IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 November 2024, accepted 9 December 2024, date of publication 23 December 2024, date of current version 14 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3521596

SURVEY

From Information Overload to Lucidity: A Survey on Leveraging GPTs for Systematic Summarization of Medical and Biomedical Artifacts

BALAMURUGAN PALANISAMY[1], ARJAB CHAKRABARTI[2], ANUSHKA SINGH[2], VIKAS HASSIJA[2], G. S. S. CHALAPATHI[1], (Senior Member, IEEE), AND AMIT SINGH[3]

[1]Department of Electrical and Electronics Engineering, BITS Pilani, Pilani Campus, Pilani 333031, India
[2]School of Computer Engineering, KIIT, Bhubaneswar 751024, India
[3]Department of Mechanical Engineering, BITS Pilani, Pilani Campus, Pilani 333031, India

Corresponding author: Amit Singh (amit.singh@pilani.bits-pilani.ac.in)

This research was funded by the Birla Institute of Technology and Science under its Open Access (OA) publishing support scheme.

ABSTRACT In medical research, the rapid proliferation of condition-specific studies has led to an information overload, making it challenging for researchers and practitioners to stay abreast of the latest findings. This paper presents a comprehensive survey on leveraging Generative Pretrained Transformers (GPTs) to summarize medical and biomedical artifacts systematically. We delve into the current applications of GPTs in this domain, discussing their role in understanding and summarizing research papers, medical dialogues, and medical records. Through a comparative analysis of recent studies and methodologies, we highlight the effectiveness of GPTs in distilling complex medical information into concise, understandable summaries. Our survey underscores the potential of GPTs as a tool for navigating the information overload in medical research and bringing clarity to healthcare professionals. This transformation will enhance patient care and outcomes, such as improving the accessibility and comprehensibility of medical research, assisting in rapid information retrieval, and facilitating the summarization of complex medical studies for broader audiences.

INDEX TERMS Biomedical, ChatGPT, Generative Pretrained Transformers, healthcare, medical, natural language processing, summarization.

I. INTRODUCTION

In the rapidly evolving medical research landscape, the proliferation of scholarly articles presents both an opportunity and a challenge. The domain of medical research is dynamic and ever-evolving, with new research findings and advancements emerging at an unprecedented rate [1]. This rapid pace of development is fuelled by the global collaborative efforts of researchers, clinicians, and academics, who contribute to the vast body of medical literature. While this wealth of knowledge is undoubtedly a boon for advancing medical science, it also presents a significant challenge. The sheer volume of research articles, case studies, clinical trials, and reviews can be overwhelming, even for the most diligent medical professionals [2], [3]. This challenge is further compounded when focusing on medical evidence, diagnosis, and treatment because comprehensive and up-to-date knowledge is paramount [4].

The medical domain is overloaded with information. Consequently, it is required to get precise information. The need for an organized methodology to condense and summarize medical information is unquestionable [5]. Such a methodology would expedite the literature review process and safeguard against the inadvertent exclusion of pivotal studies [6]. It promises to give healthcare professionals the means to deftly traverse the expansive ocean of data, procuring pertinent insights and remaining informed about the burgeoning developments within their specialties.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

