Empowering Education with AI: Automating Content Generation through Large Language Models

Andrew Emanuel Attard* and Alexiei Dingli

Department of Artificial Intelligence, Faculty of ICT, University of Malta, Msida, Malta
Email: andrew.attard.18@um.edu.mt (A.E.A.); alexiei.dingli@um.edu.mt (A.D.)
*Corresponding author

Manuscript received February 10, 2025; revised March 4, 2025; accepted May 29, 2025; published September 18, 2025

***Abstract*—Advances in educational technology are reshaping learning, with Intelligent Tutoring Systems (ITS) offering personalized education. However, creating high-quality, adaptive content remains a significant challenge for educators, requiring substantial time and effort. This research explores the potential of Large Language Models (LLMs), such as GPT-4, to automate content generation, addressing these challenges and enhancing educational efficiency. LLMs, leveraging deep learning and transformer architectures, are capable of generating human-like, contextually relevant text. By fine-tuning these models, this study investigates their application in producing diverse educational materials, including lesson plans, quizzes, and study guides. The system employs prompt engineering to ensure adaptability and alignment with learner needs. Evaluation results demonstrate promising outcomes. Content generated by the system achieved a 96% accuracy rate, overcoming common issues like hallucination, while surveys indicate an 85% likelihood of educator adoption. These findings underscore the potential of AI-powered tools to reduce the workload for educators, enabling them to focus on meaningful student engagement and tailored teaching strategies.**

***Keywords*—educational technology, intelligent tutoring systems, large language models, automated content generation, personalized learning**

I. Introduction

Technology integration in education has created new opportunities to improve the processes of instruction and learning. Among these, Intelligent Tutoring Systems (ITS) have become well-known as a means of providing tailored and flexible educational experiences. ITS are made to give each student an individualized education that takes into account their particular needs and learning preferences. Through the use of Artificial Intelligence (AI), ITS can track student progress, pinpoint areas in need of development, and modify instructional methods accordingly. However, the provision of excellent and flexible teaching resources is crucial to the efficacy of ITS. For educators, producing such content is a time-consuming process that presents a big difficulty because they have to juggle it with other duties like teaching, interacting with students, and assessment.

Simultaneously, AI has made significant strides, especially in the creation of Large Language Models (LLMs). Models like GPT-4 have proven to be incredibly effective in processing and producing human-like language, which makes them useful tools in a variety of fields, including education. By automating the preparation of tests, lesson plans, study guides, and other educational resources, these models have the potential to completely transform the content creation process for ITS. LLMs can free up educators to concentrate on developing stronger relationships with their students and designing more engaging learning environments by cutting down on the time and effort needed to create such content.

The use of LLMs to automate content creation within ITS is examined in this study. The ability of these models to provide precise, contextually relevant, and adaptive information is one of the technical and practical aspects of implementing them in education that are examined. The study assesses how well LLMs operate in practical settings and determines whether they may lessen the workload of teachers without sacrificing or even raising the standard of instruction. It also looks at the wider effects of using AI-driven solutions in education, including how they affect accessibility, inclusion, and scalability.

However, despite the promise that LLMs hold for transforming educational practices, their integration is not without significant challenges. The application of LLMs in education presents significant issues notwithstanding their potential. To reach their full potential, ethical issues like the possibility of bias in AI-generated material and practical difficulties like protecting data security and privacy must be resolved. To make sure that technology is an empowering tool rather than a substitute, the role of teachers in the AI-assisted classroom must also be clearly defined. By tackling these issues, this study hopes to add to the expanding corpus of research on AI in education and offer perspectives on how technology might be successfully incorporated to improve learning results.

Beyond education, deep learning is rapidly emerging as a transformative force across many critical domains, driving innovation and offering solutions to complex challenges. Deep learning rapidly emerges as a transformative force across many critical domains, driving innovation and offering solutions to complex challenges. In the realm of energy systems, for example, advanced forecasting models such as TriChronoNet [1] are demonstrating significant improvements in electricity price prediction, enabling more efficient energy management and resource allocation. Intelligent transportation systems are also undergoing a revolution, with models like CCTP-Net facilitating more accurate multimodal trajectory prediction for connected vehicles [2]. By incorporating factors such as human driving cognition and causal reasoning, these models contribute to enhanced safety and efficiency in increasingly complex urban environments. Cybersecurity is another area where deep learning is making substantial contributions. Federated learning approaches, exemplified by FedGAT, offer a promising avenue for network attack detection, enabling collaborative model training while preserving the privacy of sensitive data. This is particularly crucial in protecting

doi: 10.18178/ijiet.2025.15.9.2401

critical infrastructure and sensitive information [3].

In addition to these sectors, deep learning is also driving significant advancements in fields such as remote sensing, healthcare, finance, and environmental monitoring. The field of remote sensing and Earth observation is benefiting from modality fusion techniques, where models like MFViT effectively integrate data from heterogeneous sources such as hyperspectral imagery and LiDAR. This integration leads to superior classification accuracy and enables a more comprehensive analysis of the Earth' s surface, aiding in environmental monitoring and resource management. Healthcare is witnessing a paradigm shift, with deep learning applications spanning medical imaging, diagnostics, genomics, drug discovery, and patient monitoring. These technologies are facilitating more accurate and timely diagnoses, accelerating the development of new drugs and therapies, and enabling personalized patient care. In the financial sector, deep learning is being leveraged for algorithmic trading, fraud detection, credit scoring, risk management, and market forecasting. These applications enable more sophisticated financial analysis, improved decision-making, and enhanced risk mitigation. Finally, environmental monitoring is being enhanced through the application of deep learning to climate modelling, pollution analysis, biodiversity conservation, land cover classification, and precision agriculture. These tools provide more accurate and timely insights into environmental changes, supporting efforts to address climate change and promote sustainability [4].

These diverse applications, driven by continuous advancements in model architectures and training methodologies, underscore the pervasive impact of deep learning and its potential to reshape various fields. As researchers, it is imperative to have a comprehensive understanding of these wide-ranging applications to effectively contextualize our work and harness the full potential of this transformative technology.

The purpose of this study is to show how LLMs can change the educational landscape when applied carefully and morally. They could strengthen individualized learning, empower teachers, and make educational systems more accessible and inclusive. This work establishes the groundwork for future research into how AI and educators may work together to build a future where everyone has access to high-quality education.

