



ChatGPT and Large Language Models: Unraveling Multifaceted Applications, Hallucinations, and Knowledge Extraction

Hozan K. Hamarashid¹✉ Lina Tofiq Karim,² Danial Abdulkareem Muhammed³

¹ Computer Science Institute, Sulaimani Polytechnic University, Sulaymaniyah, Iraq

² Computer Department, Private Computer Institute of Sulaimani, Sulaymaniyah, Iraq

³ Computer Department, College of Science, University of Sulaimani, Sulaymaniyah, Iraq

DOI: <https://doi.org/10.15294/ijcets.v11i1.54877>

Article History

Received : 6 August 2023

Accepted : 1 October 2023

Published : 30 November 2023

Keywords

ChatGPT, hallucinations, Large language model, natural language understanding, transfer model

Abstrak

Kemajuan baru-baru ini dalam kecerdasan buatan telah melahirkan model bahasa besar (LLM), terutama yang mampu memahami pembelajaran mendalam dan pemrosesan bahasa manusia. Model-model inovatif ini, yang diwakili oleh ChatGPT, memanfaatkan algoritma transformer yang mampu untuk memahami dan menghasilkan teks seperti bahasa manusia. Meskipun memiliki kemampuan, LLM rentan terhadap halusinasi, yakni menyajikan ketidakakuratan sebagai fakta. Menggali pengetahuan yang terdapat pada model-model tersebut sangat menarik. Artikel ilmiah ini menggali penggunaan serbaguna ChatGPT sebagai model bahasa besar, mengeksplorasi kemampuannya untuk menghasilkan ringkasan singkat dari teks yang luas dan mengatasi fenomena halusinasi. Selain itu, penelitian ini menyajikan kemampuan ChatGPT dalam mengekstrak informasi penting dari kumpulan data besar, dalam hal ini mengkaji sepuluh artikel yang paling sering dikutip dalam ranah kecerdasan buatan. Artikel ilmiah ini bukan hanya menggambarkan penerapan praktis kecerdasan buatan dalam penelitian, melainkan juga menekankan pentingnya memanfaatkan model-model ini secara efektif, khususnya dalam konteks melakukan tinjauan literatur yang akurat.

Abstract

Recent strides in artificial intelligence have given rise to large language models (LLMs), particularly those adept at deep learning and human language processing. These innovative models, exemplified by ChatGPT, leverage robust transformer algorithms to comprehend and generate text like human language. Despite their prowess, LLMs are prone to hallucinations, presenting inaccuracies as facts. Unraveling the retained knowledge within these models proves challenging. This paper delves into the multifaceted use of ChatGPT as a large language model, exploring its capability to generate concise summaries from extensive text and addressing the phenomenon of hallucinations. Furthermore, the study showcases ChatGPT's proficiency in extracting crucial information from large datasets, such as revealing the top ten cited papers in the realm of artificial intelligence. The manuscript not only illustrates the practical application of artificial intelligence in research but also emphasizes the importance of utilizing these models effectively, particularly in the context of conducting precise literature reviews.

✉ Corresponding author :

Address: Computer Science Building 3rd Floor, Computer Science Institute, Sulaimani Polytechnic University, Sulaimani, Iraq, 46001
E-mail: hozan.khalid@spu.edu.iq

INTRODUCTION

Despite being a relatively new technology, large language models have already made a significant impact in the field of natural language processing. They have been applied to create new chatbots and virtual assistants capable of comprehending challenging questions and providing responses that are more human-like (Lund et al., 2023). Additionally, they have been employed to generate fresh content, including news pieces, product descriptions, and even entire books. GPT, one of the largest and most powerful language models ever created, is among the most outstanding. It comprises 175 billion parameters and has been trained on a vast dataset of over 45 terabytes of text data.

GPT has been utilized for a wide range of applications, from chatbots and virtual assistants to creative writing and even coding (Bryant, 2023). The use of large language models has revolutionized the field of artificial intelligence and has numerous applications. Among these models, GPT stands out as a robust instrument that has gained widespread acceptance. Numerous applications of ChatGPT have proven effective, such as chatbots, content creation, language translation, personalized suggestions, and even medical diagnosis and therapy. Due to its ability to generate human-like responses, comprehend natural language, and adapt to various circumstances, it has found success in several applications (Lund et al., 2023). ChatGPT, as a large language model (LLM), proves to be an effective tool for natural language processing (NLP) due to its adaptability and precision. However, it does have certain drawbacks, namely the propensity for biased responses and the potential to reinforce negative linguistic habits.

