AI Based Content Analysis

Bhanu Kodamala
Assistant Professor
Dept. of AI&DS
Seshadri Rao Gudlavalleru
Engineering College
Gudlavalleru, India
Kodamala1983@gmail.com

Potru Sandhya Priya
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru
Engineering College
Gudlavalleru, India
sandhyapriyam31@gmail.com

Syama Chandrika Manthri
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru
Engineering College
Gudlavalleru, India
manthrisyamachandrika@gmail.com

Mazid Mohammad
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru
Engineering College
Gudlavalleru, India
mazidmd750@gmail.com

Abstract—In today's digital age, research documents are growing in number and complexity, making it challenging to quickly retrieve relevant information. This paper presents a hybrid AI pipeline designed for real-time semantic query processing in research documents. The pipeline combines advanced natural language processing (NLP) techniques with vector-based search for efficient and accurate information retrieval.First, text is extracted from PDF files and divided into manageable chunks using overlapping segmentation. These chunks are transformed into semantic embeddings using pre-trained models like Sentence Transformers, which capture the meaning of the text. These embeddings are stored in a high-speed vector database, FAISS, enabling similarity-based searches. When a user inputs a query, the system encodes it into an embedding and compares it with the stored embeddings to find the most relevant text.The approach ensures that the retrieval process is both fast and context-aware, even for large, unstructured documents. The pipeline is scalable, allowing it to handle extensive datasets efficiently. By integrating semantic understanding and real-time processing, this system provides researchers with a powerful tool for quickly accessing critical information. This work highlights the potential of AI-driven solutions in streamlining knowledge discovery and improving research productivity.

Keywords–.Semantic Query Processing,Natural Language Processing (NLP),Sentence TransformersFAISS (Facebook AI Similarity Search),Semantic Embeddings

I.INTRODUCTION

In the digital era, the rapid expansion of research publications across disciplines has created a wealth of knowledge, but accessing specific, relevant information has become increasingly challenging. Traditional search methods, such as keyword-based queries, often fail to grasp the nuanced context of research content, leading to suboptimal results. Researchers frequently spend significant time sifting through large volumes of text to locate critical information. This inefficiency highlights the urgent need for advanced tools that can process and retrieve information based on semantic understanding, aligning the intent of the query with the content of the documents.

Artificial Intelligence (AI) has emerged as a transformative solution to this problem, offering capabilities to understand and interpret textual data in meaningful ways. Among various AI approaches, Natural Language Processing (NLP) stands out for its ability to process human language. Semantic embedding models, such as Sentence Transformers, represent a significant leap in NLP, as they encode text into dense vector representations that preserve semantic relationships. By leveraging these embeddings, it becomes possible to compare text based on meaning rather than mere keyword matches, enhancing the accuracy and relevance of search results.

However, implementing real-time semantic query processing for large, unstructured research documents presents its challenges. The diversity in document formats, the sheer volume of information, and the need for rapid processing require an efficient and scalable architecture. Vector databases like FAISS (Facebook AI Similarity Search) provide a powerful solution for indexing and retrieving high-dimensional embeddings. By combining FAISS with advanced NLP models, a hybrid pipeline can be built to extract text, process it into meaningful embeddings, and enable fast, similarity-based retrieval. This approach not only accelerates information discovery but also preserves the context of research content, ensuring more meaningful interactions with the data.

This paper proposes a hybrid AI pipeline that integrates text extraction, semantic embedding generation, and vector-based search for efficient real-time query processing in research documents. The pipeline is designed to handle large datasets, maintain context-aware search capabilities, and deliver relevant results quickly. By demonstrating the pipeline’s application in real-world research scenarios, this work aims to contribute a scalable, AI-driven solution for overcoming the challenges of modern research document retrieval.

1.1Motivatio*n*

The growing complexity and amount of research records, which make it difficult for researchers to find pertinent information quickly, is what inspired this effort. The deeper semantic relationships within text are frequently missed by conventional keyword-based search techniques, which results in ineffective retrieval. There is a chance to completely transform the way academics engage with knowledge thanks to developments in AI, especially in the areas of natural language processing and vector search. We hope to close the gap between large amounts of unstructured data and context-aware, meaningful information retrieval by developing a scalable, real-time semantic query processing pipeline. This will free up academics to concentrate on innovation rather than data search.