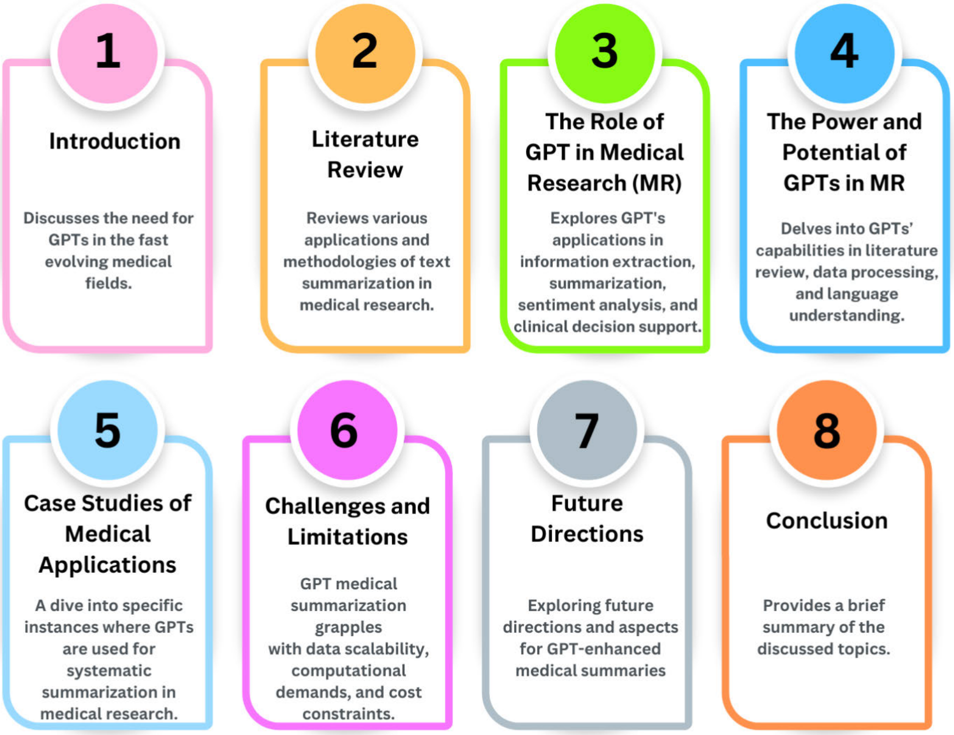

1
Introduction
Discusses the need for GPTs in the fast evolving medical fields.
2
Literature Review
Reviews various applications and methodologies of text summarization in medical research.
3
The Role of GPT in Medical Research (MR)
Explores GPT's applications in information extraction, summarization, sentiment analysis, and clinical decision support.
4
The Power and Potential of GPTs in MR
Delves into GPTs' capabilities in literature review, data processing, and language understanding.
5
Case Studies of Medical Applications
A dive into specific instances where GPTs are used for systematic summarization in medical research.
6
Challenges and Limitations
GPT medical summarization grapples with data scalability, computational demands, and cost constraints.
7
Future Directions
Exploring future directions and aspects for GPT-enhanced medical summaries
8
Conclusion
Provides a brief summary of the discussed topics.

FIGURE 1. Sections of the paper.