The study of natural language understanding (NLU), a branch of NLP, focuses on assessing how well computers can comprehend and analyze spoken language. The development of NLU systems that can comprehend and respond to human language in a more natural and human-like fashion has utilized large language models, such as GPT (Bryant, 2023). One of the most prevalent applications for NLU systems is in chatbots. These computer programs, designed to mimic human speech, are frequently employed to provide customer service, respond to inquiries, and even entertain users. However, traditional chatbots often struggle to comprehend and respond to complex questions, frequently providing irrelevant or unsatisfying answers.

Large language models, like GPT, can be

employed to build Natural Language Understanding (NLU) systems that comprehend and respond to complex questions in a more natural and human-like manner. Utilizing extensive text datasets, these algorithms can be trained to generate responses that are grammatically sound, coherent, and relevant to the user's question (Hamarashid et al., 2021). One of the advantages of these large language models is their adaptability for customization in specific domains or activities. For example, by using a dataset of support tickets and customer reviews, a chatbot designed to provide customer support for a particular product can be trained to address common questions and issues.

A powerful approach to creating Natural Language Understanding (NLU) systems that can understand and respond to human language more realistically is to leverage large language models for chatbots, such as GPT (Hamarashid et al., 2021; Liu et al., 2022). These systems can be applied to various tasks, including coding, creative writing, customer service, and question-answering.

ChatGPT is an expansion of the Large Language Model (LLM) family of machine learning models for Natural Language Processing (Hamarashid et al., 2021; Abid et al., 2021). LLMs ingest huge amounts of textual information, inferring associations between words in the text (Hamarashid et al., 2021). As computer power has improved over the past few years, these models have expanded. LLMs become more powerful as their input datasets and parameter space increase (Abid et al., 2021).

Predicting a word in a string of words is a prerequisite for training language models. The two most frequent observations of this are next-token prediction and masked language modeling. In this fundamental sequencing method, often implemented using a Long-Short-Term-Memory (LSTM) model, the model fills in the gap with the word that is statistically most likely given the context (Hamarashid et al., 2022; Hamarashid, 2021). An ML model was trained to assess texts throughout the learning phase and was rewarded for each accurate response. Additionally, a self-attention mechanism was included to boost learning efficiency by prioritizing significant combinations and giving less weight to unimportant ones.

Natural language processing models meticulously analyze vast quantities of textual data, discerning intricate relationships among words

within the text. The robust growth in the size and functionality of these machine learning models in recent years can be attributed to advancements in computational power. Large Language Models (LLMs) exhibit heightened reliability as both the input dataset and parameter space expand. The prevalent technique of next-token prediction, in language model training, involves the utilization of the Long-Short-Term-Memory (LSTM) neural network and masked-language modeling to forecast the subsequent word in a sequence (Hamarashid et al., 2022; Hamarashid, 2021). The latter method entails selecting the word that, within the given context, is statistically most likely to complete a phrase.

Within ChatGPT, the Natural Language Understanding (NLU) model plays a pivotal role in ascertaining the user's intent and extracting relevant data from their messages. For example, if a user seeks a recommendation for a restaurant, the NLU recognizes the intent as "restaurant recommendation" and extracts pertinent details, including the establishment's cuisine, location, and pricing range. Subsequently, the model generates a response based on the extracted data (Purani, 2023). To ensure the production of responses aligned with the user's intent and context, the large language model is strategically employed. In summary, the cohesive integration of the large language model and NLU results in a responsive and engaging chat experience within ChatGPT.

LITERATURE REVIEW

Large Language Models have become a crucial component in artificial intelligence, finding applications in various subjects such as machine translation, content or text generation, natural language understanding, and sentiment analysis. These models have undergone significant development recently, showcasing outstanding advancements in the field of natural language processing (NLP). This literature review provides a comprehensive overview of the evolution of massive language models, tracing their roots from early beginnings to the present (Lund et al., 2023; Bryant, 2023).

Large language models boast a long history dating back to the development of natural language processing (NLP) in the 1950s and 1960s. Rule-based systems and statistical techniques were the driving forces behind the early endeavors in this field. The Georgetown-IBM experiment, which successfully transcribed 60

Russian words into English in 1954, marked the inception of machine translation research. Despite its limitations, this experiment demonstrated the potential of applying computational tools for language processing (Lund et al., 2023).

A. Hidden Markov Models (HMMS)

Statistical techniques became more prevalent in NLP during 1980s and 1990s. For their analysis, researchers used Hidden Markov Models (HMMs) and n-gram language models to forecast the likelihood of a set of words. While HMMs described sequences of hidden states and offered a more organized approach, the n-gram model employed the co-occurrence frequencies of words in a dataset to create predictions (Bryant, 2023).