II. Introduction to Automated Content Generation in Education

Automated content generation leverages advanced computational techniques to create educational materials with minimal human intervention. It represents a transformative approach to modern education, where traditional, static content is increasingly replaced with dynamic, interactive resources tailored to individual learner needs. This technology encompasses the use of sophisticated algorithms, including LLMs, to design, generate, and adapt instructional content, significantly reducing the workload on educators while enhancing learning outcomes.

This shift toward automated, adaptive content highlights the growing importance of moving beyond static resources to meet the evolving expectations of learners. In today' s fast-paced educational landscape, the transformation from static to dynamic content is critical for maintaining student engagement and improving comprehension. Static content, such as pre-written textbooks and unchanging lecture slides, often fails to capture the attention of learners accustomed to interactive and personalised digital experiences. By integrating dynamic, AI-driven materials—such as adaptive quizzes, personalised study guides, and multimedia-rich lessons—educators can foster a more engaging and effective learning environment. Dynamic content not only aligns with diverse learning styles but also provides real-time feedback, ensuring that students remain actively involved in their educational journey.

Central to this evolution is the rapid adoption of AI technologies, particularly Large Language Models (LLMs) like GPT, T5, and BERT, which have revolutionized the creation and personalization of educational content. These models are capable of understanding and generating human-like text, enabling the automated creation of customised lesson plans, assessments, and even entire curricula. A noteworthy trend is the integration of these models within ITS, which offer personalised guidance to students based on their progress and preferences. Furthermore, the use of AI tools to enhance multilingual education, create inclusive resources for learners with disabilities, and streamline teacher workflows underscores the broad applicability and transformative potential of LLMs in reshaping education [5–7].

As these technologies become increasingly embedded in educational practice, their influence extends beyond content creation to fundamentally reshape how knowledge is delivered and experienced. As education continues to evolve, the convergence of AI and automated content generation stands poised to redefine how knowledge is imparted, accessed, and experienced. This paper delves into the pivotal role of automated content generation in fostering a more interactive, inclusive, and impactful educational paradigm.

III. Intelligent Tutoring Systems and Their Role in Education

ITS are AI-powered educational tools designed to deliver personalised learning experiences that adapt to individual student needs. Unlike traditional approaches that adopt a one-size-fits-all strategy, ITS dynamically adjusts instructional content and feedback based on real-time data collected during the learning process [8]. By offering tailored support and resources, these systems aim to improve student engagement, comprehension, and retention. Their ability to address diverse learning styles and pace variations makes ITS particularly valuable in modern educational contexts, ensuring that learners of all levels can benefit from customised guidance.

To achieve this level of personalization and adaptability, ITS rely on a set of core components that work in concert to support each learner' s unique journey. At the core of ITS are four key components that work together to facilitate personalised education and automated content generation. The student model acts as a dynamic representation of the learner' s progress, capturing data on their knowledge, preferences, and performance. This information helps the system identify gaps and tailor content accordingly. The tutoring strategy defines how the system interacts with the student, adapting its instructional methods, feedback, and content delivery to the learner' s needs. For instance, a struggling student might receive simpler explanations or additional exercises to master a concept. The content library serves as a repository of educational resources, including multimedia, quizzes, and interactive exercises, which the ITS uses to generate customised materials. Finally, the user interface (UI) provides an intuitive and engaging platform for students to interact with the system, ensuring accessibility and ease of use.

These foundational components are brought to life in a variety of ITS implementations, each demonstrating unique approaches and outcomes in educational practice. Several examples highlight the effectiveness of ITS in educational settings. LEIA is an ITS focused on improving students' algebraic skills. It uses personalised feedback and adaptive problem-solving exercises tailored to individual proficiency levels. While effective in its domain, LEIA is limited by its narrow focus on mathematics. MathBot, a conversational AI chatbot, assists learners by providing immediate feedback and step-by-step guidance for solving mathematical problems [9]. Its strength lies in error correction and motivational support, though its reliance on pre-defined conversation paths can lead to inflexibility. SeisTutor, on the other hand, is an ITS for geosciences education that combines adaptive learning strategies with a robust content library [10]. It measures student performance through quizzes and adjusts content delivery based on individual needs, demonstrating the potential of ITS to adapt to diverse subjects and learning contexts.

While these examples underscore the potential of ITS to transform education, they also reveal persistent challenges that must be addressed for broader adoption and sustained impact. Despite their strengths, ITS faces challenges in content production. Automated generation of personalised educational materials reduces the workload on educators and enables scalability, but it must ensure the accuracy and contextual relevance of the content. Limitations such as biases in training data and the high cost of development and maintenance also hinder the widespread adoption of ITS. Nevertheless, advancements in AI technologies and content generation hold promise for overcoming these challenges, paving the way for more effective, inclusive, and scalable educational solutions. ITS exemplifies how AI can revolutionise education by personalising learning experiences and addressing individual needs.

IV. LLMs and Educational Applications

LLMs represent a significant advancement in AI, with applications ranging from text generation to complex problem-solving. Models such as GPT-4, T5, and BERT are built on the transformer architecture, a groundbreaking neural network design introduced in 2017 [11]. At the core of transformers is the attention mechanism, which allows the model to process input data non-sequentially. This capability enables LLMs to understand and generate text with contextual relevance, capturing intricate relationships between words, phrases, and sentences. Transformers utilise self-attention to focus on different parts of the input simultaneously, making them particularly adept at tasks like summarization, translation, and question-answering. Combined with vast amounts of pretraining data, these models exhibit a remarkable ability to produce human-like text and perform sophisticated language tasks [12].

These advanced linguistic and contextual capabilities make LLMs especially valuable in educational settings, where they can be leveraged to support a range of pedagogical approaches and learning needs. LLMs, with their capacity to understand and generate human-like language, can support constructivism by providing tailored feedback and diverse learning scenarios, and connectivism by fostering networked learning communities. Additionally, they can aid in differentiated instruction by generating varied learning materials and promoting Universal Design for Learning (UDL) by offering flexible content formats. Furthermore, LLMs can power adaptive learning systems by dynamically adjusting content to suit individual student needs, ultimately providing a comprehensive tool to create personalized and effective educational experiences.