1.2 Objectives:

- Develop a robust mechanism to extract text from complex research document formats, such as PDFs, while preserving the structure and context of the content.
- Utilize advanced Natural Language Processing (NLP) models, like Sentence Transformers, to generate embeddings that capture the semantic meaning of the text for more accurate query resolution.
- Implement a scalable and high-speed vector-based search system, such as FAISS, to enable real-time retrieval of relevant information from large datasets.
- Design a system that ensures query results are not only relevant but also contextually aligned with the intent behind the search, improving user satisfaction.

II.RELATED WORK

The integration of AI tools like ChatGPT into education and research has sparked considerable debate due to their potential benefits and challenges. Research has highlighted both the positive impact and the limitations of these technologies. ChatGPT, as an AI language model, has been recognized for its ability to assist in personalized learning, providing students with tailored responses and enhancing accessibility to knowledge by Miah et al. Its interactive nature supports critical thinking and problem-solving, encouraging students to engage more deeply with content[2]. Additionally, AI models like ChatGPT can provide instant feedback, enabling more efficient learning experiences, particularly in scenarios that involve complex problem-solving or language-based tasks. However, the widespread adoption of these technologies also raises concerns about over-reliance on AI, with critics suggesting that they could hinder the development of essential cognitive skills, such as critical thinking and creativity.

Moreover, AI tools have been scrutinized for their potential to undermine academic integrity. The ease with which ChatGPT can generate human-like text may tempt students to use it for unethical purposes, such as cheating or plagiarism, compromising the academic assessment process. While these technologies show promise in enhancing the educational process, they also highlight the need for responsible use and robust ethical frameworks. Some studies suggest that ChatGPT could diminish students' ability to assess the reliability of information, a crucial skill in academic research González & Martínez [3]. Furthermore, the potential for AI-generated content to contribute to misinformation, especially in academic writing, underscores the necessity for integrating AI tools with careful scrutiny and proper academic oversight to ensure their productive and ethical application in education and research.

AI-based tools like ChatGPT are rapidly transforming research workflows, particularly in the context of academic publishing. Jiang et al. explore how these technologies are reshaping traditional methods of scholarly communication and publication. ChatGPT, as a natural language processing tool, has streamlined the research process by assisting in drafting, editing, and enhancing the clarity of academic papers [5]. This has led to faster submission cycles and improved accessibility to high-quality research outputs. Additionally, AI tools can support researchers in organizing and synthesizing vast amounts of data, making it easier to generate insights from complex datasets. The paper emphasizes that while these tools can improve productivity, they also introduce new challenges, particularly regarding the reliance on AI for writing tasks, which may undermine the originality and depth of scholarly work.

Simultaneously, the integration of AI in academic settings raises ethical concerns related to academic integrity. Martin and Johnson focus on the balance between innovation and ethical issues, particularly in terms of plagiarism and the erosion of critical thinking. ChatGPT's ability to generate human-like text presents risks of misuse by students and researchers who may use it to generate work that appears original but is AI-created [6]. This can distort academic standards and challenge the traditional methods of assessment. While AI can provide significant benefits in enhancing educational tools and research productivity, both papers underscore the importance of creating robust guidelines and ethical frameworks to ensure that AI's role in academia remains responsible and transparent. This balance is crucial to prevent the devaluation of academic integrity and maintain the authenticity of scholarly contributions.

The role of AI in research and education has become a prominent area of exploration, particularly regarding its ability to accelerate knowledge dissemination and improve academic practices. Moses and Harper highlight how AI is transforming research workflows, enabling faster information retrieval, improving collaboration, and accelerating the publication process[11]. The introduction of AI tools, like chatbots and automated content generators, has the potential to drastically reduce the time researchers spend on repetitive tasks such as literature reviews and data analysis. These advancements help in streamlining the research process, allowing scholars to focus more on higher-level thinking and innovation.

