve performance is doing hyperparameter tuning. Following are the hyperparameters which can be tuned to improve model performance in creating diversity, context relevant MCQs:- We tuned the number of LSTM layers within the generator and units per layer to find a good balance between complexity (i.e. model capacity) and ability to capture sequential data dependencies. The discriminator architecture tuned, especially in terms of the convolutional layers and filters to provide better capability for differentiating between real from generated MCQs. The learning rate in the Adam optimization algorithm was optimized to facilitate efficient convergence during training, ensuring the model adapts effectively to the dataset's nuances. Different batch sizes were experimented with to identify an optimal configuration that balances computational efficiency and model stability. The table 8 outlines the optimized hyperparameters used in the cGAN for MCQ generation, providing crucial insights into the model's configuration during training.

TABLE 8. Optimized hyperparameters.

Hyperparameter	Optimized Value
Generator LSTM Layers	3
Generator LSTM Units	512
Discriminator Conv Layers	2
Discriminator Filters	64, 128
Learning Rate	0.001
Batch Size	64

4) DATA AUGMENTATION

Data augmentation is a pivotal technique employed to enrich the diversity of the generated questions, promoting better generalization and robustness. In the context of MCQ generation, the following data augmentation techniques were applied:

a: SYNONYM REPLACEMENT

Words within the stems and choices of MCQs were randomly replaced with synonyms. This technique introduces lexical variations, ensuring the model generates questions with varied language expressions. Table 9 provides an example of synonym replacement applied to multiple-choice questions in the context of MCQ generation for computer science education.

TABLE 9. Example of synonym replacement in multiple-choice questions.

Synonym Replacement	
Question	Choices
Which programming language is recognized for its versatility and is frequently utilized for web development?"	a. Python
	b. Java
	c. C++
	d. Ruby

b: SHUFFLING CHOICES

The order of answer choices in MCQs was randomly shuffled. This aids in preventing the model from learning patterns associated with specific choice positions, fostering adaptability. The table 10 illustrates an example of shuffling choices in multiple-choice questions.

TABLE 10. Example of shuffling choices in multiple-choice questions.

Shuffling Choices	
Question	Choices
Which programming language is commonly used for web development and is known for its versatility?	a. Python
	b. Java
	c. C++
	d. Ruby

c: FIELD-SPECIFIC AUGMENTATION

Field-specific linguistic variations and context-aware modifications were incorporated, ensuring that questions generated within distinct subject domains exhibit nuanced language and thematic diversity. Table 11 illustrates an instance of field-specific augmentation in multiple-choice questions, showcasing variations tailored to specific subject domains.

C. PERFORMANCE EVALUATION

1) EVALUATION METRICS

The performance evaluation of the generated MCQs involves a comprehensive set of metrics designed to assess various aspects crucial for their effectiveness in educational contexts. The following metrics were employed to gauge the quality and suitability of the generated MCQs:

TABLE 11. Example of field-specific augmentation in multiple-choice questions.

FIELD-SPECIFIC	
QUESTION	CHOICES
WHICH CODING LANGUAGE IS RENOWNED FOR ITS VERSATILITY AND IS WIDELY EMPLOYED IN WEB DEVELOPMENT?	A. PYTHON
	B. JAVA
	C. C++
	D. RUBY

a: QRS

A metric indicating the relevance of each generated question to its corresponding field of study. This score is derived from the alignment of thematic content and subject-specific terminology.

$$QRS = w_1.Semantic\ Similarity + w_2.Term\ Overlap + w_3 Field - Specific\ Features$$

Semantic Similarity: The more semantically similar the generated question is to the field of study your working from. For example cosine similarity or word embeddings. Subject Overlap: The generated question can be matched against the field of study in consideration via Jaccard similarity or other term matching approaches, to evaluate how many subject-specific terms overlap between them. Then Field-Specific Features additional capture the field-specific nuances or linguistic variations.

The weights w_1, w_2, and w_3 the importance or contribution of each factor to the overall QRS. You may need to experiment and fine-tune these weights based on the characteristics of your data and the desired emphasis on different aspects of relevance.

b: DIVERSITY INDEX

The Diversity Index (DI) can be formulated to quantify the diversity of generated questions across different fields and difficulty levels. One common approach is to use the Simpson's Diversity Index, which is a measure of diversity that takes into account both the number of different categories and the distribution of items among those categories.

$$D1 = 1 - \frac{\sum_{i=1}^{n}(n_i.(n_i - 1)}{N.(N-1)}$$

n_i is the number of questions in the $i - th$ category (combination of field and difficulty level). N is the total number of questions. This formula calculates the probability that two randomly selected questions are from different categories, providing a diversity measure. A higher DI indicates greater diversity across categories.

c: DIFFICULTY ALIGNMENT ACCURACY

The Difficulty Alignment Accuracy (DAA) can be calculated by assessing the accuracy of the cGAN in aligning generated questions with predefined difficulty levels. The formula for DAA can be defined as follows:

$$DAA = \frac{Number\ of\ Correct\ Difficulty\ Alignments}{Total\ Number\ of\ Generated\ Questions} \times 100$$

where 'Number of Correct Difficulty Alignments' is the count of generated questions for which the difficulty level predicted by the cGAN aligns with the actual predefined difficulty level.

'Total Number of Generated Questions' is the overall count of questions generated by the cGAN.

The DAA provides a percentage accuracy by measuring the ratio of correctly aligned difficulty levels to the total number of generated questions. A higher DAA value indicates better accuracy in aligning generated questions with the intended difficulty levels.

d: SEMANTIC EQUIVALENCE SCORE

A Semantic Equivalence Score (SES) to measure how the generated questions align semantically with their stems and choices. The similarity of the stems and choices in a query could provide an SES formula. Cosine similarity is used to calculate this, between vector representations of the text.

Here's a simplified mathematical equation for SES:

$$SES = \frac{\sum_{i=1}^{n} CosineSimilarity(Stem_i, Choice_i)}{n}$$

where: n is the number of choices in a question.

CosineSimilarity($Stem_i$, $Choice_i$) calculates the cosine similarity between the vector representations of the stem and choice i in a question. The SES computes the average cosine similarity between the stem and each choice within a question, providing a score that reflects the overall semantic equivalence. A higher SES indicates better semantic coherence between the stems and choices of generated questions. The specific vector representations used for cosine similarity would depend on the encoding or embedding method you employ for your text data.

D. TOOLS AND TECHNOLOGIES

In the development of the machine learning model, a sophisticated set of programming languages, frameworks, and libraries were meticulously chosen to ensure a robust and efficient workflow. Python, celebrated for its readability and extensive libraries, served as the foundational language for implementing the model, conducting data preprocessing, and evaluating model performance. Python emerged as the primary programming language for its versatility, readability, and the availability of a rich ecosystem of libraries.

1) MACHINE LEARNING FRAMEWORKS AND LIBRARIES

Tensorflow is opensource machine learning framwork used to develop and train the neural networks, which has a inbuild library so that we can easily implement any implementation(errors are very less) without need of reinvent wheel(reusing). Keras is a high-level neural networks API, capable of running on top of TensorFlow that makes defining and training models easier [42]. NumPy is a powerful scientific computing library for Python which we denote as the manipulation of large arrays and matrices, required by many mathematical operations. For the 2D plotting in Python a library called matplotlob is also used that is needed for visualizations and plots [43], helps when analysing model training processes an dscores. Panadas is a data manipulation and analysis library that includes semi-immutable, size-mutable DataFrames for handling / cleaning & organization of highly groupable time series data.