To fully appreciate the transformative potential of LLMs in education, it is important to contrast their capabilities with those of traditional content generation methods. Traditional content generation methods, such as template-based and semantic approaches, offer structured ways to create educational materials but come with notable limitations compared to AI-driven methods. Template-based approaches rely on predefined structures where content is inserted into fixed formats. This ensures consistency and reliability but lacks adaptability, making it difficult to generate diverse or personalized learning experiences. Semantic approaches, on the other hand, use rule-based systems to generate content based on meaning and context. While this allows for greater flexibility than templates, it requires extensive domain knowledge to develop effective rules, limiting scalability and making adaptation to new subjects or learner needs more complex.

In contrast to these traditional approaches, LLMs (Large Language Models) introduce a paradigm shift in content generation by leveraging vast amounts of data and deep learning techniques to dynamically generate educational content. Unlike template-based methods, LLMs can produce highly diverse and adaptable learning materials without manual structuring. Compared to semantic approaches, they do not rely on manually crafted rules but rather learn patterns from large datasets, enabling them to understand context and generate nuanced explanations. However, LLMs also pose challenges, such as potential inaccuracies (hallucinations) and the need for careful validation. Despite this, their ability to personalize learning materials, provide instant feedback, and generate contextually relevant content positions them as a powerful tool for modern education. For instance, an AI-powered tutoring system can tailor explanations based on a student' s proficiency level, a capability that traditional methods struggle to achieve efficiently. By combining LLMs with structured template-based approaches or rule-based semantics, educators can create more robust and adaptive learning environments.

These capabilities are increasingly being realized in practical educational applications, where LLMs like GPT-4 are already transforming the way content is created and delivered. In the field of education, LLMs such as GPT-4 have been instrumental in automating the creation of educational content. These models can generate customised lesson plans, quizzes, study guides, and even full curricula based on specific prompts. For instance, GPT-4 has been used to create personalised instructional materials aligned with Bloom' s Taxonomy, catering to diverse learner levels and cognitive objectives. This ability to scale content production while maintaining personalisation makes LLMs invaluable in modern education. However, their deployment is not without challenges. A well-documented issue is "hallucination," where the model generates plausible-sounding but inaccurate or fabricated information. Such occurrences can compromise the reliability of AI-generated materials. Additionally, ethical concerns arise regarding bias in training data, the potential for misuse, and the need for transparency in AI-generated content [13].

Given these opportunities and challenges, it is essential to rigorously evaluate the effectiveness and reliability of automated content generation systems in educational settings. The evaluation of the automated content generation system included quantitative measures of accuracy and qualitative assessments of educator adoption. Accuracy was defined as the extent to which the generated content was factually correct, meaning it was free from errors and consistent with established knowledge. Educators were asked to evaluate the generated content, and the accuracy rate was calculated based on these evaluations. To assess educator adoption, a binary (yes/no) question was included in the survey, asking educators whether they would adopt the system for their teaching practices. The percentage of positive responses was used to determine the likelihood of educator adoption.

The results of this evaluation highlight both the promise and the practical impact of LLM-driven content generation in real-world educational contexts. Despite these challenges, the potential of LLMs in education is immense, as demonstrated in our research on transforming static to dynamic content using AI. In evaluations, GPT-based systems achieved an impressive 96% accuracy in generating educational content, addressing key requirements for relevance and correctness. Moreover, these models have shown an adoption interest exceeding 85% among educators, underscoring their scalability and acceptance in educational environments. The ability to fine-tune LLMs for specific educational tasks ensures adaptability across subjects and learner demographics, significantly reducing the workload on educators while enhancing the quality of learning resources.

These findings align with and are further contextualized by recent systematic reviews, which provide a comprehensive overview of LLM applications and highlight ongoing challenges in the field. A study conducted by Yan *et al*. [14] conducted a systematic review of 118 peer-reviewed papers to explore the use of LLMs in automating educational tasks and to identify the practical and ethical challenges associated with their implementation. The review revealed 53 different applications of LLMs in education, categorised into nine main areas: profiling/labelling, detection, grading, teaching support, prediction, knowledge representation, feedback, content generation, and recommendation. The study also identified several practical and ethical challenges, including low technological readiness, lack of replicability and transparency and insufficient privacy and beneficence considerations.

In conclusion, LLMs like GPT-4, T5, and BERT are reshaping the educational landscape by offering scalable, accurate, and personalised content generation capabilities. While issues like hallucination and ethical considerations require careful mitigation, the benefits of leveraging these models in education far outweigh their challenges. As advancements in AI continue, LLMs are poised to play an even greater role in creating inclusive, efficient, and impactful educational experiences.

V. Approaches to Content Generation

Content generation for educational purposes has evolved significantly over the years, ranging from traditional manual methods to advanced AI-driven techniques. Existing methods for generating educational materials can be broadly categorised into template-based, semantic, and AI-driven approaches, each offering distinct advantages and limitations.

Among these, template-based methods have been widely used for their efficiency and reliability, particularly in domains such as mathematics. Template-based methods rely on predefined structures to generate educational content. A notable example is the procedural generation of math problems, where templates define the constraints and structure of problems [15]. These systems first create abstract mathematical equations, followed by text generation to frame the problems in a natural language. Additionally, distractors—incorrect options designed to test critical thinking—are generated using rule-based algorithms. While these methods are efficient and ensure grammatical accuracy, they often lack adaptability and creativity, producing content that may feel repetitive or rigid.

To address some of these limitations, semantic approaches have emerged, focusing on enriching educational content with metadata and fostering greater interconnectivity between learning materials. Semantic approaches focus on enriching educational content with metadata, enabling dynamic relationships between learning materials. Systems like eLearning Objects (eLOs) incorporate metadata to describe technical and textual information, allowing content to be linked and reused across contexts. For example, the Hypermedia Learning Object System (HyLO) automatically captures raw data, processes it, and integrates metadata to classify and relate content [16]. Such systems enable efficient organisation and retrieval of learning materials, fostering a more interconnected and accessible learning environment. However, the dependency on metadata design and the lack of flexibility in adapting content to unique learner needs can limit their impact.