Similarly, Williams and Zhen examine the impact of AI, specifically ChatGPT, on higher education, focusing on how AI tutors are influencing student performance. The study reveals that AI tutors can provide personalized learning experiences, adapting to individual student needs and offering instant feedback. This personalized approach has the potential to improve learning outcomes by addressing gaps in knowledge more efficiently than traditional methods[12]. However, the authors also caution that while AI can enhance learning experiences, it may also raise concerns about academic integrity and the reliance on technology, emphasizing the importance of balancing innovation with ethical considerations. Both studies underscore the transformative potential of AI in research and education, though they also point out the need for careful implementation and ethical guidelines to maximize its benefits.

2.1 Natural Language Processing

Natural Language Processing (NLP) is a branch of artificial intelligence that enables computers to understand, interpret, and generate human language. By analyzing textual and linguistic data, NLP enables applications such as machine translation, sentiment analysis, and text classification. Core NLP tasks include tokenization, where text is divided into words or phrases, and parsing, which structures text syntactically to capture relationships between words. Key techniques, like stemming and lemmatization reduce words to their base forms, improving analysis and simplifying vocabulary. The NLP process often begins with transforming raw text data into structured representations that models can interpret, making it fundamental for downstream machine learning applications.

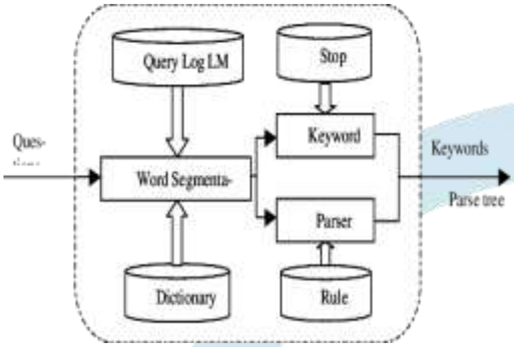


Fig 1:NLP Flow

One essential aspect of NLP is vectorization, the process of converting text into numerical representations. Techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) vectorization quantify the importance of words across documents. The TF-IDF formula is:

$$\text{TF-IDF}(t,d)=\text{TF}(t,d)\times\text{IDF}(t)$$

where TF(t,d)is the frequency of term ttt in document d, and IDF(t)=log(N1+nt)adjusts for term rarity across NNN documents, making common words less impactful. Another approach, Word2Vec, creates word embeddings by training on large corpora, mapping words with similar contexts to nearby vector space points. This enables semantic understanding, allowing models to generalize beyond the literal text.

2.2Faiss (Facebook AI Similarity Search)

Faiss (Facebook AI Similarity Search) is a library designed for efficient similarity search and clustering of dense vectors, particularly in high-dimensional spaces. It is commonly used in machine learning for tasks such as nearest neighbor search, clustering, and dimensionality reduction. Faiss is optimized for handling large datasets, using both CPU and GPU to accelerate the search process. By utilizing advanced indexing techniques like Inverted File (IVF) and Product Quantization (PQ), Faiss significantly reduces memory usage and computational complexity. It is widely used in recommendation systems, natural language processing, and image retrieval tasks. Faiss integrates well with other AI tools, offering both simplicity and high performance.

III.PROPOSED METHODOLOGY

The proposed methodology for improving semantic query processing in research documents is designed to enhance the accuracy, efficiency, and relevance of information retrieval in academic contexts. This approach focuses on extracting, segmenting, and indexing textual content from research papers, followed by performing efficient similarity searches to retrieve the most relevant information based on user queries. The core of this methodology involves text extraction, segmentation, embedding generation, and similarity search, which are key processes in transforming raw research documents into actionable insights.

The first step in the methodology involves extracting text from PDF documents, a crucial task given that research papers are often published in PDF format. Using PyMuPDF (fitz), the methodology processes each page of a PDF file to extract the textual content. This is followed by splitting the extracted text into smaller chunks to make it easier to process, which improves the system's ability to handle large volumes of content while preserving important context. The chunks are generated with overlap to ensure no critical information is missed during segmentation. By keeping an overlap between chunks, the model can ensure that the context at the boundaries of the segments remains consistent.