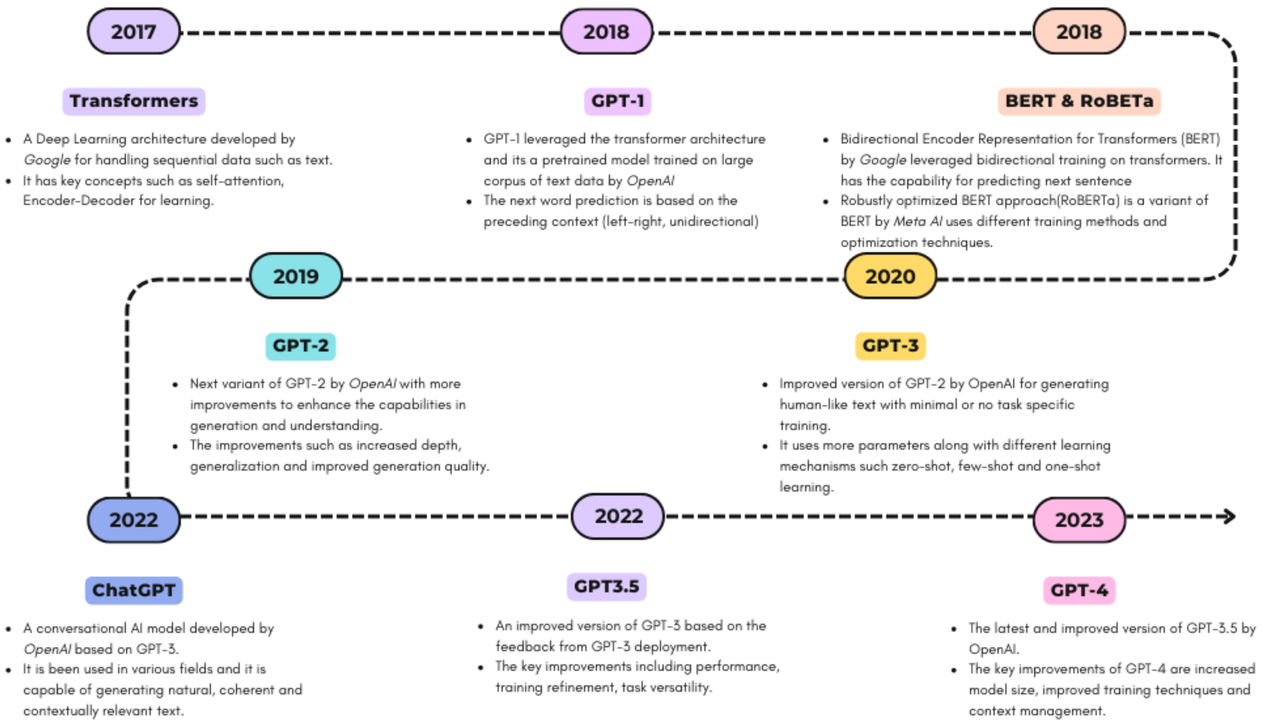
Within this milieu, Generative Pretrained Transformers (GPTs) have surfaced as a beacon of potential. These sophisticated linguistic models are grounded in the twin foundations of Machine Learning (ML) and Natural Language Processing (NLP). GPTs have demonstrated their ability to understand and generate texts as humans [7]. Their applications have spanned in various domains, ranging from the automation of client services to the genesis of content, and presently, they are on the verge of creating a paradigm shift in the management of medical literature [8].

The advancements in NLP technology leverage state-of-the-art models such as GPTs to condense medical research on specific illnesses systematically. We will undertake a detailed exploration of the workings of these models, examining how they process vast datasets to understand and summarize text effectively. The scrutiny will extend to their extant deployments within the medical research sphere and highlight instances where they are successfully employed. Concurrently, we will navigate the potential pitfalls and challenges inherent in the systematic summarisation endeavors of GPTs, such as questions of veracity, partiality, and the ethical ramifications therein.

Progressing beyond a mere assessment of the current landscape, this paper will foresee the transformative prospects of GPTs in the domain of medical literature examination. Our exploration will be twofold. Firstly, contemplate the avenues for their enhancement and refinement to augment precision and efficiency. Secondly, contemplate their expansive impact on the medical profession and healthcare in a reflective and forward-looking manner. The primary contributions of this paper are described as below:

1) This paper reviews various applications and methodologies of text summarization in medical research
2) The role of GPT in medical research for information extraction, summarization, sentiment analysis and clinical decision support are explored.
3) The paper delves deep into the capabilities of GPTs in literature review, data pre-processing and language understanding.
4) Four case studies related to medical applications are discussed.
5) The paper also briefly describes about various challenges and limitations of the existing GPTs.
6) Explored the future directions and aspects of GPTs are explored.

The organization of the paper is depicted in Fig. 1. This paper is divided into eight sections. Section I gives a brief introduction and motivation of this paper. Some of the already existing works on summarization using GPT are described in Section II. Section III explores the applications of GPT in information extraction, summarization, sentiment analysis, and clinical decision support. The power and potential of GPTs are explained in Section IV. Different case studies that use GPT for systematic summarization of medication research are described in Section V. Section VI lists various challenges and limitations of existing GPT models. The



Transformers

- A Deep Learning architecture developed by *Google* for handling sequential data such as text.
- It has key concepts such as self-attention, Encoder-Decoder for learning.