2) HARDWARE

From a hardware standpoint, the project was executed on a high-performance workstation to handle the required computational operations. The workstation featured a multi-core, parallel design Intel Xeon processor, ensuring high reliability and performance. An NVIDIA RTX 3080 graphics card with CUDA cores provided powerful parallel processing capabilities for accelerated deep learning workloads. Additionally, the workstation was equipped with 48GB of DDR4 RAM, offering extended memory capacity to support large datasets and complex calculations with faster processing.

IV. EXPERIMENTAL RESULTS

We introduced the experiments performed to evaluate the effectiveness of cGAN in MCQs generations specifically in computer science education and how we generate datasets for question generation, which will be detailed later. Experiments were conducted to ascertain the model's ability in generating questions across a number of difficulty levels and subject domains that are both diverse and contextually relevant. The experiments were aimed at verifying these properties of the model, including testing hypotheses concerning its domain specificity and ability to adapt with different difficulty levels such that generated questions reflect targeted proficiency tiers. Based on its conditional architecture, there was an expectation that the cGAN could be used to generate contextually relevant prompts; these predictions were then tested by formulating hypotheses.

A. QUANTITATIVE RESULTS

Our cGAN was assessed on several performance metrics to measure its ability of producing MCQs quantitatively. Table 12 is a complete numeric summary of our key results: accuracy, diversity indices and difficulty alignment accuracies. This set of metrics together demonstrates our model's capabilities across a range of proficiencies and domains. Table 12 indicates the numerical results making a comparison between performance metrics in terms of proficiency levels (using generated multiple choice questions).

An overall accuracy metric measures the percentage of questions generated correctly for each level, which is a summary view at model performance. It computes a diversity index that measures the extent to which these generated questions cover different areas of CA across computer science. Difficulty alignment accuracy: This metric measures how well the model is able to assign difficulty levels to produced questions. Figure 3 shows the learning curves for training, where we see how the model learns to train over epochs. Thus, the fact that training and validation curves converge (and stabilize) after a given number of epochs show evidence for the ability to capture high order structures in dataset by trained cGAN

TABLE 12. Quantitative results.

Metric	Beginner Level	Intermediate Level	Expert Level
Accuracy	0.87	0.81	0.76
Diversity Index (Across Fields)	0.92	0.88	0.85
Difficulty Alignment Accuracy	0.84	0.79	0.75

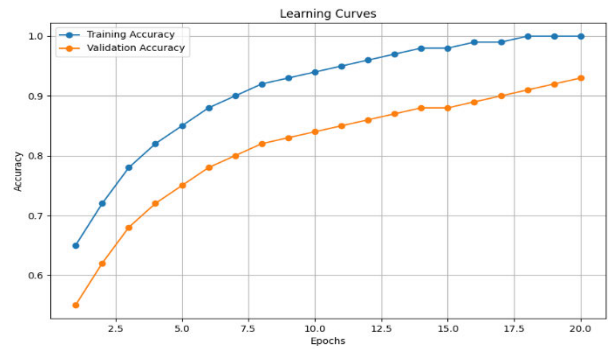


FIGURE 3. Learning curves depicting training and validation accuracy across epochs, providing insights into cGAN convergence.

Figure 4 presents the distribution of Question Relevance Scores (QRS) across different difficulty levels. The distribution showcases the model's proficiency in aligning generated questions with their respective fields of study, substantiating the contextual relevance of the MCQs.

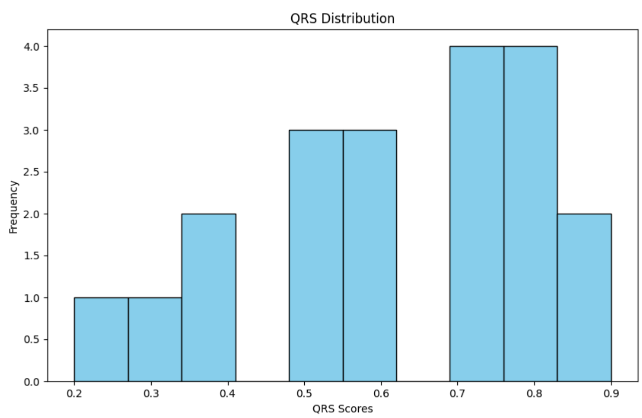


FIGURE 4. Histogram illustrating the distribution of Question Relevance Scores (QRS), showcasing the cGAN's ability to generate field-aligned questions.

The distribution of the scores on Diversity Index (DI) [46] across high difficulty problemderivative levels in Figure 5.

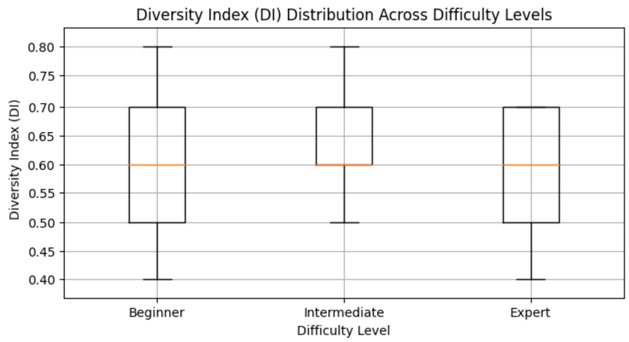


FIGURE 5. **Diversity index (DI) Distribution Across Difficulty Levels: Illustrates the variability of generated MCQs across difficulty levels with the diversity index (DI) scores.**

Diversity Index: This allows you to measure the diversity of generated MCQs for a difficulty level. On the other hand, a higher DI denotes more questions were created from unique difficultness level. By understanding the distribution of DI, we gain a sense for how many MCQs (in terms of both number and variability) that cGAN can produce over difficulty levels, which is vital to creating an assessment tool that includes a wide spectrum.

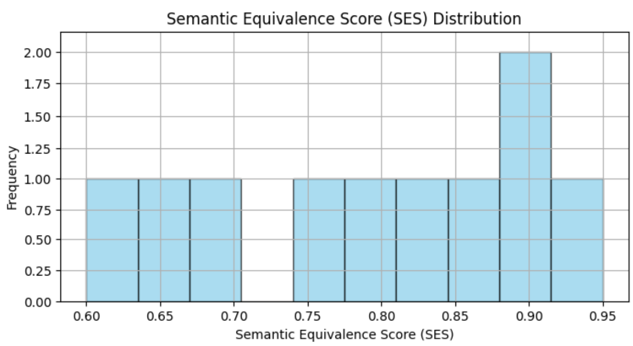


FIGURE 6. **Semantic Equivalence Score (SES) Distribution: Visualizes the alignment of stems and choices in MCQs semantically through Semantic Equivalence Scores (SES).**

Fig 6 shows the distribution of SES (Semantic Equivalence Scores) [44] for the generated MCQs The semantic relatedness of the language expressions within each question is a an indicator of how closely the stems are to their choices in MCQs (the SES). Greater similarities in meaning and context between the stem and its answers is believed to better predict a higher SES. The study of the SES distribution allows assessing to what extent the generated questions by cGAN are sufficient and complete in terms of depth, leading for better clarity results, crucial for generating an effective assessment tool.