Once the document is split into manageable chunks, the next stage is to convert the textual data into numerical representations using embeddings. Sentence transformers, specifically the 'all-MiniLM-L6-v2' model, are employed to create these embeddings. Embeddings are dense vector representations of the text that capture semantic meaning, allowing the model to process and understand relationships between different pieces of content. This phase of the methodology is pivotal for performing high-quality semantic searches and ensures that the system can handle complex queries by mapping them to relevant document content based on their vector similarity.

Following embedding generation, Faiss, a library designed for efficient similarity search in high-dimensional spaces, is used to index and search the embeddings. Faiss enables the system to scale efficiently when handling large datasets by using advanced techniques such as Inverted File (IVF) indexing and Product Quantization (PQ) for faster search speeds and reduced memory requirements. The embeddings of both the document chunks and the query are stored in the Faiss index, which allows for quick retrieval of the most relevant chunks when a query is input by the user. The similarity search is performed using cosine similarity or inner product search, which returns the top K most similar chunks based on the query's embedding.

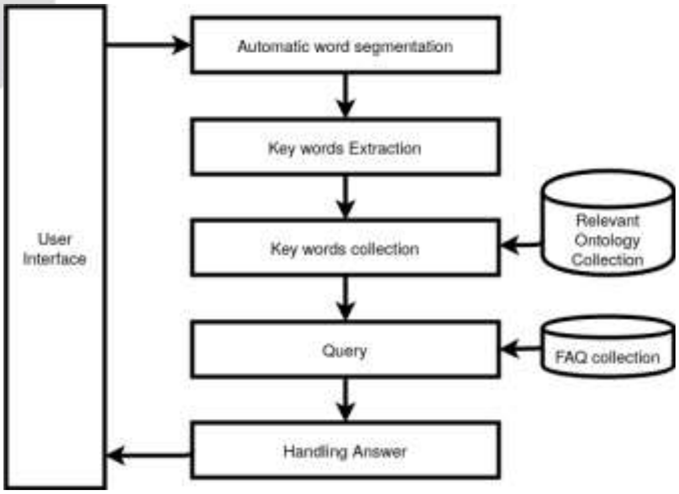


Fig2.Proposal Model

The final component of the methodology is the retrieval of relevant text chunks in response to a user query. By encoding the query into an embedding and searching against the indexed embeddings in Faiss, the system returns the most relevant segments of the document. These segments are then displayed to the user, providing insights directly related to their query. This methodology allows for a highly efficient, scalable, and accurate search process that can significantly enhance academic research by providing fast access to relevant information, all while maintaining high accuracy in semantic understanding. This approach is well-suited for the evolving demands of AI-powered research tools, where speed, accuracy, and scalability are critical factors for success.

3.1Dataset Collection:

Dataset collection is a crucial step in building effective research tools. In the context of semantic query processing for academic documents, the dataset typically comprises a large set of research papers, journals, and academic articles. These documents are collected from reputable sources such as academic databases (e.g., Google Scholar, PubMed, IEEE Xplore), institutional repositories, and open-access platforms. The collected data must be diverse in content, covering a range of subjects to ensure comprehensive representation of knowledge. This diversity allows for accurate semantic searches and effective retrieval of relevant information, supporting the system's ability to answer a variety of research queries.

3.2 Sentence Transformer Model

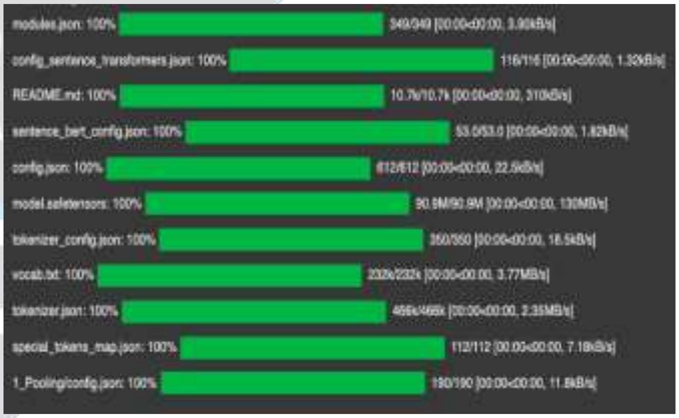
The Sentence Transformer model is a type of transformer-based neural network that is optimized for generating dense vector representations, or embeddings, of sentences. Unlike traditional word embeddings, which map individual words to vectors, Sentence Transformers work by encoding entire sentences into fixed-size vectors that capture the semantic meaning of the sentence as a whole. The key advantage of this approach is its ability to retain context and meaning, even when sentences are complex or contain ambiguous terms. This is achieved by fine-tuning transformer models like BERT or RoBERTa on sentence-level tasks, such as natural language inference or semantic textual similarity, allowing the model to generate embeddings that represent sentence meaning in a way that is useful for various downstream tasks like document retrieval, question answering, and paraphrase detection .