GPT-1

- GPT-1 leveraged the transformer architecture and its a pretrained model trained on large corpus of text data by *OpenAI*
- The next word prediction is based on the preceding context (left-right, unidirectional)

BERT & RoBETa

- Bidirectional Encoder Representation for Transformers (BERT) by *Google* leveraged bidirectional training on transformers. It has the capability for predicting next sentence
- Robustly optimized BERT approach(RoBERTa) is a variant of BERT by *Meta AI* uses different training methods and optimization techniques.

GPT-2

- Next variant of GPT-2 by *OpenAI* with more improvements to enhance the capabilities in generation and understanding.
- The improvements such as increased depth, generalization and improved generation quality.

GPT-3

- Improved version of GPT-2 by OpenAI for generating human-like text with minimal or no task specific training.
- It uses more parameters along with different learning mechanisms such zero-shot, few-shot and one-shot learning.

ChatGPT

- A conversational AI model developed by *OpenAI* based on GPT-3.
- It is been used in various fields and it is capable of generating natural, coherent and contextually relevant text.

GPT3.5

- An improved version of GPT-3 based on the feedback from GPT-3 deployment.
- The key improvements including performance, training refinement, task versatility.

GPT-4

- The latest and improved version of GPT-3.5 by OpenAI.
- The key improvements of GPT-4 are increased model size, improved training techniques and context management.

FIGURE 2. Timeline and evolution of GPT.

future directions are explained in Section VII, and finally, Section VIII concludes the study by providing a summary.

II. LITERATURE REVIEW

The development of Transformer architecture [9] by Google has tremendously changed the realm of Natural Language Processing. The transformer has improved the efficiency of NLP tasks such as classification, translation, language modeling, etc. The transformer use a self-attention operation, which is a weighted sum average of all the input vectors, and every input vector is mapped to query, key and value elements. The transformer has two functional blocks, viz., encoder and decoder. The input vectors are fed into the encoder block, and the decoder block generates output probabilities.

In 2018, GPT (Generative Pretrained Transformers) was developed by OpenAI, which was pretrained on a large-scale dataset using a self-supervised learning method [10]. The introduction of GPT has simplified most of the NLP tasks. GPT models were designed flexibly to fine-tune the model with task or domain-specific data. Since then, OpenAI has been constantly working on improving GPT architecture and has released different versions of GPT, viz., GPT2, GPT3, GPT3.5, and GPT4. The evolution of the generative models is shown in Fig. 2.

The application of GPTs in the medical field, particularly in the systematic summarization of medical and biomedical artifacts, has been the subject of several recent studies. Batra et al. [11] leveraged the power of recent advances in pre-trained and transformer-based NLP models to perform text summarization over the COVID-19 Public Media Dataset. They analyzed and compared the results of BERT, GPT-2, XLNet, BART, and T5, which are among the most popular extractive and abstractive summarization models. The study found that BERT, a transformer autoencoder, outperformed the other models in SARS-CoV-2 news summarization. As a result, they utilized BERT in their web application "CoVShorts" to summarize COVID-19 articles. The application serves the public by providing brief, concise, and to-the-point summaries quickly, helping them stay up-to-date with all the information related to COVID-19.

The research conducted by Zhu et al. [14] introduces three large, deidentified medical text datasets: DISCHARGE, ECHO, and RADIOLOGY, sourced from MIMIC-III [17], containing varying numbers of report-summary pairs. For automated abstractive summarization of these datasets, the study employed pre-trained encoder-decoder language models such as BERT2BERT, BERTShare, RoBERTaShare, Pegasus, ProphetNet, T5-large, BART, and GSUM. Notably, the BART model was enhanced by incorporating sampled summaries from the training set as contextual guidance, which bolstered both encoding and decoding processes. This novel approach led to improved ROUGE scores and BERTScore in the experiments.

The work by Cai et al. [15] emphasizes the importance of effectively generating the "impression" section from

TABLE 1. Summary of literature review.

