Grenze International Journal of Engineering and Technology, January Issue

Enhancing Educational Interactivity: Harnessing Multimodal AI for Dynamic Question Formulation

Dr J Uma Maheswari[1], S Shiva Chidambaram[2] and Dr S Brindha[3]
[1-3]SCOPE, Vellore Institute of Technology, Chennai, India

***Abstract*—This study looks into the combination of Multimodal Artificial Intelligence (AI) to improve interaction and promote the active generation of questions. It is based on the need for custom-made and interesting experiences, focusing on where AI, multimodal learning, and creating questions spontaneously meet. This research plans to make a new system that can streamline this process of question generation. It wants to fill up holes in personalized learning feedback methods as well as assessment ways. The study uses a review of literature, methodological frameworks and interpretative analyses to reveal the effectiveness of this suggested multimodal AI system. It stresses on ethical aspects and user responses, adding to academic discussion about AI incorporation and provides an all-encompassing way to improve interaction and efficacy.**

***Index Terms*— Artificial Intelligence (AI), multimodal, methodological framework, personalized learning.**

I. INTRODUCTION

In the field of education, one big struggle is to give each student a learning experience that fits them and find good ways for evaluation. In this research, we look at transformer models and how they might change education by dealing with these difficulties. The goal of this study is to add valuable thoughts on using high-level technologies in education for making it better, flexible, and inclusive.

Typical ways to create educational content sometimes find it hard to satisfy the various requirements of learners. Influenced by the extensive past of AI uses in education, this project combines advanced procedures like extracting keywords, identifying senses, clarifying ambiguity and producing questions. The aim is to close the distance between usual content making methods and what these new adaptable learning places need.

Basically, this project is an investigation in the developing area of Multimodal AI. It will likely change how we interact with education. By smoothly combining Artificial Intelligence along with multimodal features, the study hopes to give learners tailored and interesting experiences that also adapt to their unique requirements. This research is not just a way to solve technological problems; it also aims for significant changes in teaching methods by providing an inclusive structure that improves how efficiently education can be delivered.

A key part of this study is creating and using a system controlled by AI for making multiple-choice questions (MCQs). These types of questions have many benefits in both testing and learning, like quickly measuring knowledge about various subjects, easy to give out tests, and fairness when it comes to grading. The main part or "stem" in MCQ is where the question or problem is shown. It's followed by options which give possible answers for person taking test to pick correct one from them. This study wants to make the process of creating educational content easier by giving teachers a strong tool for making multiple-choice questions that doesn't

Grenze ID: 01.GIJET.11.1.536

need them to know AI deeply.
This study holds importance in the field of technology and beyond. The Multimodal AI system it suggests is a new way for education interactions that aim to give customized learning experiences that are interesting and keep engaging learners. As this research works to influence how technology for education develops, it hopes to encourage beneficial change that improves the experience of learning for learners, teachers and institutions.

II. Literature Review

A. Personalized Learning And Adaptive Feedback

Thomas K.F. Chiu's qualitative study on the impact of generative AI, particularly ChatGPT, in higher education reveals valuable insights into student perceptions and experiences with AI tools. Through thematic analysis and data collection from workshops, questionnaires, and focus group discussions, the study uncovers concerns and expectations regarding learning outcomes, pedagogy, and assessment in the context of AI integration. Findings suggest that students express apprehensions about the influence of ChatGPT on exam performance, highlighting the need for a balance between summarization and cognitive engagement in learning. Pedagogical implications emphasize the role of interdisciplinary teaching and hands-on activities facilitated by AI tools, promoting a holistic approach to knowledge integration. Assessment strategies are seen to shift towards emphasizing originality and aligning assessments with future workforce requirements driven by GenAI. However, the study acknowledges limitations such as cultural variations in perceptions, the need for instructor perspectives, and the evolving nature of experiences with AI tools in education. By addressing these limitations and building on the research findings, future studies can further explore the transformative potential of GenAI in higher education, preparing students for the AI-powered future workforce.
W. Tam et al., authors of the paper, undertook an exhaustive examination into integrating AI-Chatbot technology in nursing education: specifically focusing on the advantages that ChatGPT offers for augmenting learning experiences among nursing students. Simultaneously–they acknowledged and wrestled with ethical considerations as well as challenges tied to its deployment; their research underscored not only ChatGPT's rapid user expansion but also its proficiency in participating contextually within dialogues spanning multiple knowledge domains. The study pinpointed ChatGPT's strengths: it excels in generating and revising text content; delivering personalized self-paced learning experiences--even simulating real-world patient encounters for students. Nonetheless, we must acknowledge some limitations: non-deterministic responses, a dependence on pre-2021 data – and the looming potential of plagiarism via AI-generated materials. Undoubtedly crucial is this fact--educators must guide their charges judiciously using AI-Chatbot technology; not only will critical thinking flourish but also problem-solving skills will significantly enhance. We cannot ignore, however: there remains an urgent necessity for further investigation into the impact of such advancements within nursing education.