To overcome these constraints and offer even greater adaptability and personalization, AI-driven approaches have become the new frontier in educational content generation. AI-driven approaches represent the forefront of content generation, leveraging advanced machine learning models to create personalised and scalable educational materials. By fine-tuning models like GPT-4 for specific educational contexts, AI systems can generate quizzes, lesson plans, and instructional materials that are tailored to the preferences and abilities of individual learners [17]. These models are trained on extensive datasets and refined to address educational requirements, ensuring both relevance and adaptability.

Unlike traditional methods, AI-driven systems excel in producing diverse and contextually accurate content at scale. However, challenges such as the potential for hallucination, biases in training data, and ethical considerations require careful management to ensure quality and fairness.

Given these advancements and challenges, it is important to compare AI-powered solutions with traditional content generation methods to fully understand their respective strengths and limitations. When comparing traditional methods with AI-powered solutions, significant differences emerge in terms of efficiency, scalability, and quality. Template-based and semantic methods are well-suited for tasks requiring structured and consistent content but often lack the flexibility and depth needed for personalised education. In contrast, AI-driven approaches offer unparalleled scalability and customisation, enabling educators to generate high-quality materials quickly and efficiently. Despite the higher initial cost of developing AI systems, their long-term benefits in terms of reduced workload and enhanced learner engagement make them a transformative tool in modern education. As technology continues to advance, AI-driven approaches are poised to further revolutionise content generation, bridging gaps in accessibility and personalisation while complementing traditional methodologies.

VI. Challenges in Automating Content Generation

Automating content generation holds immense potential to transform education, but it also presents several challenges that educators and institutions must address to fully realise its benefits. One significant hurdle is the lack of expertise among educators in handling AI systems. Many teachers are not well-versed in the technical intricacies of these tools, making it difficult for them to integrate automated systems into their teaching workflows effectively [18]. This knowledge gap often leads to underutilisation of the technology or reliance on simplified functionalities, which diminishes the potential impact of such systems.

Beyond technical proficiency, there are also deeper ethical and social concerns that must be considered. Another critical issue is the potential for biases in AI-generated content. AI systems are trained on large datasets that may inadvertently include societal or cultural biases. These biases can manifest in the content produced, potentially perpetuating stereotypes or presenting skewed perspectives [19]. In an educational setting, such biases are particularly concerning as they can misinform students or marginalise certain groups, undermining the inclusivity and fairness that education aims to promote.

In addition to concerns about bias, the accuracy and reliability of AI-generated materials present further challenges for educational adoption. While state-of-the-art models such as GPT-4 have demonstrated high levels of accuracy, occasional errors—referred to as hallucinations—remain a challenge. These inaccuracies can erode trust in the technology and place an additional burden on educators, who must review and verify the content before it is used in classrooms. This issue is especially problematic in subjects requiring precise information, such as science or mathematics, where even minor errors can lead to significant misconceptions.

Beyond technical and content-related challenges, broader ethical considerations must also be taken into account when adopting automated content generation in education. One major concern is the risk of reinforcing existing inequalities in education due to disparities in access to technology. Students and schools in underprivileged areas often lack the infrastructure and resources required to utilise AI systems, widening the digital divide [20]. This imbalance threatens to create a two-tiered education system where only those with access to advanced tools can benefit from personalised and dynamic learning experiences, leaving others at a disadvantage. Addressing these ethical challenges requires not only technical solutions but also policy interventions to ensure equitable access to educational technology.

Taken together, these multifaceted issues highlight the importance of a holistic approach to implementing automated content generation in education. In summary, while automated content generation offers transformative potential, its adoption is hindered by technical, ethical, and practical challenges. Overcoming these barriers will require a concerted effort to provide educators with the necessary training, develop unbiased and reliable AI systems, and implement strategies to bridge the digital divide. Only then can the promise of automated content generation be fully realised in creating an inclusive and equitable educational landscape.

VII. System Design and Development

The methodology for this research focuses on designing and evaluating a system capable of automating content generation in educational contexts. The approach integrates state-of-the-art AI technologies, specifically LLMs, to address the challenges of scalability, accuracy, and personalisation. The process is divided into distinct phases, each targeting specific research objectives.

Central to this methodology is the development of a robust system architecture that harnesses the capabilities of advanced LLMs for educational use. The system for automating content generation was meticulously designed to utilise advanced LLMs, such as GPT-4, tailored specifically for educational applications. The architecture was developed to balance computational efficiency, accuracy, and adaptability, ensuring it could meet the diverse needs of educators and students across various educational contexts.

A key component of this architecture (illustrated in Fig. 1 and Fig. 2) is the back-end engine, which drives the system' s ability to process inputs and generate high-quality educational content. At the core of the system is the back-end engine, which serves as the computational foundation for processing inputs and generating educational content. This engine employs Natural Language Processing (NLP) techniques to interpret prompts and produce relevant outputs. A critical aspect of this process is prompt engineering, where structured templates guide the model to create specific types of content, such as lesson plans, quizzes, or study materials. The prompt design process of the system was a process of iterative experimentation where prompts were created based on the tasks assigned to the system, then the prompts were iteratively refined to yield better results, this iterative process showed that it had an impact on the output of the LLM.

Building on this foundation, the system further enhances content quality through targeted fine-tuning and robust validation mechanisms. Fine-tuning the LLM was a key step, involving the use of domain-specific datasets that included instructional materials and Bloom' s Taxonomy-aligned resources. This fine-tuning allowed the model to generate content with a tone and structure that aligns with pedagogical best practices. Additionally, the back-end incorporates content validation mechanisms including the use of a human-in-the-loop to identify and filter out inaccuracies, ensuring that the materials produced meet quality standards.

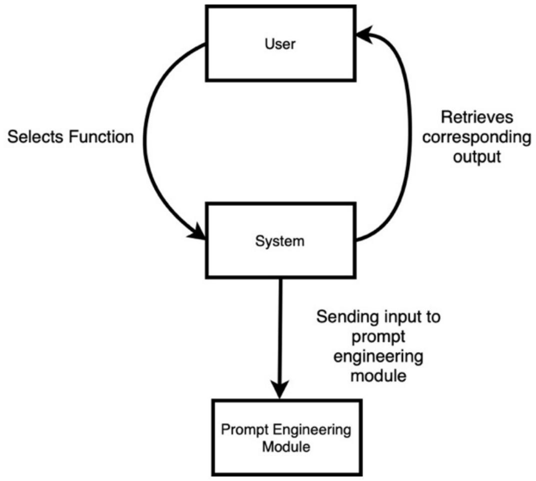

Fig. 1. System processing user input.