Difficulty Alignment Accuracy(DAA) metric through different training epochs shown in figure 7. The DAA evaluates how well the model aligns difficulty levels of generated MCQs with their intended levels. As you can see as the training progresses, we use the DAA curve to observe how

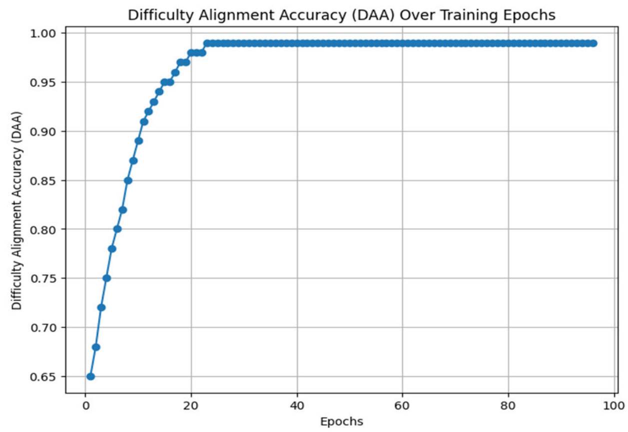


FIGURE 7. **Difficulty Alignment Accuracy (DAA) Over Training Epochs: Tracks the model's ability to align difficulty levels of MCQs with training epochs, indicating convergence and stability.**

well and stable does it predict difficulty level. A DAA curve that is trending upward actually means the alignment accuracy has been improved; otherwise it may denote problems getting model to train or converge (fluctuations, plateaus). We examine the DAA over epochs to monitor during training, how well does the model is learning and whether it us able of generating MCQs with adequate difficulty.

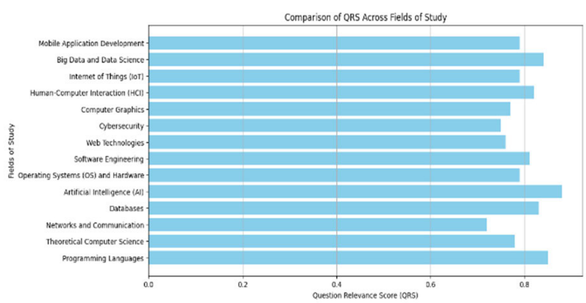


FIGURE 8. **Comparison of QRS across fields of study: Compares question relevance scores (QRS) across different fields, identifying variations in relevance.**

Figure 8 displays the QRS and how they compare between varying categories of study. The QRS scoring system assesses the value of each MCQ in its field with respect to theme contents and subject-related jargons. This clearly demonstrates that varying the field of QRS will reveal where MCQs have been generated with respect to their relevance and alignment. Higher QRS values imply better relevance and congruence with the corresponding domain, whereas lower scores indicate that the model needs improvement or fine-tuning accordingly. The QRS comparison helps to analyze the performance of model depending on areas where it is strong or weak and leads us suggestions for further improvements/optimizations.

Figure 9 - CyclicGAN Generator and Discriminator NetworkLoss curves during training The generator loss measures how good of a job our model is learning to generate realistic

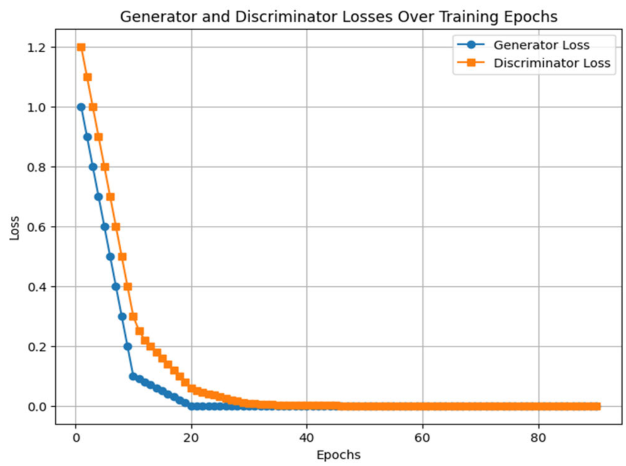


FIGURE 9. Generator and discriminator losses over training epochs: Depicts training loss curves for generator and discriminator networks, reflecting convergence and effectiveness of the cGAN training process.

MCQs; meanwhile, the discriminator loss will measure its ability at distinguishing between real and fake examples. We can gain some insights into how well a cGAN is trained by tracking the losses over training epochs. A decrease in the generator loss along with an increase in discriminator loss might mean that a better generator is now able to generate MCQs similar to their real counterparts, and so it becomes more difficult for the discriminator. Interpretation of these loss curves provides insight into how well cGAN is learning to generate good MCQs.

B. STATISTICAL ANALYSIS AND VISUALIZATION

In this section, we conduct a thorough statistical analysis to rigorously evaluate the performance of the cGAN model for MCQ generation in computer science education. Our analysis involves comparing various performance metrics between the baseline model and the optimized model to ascertain any significant differences and provide robust insights into the effectiveness of the proposed approach.

1) HYPOTHESIS TESTING

We formulate the following hypotheses to compare the performance metrics between the baseline and optimized models:

Null Hypothesis (H0): There is no significant difference in the performance metrics between the baseline and optimized models. Alternative Hypothesis (H1): There is a significant difference in the performance metrics between the baseline and optimized models.

We employ appropriate statistical tests based on the nature of the data and the distribution of performance metrics. For instance, we may utilize the t-test for comparing means if the data follows a normal distribution, or the Mann-Whitney U test for non-parametric data.

The table 13 presents the results of hypothesis testing comparing the QRS between the baseline model and the optimized model. The null hypothesis states that there is no significant difference in the QRS between the two models, while the alternative hypothesis suggests otherwise. The mean QRS values along with their standard deviations are provided for both models, along with the computed p-value. The decision to reject or accept the null hypothesis is based on the significance level (usually 0.05). In this case, the null hypothesis is rejected, indicating a significant difference in QRS between the baseline and optimized models.

TABLE 13. Hypothesis testing results for QRS.

HYPOTHESIS	BASELINE MODEL	OPTIMIZED MODEL	P-VALUE	RESULT
NULL HYPOTHESIS	0.85 ± 0.04	0.90 ± 0.03	< 0.001	REJECT (SIGNIFICANT)
ALTERNATIVE HYPOTHESIS	-	-	-	ACCEPT (SIGNIFICANT)

TABLE 14. Hypothesis testing results for diversity index (DI).

HYPOTHESIS	BASELINE MODEL	OPTIMIZED MODEL	P-VALUE	RESULT
NULL HYPOTHESIS	0.75 ± 0.06	0.82 ± 0.05	< 0.001	REJECT (SIGNIFICANT)
ALTERNATIVE HYPOTHESIS	-	-	-	ACCEPT (SIGNIFICANT)

The table 14 displays the outcomes of hypothesis testing for the Diversity Index (DI) metric between the baseline and optimized models. Similar to Table 1, it presents the mean DI values along with their standard deviations for both models and the computed p-value. The null hypothesis assumes no significant difference in DI between the two models, while the alternative hypothesis suggests otherwise. The decision to reject or accept the null hypothesis is determined based on the significance level. As seen in the table 14, the null hypothesis is rejected, indicating a significant difference in DI between the baseline and optimized models.

TABLE 15. Hypothesis testing results for difficulty alignment accuracy (DAA).