The following equation represents the HMM: $Q = q_1q_2 \dots q_N$ a set of N states $A = a_{11} \dots a_{ij} \dots a_{NN}$ a transition probability matrix A , each a_{ij} representing the probability of moving from state i to state j , s.t. $\sum_{j=1}^N a_{ij} = 1 \forall i$, $O = o_1o_2 \dots o_T$ a sequence of T observations, each one drawn from a vocabulary $V = v_1, v_2, \dots, v_V$, $B = b_i(o_t)$ a sequence of observation likelihoods, also known as emission probabilities, each expressing the probability of an observation o_t being generated from a state i , $\pi = \pi_1, \pi_2, \dots, \pi_N$ an initial probability distribution over states. π_i is the probability that the Markov chain will start in state i . Some states j may have $\pi_j = 0$, meaning that they cannot be initial states. Also, $\sum_{i=1}^N \pi_i = 1$.

Markov Assumption equation is:

$$P(q_i | q_1 \dots q_{i-1}) = P(q_i | q_{i-1}) \text{-----(1)}$$

B. The Development of Neural Networks

The backpropagation technique, which made it possible to train multi-layer neural networks more efficiently, was the driving force for the rebirth of neural networks at the turn of the century. Bengio et al.'s ground-breaking 2001 proposal of feed-forward neural networks for language modelling established the principles of deep learning in NLP. However, these models remained relatively small compared to those created subsequently due to computing constraints (Bryant, 2023; Hamarashid et al., 2022).

There are three steps to perform in any neural network:

1. We take the input variables and use the linear combination equation:

$$Z = W_0 + W_1X_1 + W_2X_2 + \dots + W_nX_n \text{---(2)}$$

to compute the output or the predicted Y values, called the Ypred.

2. Calculate the loss or the error term. The error term is the deviation of the actual values from the predicted values.
3. Minimize the loss function or the error term.

C. Word Embedding

Through the Word2Vec model, which represents words as continuous vectors in a high-dimensional space, the idea of word embeddings was introduced in 2013. Significant improvements in NLP tasks were made as a result of this method's superior ability to capture the semantic links between words compared to earlier approaches (Bryant, 2023; Hamarashid, 2021). The equation of word embedding is illustrating in equation 3.

$$P(W_o|W_i) \rightarrow \text{probability of getting a context word } W_o \text{ given the input word } W_i \text{ (3)}$$

D. Developing Long Short-Term Memory (LSTM) From Recurrent Neural Networks (RNNs)

Language modeling was a natural application for RNNs because they could handle data sequences. However, due to the vanishing gradient issue, they encountered difficulties with long-range dependencies. This problem was addressed by the Long Short-Term Memory (LSTM) architecture, introduced in 1997. For NLP applications like sentiment analysis and machine translation, LSTMs have become a prominent option (Bryant, 2023; Hamarashid et al., 2022).

The complete equations for LSTM phases are addressed in equations 4 to 9:

$$f_t = \sigma(w_f [h_{(t-1)}, x_t] + b_f) \text{ (4)}$$

$$i_t = \sigma(w_i [h_{(t-1)}, x_t] + b_i) \text{ (5)}$$

$$\check{c}_t = \tanh(w_c [h_{(t-1)}, x_t] + b_c) \text{ (6)}$$

$$c_t = f_t * c_{(t-1)} + i_t * \check{c}_t \text{ (7)}$$

$$o_t = \sigma(w_o [h_{(t-1)}, x_t] + b_o) \text{ (8)}$$

$$h_t = o_t * \tanh(c_t) \text{ (9)}$$

E. The Architecture of Transformers

The Transformer architecture, introduced in 2017, revolutionized natural language processing by leveraging self-attention techniques instead of the recurrent layers commonly found

in RNNs. This innovative approach significantly decreased model training times and allowed for the parallel processing of sequences. The Transformer's impact was profound, paving the way for the development of numerous large-scale language models in the subsequent years (Uszkoreit, 2017).