B. Engaging Educational Experiences

Nikahat Mulla et al. actively explore the integration of genetic algorithms into a topic-aware transformer-based framework for conversational question generation in their research. They initially introduce two variants of T5 models: one devoid of topics; and the second, incorporating a topic-based classifier – subsequently comparing their respective performances. Optimizing the models with a specific-parameter genetic algorithm achieves an improved quality of question generation. The study establishes that, particularly under the influence of optimization through genetic algorithms, the topic-aware model generates questions possessing superior coherence and relevance in comparison to its topic-unaware counterpart. Furthermore, on a bespoke dataset comprising Java passages, the authors diligently evaluated their approach - revealing substantial improvements in question quality. The research offers promising results; however, scalability issues with larger datasets and the necessity for additional exploration of diverse evolutionary algorithms for optimization might pose potential limitations.
In their paper by Yun Dai et al. they delve into the evolving landscape of education technology, specifically focusing on the shift from learning via AI to learning with AI, emphasizing the importance of AI literacy and assessment literacy among educators & students. The authors highlight the need of fine-tuning ChatGPT for domain-specific educational use and stress the significance of interdisciplinary collaboration in advancing AI-driven educational systems. They also discuss the changing expectations of students' skillsets and competencies in the era of AI integration in education. While the research provides valuable insights into the implications of ChatGPT in education, it also acknowledges the challenges related to academic integrity, ethical considerations, and the potential erosion to creativity and critical thinking skills in students. The paper calls for further research and dialogue to fully comprehend the impact of generative AI tools like ChatGPT in educational settings.

C. Assessment And Question Generation

In their comprehensive research endeavor, Ricardo Rodriguez-Torrealba, Eva Garcia-Lopez, and Antonio Garcia-Cabot delved into the intricate realm of automated question generation using fine-tuned T5 models. With a keen focus on Question Generation (QG), Question Answering (QA), and Distractor Generation (DG), the authors meticulously crafted a sophisticated processing pipeline centered around pre-trained T5 language models. By ingeniously implementing a Text-To-Text approach for DG, the study aimed to revolutionize the creation of educational assessments tailored for e-learning environments. The outcomes of this innovative exploration not only showcased the remarkable efficacy of T5 models in seamlessly generating multiple-choice questions but also shed light on the vast potential of the Text-To-Text paradigm across a myriad of tasks. However, a critical observation surfaced during the study, revealing a tendency in the generated questions to prioritize assessing retention over fostering deeper comprehension. This insightful revelation underscores the need for further refinement in automated question generation systems to ensure a more holistic evaluation of students' cognitive engagement and critical thinking abilities in educational settings.

In their study, conducted by Kate Pearce, Tiffany Zhan, Aneesh Komanduri, and Justin Zhan from the Department of Computer Science and Computer Engineering at the University of Arkansas and the Army Educational Outreach Program UNITE, the authors investigate the performance of various pre-trained language models on extractive question answering tasks. They train and fine-tune these models on multiple datasets of varying difficulty levels to assess their generalizability. Additionally, the authors propose a novel architecture, BERT-BiLSTM, and compare its effectiveness with other models. The outcomes reveal that RoBERTa and BART models show much better performance across all datasets, with the BERT-BiLSTM model showing improvement over the baseline BERT model. The study highlights the importance of bidirectionality in enhancing model performance and suggests potential avenues for future research, such as incorporating additional attention mechanisms for further improvements.