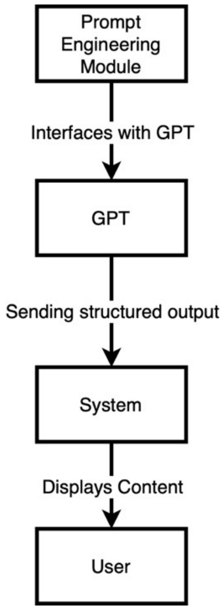

Fig. 2. System processing user input.

To ensure these advanced capabilities are accessible and practical for educators, the system is complemented by a thoughtfully designed front-end interface. The interface allows users to input prompts, customise parameters such as subject matter and difficulty level, and view generated content in real time. Educators can refine the content as needed, ensuring it aligns with their specific instructional goals. Features such as export options and integration capabilities with Learning Management Systems (LMS) make the interface highly practical and adaptable to various teaching environments. These design choices prioritise accessibility, enabling educators with varying levels of technical expertise to effectively use the system.

Equally important to the system' s usability is its ability to adapt and scale with emerging needs and technologies. The development process also emphasised scalability and adaptability. The system' s modular architecture allows for the integration of additional datasets and updates, ensuring that it can evolve alongside advancements in education and AI. For example, future iterations of the system can incorporate improved algorithms for content validation or extend support to additional languages, further enhancing its versatility. This scalability ensures that the system remains a valuable tool in diverse educational settings.

In addition to adaptability, seamless integration with existing educational tools was a major priority during development. Application Programming Interfaces (APIs) were created to enable interoperability with platforms such as LMS and ITS furthermore content can be directly embedded into such systems directly. This integration capability allows educators to embed AI-generated content directly into their teaching workflows, reducing manual effort and enhancing efficiency. By fitting into established systems, the tool facilitates the adoption of AI-driven content generation without disrupting traditional teaching practices.

To ensure that these integration features translated into real-world effectiveness, the system underwent rigorous testing and iterative refinement. Feedback from educators and subject matter experts was integral to the iterative refinement process. For instance, adjustments were made to address challenges like generating content for niche subjects or accommodating varying learning styles. These refinements resulted in a tool that balances technical sophistication with practical usability.

In summary, the system design and development process combined cutting-edge AI technologies with a focus on user-centric design. By prioritising accuracy, scalability, and ease of use, the system provides a transformative solution for educational content generation. It empowers educators to deliver high-quality, personalised learning experiences while addressing the challenges of efficiency and scalability in modern education.

VIII. Results and Discussion

The evaluation of the automated content generation system provided comprehensive insights into its performance, strengths, and areas requiring improvement. By combining quantitative metrics and qualitative feedback from educators and subject matter experts, the analysis highlighted the potential of the system to revolutionise content creation in education while underscoring the challenges inherent in adopting such advanced technologies. The evaluation was conducted amongst Maltese & International educators with a variety of educational backgrounds via a survey.

Understanding the demographic profile of survey participants is essential for contextualising these findings and assessing the system' s relevance across diverse educational settings. The survey results revealed distinct demographic characteristics among the participating educators, who represented both Maltese and international teaching professionals. Approximately 68% of the respondents were from Malta. Within this Maltese group, the gender distribution, based on 2021 national census data, showed that around 30% were male and 70% were female. The age of the Maltese respondents indicated a relatively young demographic, with only about 23% being 50 years of age or older. Regarding educational attainment, a quarter of the Maltese educators held tertiary-level degrees, reflecting a trend towards formal qualifications within the teaching profession.

To place these findings in a broader context, it is useful to compare them with demographic patterns observed among educators globally. Considering the global educator demographic, and under the assumption that European Union statistics offer a reasonable representation, a similar gender distribution was observed, with roughly 75% identifying as female and 25% as male. However, the global teaching workforce presented a more pronounced ageing trend, with approximately 38% of educators being 50 years or older. These demographic trends suggest a consistent pattern in gender balance between Maltese and international educators, while the global teaching population exhibits a broader age range compared to their Maltese counterparts.

With this demographic context in mind, it is particularly noteworthy to consider how the system performed across such a diverse group of educators. One of the most significant findings was the system’ s high accuracy rate in generating educational materials. Across all tested scenarios, the model achieved a 96% accuracy rate, particularly excelling in producing content aligned with Bloom’ s Taxonomy (Fig. 3 and Fig. 4). This demonstrates the system’ s capability to adapt to varying cognitive levels and learning objectives, offering educators content that is not only relevant but also structured to support progressive learning. For instance, lesson plans and quizzes generated for subjects such as mathematics and history were found to be pedagogically sound and contextually appropriate. Educators noted that these outputs met the curriculum standards and effectively addressed their students’ learning needs.

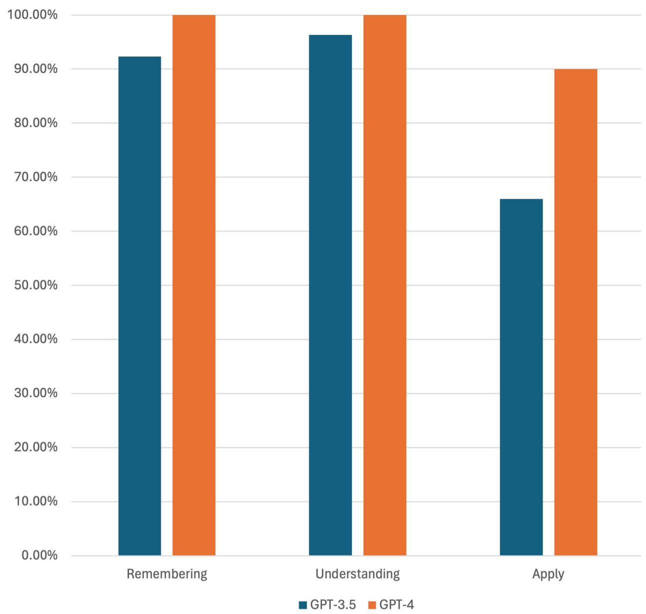

Fig. 3. Base model performance on bloom’ s taxonomy remembering, understanding & apply categories (GPT-3.5 blue, GPT-4 orange).