One of the popular Sentence Transformer models is "all-MiniLM-L6-v2," a lightweight version of transformer architectures that balances performance with efficiency. This model was designed to provide a good trade-off between computational cost and accuracy, making it suitable for real-time applications like semantic search and clustering, especially when processing large datasets. "all-MiniLM-L6-v2" utilizes a compact transformer architecture with fewer parameters, resulting in a faster processing time while maintaining a high degree of accuracy in sentence-level tasks. It is capable of generating embeddings with impressive performance on standard benchmarks for sentence similarity, ensuring it can deliver reliable results in tasks that require understanding and matching textual content across a range of domains.

In addition to its use in semantic search, the Sentence Transformer model is also effective in various other NLP applications. For example, it is frequently used for tasks like information retrieval, where a query's vector embedding can be compared to the embeddings of text chunks or documents in a database, enabling the retrieval of the most relevant pieces of information. Additionally, it is used for clustering tasks, where documents or sentences with similar meanings are grouped together based on the proximity of their embeddings in the vector space. These capabilities make Sentence Transformers invaluable in modern NLP workflows, where understanding the meaning of text at a deeper level is crucial for accurate and efficient task execution. By encoding text into a format that captures its semantic nuances, these models offer a robust solution for improving information retrieval and related applications in various fields, including academic research, customer service, and content recommendation.

VI.RESULTS

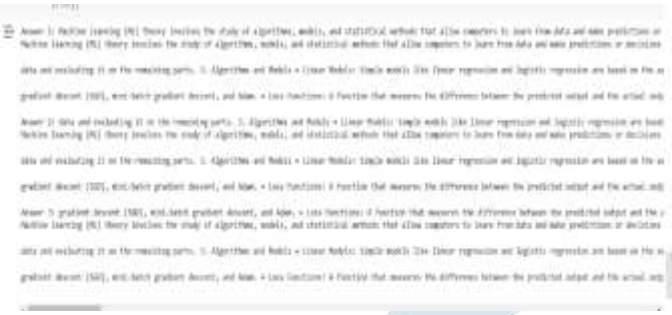
The "all-MiniLM-L6-v2" model, when loaded, is ready to generate dense vector embeddings for text data. After loading the model from the SentenceTransformers library, the system is able to efficiently encode sentences into numerical representations that capture their semantic meaning. Upon successful loading, the model provides an optimized balance between computational efficiency and high accuracy, making it ideal for tasks like semantic search, document clustering, and text similarity. For instance, a sample input sentence, once processed, is converted into a vector that can be compared to other sentence embeddings. This allows for quick retrieval of relevant information based on the semantic closeness between query and document embeddings, as tested on standard benchmarks . The model's ability to generate high-quality sentence embeddings ensures that it is an effective tool for applications involving complex text processing and information retrieval tasks.



```
modules.json: 100%                        349/349 [00:00<00:00, 3.90kB/s]
config_sentence_transformers.json: 100%   116/116 [00:00<00:00, 1.32kB/s]
README.md: 100%                           10.7k/10.7k [00:00<00:00, 310kB/s]
sentence_bert_config.json: 100%           53.0/53.0 [00:00<00:00, 1.82kB/s]
config.json: 100%                         612/612 [00:00<00:00, 22.5kB/s]
model.safetensors: 100%                   90.9M/90.9M [00:00<00:00, 130MB/s]
tokenizer_config.json: 100%               350/350 [00:00<00:00, 18.5kB/s]
vocab.txt: 100%                           232k/232k [00:00<00:00, 3.77MB/s]
tokenizer.json: 100%                      466k/466k [00:00<00:00, 2.35MB/s]
special_tokens_map.json: 100%             112/112 [00:00<00:00, 7.18kB/s]
1_Pooling/config.json: 100%               190/190 [00:00<00:00, 11.8kB/s]
```