D. Ethical Considerations And Faculty Readiness

Pauletta Irwin, Donovan Jones, and Shanna Fealy's comprehensive review delves into the implications of ChatGPT and artificial intelligence in nursing and midwifery education. Meticulously examining the integration of AI technologies—specifically ChatGPT—in higher education settings; they highlight both its opportunities and challenges. Through synthesizing insights from various studies: it becomes clear that AI has potential to enhance personalized learning; facilitate authentic assessments –even improve student engagement–thus underscoring its versatile efficacy. The authors, too, emphasized the urgent issues of academic integrity and plagiarism; they further highlighted the sluggish uptake of AI in educational institutions. While AI integration yields promising results--a fact not overlooked by them--it is crucial to consider ethical implications with care: data quality limitations also warrant reflection. Moreover—adequate resource allocation for supporting AI implementation within education remains an absolute necessity.

Kevin Peyton and Saritha Unnikrishnan conducted a comprehensive literature review in their study, aiming to analyze existing approaches in natural language understanding (NLU) accuracy within the context of chatbot platforms. Popular platforms such as Dialogflow, LUIS, and QnA Maker underwent comparison with a state-of-the-art SBERT model; this was an effort towards determining the most effective method for answering FAQs from prospective online students. The researchers followed an involved methodology: they created a specific dataset – focused on educational queries; provided explicit details about the frameworks and models utilized – thereby facilitating reproducibility of their results; finally, through rigorous comparative analysis– generated insightful conclusions emerged. The SBERT model demonstrated superior performance in precision, recall, and F1-score compared to traditional feedforward models and existing platforms; indeed, the results were remarkable. The study did acknowledge limitations: a small dataset size posed an issue—highlighting the need for future investigations with larger datasets—and newer pre-trained models demanded further scrutiny. Emphasizing on AI ecosystems' continuously evolving nature, authors underscored potentiality towards enhancing NLU accuracy for chatbot applications within education sector: a testament to their commitment for perpetual advancements.

III. Methodology

A. Research Approach

Identifying gaps in literature underscores the need for innovative solutions to enhance educational interactivity with AI-driven technologies. This study focuses on developing an AI-propelled system for dynamic question formulation in educational contexts. Employing experimental prototyping aligns with recognized gaps in domain-specific research, aiming to explore how AI can generate domain-specific multiple-choice questions (MCQs).

Adopting user-centered design principles addresses the faculty training gap, engaging educators in designing and developing the AI-driven system to meet end-users' needs. The iterative development process refines the system based on user feedback, aiming to foster active learning and critical thinking among students. Ultimately, this research approach aims to tackle primary challenges in education by implementing an AI-driven system that enhances educational interactivity and fosters meaningful learning experiences.

B. Hardware And Software Requirements

Hardware Specifications: The research setting includes a laptop equipped with an Intel Core i5 processor, 8GB of RAM, and a Solid State Drive (SSD). The specified hardware specification provides sufficient computational power to handle the complex algorithms and computations required for generating high-quality multiple-choice questions (MCQs). Moreover significant computational power is not necessary as the system is being executed in Google Colab, a cloud-based coding environment.

Software Environment: The code implementation and execution of the system are carried out using Google Colab notebooks. The main advantage of Google Colab is that anyone can easily collaborate on a complex project without actually having a powerful computer. Additionally, Google Colab supports the integration of popular machine learning libraries and frameworks, such as TensorFlow and PyTorch, facilitating the implementation of AI models for question generation and processing.

C. Modules In The Project

YAKE for Keyword Extraction

YAKE (Yet Another Keyword Extractor) emerged as the top option for my project. Known for its efficiency and flexibility, YAKE uses a mix of statistical analysis and machine learning to find keywords in different kinds of text datasets. The ability of YAKE to learn on its own and generalize patterns without needing manual adjustments makes it suitable for making dynamic questions using the extracted keywords. YAKE utilizes semantic relationships to capture the exact underlying concepts and themes in educational content.

The choice to include YAKE in the project was guided by its strong features, flexibility, and ready-made models. Using already existing resources and libraries made development simpler for me. I could concentrate more on customizing parameters to fit the needs of this particular project. YAKE's ability to effectively extract significant keywords in pairs of two words each was enhanced by utilizing the sense identification function from sense2vec to ensure chosen keywords are pertinent and coherent. So, YAKE was a key part in making the keyword extraction process easier and helping to produce questions that are relevant to the context of educational material.