In addition to its impressive accuracy, the system’ s ability to personalise content for individual learners stood out as a major advantage. By leveraging advanced prompt engineering and fine-tuning techniques, the system was able to generate customised materials tailored to specific learner profiles. Educators highlighted the system’ s ability to adapt to diverse instructional needs, such as creating content for visual learners or designing materials for students with specific cognitive challenges. This feature was particularly valued in classroom settings where differentiated instruction is essential for ensuring that all students progress effectively. The flexibility to specify parameters such as subject, difficulty level, and learning objectives further enhanced the relevance and usability of the generated content.

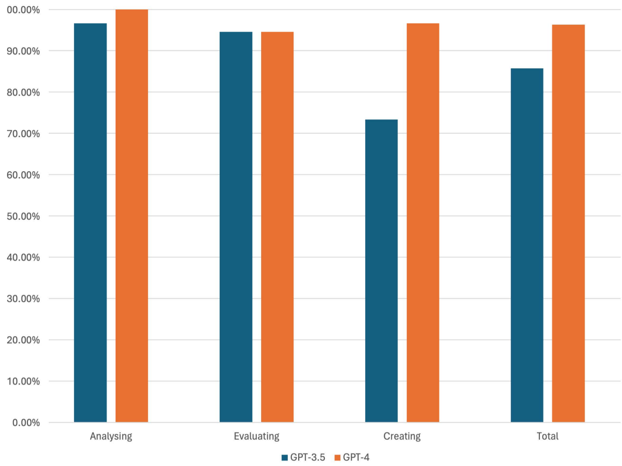

Fig. 4. Base model performance on bloom’ s taxonomy analysing, evaluating & creating categories (GPT-3.5 blue, GPT-4 orange).

Alongside these instructional benefits, educators also noted significant improvements in their own workflow efficiency. Feedback from educators also underscored the system’ s potential to improve teaching efficiency. Many participants reported significant reductions in the time and effort required to create lesson plans, quizzes, and study guides. The real-time content generation feature was particularly praised, as it enabled educators to generate and refine materials on the fly. This efficiency allowed teachers to redirect their efforts towards more interactive and student-focused activities, such as discussions and one-on-one support. Despite these benefits, some educators noted a learning curve in optimising the system’ s outputs, particularly in structuring prompts to achieve desired results. This feedback suggests that additional training resources or a library of pre-built prompt templates could further improve the user experience.

However, despite these notable advantages, the evaluation also revealed certain limitations, especially in the context of niche or highly specialised topics. While the system’ s performance was impressive, some limitations were observed, particularly concerning the accuracy and reliability of outputs in niche or highly specialised topics. In these cases, occasional errors or incomplete content were reported, requiring manual review and correction. These instances highlight the broader challenge of addressing “hallucination” in large language models, a phenomenon where outputs may appear plausible but lack factual accuracy. For example, in a few instances, the system generated historical dates that were inconsistent with facts. Although these occurrences were infrequent, they emphasise the need for robust post-generation validation mechanisms and continued refinement of the training datasets.

Beyond technical limitations, ethical considerations also featured prominently in the discussions surrounding the system’ s deployment. Educators raised concerns about potential biases in the generated content, stemming from biases present in the training data. Ensuring that outputs are inclusive and free from stereotypes remains a critical area for improvement. Biases were mitigated by having human annotators manually review training data. Another ethical challenge is the disparity in access to such advanced technologies. Participants pointed out that schools in underprivileged areas often lack the infrastructure and resources needed to utilise AI-driven systems, potentially exacerbating educational inequalities. Addressing this issue requires collaborative efforts among policymakers, technology developers, and educators to ensure equitable access and support for integrating these tools into diverse educational settings.

Despite these challenges, the system’ s scalability and adaptability offer promising avenues for expanding access and meeting diverse educational needs. The scalability of the system was one of its most significant advantages. Its modular architecture enables the integration of additional datasets and functionalities, ensuring that it can evolve with advancements in AI and educational needs. This adaptability was evident in the system’ s ability to handle a wide range of tasks, from generating introductory materials for primary school students to crafting advanced quizzes for higher education learners. The capacity to scale across different educational levels and subjects makes the system a sustainable solution for addressing the growing demand for personalised and high-quality educational resources.

This flexibility was further reflected in the system’ s successful application across diverse teaching contexts. Educators praised the AI-generated content and have shown a willingness to continue using such a system should it become integrated in a Learning Management System (LMS). These results validate the system’ s potential to complement traditional teaching methods and enhance the overall learning experience. The advantage of the modularity of such a system is that it can be integrated within various LMSs owing this due to the system being standalone and running via an API.

In conclusion, the evaluation highlights the system’ s transformative potential in automating content generation for education. Its strengths in accuracy, scalability, and personalisation position it as a valuable tool for modern teaching practices. However, addressing challenges such as occasional inaccuracies, ethical considerations, and accessibility disparities will be crucial for ensuring its broader adoption and effectiveness. Future work will focus on refining the system’ s validation processes, expanding its training datasets, and enhancing its integration with existing educational infrastructure. By addressing these challenges, the system can continue to empower educators and contribute to a more inclusive and efficient educational landscape.

IX. CONCLUSION

The advancements in automated content generation hold the potential to revolutionise education by addressing critical challenges faced by educators, such as the need for scalable, personalised, and high-quality instructional materials. The system developed and evaluated in this research demonstrated significant promise, achieving high accuracy and relevance in generating content across diverse educational contexts. By leveraging state-of-the-art LLMs like GPT-4, the system proved capable of transforming static, one-size-fits-all resources into dynamic and interactive learning tools tailored to individual learner needs. This capability not only reduces the workload for educators but also enhances the overall learning experience by fostering engagement and adaptability.

Building on these promising results, this research advances the state of the art in educational technology by shifting from traditional, inflexible content generation methods to a dynamic, AI-driven approach. The developed system leverages LLMs to create personalized and adaptable learning materials across various content types and cognitive levels, enhancing efficiency and scalability. Furthermore, it addresses key challenges like AI "hallucination" and ethical considerations, promoting a more responsible and equitable implementation of AI in education.