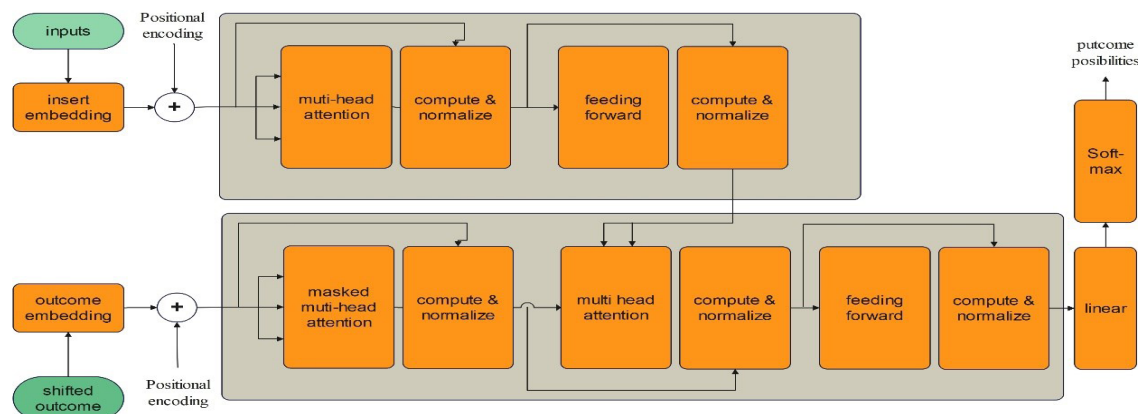
This groundbreaking architecture not only addressed the limitations associated with traditional recurrent layers but also offered enhanced efficiency in capturing intricate dependencies within sequences. Its success marked a paradigm shift in language modeling, enabling the creation of more sophisticated and context-aware models that excel in various NLP tasks. The versatility of the Transformer architecture extended beyond just language models, influencing a broad spectrum of applications in artificial intelligence and machine learning. Its scalability, coupled with its ability to capture long-range dependencies effectively, made it a cornerstone in the evolution of modern language processing techniques.

F. Hidden Markov Models (HMMS)

The GPT series from OpenAI made its debut in 2018 with GPT, followed by GPT-2 in 2019 and GPT-3 in 2020. The Transformer architecture forms the bedrock for the GPT models, conceptualized as generative language models. Utilizing unsupervised learning on extensive text datasets, they generate coherent and contextually relevant text, evolving in size and capability throughout the series (Uszkoreit, 2017). With 1.5 billion parameters, GPT-2 faced concerns that led to a cautious approach to its full release. Despite these worries, OpenAI eventually released the entire model, accompanied by guidelines for responsible use and study.

The release of GPT-3 in 2020 marked a substantial leap, boasting 175 billion parameters. This model exhibited remarkable advancements in both size and performance, unlocking applications in content creation, translation, summarization, and various other fields. However, the growth in size and capabilities of these massive language models has raised ethical concerns. Issues such as bias in training data, environmental impact, and the potential for generating inaccurate or harmful information have come to the forefront. Recognizing the importance of responsible AI development, organizations like OpenAI are actively addressing these concerns. They are committed to mitigating biases, mini-

Figure 1 Methodology



mizing environmental impact, and ensuring the secure deployment of these models. The subsequent discussion will delve into the methodology employed to achieve these goals.

METHOD

In this section, the methodology of the research paper is addressed. To execute the methodology steps, various tasks need to be accomplished. Figure 1 illustrates the methodology.

As depicted in Figure 1, the term "autoregressive generation" refers to this process, illustrating how a large language model operates. Notably, such models, featuring multiple layers in both encoder and decoder, excel in handling lengthy input texts, large context vectors, and intricate ideas. Recent strides in deep learning models capable of processing human languages are recognized as large language models (LLMs), utilized in innovative ways. A well-trained deep-learning model, commonly referred to as a large language model, comprehends and generates text akin to human expression, with the intricate workings managed by a powerful transformer model.

In the context of LLMs, hallucination occurs when the model produces false, illogical, or unreal language, not specifying its response source. Unlike databases or search engines, LLMs extrapolate from provided prompts, generating text correlated with the prompt, even if not always backed by training data. To comprehend hallucination using text, a two-letter bigrams Markov model can be created. Extracting a lengthy passage, generating a table of adjacent letter pairs, and counting entries can reveal statistical patterns. Despite the model's complexity, hallucination in LLMs is not significantly more intricate. It results from a lack of contextual awareness, necessitating the abstraction of training data and prompts, potentially leading to infor-

mation loss. Additionally, noise in the training data may introduce skewed statistical patterns, causing unexpected model behavior. Figure 2 outlines the steps of conducting the research paper.

A. Utilizing Hallucinations

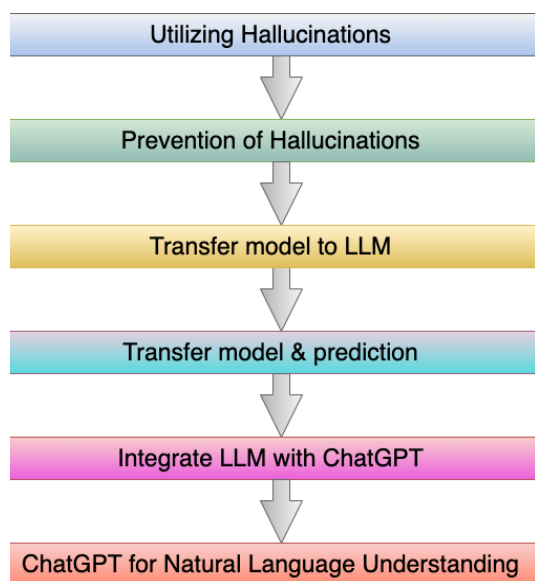
Researchers might consider hallucinations as a characteristic of extensive language models, envisioning them as a means to enhance creativity (Rebuffel et al., 2022). For example, when soliciting a fantasy narrative plot from ChatGPT or another large language model, the goal is to encourage the generation of entirely new characters, settings, and plots rather than recycling existing ones. This is achievable only if the models refrain from referencing the data on which they were trained. Hallucinations may also be desirable when aiming for diversity, such as when seeking suggestions or ideas. In essence, it is akin to prompting the models to generate novel concepts.