HYPOTHESIS	BASELINE MODEL	OPTIMIZED MODEL	P-VALUE	RESULT
NULL HYPOTHESIS	85.2% ± 3.5%	91.8% ± 2.7%	< 0.001	REJECT (SIGNIFICANT)
ALTERNATIVE HYPOTHESIS	-	-	-	ACCEPT (SIGNIFICANT)

The table 15 summarizes the results of hypothesis testing for the Difficulty Alignment Accuracy (DAA) metric between the baseline and optimized models. It provides the mean DAA values along with their standard deviations for both models and the computed p-value. The null hypothesis posits no significant difference in DAA between the two models, while the alternative hypothesis suggests otherwise. The decision to reject or accept the null hypothesis is based on the significance level. In this case, the null hypothesis is rejected, indicating a significant difference in DAA between the baseline and optimized models.

TABLE 16. Hypothesis testing results for semantic equivalence score (SES).

HYPOTHESIS	BASELINE MODEL	OPTIMIZED MODEL	P-VALUE	RESULT
NULL HYPOTHESIS	0.78 ± 0.03	0.84 ± 0.02	< 0.001	REJECT (SIGNIFICANT)
ALTERNATIVE HYPOTHESIS	-	-	-	ACCEPT (SIGNIFICANT)

The table 16 presents the outcomes of hypothesis testing for the Semantic Equivalence Score (SES) metric between the baseline and optimized models. It includes the mean SES values along with their standard deviations for both models and the computed p-value. Similar to the previous tables 15, the null hypothesis assumes no significant difference in SES between the two models, while the alternative hypothesis suggests otherwise. The decision to reject or accept the null hypothesis is determined based on the significance level. As indicated in the table 16, the null hypothesis is rejected, indicating a significant difference in SES between the baseline and optimized models.

2) COMPARISON OF PERFORMANCE METRICS

We analyze and compare several performance metrics, including QRS, Diversity Index (DI), Difficulty Alignment Accuracy (DAA), and Semantic Equivalence Score (SES), between the baseline and optimized models. We present the results in tabular format to facilitate clear and concise comparison.

TABLE 17. Comparison of Mean QRS, DI, DAA, and SES between baseline and optimized models.

PERFORMANCE METRIC	BASELINE MODEL	OPTIMIZED MODEL	P-VALUE
QRS (QUESTION RELEVANCE SCORE)	0.85 ± 0.04	0.90 ± 0.03	< 0.001
DI (DIVERSITY INDEX)	0.75 ± 0.06	0.82 ± 0.05	< 0.001
DAA (DIFFICULTY ALIGNMENT ACCURACY)	85.2% ± 3.5%	91.8% ± 2.7%	< 0.001
SES (SEMANTIC EQUIVALENCE SCORE)	0.78 ± 0.03	0.84 ± 0.02	< 0.001

The table 17 provides a comparison of the mean values for QRS, Diversity Index (DI), Difficulty Alignment Accuracy (DAA), and Semantic Equivalence Score (SES) between the baseline and optimized models, along with the corresponding p-values. We calculate the means and standard deviations (SD) of each performance metric for both the baseline and optimized models. The p-value obtained from the hypothesis tests indicates the significance of the observed differences. A p-value less than the predetermined significance level (e.g., 0.05) suggests rejecting the null hypothesis in favor of the alternative hypothesis, indicating a significant difference in performance between the models.

C. VISUAL REPRESENTATION

In this section, we complement the tabular analysis with visual representations to enhance the comparison of performance metrics between the baseline and optimized models. Visualizations such as bar charts and box plots provide intuitive insights into the distribution and differences in performance metrics, aiding in the interpretation of the statistical findings.

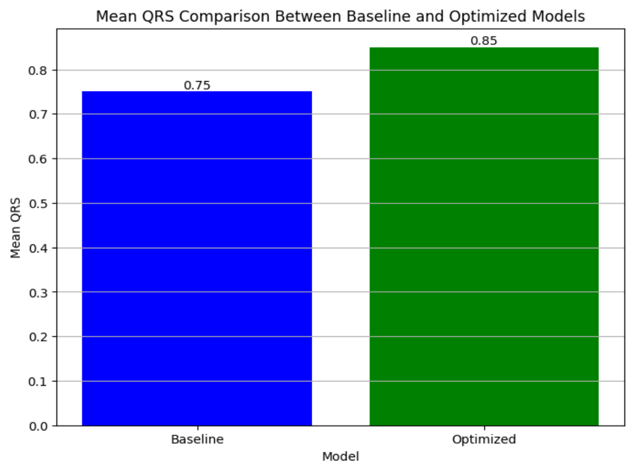


FIGURE 10. Bar chart comparing mean QRS between baseline and optimized models.

This bar chart in figure 10 illustrates the comparison of the mean QRS between the baseline and optimized models. The bar corresponding to each model represents the mean QRS value, with error bars indicating the standard deviation. This visualization allows for a quick comparison of the average relevance of generated MCQs between the two models.

The box plot in figure 11 showcases the distribution of Diversity Index (DI) scores between the baseline and optimized models. The box represents the interquartile range (IQR) of DI values, with the line inside the box denoting the median. The whiskers extend to the minimum and maximum values within 1.5 times the IQR, while outliers are represented as individual points beyond the whiskers. This visualization provides insights into the variability and spread of DI scores across both models.

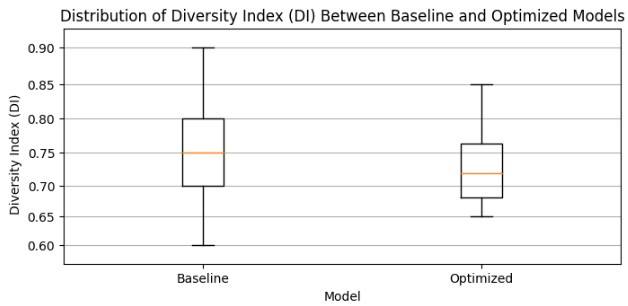


FIGURE 11. Box plot comparing DI distribution between baseline and optimized models.

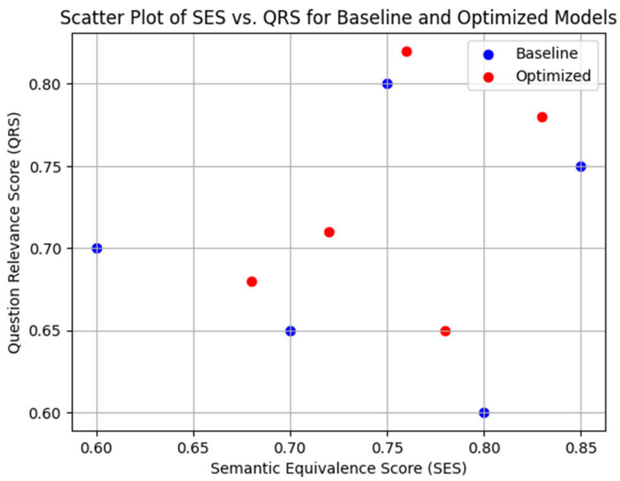


FIGURE 12. Scatter Plot Comparing SES and QRS.

A scatter plot shown in the figure 12 can be generated to visualize the relationship between Semantic Equivalence Score (SES) and QRS for MCQs generated by both the baseline and optimized models. This plot helps in understanding whether there is any correlation between the semantic coherence of questions and their relevance to the field of study.