To fully appreciate the significance of this advancement, it is useful to contrast it with the limitations of traditional AI-driven educational tools. Traditional AI-driven educational tools, such as intelligent tutoring systems and automated quiz generators, often rely on rule-based frameworks or limited machine learning models that require extensive manual intervention. While these approaches offer structured learning experiences, they lack the adaptability needed to generate diverse, contextually rich content at scale. In contrast, this study explores how LLMs revolutionize content generation by dynamically creating personalized and context-aware learning materials without predefined templates or rigid rule sets. Unlike existing AI-driven tools, which typically focus on specific tasks such as grading or multiple-choice question generation, LLMs can generate a wide range of content types—including explanations, interactive exercises, and personalized feedback—while continuously adapting to student needs. This flexibility positions LLMs as a transformative tool in education, offering a novel approach that surpasses conventional AI-driven systems in terms of scalability, personalization, and contextual understanding.

However, realizing the full potential of LLMs in education also brings forth new challenges, particularly when scaling their application across diverse educational levels and subject areas. While the system demonstrates strong potential for automating content generation, challenges may arise when scaling its application across diverse educational levels and subjects. Generating content that is appropriately tailored to the specific cognitive development of students at different levels (e.g., primary, secondary, and higher education) requires careful consideration of pedagogical approaches and content complexity. Additionally, adapting the system to the nuances of various subjects, each with its specific terminology, concepts, and learning objectives, poses a significant challenge. Ensuring consistent accuracy, relevance, and effectiveness across all these contexts will necessitate ongoing refinement of the LLMs, extensive fine-tuning with subject-specific data, and robust validation mechanisms.

Despite these challenges, the research yielded particularly encouraging results in the system’ s ability to generate materials aligned with Bloom’ s Taxonomy, demonstrating its effectiveness in supporting a wide range of cognitive objectives and learner needs. The combination of fine-tuning, prompt engineering, and validation mechanisms ensured that the outputs were both relevant and accurate. The high adoption interest among educators underscores the practicality of such systems in modern classrooms. However, challenges such as occasional inaccuracies, biases in generated content, and the ethical implications of AI adoption in education remain areas of concern. These limitations highlight the importance of ongoing refinement and critical evaluation to ensure that the technology not only meets but exceeds educational standards.

The success of this system also opens the door to broader discussions about the role of AI in education. Beyond content generation, AI-driven tools have the potential to redefine teaching methodologies, assessment techniques, and even curriculum design. The ability to generate inclusive and adaptive materials aligns with the global push towards equity in education. However, the risks of technology exacerbating existing disparities, particularly in under-resourced areas, cannot be overlooked. Ensuring that automated systems are accessible to all educators and learners, regardless of their technological infrastructure, will be essential for fostering an inclusive educational ecosystem.

Looking forward, future work will focus on addressing the challenges identified in this study while expanding the capabilities of the system. One key area of development will be the refinement of validation mechanisms to further reduce inaccuracies and instances of hallucination. This could involve incorporating techniques such as Retrieval-Augmented Generation (RAG) to ground the LLM' s output in verifiable facts, implementing multi-source verification to automatically check generated content against reliable sources, or utilizing reinforcement learning from human feedback (RLHF) to fine-tune the model for improved accuracy. Additionally, improving the system' s ability to handle niche and highly specialised topics will be critical for broadening its applicability in higher education and professional training contexts. Potential solutions include fine-tuning the LLM with domain-specific datasets, employing a hybrid approach that integrates the LLM with external knowledge bases, or developing prompt engineering strategies in collaboration with subject matter experts to guide content generation in these specialized areas.

Ethical considerations will also play a central role in future iterations of the system. Mitigating biases in AI-generated content will require the development of more robust training protocols and regular audits of the datasets used. Collaboration with educators, researchers, and ethicists will be essential to ensure that the system promotes inclusivity and avoids reinforcing stereotypes or disparities. Transparency features, such as detailed logs of content generation processes, can further build trust among users by allowing them to understand and verify the sources and logic behind the outputs.

Another avenue for future research will be enhancing the accessibility and usability of the system. This includes developing lightweight versions of the system that can function effectively in low-resource environments, thereby bridging the digital divide. Providing extensive training and support materials for educators, including pre-designed prompts and best practice guides, will also help overcome the initial learning curve and encourage broader adoption. Integrating the system more seamlessly with existing educational platforms, such as LMS and ITS, will ensure that it becomes an integral part of the teaching and learning process.

In moving towards practical implementation, it is crucial to emphasize that AI-generated content should serve as a tool to augment, not replace, educators' expertise. Educators should adopt a curation and adaptation approach, carefully reviewing and modifying AI-generated materials to align with specific learning objectives and student needs. AI tools can be leveraged to streamline lesson planning, generate diverse assessment items, and create differentiated materials, ultimately freeing educators to focus on personalized instruction and student engagement. Furthermore, seamless integration with LMS platforms can facilitate efficient workflow integration. However, educators must remain mindful of ethical considerations, including the potential for bias in AI-generated content and the importance of promoting academic integrity.

The long-term educational impacts and sustainability of AI-driven content generation warrant careful consideration. While the technology offers substantial benefits in terms of efficiency and personalization, its widespread and sustained use necessitates a focus on several key factors. Firstly, continuous investment in the development and maintenance of these systems is crucial to ensure they remain up-to-date, accurate, and aligned with evolving pedagogical practices. Secondly, ongoing research is needed to fully understand the long-term effects on student learning outcomes, critical thinking skills, and engagement.

Future development will also focus on a multi-phase approach to advance AI-driven content generation in education. This includes enhancing reliability and accuracy through refined validation mechanisms (hybrid frameworks, RAG, RLHF), advanced hallucination mitigation, and the curation of expanded and diversified datasets, necessitating strong interdisciplinary collaborations. Subsequent phases will emphasize enhanced adaptability and personalization via multimodal content generation, improved adaptive learning, and a focus on accessibility and inclusivity, leveraging technological advancements in LLM architecture. Underpinning all development is an ongoing commitment to ethical and responsible AI implementation, addressing bias, ensuring transparency, empowering educators, and conducting long-term impact studies to secure a sustainable and transformative future for AI in education.