Researchers seek variations on known concepts and those that may be latent in the training data. The use of hallucinations can open up numerous alternatives and possibilities. Language models often incorporate a temperature parameter. Unlike adjusting it through the online interface, in ChatGPT, the temperature can be modified using the API, serving as a randomization factor. As the temperature increases, the likelihood of hallucinations becoming more prevalent also rises.

B. Mitigating Hallucinations

Language models are neither databases nor search engines. It is impossible to entirely eliminate hallucinations. It is frustrating when models produce text with subtle typos. If the corrupted training data is identified as the cau-

Figure 2 The steps of the research methodology



se of hallucinations, the data can be cleaned up, and the model can be retrained. However, the majority of models are too large for self-paced training, and on common hardware, tweaking an existing model can be challenging.

Human intervention in the outcome, along with requesting the model to regenerate if a significant error occurs, may be the most effective forms of mitigation. Controlled creation is another method to prevent hallucinations. This involves providing the model with sufficient information and setting limitations in the prompt. As a result, the model has limited freedom for hallucinations. The purpose of prompt engineering is to describe the role and situation to the model, directing the generation process and preventing it from having unrestricted hallucinations.

C. The Improvement from Transfer Model to Large Language Model

Text is perceived as a collection of words by humans. Word groups are arranged into sentences, and documents consist of sections, chapters, and paragraphs. However, for computers, text is merely a series of characters. A model based on recurrent neural networks can be created to enable robots to understand text. This model processes all input text and produces an output by handling one word or character at a time. Occasionally, this model may forget what happened at the beginning of the run until the conclusion of the series is reached.

A transformer model was introduced

in a study titled "Attention is All You Need" (Wang and Chen, 2023), based on the attention mechanism. The attention mechanism allows the model to perceive the entire phrase or paragraph at once, unlike recurrent neural networks, which process one word at a time. This capability enables the transformer model to better comprehend the context of each word. Transformers serve as the foundation for many modern language processing models.

Before processing with a transformer model, it is necessary to tokenize a text input into a series of words. These tokens are then encoded as integers, creating embeddings—vector-space representations of the tokens that preserve their meaning. The transformer’s encoder converts the embeddings of each token into a context vector. An illustration of a text string, its tokenization, and vector embedding can be observed.

For instance, the term "rapid way" in the text is tokenized into "rapid" and "way," noting that tokenization can also be performed using sub-words. The context vector, similar to the substance of the entire input, is used by the transformer decoder to produce output based on hints. For example, you can provide the first input as a hint to the transformer decoder and let it generate the term that naturally follows. This process can be repeated, starting with a leading sentence, to generate a full paragraph using the same decoder, where the hint is the previously generated next-word.

D. Transfer Model Prediction

Recurrent neural networks can reasonably anticipate the following word in a text. In addition to the redundancy in language, there are restrictions in human language (i.e., grammar) that limit the usage of words in various contexts within a phrase. The English language consists of 26 letters, not including spaces. However, "Prediction and Entropy of Printed English" reveals an entropy of just 2.1 bits per letter. If letters were used arbitrarily, the entropy would be 4.8 bits, making it simpler to anticipate what would follow next in a human language document. Transformer models, in particular, are effective at making such predictions in machine learning models. By repeating this method, a transformer model can create the full passage word for word. But what does a transformer model think about grammar?

Grammar essentially describes the way words are used in language, classifying them into

different parts of speech and requiring a specific sequence within a phrase. Despite this, listing every grammatical rule is challenging. In reality, the transformer model intuitively learns these principles through examples rather than formally storing them. The transformer model needs to be large enough for it to learn more than just grammatical rules and expand to encompass the concepts presented in those examples.

E. Integrate Large Language Model with ChatGPT

A large-scale transformer model is a massive language model, typically challenging to operate on a single machine. Consequently, the service is provided through an API or a web interface. As you might anticipate, such a large model needs to analyze a considerable amount of text before it can grasp the structures and patterns of language. For example, the GPT-3 model that supports the ChatGPT service was trained using vast volumes of text data from the internet, encompassing books, articles, websites, and various other materials. When presented with a prompt or inquiry, the model can generate coherent and contextually appropriate replies because, during training, it learns the statistical correlations between words, phrases, and sentences.