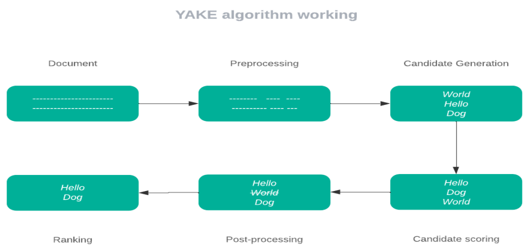

FIG – 1 Yake algorithm flowchart

Wordnet for Sense Identification

In my project, I have chosen WordNet as the main instrument to identify word meanings. WordNet has a large collection of words and synsets, which makes it very good at giving exact definitions for many different senses of words. It works well in tasks like getting key words or making questions because you can find clear and precise meanings for each word sense from this resource. The hierarchical connections in WordNet such as hypernyms and hyponyms help a lot to understand subtle variations in how words are used, making sure that identification of word senses based on context is accurate enough. WordNet's hierarchical links were very helpful in understanding word meanings and to make precise keyword extraction possible. With these connections between different senses of a word, WordNet allowed the project to reach its goal of improving interaction and understanding within learning spaces. Also, having open-source versions and tool collections for

WordNet made it easy to include it in project tasks. This helped us use its characteristics without difficulty for boosting the AI system's functionality.
To sum up, for the project's need to recognize word meanings, WordNet was chosen as the most suitable option because of its reliability, adaptability and ability to work well with other tools. The extensive word bank inside WordNet along with its hierarchical connections greatly assisted in improving keyword extraction and question formulation procedures. This eventually aided the project's aims of enhancing learning encounters and promoting better comprehension of content materials. Yet, it should be highlighted that WordNet solely manages single-word components which restricts its usage on multi-word keywords such as "Artificial Intelligence" or "Machine Learning."

LESK for Sense Disambiguation

In my project, the integration of the Lesk Algorithm played a pivotal role in disentangling the nuanced meanings of single-word keywords, a fundamental aspect of keyword extraction and question-generation tasks. Renowned for its simplicity and efficacy, the Lesk Algorithm excelled in discerning word senses by scrutinizing contextual clues and comparing them against predefined lists of potential meanings. This capability proved invaluable in educational contexts, where precise language comprehension is essential for crafting meaningful multiple-choice questions (MCQs) from extracted keywords. The success of the project hinged on the Lesk Algorithm's ability to accurately identify various word meanings, thereby enhancing the precision of keyword extraction and question formulation processes.
However, it's worth noting that the Lesk Algorithm's effectiveness is primarily optimized for single-word keywords, posing challenges when confronted with multi-word phrases or expressions. Its reliance on context analysis for individual words limits its efficacy in handling complex phrases, necessitating alternative methods or strategies for accurate interpretation within the project framework. Nonetheless, the algorithm's simplicity and efficient performance facilitated seamless integration into project tasks, allowing me to leverage its capabilities to enhance keyword selection and question generation. By focusing on refining the algorithm's features, I was able to contribute to the project's overarching goals of improving educational interactivity and content comprehension, underscoring the Lesk Algorithm's role as a reliable solution for identifying word meanings and advancing the objectives of education-oriented AI systems.

Distractors Generation

In my project, I implemented a meticulous approach to generate plausible wrong answer options, a critical component in crafting effective multiple-choice questions (MCQs) that truly gauge comprehension and critical thinking skills. For single-word keywords, I relied on WordNet, leveraging its semantic database to identify potential distractors by exploring word connections within its hierarchical structure. By meticulously navigating WordNet's network of meanings, I selected alternative choices closely related to the main word yet distinct enough to serve as plausible incorrect answers, enriching the question pool with diverse options reflective of the keyword's nuances.
Conversely, devising distractors for two-word keywords posed greater complexity due to the intricacies of multi-word phrases. To address this challenge, I turned to Sense2Vec, a model adept at discerning word meanings through contextual analysis of extensive text corpora. Utilizing Sense2Vec's ability to contextualize word embeddings within a multidimensional space, I identified words closely aligned with the semantic essence of the two-word keyword, thus generating fitting alternatives grounded in the text's subject matter. By synergizing WordNet's hierarchical analysis with Sense2Vec's contextual insights, I ensured the creation of distractors that effectively challenged students' comprehension and critical thinking abilities across diverse educational topics. Additionally, the utilization of open-source tools like WordNet and Sense2Vec streamlined the implementation process, fostering efficiency and facilitating seamless integration into the project workflow, ultimately enhancing the interactivity and educational value of the AI system.

