

ABSTRAK

Abstrak—*Automatic Item Generation* (AIG) adalah teknologi yang secara otomatis menghasilkan item penilaian seperti pertanyaan pilihan ganda atau *Multiple Choice questions* (MCQs) dan *Question answer* (QA) berbasis model komputasi. AIG memfasilitasi dan mempercepat proses pengembangan untuk item seperti pengalih, yang dikenal sebagai *Automatic Distractors Generation* (ADG) dalam MCQs. *Large Language Models* (LLM), seperti *Generative Pre-trained Transformer* (GPT), telah meningkatkan kapasitas AIG untuk menghasilkan pertanyaan berkualitas tinggi di berbagai bidang, termasuk MCQs untuk *vocabulary test*, karena kecerdasan buatan telah berkembang dengan sangat pesat. Namun tidak terlepas dari beberapa tantangan seperti pemahaman kontekstual yang tidak selalu mendalam, pembangkitan pengalih yang tidak relevan, dan halusinasi dalam menghasilkan *distractors* yang memiliki hubungan sintagmatik dalam konteks bacaan pada *reading passage*.

Untuk mengatasi tantangan ini, penelitian ini mengadopsi LLM dengan memanfaatkan beberapa teknik *prompting* seperti *Few shot prompting* dan *retrieval augmented generation* (RAG) untuk menghasilkan *distractor* yang mempertahankan hubungan sintagmatik dengan konteks bacaan yang lebih luas. Dibandingkan dengan penelitian sebelumnya, pendekatan dengan model RAG-based with COC telah menunjukkan peningkatan dengan akurasi mencapai 86,67% dalam kualitas *distractor*. Lebih lanjut, penerapan model OpenAI GPT-4.0 menunjukkan peningkatan substansial dalam akurasi *distractor* dalam menangkap makna kontekstual yang lebih luas dari konteks bacaan.

Dengan menggabungkan metode *prompting* RAG dan GPT-4.0, penelitian ini berkontribusi dalam pengembangan AI untuk pendidikan, khususnya dalam menghasilkan *distractor* otomatis yang relevan dan masuk akal pada soal kosakata berbasis bacaan. Namun, model *RAG-based with COCA* ini masih belum maksimal karena masih ditemukan *problematic distractors* seperti antonim langsung dari *target word* dan *Correct answer, synonym pair*, serta *distractor* yang tidak memiliki hubungan makna (*no relation*). Pengembangan lanjutan menggunakan model GPT-4.1 *Mini* dengan metode *RAG-Self Review Filtering* menunjukkan peningkatan reliabilitas menjadi 95%. Penelitian ini memberikan kontribusi terhadap pengembangan sistem AIG dalam bidang pendidikan, khususnya dalam pembuatan *distractor* otomatis untuk soal kosakata berbasis bacaan. Untuk penelitian selanjutnya, disarankan fokus pada pengukuran dan pengendalian tingkat kesulitan *distractor* berdasarkan kerangka CEFR (*Common European Framework of Reference for Languages*). Hal ini bertujuan agar sistem AIG tidak hanya menghasilkan *distractor* yang relevan secara semantik, tetapi juga mampu menyesuaikan tingkat kesulitan soal secara adaptif sesuai level kemampuan bahasa pembelajar.

Kata kunci—Pembangkitan *Distractor*, Soal Pilihan Ganda, Relasi Sintagmatik, Model Bahasa Besar, Retrieve Augmented Generation, Self Review *Filtering*

ABSTRACT

Abstract—Automatic Item Generation (AIG) is a technology that automatically generates assessment items such as multiple-choice questions (MCQs) and question-answer (QA) based on computational models. AIG facilitates and accelerates the development process for items such as *distractors*, known as Automatic *Distractors* Generation (ADG) in MCQs. Large Language Models (LLMs), such as the Generative Pre-trained Transformer (GPT), have significantly enhanced AIG's capacity to generate high-quality questions across various fields, including MCQs for vocabulary tests, as artificial intelligence has advanced rapidly. However, it faces several challenges, including a lack of deep contextual understanding, the generation of irrelevant distractors, and hallucinations that create distractors with *syntagmatic* relationships within the context of the reading passage.

To address these challenges, this study uses an LLM with several prompting techniques, such as few-shot prompting and retrieval-augmented generation (RAG), to generate *distractors* that maintain *syntagmatic* relationships with the broader reading context. Compared to previous research, the RAG-based approach with COCA has demonstrated an improvement, achieving an accuracy of up to 86.67% in *distractor* quality. Furthermore, the application of the OpenAI GPT-4.0 model showed a substantial increase in *distractor* accuracy in capturing broader contextual meaning from the reading context.

By combining the RAG *prompting* method and GPT-4.0, this research contributes to the development of AI for education, particularly in generating relevant and plausible automatic *distractors* for reading-based vocabulary questions. However, the RAG-based model with COCA is still not optimal because it still finds problematic distractors, such as direct antonyms of the target word and the Correct answer, synonym pairs, and distractors with no meaningful relationship (no relation). Further development using the GPT-4.1 Mini model with the RAG–Self Review *Filtering* method increased accuracy to 95%. This research contributes to the development of the AIG system in education, particularly by creating automatic *distractors* for reading-based vocabulary questions. For further research, it is recommended to focus on measuring and controlling the level of *distractor* difficulty using the CEFR (Common European Framework of Reference for Languages). This aims to ensure that the AIG system not only produces *semantically* relevant *distractors* but also adjusts the difficulty level of the questions to the learner's language ability.

Keywords, *Distractor* Generation, Multiple Choice Questions, *Syntagmatic* Relation, Large Language Model, Retrieve Augmented Generation, Self Review *Filtering*