As a result of distilling information from this extensive body of data, the GPT-3 model can understand multiple languages and possess in-depth knowledge of various topics. Consequently, it can produce text in a variety of styles. While it may be impressive that a large language model can translate, summarize texts, and respond to questions, it should not be surprising, as these models operate based on unique "grammars" corresponding to the provided text, also known as prompts.

F. Utilizing Chatgpt for Natural Language Understanding

Text should be relatively simple for ChatGPT to interpret as a language model (Logunova, 2023; Weberbos, 2023). To maximize its natural language comprehension abilities, you must be careful to prevent it from going into a trance. Hallucinations occur when you let it generate text that isn't relevant to the situation. This can be prevented by specifically stating the use of only the information provided for the task in the prompt and nothing else. Nevertheless, keep in mind that ChatGPT might not fully comprehend

your message. It wouldn't be aware if any material was taken out of context, including domain-specific acronyms and jargon. ChatGPT may also struggle to extract data if it isn't particularly clear-cut. You could try several times to regenerate the second example from above, as the text specified it as a proportion of another volume, leading to occasional difficulties for ChatGPT in locating the precise passenger volume.

RESULT AND DISCUSSION

In this section we illustrate the results of our analysis upon ten papers we mentioned earlier. Assume a meeting transcript featuring conversations among different participants. While meeting minutes should simplify understanding the outcome, the transcript would be a challenging way to monitor the discussion and see how a result is achieved. Requesting ChatGPT to create meeting minutes from a transcript is possible.

If we have the following transcript:

Good evening, everyone. I am Hozan Khalid, the Sulaimani Polytechnic University's IT Committee head. Please remember to keep your phones silent unless you are speaking. I will now make a call to call the IT committee meeting to order. Bryar, would you kindly ring the roll?

- Thank you, everyone, have a wonderful weekend
- Have a wonderful weekend
- Thank you

Although the meeting was brief, the transcript took up several pages. Using this prompt, you can condense the transcript into one page of minutes. You provided ChatGPT with the transcript and instructions on what to include in the output to ensure all necessary information was incorporated in the prompt. The results from ChatGPT are as follows:

Meeting: IT Committee Meeting
 Date : (Input Date)
 Time : (Input Time)
 Venue: (Input Venue)
 Participants:
 • Ali Muhamad (Representing Director Hozan Khalid)

- Mahmood Hasan (Representing Commissioner Bryar Ahmad)
- Lina Tofiq (Representing Commissioner Adam Amanj)
- Arkan Mahmud (Representing Speaker Azad Ali)

Agenda:

- Approval of the committee’s most recent meeting’s minutes.
- Resolution urging approval of the 125th series resolution permitting the insurance of up to 1.6 billion in fiscal 23 series D future tax-secured bonds by the Board of Directors.

Important Choices:

- Acceptance of the committee’s most recent meeting’s minutes. The minutes were looked over and overwhelmingly approved.
- Resolution urging the Board of Directors to ratify the 125th series resolution approving the issue of up to 1.6 billion of fiscal 23 series D future tax-secured bonds. The resolution was announced, and the anticipated subseries breakdown was given. The resolution received a unanimous vote of support from the committee members and was adopted.

Items for Action:

- None are referenced in the transcript.

Postponed:

- The IT committee meeting was postponed.

To emphasize a crucial point: Specific information about the meeting, such as dates, hours, and venue, is not included in the provided transcript. Given the circumstances and the organization’s needs, please fill in the blanks accordingly.

In comparison to the transcript, the minutes are much neater and simpler to read. It is evident from the minutes that this transcript encompasses two meetings planned to be one in this particular transcript. However, it might still need some polishing if you desire it in a specific format. There are several placeholders that must be filled in for information that should be in the minutes but cannot be found in the transcript. In contrast to summarizing the transcript, this task shouldn’t be too challenging. ChatGPT has

been trained to interpret language and can extract essential information, making it capable of completing this task easily. With the right prompt, you can effectively turn ChatGPT into a useful secretary by instructing it to replicate the important data in a specific format.

However, some meetings can be quite lengthy, and the transcript might exceed ChatGPT’s context length restriction because it is very long. In that situation, ChatGPT will not continue. You may want to divide the transcript into sections and have ChatGPT summarize each separately first, and then combine the results to produce the minutes.