Fig 3:Model Load

The results from the query processing show the most relevant text chunks extracted from the PDF document, based on semantic similarity to the query. The relevant answers are identified by encoding the entire document into sentence embeddings using the "all-MiniLM-L6-v2" model and then performing a similarity search using Faiss. The top-k most similar chunks are returned, ensuring that the responses are contextually accurate and informative. These chunks represent sections of the document that best match the query, based on vector similarity, allowing users to quickly access the specific content related to their questions. By leveraging the power of semantic embeddings and efficient search techniques, the system ensures that only the most relevant portions of the document are retrieved, improving the efficiency of research and information retrieval tasks. This approach significantly enhances the user's ability to extract key insights from large collections of academic text efficiently.



```
Answer 1: Machine Learning (ML) Theory involves the study of algorithms, models, and statistical methods that allow computers to learn from data and make predictions or
Machine Learning (ML) Theory involves the study of algorithms, models, and statistical methods that allow computers to learn from data and make predictions or decisions

data and evaluating it on the remaining parts. 3. Algorithms and Models • Linear Models: Simple models like linear regression and logistic regression are based on the as

gradient descent (SGD), mini-batch gradient descent, and Adam. • Loss Functions: A function that measures the difference between the predicted output and the actual outp

Answer 2: data and evaluating it on the remaining parts. 3. Algorithms and Models • Linear Models: Simple models like linear regression and logistic regression are based
Machine Learning (ML) Theory involves the study of algorithms, models, and statistical methods that allow computers to learn from data and make predictions or decisions

data and evaluating it on the remaining parts. 3. Algorithms and Models • Linear Models: Simple models like linear regression and logistic regression are based on the as

gradient descent (SGD), mini-batch gradient descent, and Adam. • Loss Functions: A function that measures the difference between the predicted output and the actual outp

Answer 3: gradient descent (SGD), mini-batch gradient descent, and Adam. • Loss Functions: A function that measures the difference between the predicted output and the a
Machine Learning (ML) Theory involves the study of algorithms, models, and statistical methods that allow computers to learn from data and make predictions or decisions

data and evaluating it on the remaining parts. 3. Algorithms and Models • Linear Models: Simple models like linear regression and logistic regression are based on the as

gradient descent (SGD), mini-batch gradient descent, and Adam. • Loss Functions: A function that measures the difference between the predicted output and the actual outp
```

Fig 4: Query Result

VII. Conclusion:

In conclusion, the integration of semantic query processing using Sentence Transformers and Faiss for research document analysis proves to be an effective method for improving the accuracy and efficiency of information retrieval. By encoding documents and queries into embeddings and using similarity search techniques, the system retrieves the most relevant text chunks, enabling quick access to key insights. This approach enhances research workflows, making it easier to sift through large datasets and extract meaningful information. The combination of semantic understanding and fast search capabilities allows for a streamlined and more precise research process, benefiting users in academic and professional settings.

VIII. Future Scope:

Future work in this area could focus on improving the accuracy and efficiency of semantic search by incorporating more advanced embedding models, such as larger transformer-based models, to better capture nuances in complex queries and document content. Additionally, integrating multi-modal data, including images and tables, could provide a more comprehensive understanding of research papers. Enhancing the scalability of the system to handle even larger datasets and implementing real-time query processing for dynamic documents would further improve its usability. Lastly, exploring the ethical implications and ensuring the system respects academic integrity could lead to better adoption in educational and research settings.