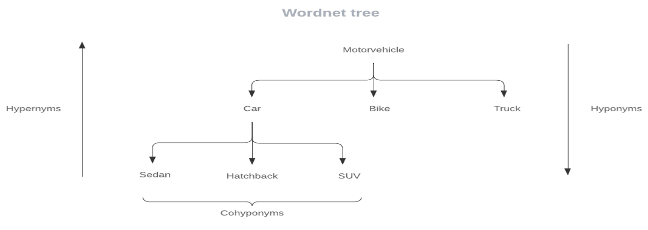

Fig – 2 Wordnet Concept Tree

T5 for Question Generation

In my project, I applied the T5 model for automatic question creation that depends on chosen keywords. T5 is an advanced language model made by Google and it performs well in tasks related to natural language. This makes it a perfect choice for creating different types of queries which are relevant to the context. When we consider the task of generating questions as a transformation from one text form into another, using T5 can provide precise queries that match with extracted keywords exactly. This improves both interaction and content's efficiency.

T5 is good at matching the input text, which helps in understanding and keeping things together. The way it uses natural language processing makes questions that look like main concepts linked to chosen keywords. This improves the quality and importance of evaluations. T5 also gives flexibility for creating different types of questions, including those with multiple choices or spaces to fill in as well as open-ended ones. Such feature helps educators make assessments more suited to their specific teaching objectives.

The use of T5 in the project workflow made question generation more efficient, helping with fast creation of initial models and repeated testing. We used already trained T5 models and Transformer resources from Hugging Face to improve the process, making sure that we deliver high-quality evaluations on time. T5's automatic question-making ability also helps make the system easily expandable; it can create all kinds of inquiries in different content areas for students with various requirements and likes.

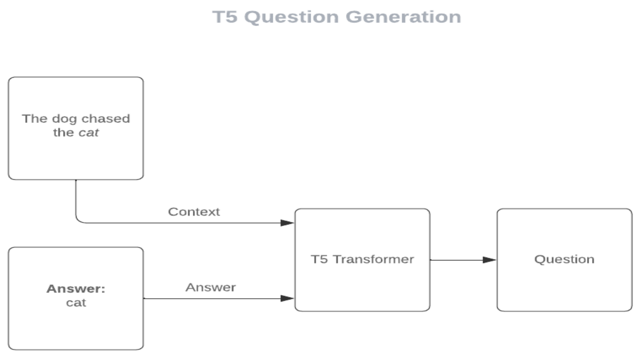

Fig – 3: Text to Text Transfer Transformer (T5) framework

Module Integration

Our project architecture is made up of different modules that make automated question creation from text content possible. Every module has a unique role to play in achieving the goal we have set, starting with YAKE for keyword extraction which forms a strong base for following stages. WordNet helps in recognizing word senses, an important part of creating questions and understanding context correctly. After this, the Lesk algorithm makes the system's understanding better by giving clarity to single-word keywords. Sense2Vec is good at dealing with multi-word phrases. Next, it goes to distractor generation where WordNet helps for single-word keywords and Sense2Vec assists with multi-word phrases.

For the last step, T5 is used to form questions. It creates queries that are specific according to keywords and distractors. This makes sure the assessments are coherent and appropriate, correctly judging understanding as well as thinking skills that are critical. The system uses separate models for every stage making it strong and flexible, showing a good way to make complex information into brief assessments.

To sum up, the project's structure smoothly puts together different parts to create an automatic method for making multiple-choice questions. By using the abilities of each model, it helps in improving accuracy, consistency and importance when creating questions. This leads to better results and involvement with educational evaluations.

D. Results and Findings

The model's performance is evaluated by employing multiple metrics to measure its effectiveness in generating high-quality questions from the provided text. The SentenceTransformer model computes a key metric – the semantic similarity score: this quantifies how close semantically the generated question to the inputted text is; it serves as an indicator of performance. Flesch-Kincaid Grade Level metric is also used--a tool designed to evaluate question readability: it estimates the educational grade level needed for text comprehension. The complexity and accessibility of questions manifest through an analysis of factors—including sentence length, syllable count, among others; this is where the Flesch-Kincaid Grade Level comes into play.