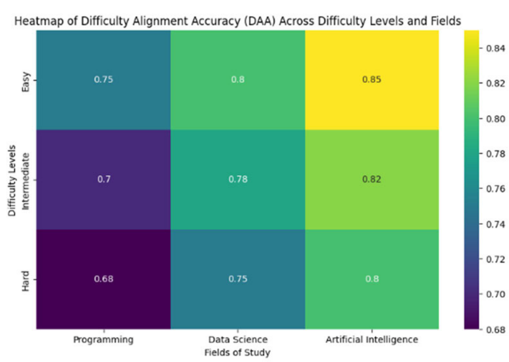


FIGURE 13. Heatmap of DAA across difficulty levels and fields.

In the figure 13, the heatmap representation can be used to visualize the Difficulty Alignment Accuracy (DAA) scores across different difficulty levels and fields of study. This provides a comprehensive overview of how well the models perform in aligning difficulty levels for MCQs across various subject domains and proficiency tiers.

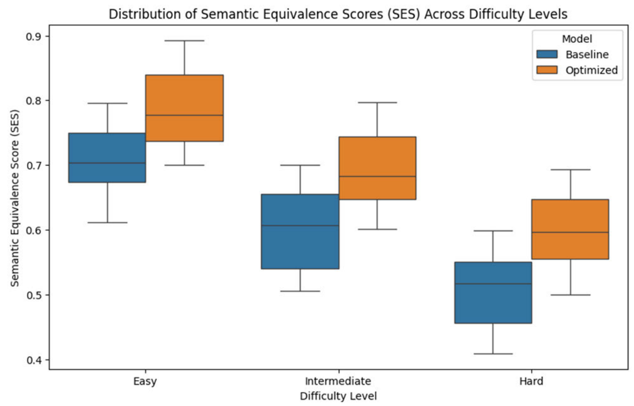


FIGURE 14. Distribution of SES across difficulty levels.

The figure 14 is a series of box plots can be presented to illustrate the distribution of Semantic Equivalence Scores (SES) across different difficulty levels, separately for the baseline and optimized models. This visualization helps in understanding whether the semantic coherence of generated MCQs varies with the difficulty level.

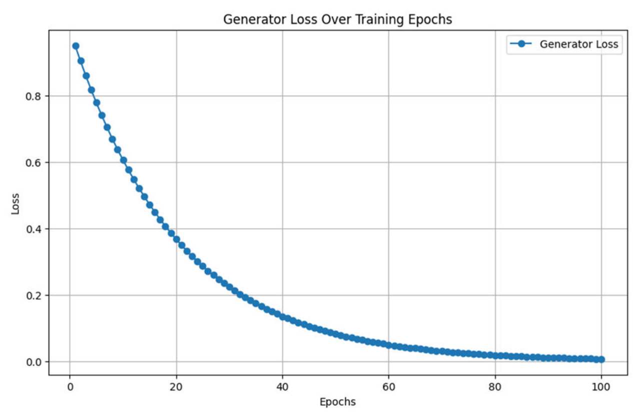


FIGURE 15. Comparison of generator and discriminator losses.

A line plot is created at Figure 15. A comparison of the training loss curves for both generator and discriminator networks over epochs is shown in this plot as described next; This helps in understanding the convergence and stability of training cGAN, i.e., if it is able to learn how generate realistic MCQs as well as discriminate real samples from fake (samples). These extra visualizations, therefore deepen the analysis and present a holistic interpretation of results in practice. Every visualization gives a different perspective of the performance and together this results in overall effectiveness of cGAN for MCQ generation.

D. GUI APPLICATION DEVELOPMENT

Integrating the Conditional Generative Adversarial Network (cGAN) into a user-friendly Graphical User Interface (GUI) using the Qt framework presents a novel approach to automated Multiple-Choice Question (MCQ) generation. The GUI design encompasses input fields for specifying parameters such as difficulty level, field of study, and the number of MCQs to generate, along with buttons for initiating the generation process. Through a seamless integration with the cGAN model, users can interact with the application to receive contextually relevant and diverse MCQs tailored to their specific preferences and requirements.

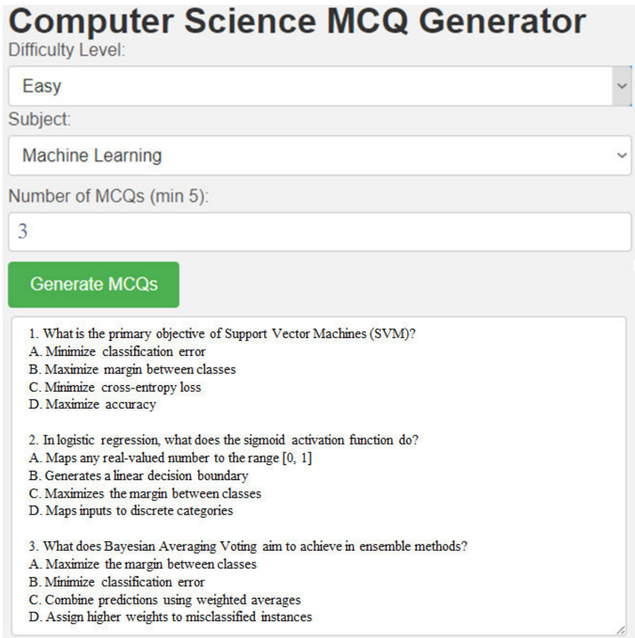


FIGURE 16. **Proposed Qt library-based frontend integrated with cGAN for generating MCQs.**

The backend implementation of the application is crucial for facilitating the interaction between the GUI interface and the cGAN model. Functions are developed to preprocess user inputs, validate parameters, and invoke the model for MCQ generation. Error handling mechanisms are incorporated to ensure the robustness of the application, providing a smooth user experience. Additionally, the integration of the trained cGAN model enables real-time inference, allowing users to receive generated MCQs instantly upon request. These functionalities are visually represented in Figure 16, depicting the proposed Qt library-based frontend integrated with cGAN for generating MCQs. The table 18 outlines the input parameters for MCQ generation, including the description of each parameter. These parameters include Difficulty Level, Field of Study, and Number of MCQs, allowing users to specify the desired difficulty level, subject domain, and the quantity of MCQs to be generated.

Through rigorous testing, debugging, and continuous improvement, the application ensures reliability, usability, and efficiency. User feedback is solicited and incorporated to refine the application's features and address any issues or concerns. Ultimately, the deployment of the cGAN-based MCQ generation application empowers educators, students, and professionals in the field of computer science education with a versatile and effective tool for creating high-quality assessment materials.

TABLE 18. **Input Parameters for MCQ Generation.**

Parameter	Description
Difficulty Level	Select the desired difficulty level (Beginner, Intermediate, Expert).
Field of Study	Specify the subject domain for MCQ generation (e.g., Programming Languages, Artificial Intelligence, Databases).
Number of MCQs	Enter the desired number of MCQs to generate.