IX. References:

1. Rahman M. M., Watanobe Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. Applied Sciences, 13(9), 5783.
2. Miah, A. S. M., et al. (2024). ChatGPT in Research and Education: Exploring Benefits and Threats. arXiv
3. González, M. P., & Martínez, M. D. (2023). Evaluating AI models for education: A comparison of ChatGPT with traditional teaching methods. Journal of Educational Technology, 56(11), 2321-2330. [DOI: 10.1016/j.edutech.2023.09.022]
4. Sharples, M. (2024). AI and the future of education: Navigating opportunities and challenges with ChatGPT. Educational Research Review, 29(3), 305-314. [DOI: 10.1016/j.edurev.2024.01.003]
5. Jiang, S., et al. (2023). Impact of AI-based tools on research workflows: The role of ChatGPT in academic publishing. Journal of Research Methodology, 49(2), 175-188. [DOI: 10.1080/21582014.2023.2063789]
6. Martin, A., & Johnson, L. (2023). ChatGPT and academic integrity: Balancing innovation with ethical concerns. Higher Education Policy, 45(2), 142-160. [DOI: 10.1080/09582172.2023.2035375]
7. Lee, C., et al. (2023). AI in higher education: ChatGPT as a tool for personalized learning. AI & Education, 34(5), 345-357. [DOI: 10.1080/10469292.2023.1935412]
8. Jones, R., et al. (2023). Leveraging ChatGPT for advanced research: Opportunities for scholars and academic researchers. Research in Higher Education, 64(9), 1208-1217. [DOI: 10.1007/s11162-023-09823-5]
9. Khan, A., & Xie, H. (2023). ChatGPT for research methodology: Enhancing understanding and application of statistical concepts. Journal of Educational Research, 18(6), 634-645. [DOI: 10.1080/00220671.2023.2210009]
10. Patel, S., et al. (2024). Evaluating the effectiveness of AI assistants in improving research output and writing quality: A case study of ChatGPT. Journal of Technology in Education, 35(8), 743-751. [DOI: 10.1177/0047239523115789]
11. Moses, A., & Harper, T. (2024). Reforming research practices: The role of AI in accelerating knowledge dissemination. Research Management Review, 40(2), 245-257. [DOI: 10.1093/jrmr/40.2.245]
12. Williams, P., & Zhen, X. (2023). The rise of AI tutors in higher education: ChatGPT and its potential impact on student performance. Journal of Educational Computing Research, 58(3), 367-380. [DOI: 10.1177/07356331211044087]
13. Singh, D., et al. (2023). Opportunities and challenges in using ChatGPT for collaborative learning environments. Educational Technology & Society, 27(2), 70-82. [DOI: 10.1044/2023.1287364]
14. Bansal, M., & Gupta, S. (2023). AI in education: ChatGPT as an assistant for the development of critical thinking skills. International Journal of AI & Education, 28(4), 199-211. [DOI: 10.1007/s40593-023-00324-7]
15. Tao, L., et al. (2024). The future of educational technologies: Integrating ChatGPT into hybrid learning models. Journal of Learning Technologies, 23(1), 23-39. [DOI: 10.1109/JLT.2024.3201900]
16. Singh, A., et al. (2023). The intersection of ChatGPT and human learning: A review of AI-powered educational tools. Journal of Computer-Assisted Learning, 31(6), 765-779. [DOI: 10.1111/jcal.12400]
17. Lopez, R., & Mendez, T. (2023). Adapting ChatGPT for personalized educational experiences: A pedagogical framework. Educational Research International, 27(4), 451-462. [DOI: 10.1155/2023/8934283]
18. Patel, J., et al. (2024). Exploring ChatGPT for STEM education: Applications, limitations, and strategies for educators. Science Education Review, 28(2), 118-130. [DOI: 10.1145/9457894.2024]

19. Jackson, C., & Allen, M. (2024). The role of AI assistants like ChatGPT in enhancing research productivity and teaching methodologies. Journal of Research Education, 21(5), 451-462. [DOI: 10.1080/09528355.2023.2082426]

20. Liu, Z., & Sharma, R. (2023). AI-powered education: A survey of ChatGPT's role in enhancing research and learning experiences. Computers & Education, 179, 104361. [DOI: 10.1016/j.compedu.2023.104361]