Author Name	Contributions	Outcome	Limitations
H. Batra *et al.* [2021] [11]	Utilized BERT for summarizing COVID-19 articles in their web application "CoVShorts".	Provided brief, concise, and to-the-point summaries quickly, helping the public stay up-to-date with all the information related to COVID-19.	Limited to COVID-19 articles relies on a specific model (BERT).
H. Zhuang *et al.* [2019] [12]	Proposed an innovative approach to generate human-readable summaries that maintain the core idea of the original text.	Outperformed previous state-of-the-art models in the CNN/Daily Mail summarization task.	The model may struggle with longer texts or more complex summarization tasks.
Sheela J. *et al.* [2021] [13]	Introduced a multi-document summarization model that employs a novel optimization algorithm - CAVIAR Sun Flower Optimization (CAV-SFO).	The proposed method obtained performance in terms of metrics such as percision and F-Messure.	The model is complex and may not be suitable for real-time applications due to computational demands.
Y. Zhu *et al.* [2023] [14]	Introduced three medical datasets from MIMIC-III and employed encoder-decoder models for summarization, notably enhancing BART with prior knowledge guidance	Leveraged diverse datasets and state-of-the-art models, achieving improved ROUGE and BERTScore results.	The study lacked comparison with other leading methods and had potential dataset limitations, which lack the model to generalize across varied medical contexts.
X. Cai *et al.* [2023] [15]	• Developed ChestXRayBERT for chest radiology. • Trained on papers and fine-tuned on X-ray reports. • Excelled on OPEN-I and MIMIC-CXR datasets.	• Improved radiology communication with specialized terminology. • Highlighted NLP's role in medical imaging.	• Limited testing scope. • Potential for overfitting to specific dataset characteristics. • This could reduce its effectiveness when applied to diverse or unseen medical imaging data
R.K.Garg *et al.* [2023] [16]	Conducted a PRISMA review on ChatGPT in medicine, covering diagnosis, treatment, and research.	Showcased ChatGPT's versatility in patient care and research, highlighting its assistant potential.	Challenges in originality, accuracy, bias, and academic writing.

the "findings" in radiology reports to foster efficient communication between radiologists and referring physicians. To ease radiologists' workload, they introduced a specialized abstractive summarization model tailored for chest radiology reports. While existing NLP advancements like BERT struggled with the domain-specific terminology of radiology, they developed a domain-centric model, ChestXRayBERT. This model was pre-trained on a collection of radiology papers and further combined with a Transformer decoder. When fine-tuned and evaluated on recognized datasets like OPEN-I [18] and MIMIC-CXR [19], ChestXRayBERT markedly outperformed other neural abstractive models. This approach underscores the potential of customizing advanced NLP tools for medical imaging, radiology, and the broader realms of biomedicine and healthcare.

Training a neural network to generate a human-readable and semantically correct summary is one of the very challenging tasks. Zhuang and Zhang [12] proposed a technique to generate summaries using a Generative Adversarial Network (GAN). The GAN architecture contains one generator and two discriminators. The generator component is used to encode the long input text. One of the discriminators learns the input representation from the text encoding, and another generator takes care of the generated summary to be semantically similar to the original text context. The GAN model was evaluated using the ROUGE score with the CNN/Daily Mail dataset [20]. Also, the proposed technique was compared with other state-of-the-art models. The evaluation results proved that the summaries generated from the GAN model are better than those of other compared models. Though the standard dataset was used in research work for summarization, this work also has great potential in medical research.

Sheela and Janet [13] introduced a multi-document summarization model that employs a novel optimization algorithm, CAVIAR Sun Flower Optimization (CAV-SFO). This is yet another work related to multi-document summarization and shall also be extended to medical research. The proposed model uses two classifiers, a GAN and a Deep Recurrent Neural Network (Deep RNN), to score sentences for summarization. The process begins with removing duplicate content using the simHash method, then scoring each sentence with a CAV-SFO-based GAN classifier. The sentences are then pre-processed, and text-based features