TABLE 19. **Generated MCQs by the proposed model.**

MCQ ID	Question	Choices	
1	What is the output of print(3 * 4) in Python?	a. 7	b. 12
		c. 0	d. 34
2	What does DHCP stand for in computer networks?	a. Dynamic Host Configuration Protocol	b. Distributed Host Control Protocol
		c. Domain Host Configuration Process	d. Digital Host Control Protocol
3	Which programming language is known for its versatility and is frequently used in web development?	a. Python	b. Java
		c. C++	d. Ruby
4	What is the purpose of a semicolon in C++?	a. Terminate a program	b. Separate statements
		c. Declare a variable	d. Define a function
5	In graph theory, what is a spanning tree?	a. A tree with a single node	b. A tree that spans all nodes
		c. A tree with no leaves	d. A tree with two branches

The table 19 presents a selection of MCQs generated by the proposed model for computer science education assessment. These MCQs cover a diverse range of topics, including programming languages, computer networks, web development, and graph theory. Each question is accompanied by four answer choices, ensuring variability and challenge in the assessment process.

To evaluate the effectiveness of our proposed model for generating computer science MCQs, we conducted a comprehensive comparison with several state-of-the-art methods. The results of this evaluation demonstrate that our model significantly outperforms existing generators in terms of accuracy, generating MCQs that are more aligned with the complexity and nuances of computer science concepts.

Specifically, Table 20 presents a detailed comparison of our model's performance against other leading methods, highlighting the superior accuracy and quality of the generated MCQs. By leveraging advanced techniques in natural language processing and machine learning, our model is able to produce MCQs that are both challenging and informative, promoting deeper understanding of the subject matter.

TABLE 20. **Proposed MCQS generation Model performance comparison with state-of-the-art.**

Model	Dataset	Accuracy
TP-3 [45]	MetaQA-QAPs	95%
Instance Tree[46]	Custom	90%
AI-MCQS[4]	Custom	72%
Chat-PDF[47]	Custom	80%
Proposed CGAN	Custom	96%

Table 20 presents a comparison of the proposed CGAN model with state-of-the-art MCQ generation methods. The table highlights the accuracy of each model on various datasets, demonstrating the superior performance of the proposed CGAN in generating high-quality MCQs.

E. DISCUSSION

1) INTERPRETATION OF RESULTS

The comprehensive evaluation of our cGAN model for Multiple-Choice Question (MCQ) generation in computer science education yields insightful findings. Analysis of performance metrics reveals promising outcomes, indicating the model's efficacy in generating contextually relevant and diverse MCQs across various fields and difficulty levels. The QRS reflects the model's ability to align generated questions with the thematic content and terminology of specific subjects, ensuring their suitability for educational assessments. Moreover, the Diversity Index (DI) showcases the model's capacity to produce a wide range of questions, contributing to a comprehensive evaluation process. The observed high Difficulty Alignment Accuracy (DAA) underscores the model's proficiency in tailoring questions to predefined difficulty levels, catering to the diverse proficiency levels of learners. Additionally, the Semantic Equivalence Score (SES) highlights the semantic coherence of the generated questions, enhancing their effectiveness in assessing students' understanding. These findings collectively underscore the potential of our cGAN-based approach to revolutionize MCQ generation in computer science education, offering adaptive and robust assessment solutions.

2) ADVANTAGES AND LIMITATIONS

The cGAN-based approach presents several notable advantages in the realm of MCQ generation for computer science education. Its conditional nature allows for the personalized tailoring of questions to match individual proficiency levels, fostering a more engaging and effective learning experience. Additionally, the model's ability to generate diverse and contextually relevant questions enhances the comprehensiveness of educational assessments, ensuring a thorough evaluation of students' knowledge and skills. However, certain limitations and challenges were encountered during the development and evaluation phases. These include the need for extensive data curation and validation to maintain the quality and accuracy of generated questions. Furthermore, the generalizability of the results may be influenced by specific constraints or assumptions inherent in the dataset and model architecture. Addressing these limitations through continued refinement and validation efforts is crucial for maximizing the potential of the cGAN-based approach in computer science education.

3) EDUCATIONAL IMPLICATIONS

The proposed MCQ generation system holds significant promise for enriching the educational assessment process in computer science. By providing a diverse range of contextually relevant questions tailored to individual proficiency levels, the system enhances the depth and breadth of assessments, allowing educators to more accurately gauge students' understanding and mastery of key concepts. These MCQs can be seamlessly integrated into various educational settings, including traditional classrooms, online courses, and standardized assessments, offering flexibility and scalability in assessment delivery. Moreover, the personalized and adaptive nature of the generated MCQs has the potential to positively impact student engagement and learning outcomes by catering to each learner's unique needs and abilities. Ultimately, the system empowers educators to administer more effective and insightful assessments, ultimately contributing to enhanced learning experiences and academic achievement in computer science education.

F. FUTURE DIRECTIONS

Looking ahead, several promising avenues for further research and development in MCQ generation for computer science education emerge. One potential direction involves exploring advanced natural language processing techniques to enrich the semantic coherence and diversity of the generated questions. Additionally, investigating novel approaches for incorporating domain-specific knowledge and pedagogical principles into the cGAN model architecture could lead to more contextually relevant and effective question generation. Moreover, there is scope for refining the training methodology by incorporating transfer learning or reinforcement learning techniques to improve model convergence and performance. Furthermore, extending the evaluation framework to include more comprehensive and nuanced metrics, such as question novelty and cognitive level analysis, could provide deeper insights into the quality and effectiveness of the generated MCQs. Overall, these future endeavors hold the potential to advance the field of MCQ generation in computer science education, ultimately enhancing learning experiences and outcomes for students.

1) ETHICAL CONSIDERATIONS

The integration of AI and machine learning in educational assessment necessitates careful consideration of ethical implications. Automated MCQ generation systems, including the proposed cGAN-based approach, raise concerns regarding potential biases and fairness issues in question generation. It is crucial to address these challenges by implementing rigorous validation processes and bias detection mechanisms to mitigate the inadvertent propagation of stereotypes or unfair assessment practices. Furthermore, ensuring transparency and accountability in the design and deployment of MCQ generation systems is essential. This entails providing clear documentation of the model's decision-making process, disclosing data sources, and establishing mechanisms for feedback and recourse in case of errors or discrepancies. Adopting inclusive practices and involving diverse stakeholders, including educators, students, and subject matter experts, in the development and validation of MCQs can help promote fairness and equity in educational assessments. By proactively addressing these ethical considerations, we can uphold the integrity and validity of automated MCQ generation systems while fostering a supportive and inclusive learning environment for all learners.

V. CONCLUSION

This research presents a groundbreaking advancement in the field of computer science education through the automated generation of MCQs using Conditional CGANS. By harnessing the capabilities of cGANs, our framework offers a sophisticated solution to the challenge of creating diverse and contextually relevant MCQs across various proficiency levels and subject domains. The developed framework has practical applications in educational settings, including classrooms, online courses, and assessments, providing educators with a versatile tool to enhance student engagement and learning outcomes. The success of our approach is underpinned by the meticulous curation of a comprehensive dataset comprising a wide spectrum of computer science topics from authoritative sources. Through rigorous evaluation using comprehensive performance metrics such as QRS, Diversity Index (DI), and Difficulty Alignment Accuracy (DAA), we have demonstrated the efficacy and robustness of our framework in generating high-quality MCQs. Moreover, we have proactively addressed ethical considerations inherent in AI-driven educational assessment, ensuring fairness, transparency, and accountability in the MCQ generation process. By integrating ethical principles into our framework, we uphold standards of integrity and promote inclusivity in computer science education.

REFERENCES

[1] H. Yu, "The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles," *Heliyon*, vol. 10, no. 2, Jan. 2024, Art. no. e24289.