A. Unstructured Information Extraction

ChatGPT can be used to extract unstructured information because it can comprehend text. Unlike lists and tables, unstructured information is found inside paragraphs of text. It is challenging to search because there isn’t any obvious pattern. To extract unstructured data and create a list, you can leverage ChatGPT’s language understanding feature. Consider the following instance:

“Coronavirus, often referred to as COVID-19, which is an infectious infection that affects the human respiratory system. Covid 19 refers to a novel coronavirus disease that was discovered in 2019. Our daily lives have been impacted by the Coronavirus. Millions of individuals have been impacted by this pandemic and are either sick or dead due to the disease. Fever, cold, cough, bone pain, and respiratory issues are the most typical signs of this viral infection. Patients with the coronavirus may also experience symptoms like fatigue, a sore throat, muscle soreness, and a loss of taste or smell in addition to these ones. Around the world, the pandemic has caused significant social and economic disruption, including the biggest global recession since the Great Depression. Supply chain instability led to widespread shortages of supplies, particularly food supplies. Pollution fell by an unprecedented amount as human activity decreased. Throughout 2020 and 2021, many jurisdictions closed all or part of their educational institutions and public spaces, and numerous events were postponed or cancelled. Political tensions have increased as a result of false information spreading through social media and the media. The pandemic has

brought up questions of racial and geographic discrimination, health fairness, and how to strike a balance between the needs of public health and the rights of individuals.” (English Summary, 2023)

Answer the following questions:

1. When coronavirus disease was discovered?
2. How many people were infected by this disease?
3. How many people has infected in Kurdistan?
4. What was the date referenced for the infected person for the first time in the Turkey?

Following is the ChatGPT outcome:

1. Coronavirus disease was discovered in 2019.
2. Millions of individuals have been impacted by this pandemic.
3. The text does not contain information on the number of infected people in Kurdistan.
4. The date referenced for the infected person for the first time in the Turkey is not mentioned in the text.

The distilled outcome is easier to extract and can be utilized in various ways. Additionally, please note that it will inform you if the delivered text does not contain the information you were looking for. This is a result of your instruction, which directed it to respond solely based on the content.

B. Conducting Literature Review With LLM

Assume that Person₁ is a researcher and a postgraduate student. He intends to conduct a review of research on artificial intelligence. To formulate research questions and assist in the literature evaluation process, he aims to identify any gaps that might exist in this field. You are about to observe Person₁ engaging in a one-on-one conversation with ChatGPT.

(Person₁): Could you please provide top ten works in artificial intelligence have received the most citations?

[ChatGPT]: I can give you a general overview of several key works in the area that received a lot of citations as of the knowledge

cutoff date. The following papers have made significant contributions to the comprehension and advancement of artificial intelligence:

1. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models Using AlphaFold to augment protein structure database coverage.
2. ColabFold: making protein folding accessible to all an open-source and efficient protein folding model.
3. Hierarchical Text-Conditional Image Generation with CLIP Latents DALL-E 2, complex prompted image generation that left most in awe.
4. A ConvNet for the 2020s A successful modernization of CNNs at a time of boom for Transformers in Computer Vision.
5. PaLM: Scaling Language Modeling with Pathways Google’s mammoth 540B Large Language Model, a new MLOps infrastructure, and how it performs.
6. Highly accurate protein structure prediction with Alpha Fold, a breakthrough in protein structure prediction using Deep Learning.
7. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows A robust variant of Transformers for Vision.
8. Learning Transferable Visual Models from Natural Language Supervision CLIP, image-text pairs at scale to learn joint image-text representations in a self-supervised fashion
9. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Famous position paper is very critical of the trend of ever-growing language models, highlighting their limitations and dangers.
10. Emerging Properties in Self-Supervised Vision Transformers DINO, showing how self-supervision on images led to the emergence of some sort of proto-object segmentation in Transformers.

Compiling publications with significant citations and influence is a time-consuming task, particularly when data is scattered across various sources. However, integrating ChatGPT as a research assistant can significantly stream-

line this process. Muhammad, having acquired a list of the top 10 most-cited papers, now seeks to identify potential research gaps to form the basis of his thesis.

While ChatGPT proves to be a valuable tool, it is imperative to underscore that high-quality research necessitates human involvement, drawing upon the researcher's knowledge and expertise in the field. Key takeaways from this chapter include the notion that ChatGPT can enhance research efforts, enabling researchers to focus more on the substantive aspects of their work. Examples and sample prompts are provided to illustrate its utility across various stages of the research process. To ensure ChatGPT produces relevant and valuable responses, researchers are advised to meticulously craft their prompts, providing ample detail and context. Researchers bear the responsibility for the accuracy and integrity of their findings, adhering to ethical guidelines when utilizing AI for research assistance.

Consequently, the integration of artificial intelligence, exemplified by employing ChatGPT, has alleviated and resolved several challenges. Notably, interactions involving straightforward questions and answers with ChatGPT have proven effective. Moreover, the application extends to requesting literature reviews on diverse subjects, with a specific case study focusing on artificial intelligence in this research. Additionally, the retrieval and summarization of pertinent information have been rendered beneficial through this approach.