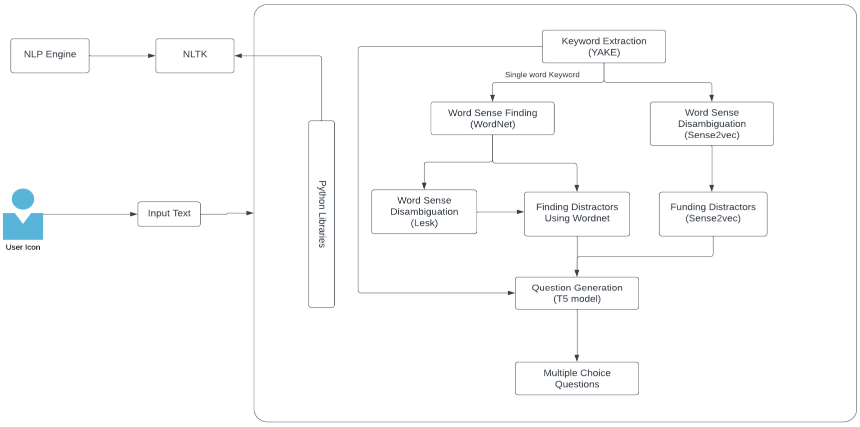

Fig – 4: Modules Integration

By analyzing semantic similarity and readability metrics, we glean valuable insights into our question generation model's performance: these quantitative evaluations--assessing relevance and complexity in a precise manner--illuminate areas for potential improvement. As our approach evolves; as enhancements are continually introduced into the framework of our model – leveraging these very same metrics becomes vital: it fuels essential optimizations that guarantee an output of high-quality questions–immaculately tailored to each specific text provided.

We finally tabulate the results that are obtained when tested with four different textual contexts, to see how the systems performs. The above mentioned metrics are then averaged for each context to gauge the overall performance. The results obtained are as follows.

Context	**Average Kinclaid Grade Level**	**Average Similarity Score**
Context - 1	8.35	0.68
Context – 2	8.3	0.78
Context – 3	8.1	0.65
Context - 4	8.275	0.6

When we look at the table given for testing the question generation system, it shows quite good results. The average Kincaid Grade Level shows high readability levels in a steady manner of 8.1 to 8.35 across different settings or contexts. This means that the questions made fit well for an audience who has advanced understanding skills and matches nicely with planned educational levels too. Moreover, ordinary similarity ratings between 0.6 and 0.78 show a solid match of created questions with their related contexts. This demonstrates the ability of the system to understand content essence well. The significant uniformity in readability and similarity scores indicates that the system is dependable, offering consistent high-grade output for different input content types. All in all, the system does a good job at creating questions that are easy to read and pertinent – it shows promise as an effective educational aid tool for users who require such assistance.

REFERENCES

[1] Chiu, Thomas KF. "Future research recommendations for transforming higher education with generative AI." Computers and Education: Artificial Intelligence (2023): 100197.

[2] Pearce, Kate, et al. "A comparative study of transformer-based language models on extractive question answering." arXiv preprint arXiv:2110.03142 (2021).

[3] Nikahat Mulla, Prachi Gharpure,. "Genetic Algorithm Optimized Topic-aware Transformer-Based Framework for Conversational Question Generation" ,Procedia Computer Science,Volume 230

[4] Tam, Wilson, et al. "Nursing education in the age of artificial intelligence powered Chatbots (AI-Chatbots): Are we ready yet?." Nurse Education Today 129 (2023): 105917.

[5] Irwin, Pauletta, Donovan Jones, and Shanna Fealy. "What is ChatGPT and what do we do with it? Implications of the age of AI for nursing and midwifery practice and education: An editorial." Nurse Education Today 127 (2023): 105835.

[6] Peyton, Kevin, and Saritha Unnikrishnan. "A comparison of chatbot platforms with the state-of-the-art sentence BERT for answering online student FAQs." Results in Engineering 17 (2023): 100856.
[7] Dai, Yun, Ang Liu, and Cher Ping Lim. "Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education." (2023).
[8] Rodriguez-Torrealba, Ricardo, Eva Garcia-Lopez, and Antonio Garcia-Cabot. "End-to-End generation of Multiple-Choice questions using Text-to-Text transfer Transformer models." Expert Systems with Applications 208 (2022): 118258.