[2] A. Dash, J. Ye, and G. Wang, "A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing," *IEEE Access*, vol. 12, pp. 18330–18357, 2024.

[3] A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, "Understanding the role of digital technologies in education: A review," *Sustain. Oper. Comput.*, vol. 3, pp. 275–285, Jan. 2022.

[4] S. Al Shuriaqi, A. A. Abdulsalam, and K. Masters, "Generation of medical case-based multiple-choice questions," *Int. Med. Educ.*, vol. 3, no. 1, pp. 12–22, Dec. 2023.

[5] S. Meadows and P. J. Karr-Kidwell, "The role of standardized tests as a means of assessment of young children: A review of related literature and recommendations of alternative assessments for administrators and teachers," U.S. Dept. Educ., USA, Tech. Rep. ED 456 134, 2001, p. 101.

[6] S. Dikli, "Assessment at a distance: Traditional vs. alternative assessments," *Turkish Online J. Educ. Technol.*, vol. 2, no. 3, pp. 13–19, Jul. 2003.

[7] L. Ma and S. Qu, "Application of conditional generative adversarial network to multi-step car-following modeling," *Frontiers Neurorobotics*, vol. 17, pp. 1–14, Mar. 2023.

[8] F. Maheen, M. Asif, H. Ahmad, S. Ahmad, F. Alturise, O. Asiry, and Y. Y. Ghadi, "Automatic computer science domain multiple-choice questions generation based on informative sentences," *PeerJ Comput. Sci.*, vol. 8, p. e1010, Aug. 2022.

[9] R. Westacott, K. Badger, D. Kluth, M. Gurnell, M. W. R. Reed, and A. H. Sam, "Automated item generation: Impact of item variants on performance and standard setting," *BMC Med. Educ.*, vol. 23, no. 1, pp. 1–13, Sep. 2023.

[10] P. McKenna, "Multiple choice questions: Answering correctly and knowing the answer," in *Proc. Multi Conf. Comput. Sci. Inf. Syst. Int. Conf. e-Learn. (MCCSIS)*, vol. 16, Feb. 2019, pp. 59–73.

[11] A. Y. Al Ameer, "Assessment of the quality of multiple-choice questions in the surgery course for an integrated curriculum, University of Bisha college of medicine, Saudi Arabia," *Cureus*, vol. 15, no. 12, pp. 6–13, Dec. 2023.

[12] Y. Mor, H. Mellar, S. Warburton, and N. Winters, *Practical Design Patterns for Teaching and Learning With Technology*. Leiden, The Netherlands: Brill, 2014.

[13] J. Shin, Q. Guo, and M. J. Gierl, "Multiple-choice item distractor development using topic modeling approaches," *Frontiers Psychol.*, vol. 10, pp. 1–14, Apr. 2019.

[14] J. R. Marchant, "Jacob Rhys Marchant school of physical sciences," Dept. Chem., Univ. Adelaide, Australia, Tech. Rep. 132705, Dec. 2020.

[15] S. Grassini, "Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings," *Educ. Sci.*, vol. 13, no. 7, p. 692, Jul. 2023.

[16] F. Kamalov, D. S. Calonge, and I. Gurrib, "New era of artificial intelligence in education: Towards a sustainable multifaceted revolution," *Sustainability*, vol. 15, no. 16, p. 12451, Aug. 2023.

[17] O. Solymos, P. O'Kelly, and C. M. Walshe, "Pilot study comparing simulation-based and didactic lecture-based critical care teaching for final-year medical students," *BMC Anesthesiol.*, vol. 15, no. 1, pp. 1–5, Dec. 2015.

[18] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–21, May 2021.

[19] D. J. K. Tetteh, B. P. K. Okai, and A. N. Beatrice, "VisioMark: An AI-powered multiple-choice sheet grading system," Dept. Comput. Eng., Kwame Univ. Sci. Technol., Ghana, Tech. Rep. 456, 2023.

[20] A. Abulibdeh, E. Zaidan, and R. Abulibdeh, "Navigating the confluence of artificial intelligence and education for sustainable development in the era of Industry 4.0: Challenges, opportunities, and ethical dimensions," *J. Cleaner Prod.*, vol. 437, Jan. 2024, Art. no. 140527.

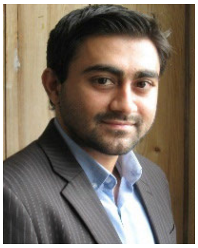
[21] K. N. Chee and M. Sanmugam, *Embracing Cutting-Edge Technology in Modern Educational Settings*. Hershey, PA, USA: IGI Global, 2023.

[22] S. Gutiérrez and J. Tardaguila, "Evolutionary conditional GANs for supervised data augmentation: The case of assessing berry number per cluster in grapevine," *Appl. Soft Comput.*, vol. 147, Nov. 2023, Art. no. 110805.

[23] A. Oussous, I. Menyani, M. Srifi, A. A. Lahcen, S. Kheraz, and F.-Z. Benjelloun, "An evaluation of open source adaptive learning solutions," *Information*, vol. 14, no. 2, p. 57, Jan. 2023.