Exploring the use of multimodal AI models that incorporate visual, auditory, and textual data represents another exciting direction for future work. By integrating these capabilities, the system could generate richer, more interactive educational experiences, such as video-based tutorials, gamified learning modules, or augmented reality lessons. These innovations would further enhance student engagement and create opportunities for more immersive and impactful education.

One critical research question is: How can we improve the reliability of AI-generated educational content to minimize inaccuracies? Potential methodologies could involve the development of hybrid validation frameworks that combine automated fact-checking with human expert reviews.

Additionally, exploring techniques such as Retrieval-Augmented Generation (RAG) or Reinforcement Learning from Human Feedback (RLHF) could enhance the accuracy and contextual relevance of AI-generated materials.

Another important avenue for future work is the expansion and diversification of training datasets to improve the adaptability of AI-driven educational tools across different subjects and learning levels. Research could focus on curating high-quality, domain-specific datasets that minimize bias and enhance the contextual depth of generated content. Additionally, integrating real-world student interactions as feedback loops into training pipelines could help LLMs refine their responses based on actual learning outcomes. By addressing these research questions through targeted methodologies, future work can strengthen the reliability and effectiveness of AI-powered educational tools.

Furthermore potential mitigation strategies and ethical guidelines for LLMs used in education should be explored further, such as bias detection and correction mechanisms, adversarial debiasing mechanism and data validation techniques such as fact-checking APIs.

In conclusion, the research presented here marks a significant step forward in the field of automated content generation for education. While the system demonstrated exceptional potential, it also revealed areas that require continued attention and innovation. By addressing these challenges and expanding the system' s capabilities, future developments can ensure that AI-driven tools not only meet the immediate needs of educators but also contribute to a more equitable, efficient, and transformative educational landscape. The promise of automated content generation lies not just in its ability to scale and personalise learning but also in its capacity to empower educators and learners alike, fostering a future where technology and education work hand in hand to unlock human potential.

Conflict of Interest

The authors declare no conflict of interest

Author Contributions

Mr. Attard was responsible for overseeing the development of the system, collecting the data, and writing the paper. Throughout the process, Dr. Dingli offered valuable feedback on various aspects of the work and contributed to the initial conceptualization of the project. All authors had approved the final version.

References

[1] M. He, W. Jiang, and W. Gu, "TriChronoNet: Advancing electricity price prediction with multi-module fusion," *Applied Energy*, vol. 371, 123626, 2024.

[2] Y. Lu, W. Wang, R. Bai, S. Zhou, L. Garg, A. K. Bashir, W. Jiang, and X. Hu, "Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cities," *Information Fusion*, vol. 114, 102682, 2025.

[3] W. Jianping, Q. Guangqiu, W. Chunming, W. Jiang, and J. Jiahe, "Federated learning for network attack detection using attention-based graph neural networks," *Scientific Reports*, vol. 14, no. 1, 19088, 2024.

[4] B. Yang, X. Wang, Y. Xing, C. Cheng, W. Jiang, and Q. Feng, "Modality fusion vision transformer for hyperspectral and LiDAR data collaborative classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 2024.

[5] B. Dong, J. Bai, T. Xu and Y. Zhou, "Large language models in education: A systematic review," in *Proc. 2024 6th International Conference on Computer Science and Technologies in Education (CSTE)*, Xi' an, China, 2024, pp. 131–134. doi: 10.1109/CSTE62025.2024.00031

[6] Y. Walter, "Embracing the future of artificial intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 15, 2024. https://doi.org/10.1186/s41239-024-00448-3

[7] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023. https://doi.org/10.3390/educsci13121216

[8] A. El Hadbi, E. K. Hachem, H. Bourray, M. H. Rziki, and H. Oubalahcen, "Review study on the adaptive and personalized intelligent tutoring systems in e-learning," in *Proc. International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD' 2023)*, vol. 931, Cham, Switzerland: Springer, 2024. https://doi.org/10.1007/978-3-031-54288-6_5

[9] J. Grossman, Z. Lin, H. Sheng, J. T. Wei, J. J. Williams, and S. Goel, "MathBot: Transforming online resources for learning math into conversational interactions," *AAAI 2019 Story-Enabled Intelligence*, 2019.

[10] N. Singh, V. K. Gunjan, A. K. Mishra, R. K. Mishra, and N. Nawaz, "SeisTutor: A custom-tailored intelligent tutoring system and sustainable education," *Sustainability*, vol. 14, no. 7, p. 4167, 2022. https://doi.org/10.3390/su14074167

[11] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[12] P. Kaur *et al.*, "From text to transformation: A comprehensive review of large language models' versatility," arXiv preprint arXiv:2402.16142, 2024.

[13] R. Yu *et al.*, "The life cycle of large language models in education: A framework for understanding sources of bias," *Br. J. Educ. Technol.*, vol. 55, no. 4, pp. 1234–1250, 2024.

[14] L. Yan *et al.*, "Practical and ethical challenges of large language models in education: A systematic scoping review," *Br. J. Educ. Technol.*, vol. 55, no. 1, pp. 90–112, 2024.

[15] Y. Xie *et al.*, "Procedural generation of problems for elementary math education," *Int. J. Serious Games*, vol. 8, no. 2, pp. 21–38, 2021.

[16] M. Engelhardt *et al.*, "Hypermedia learning objects system - On the way to a semantic educational web," arXiv, preprint cs/0408004, 2004. https://arxiv.org/abs/cs/0408004

[17] V. Liyanage and S. Ranathunga, "A multi-language platform for generating algebraic mathematical word problems," arXiv, preprint 1912.01110, 2019. https://arxiv.org/abs/1912.01110

[18] M. L. Owoc, A. Sawicka, and P. Weichbroth, "Artificial intelligence technologies in education: Benefits, challenges and strategies of implementation," arXiv, preprint 2102.09365, 2021. https://arxiv.org/abs/2102.09365

[19] J. Lee *et al.*, "The life cycle of large language models: A review of biases in education," arXiv, preprint 2407.11203, 2024. https://arxiv.org/abs/2407.11203

[20] *Artificial Intelligence and the Future of Teaching and Learning*, U.S. Department of Education, 2023.