The findings of this research bear significant implications for the field of educational technology, encompassing both studies and practices. By utilizing ChatGPT as a research assistant and leveraging its capabilities to streamline tasks such as literature reviews and information retrieval, researchers in educational technology stand to benefit from enhanced efficiency and productivity. The integration of artificial intelligence, as demonstrated in this research, opens avenues for the development of innovative tools and methodologies that could revolutionize educational technology practices. Furthermore, the emphasis on crafting careful prompts and ensuring ethical guidelines are followed establishes a framework for responsible and effective utilization of AI in educational technology research. This research contributes to the ongoing discourse in the field, offering insights that may shape future developments and applications, ultimately fostering advancements

in educational technology research and practice.

CONCLUSION

In conclusion, this paper illuminates the utilization of a large language model (LLM) to retrieve essential information from extensive text databases. Additionally, the use of hallucination in LLM has been explored. Moreover, a transfer model was incorporated for prediction purposes. Subsequently, LLM was integrated to extract and summarize textual content. The evaluation of ChatGPT for natural language understanding focused on extracting information on various topics, including retrieving the top ten references in artificial intelligence. Furthermore, a literature review was conducted to import and summarize the most crucial texts pertaining to the requested literature review. While this approach aids users in efficiently retrieving texts and information, it is advised that users verify the reliability of the retrieved texts.

REFERENCES

- Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-muslim bias in large language models. *In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 298-306).
- Bryant, S. (2023). Assessing GPT-4's role as a co-collaborator in scientific research: a case study analyzing Einstein's special theory of relativity. *Discov Artif Intell* 3, 26. <https://doi.org/10.1007/s44163-023-00075-3>
- English Summary, (2023), 2 Minute Speech On Covid-19 In English. Available at: https://english-summary.com/2-minute-speech-on-covid-19-in-english/#google_vignette. Accessed on (10/3/2023)
- Hamarashid, H.K, Qader, S., A. Saeed, S., Hassan, B., & A. Ali, N. (2022, June 30). Machine Learning Algorithms Evaluation Methods by Utilizing R. *UKH Journal of Science and Engineering*, 6(1), 1-11. <https://doi.org/https://doi.org/10.25079/ukhjse.v6n1y2022.p1-11>
- Hamarashid, H.K. (2021). Modified Long Short-Term memory and Utilizing in Building sequential model. *International Journal of Multidisciplinary and Current Research*, 2021/04/Paper2207-211.
- Hamarashid, H.K., Saeed, S.A. & Rashid, T.A. (2021). Next word prediction based on the N-gram model for Kurdish Sorani and Kurmanji. *Neural Comput & Applic* 33, 4547-4566. <https://doi.org/10.1007/s00521-020-05245-3>
- Hamarashid, H.K., Saeed, S.A. & Rashid, T.A. (2022). A comprehensive review and evaluation on text predictive and entertainment systems. *Soft Comput* 26, 1541-1562 (2022). <https://doi.org/10.1007/s00521-020-05245-3>

- org/10.1007/s00500-021-06691-4
- Liu, N., Wang, Z., Baraniuk, R., & Lan, A. (2022, December). GPT-based Open-ended Knowledge Tracing. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 3849-3862).
- Logunova, I. (2023), Inside ChatGPT's Brain: Large Language Models. Available at: <https://se-rokell.io/blog/language-models-behind-chat-gpt>. Accessed on (12/3/2023)
- Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science & Technology*, 74(5), 570-581. <https://doi.org/10.1002/asi.24750>
- Purani, A. (2023), What is ChatGPT?. Available at: <https://www.sevensquaretech.com/what-is-chatgpt>. Accessed on (5/2/2023)
- Rebuffel, C., Roberti, M., Soulier, L. Scoutheeten, G., Cancelliere, R., & Gallinari, P. (2022). Controlling hallucinations at word level in data-to-text generation. *Data Min Knowl Disc* 36, 318-354 (2022). <https://doi.org/10.1007/s10618-021-00801-4>
- Uszkoreit, J. (2017). Transformer: A Novel Neural Network Architecture for Language Understanding, a google research. Available at: <https://blog.research.google/2017/08/transformer-novel-neural-network.html>. Accessed on (18/3/2023)
- Wang, J., Chen, Y. (2023). Transfer Learning for Natural Language Processing. In: Introduction to Transfer Learning. Machine Learning: Foundations, Methodologies, and Applications. Springer, Singapore. https://doi.org/10.1007/978-981-19-7584-4_16
- Weberbos, B. (2023), Conversation with ChatGPT. Available at: <https://www.patentkinetics.com/>. Accessed on (6/3/2023)