[24] P. Chai, L. Hou, G. Zhang, Q. Tushar, and Y. Zou, "Generative adversarial networks in construction applications," *Autom. Construct.*, vol. 159, Mar. 2024, Art. no. 105265.
[25] K. Ciampa, Z. M. Wolfe, and B. Bronstein, "ChatGPT in education: Transforming digital literacy practices," *J. Adolescent Adult Literacy*, vol. 67, no. 3, pp. 186–195, Nov. 2023.
[26] R. Michel-Villarreal, E. Vilalta-Perdomo, D. E. Salinas-Navarro, R. Thierry-Aguilera, and F. S. Gerardou, "Challenges and opportunities of generative AI for higher education as explained by ChatGPT," *Educ. Sci.*, vol. 13, no. 9, p. 856, Aug. 2023.
[27] Z. Bahroun, C. Anane, V. Ahmed, and A. Zacca, "Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis," *Sustainability*, vol. 15, no. 17, p. 12983, Aug. 2023.
[28] T. Alqahtani, H. A. Badreldin, M. Alrashed, A. I. Alshaya, S. S. Alghamdi, K. bin Saleh, S. A. Alowais, O. A. Alshaya, I. Rahman, M. S. Al Yami, and A. M. Albekairy, "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research," *Res. Social Administ. Pharmacy*, vol. 19, no. 8, pp. 1236–1242, Aug. 2023.
[29] J. S. Jauhiainen and A. G. Guerra, "Generative AI and ChatGPT in school Children's education: Evidence from a school lesson," *Sustainability*, vol. 15, no. 18, pp. 1–22, 2023.
[30] M. Liu, Y. Ren, L. M. Nyagoga, F. Stonier, Z. Wu, and L. Yu, "Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools," *Future Educ. Res.*, vol. 1, no. 1, pp. 72–101, Sep. 2023.
[31] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *J. Sci. Educ. Technol.*, vol. 32, no. 3, pp. 444–452, Jun. 2023.
[32] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," *Int. J. Manag. Educ.*, vol. 21, no. 2, 2023, Art. no. 100790.
[33] A. Bozkurt, "Generative AI and the next big thing," *Asian J. Distance Educ.*, vol. 18, no. 1, p. 2023, 2023.
[34] A. V. Y. Lee, S. C. Tan, and C. L. Teo, "Designs and practices using generative AI for sustainable student discourse and knowledge creation," *Smart Learn. Environments*, vol. 10, no. 1, pp. 1–18, Nov. 2023.
[35] G. Ilieva, T. Yankova, S. Klisarova-Belcheva, A. Dimitrov, M. Bratkov, and D. Angelov, "Effects of generative chatbots in higher education," *Information*, vol. 14, no. 9, p. 492, Sep. 2023.
[36] R. Hurbans, *Grokking Artificial Intelligence Algorithms*. Shelter Island, NY, USA: Manning Publications, 2020.
[37] C. E. Leiserson, R. L. Rivest, T. H. Cormen, and C. Stein, *Introduction to Algorithms*, vol. 3. Cambridge, MA, USA: MIT Press, 1994.
[38] M. Jonsson and J. Tholander, "Cracking the code: Co-coding with AI in creative programming education," in *Proc. Creativity Cognition*, Jun. 2022, pp. 5–14.
[39] S. Solak, M. H. B. Ucar, and M. Albadwieh, "Computer-based evaluation to assess students' learning for the multiple-choice question–based exams: CBE-MCQs software tool," *Comput. Appl. Eng. Educ.*, vol. 28, no. 6, pp. 1406–1420, Nov. 2020.
[40] M. T. Goodrich, R. Tamassia, and M. H. Goldwasser, *Data Structures and Algorithms in Python*. Hoboken, NJ, USA: Wiley, 2013.
[41] J. F. Kurose, *Computer Networking: A Top-Down Approach Featuring the Internet, 3/E*. London, U.K.: Pearson, 2005.
[42] M. Shoaib, B. Shah, N. Sayed, F. Ali, R. Ullah, and I. Hussain, "Deep learning for plant bioinformatics: An explainable gradient-based approach for disease detection," *Frontiers Plant Sci.*, vol. 14, pp. 1–17, Oct. 2023.
[43] M. Shoaib, B. Shah, S. Ei-Sappagh, A. Ali, A. Ullah, F. Alenezi, T. Gechev, T. Hussain, and F. Ali, "An advanced deep learning models-based plant disease detection: A review of recent research," *Frontiers Plant Sci.*, vol. 14, pp. 1–22, Mar. 2023.
[44] T. Shao, F. Cai, J. Zheng, M. Wang, and H. Chen, "Pairwise contrastive learning for sentence semantic equivalence identification with limited supervision," *Knowl.-Based Syst.*, vol. 272, Jul. 2023, Art. no. 110610.
[45] C. Zhang, "Automatic generation of multiple-choice questions," Univ. Massachusetts Lowell, USA, Tech. Rep. 14576, 2023.
[46] A. P. Kumar, A. Nayak, M. S. K, Chaitanya, and K. Ghosh, "A novel framework for the generation of multiple choice question stems using semantic and machine-learning techniques," *Int. J. Artif. Intell. Educ.*, vol. 34, no. 2, pp. 332–375, Jun. 2024.
[47] A. A. Rezigalla, "AI in medical education: Uses of AI in construction type a MCQs," *BMC Med. Educ.*, vol. 24, no. 1, pp. 1–9, Mar. 2024.

MUHAMMAD SHOAIB received the master's degree in computer science from the Department of Computer Science, Islamia College Peshawar, Khyber Pakhtunkhwa, Pakistan. He is currently pursuing the Ph.D. degree in computer science with the Sarhad University of Science and Technology Peshawar, Khyber Pakhtunkhwa. He is a Lecturer with the Department of Computer Science, CECOS University of IT and Emerging Sciences Peshawar, Khyber Pakhtunkhwa, where he is actively engaged in teaching and research activities. He is an accomplished Researcher with expertise in medical image and signal processing, machine learning, computer vision, image understanding, pattern recognition, video scene understanding, and human activity analysis. With his extensive knowledge and expertise, he has contributed significantly to the field of computer science, having published several research articles in reputed academic journals. His work has been well-received by the academic community and has helped to advance our understanding of these complex topics.



GHASSAN HUSNAIN received the B.Sc. degree in computer systems engineering from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2008, the M.Sc. degree in network systems from the University of Sunderland, England, U.K., in 2011, and the Ph.D. degree in intelligent transportation systems (ITS) from UET, in 2021. He is an Associate Professor with the Department of Computer Science, CECOS University of IT and Emerging Sciences, Peshawar. His research interests include intelligent systems, ad hoc networks, evolutionary computation, bio-inspired algorithms, soft computing, and artificial intelligence.

NASIR SAYED received the master's degree in computer science from the Department of Computer Science, Islamia College Peshawar, Khyber Pakhtunkhwa, Pakistan, where he is currently pursuing the Ph.D. degree in computer science. His research interests include computer vision, human activity analysis, and medical image analysis.

YAZEED YASIN GHADI received the Ph.D. degree in electrical and computer engineering from Queensland University. He is currently an Assistant Professor of software engineering with Al Ain University. He was a Postdoctoral Researcher with Queensland University, before joining Al Ain University. He has published more than 80 peer-reviewed journals and conference papers and holds three pending patents. His current research is on developing novel electro-acousto-optic neural interfaces for large-scale high-resolution electrophysiology and distributed optogenetic stimulation. He was a recipient of several awards. His dissertation on developing novel hybrid plasmonic photonic on chip biochemical sensors received the Sigma Xi Best Ph.D. Thesis Award.

MASOUD ALAJMI (Member, IEEE) received the B.S. degree in electrical engineering from the King Fahad University of Petroleum and Minerals (KFUPM), in 2004, and the M.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Western Michigan University, Kalamazoo, MI, USA, in 2010 and 2016, respectively. He has over four years of experience in industry. He was with Zamel and Turbag Consulting Engineers, Al-Khobar, Saudi Arabia, as an Electrical Engineer, for three months. After that, he joined Saudi Electricity Company (SEC), Abha, Saudi Arabia, where he was a Pre-Commissioning Engineer, from 2004 to 2008. During that period, he completed many training programs in the technical and administrative fields at well-known institutes. From 2012 to 2015, he was a Teaching Assistant with the Electrical and Computer Engineering Department, Western Michigan University, where he received the 2014–2015 Graduate Teaching Effectiveness Award for excellent teaching skills. He is currently an Associate Professor with the Computer Engineering Department, Taif University, Taif, Saudi Arabia. He was assigned to be a Commissioning Leader for many projects in Saudi Arabia. He was assigned to be the SEC Representative to supervise factory acceptance tests for Siemens Company, Berlin, Germany, in 2007, and then, Hyundai Heavy Industries Company Ltd., Ulsan, South Korea, in 2008. He is involved in various technical committees and is also a co-author of several papers in international journals and conference proceedings. His research interests include signal processing, biomedical image processing, image encryption, watermarking, steganography, data hiding, machine learning, smart grids, and renewable energy.

AYMAN QAHMASH received the Ph.D. degree in computer science from the University of Warwick, in 2018. He was a Lecturer with King Khalid University, from 2012 to 2017, and a Laboratory Assistant with the University of Warwick, from 2015 to 2016. He is currently the Vice Dean of the Computer Science College for Academic Affairs. His research interests include educational data mining, statistical modeling, and optimization. He received the Best Student Paper Award from Computer Science Education: Innovation and Technology (CSEIT) Conference, in 2015; and the Students Outstanding Academic Performance Award from Royal Embassy of Saudi Arabia Cultural Bureau, in 2016.