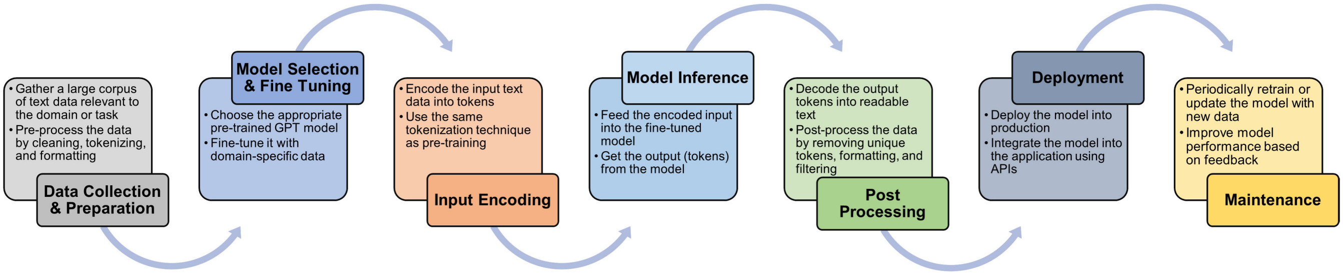

Gather a large corpus of text data relevant to the domain or task
Pre-process the data by cleaning, tokenizing, and formatting
Data Collection & Preparation
Model Selection & Fine Tuning
Choose the appropriate pre-trained GPT model
Fine-tune it with domain-specific data
Encode the input text data into tokens
Use the same tokenization technique as pre-training
Input Encoding
Model Inference
Feed the encoded input into the fine-tuned model
Get the output (tokens) from the model
Decode the output tokens into readable text
Post-process the data by removing unique tokens, formatting, and filtering
Post Processing
Deployment
Deploy the model into production
Integrate the model into the application using APIs
Periodically retrain or update the model with new data
Improve model performance based on feedback
Maintenance

FIGURE 3. GPT workflow - describing various stages of pipeline to work with GPT.

are extracted for scoring by a CAV-SFO-based Deep RNN. The outcomes from both classifiers are integrated, and a final evaluation is calculated based on a multi-document summarization scale.

The research by Garg et al. [16] delves into the capabilities and implications of ChatGPT, a state-of-the-art AI tool from OpenAI, within the medical domain. Addressing the ever-evolving landscape of patient diagnosis, treatment, and medical research, they embarked on a comprehensive systematic review using PRISMA standards [21]. Their exhaustive analysis of 118 selected publications illuminated ChatGPT's multifaceted role, spanning patient interactions to research facilitation. However, they also highlighted the inherent challenges of deploying such AI tools, like concerns over originality, accuracy, and bias, especially in academic writing. Their findings, particularly around the limitations of ChatGPT (GPT- 3.5) in healthcare settings, signal a note of caution. Nevertheless, they recognize its potential as a clinical assistant, emphasizing its transformative impact in research and scholarly endeavors.

The related works demonstrate the diverse applications and methodologies employed in the field of text summarization from different forms of data using advanced transformer models. The literature survey presented in Section II is summarized in Table. 1. While these works have made significant strides in the field, there is still room for exploration and improvement, particularly in the areas of multi-document summarization and the generation of more human-like summaries.

III. THE ROLE OF GPT IN MEDICAL RESEARCH

The domain of GPTs has witnessed substantial strides in innovation over the past few years, a period marked by the advent of progressively intricate GPT models. These models can understand, generate, and even imitate the intricate details of human speech [22]. The propulsion behind these strides has been the confluence of the following factors [23]:

(i) The proliferation of expansive corpora of textual data
(ii) The advent of formidable computational infrastructures
(iii) The iterative refinement of machine learning paradigms

GPT models work by using a deep learning architecture called the Transformer [9]. They process input text through multiple layers of attention mechanisms, which allow the model to focus on different parts of the text while generating predictions for the next word. This is done using a process called "self-attention," where each word's context is considered relative to every other word in the sequence. The model is trained to minimize prediction errors across vast amounts of text data, learning patterns, syntax, and context to generate coherent and contextually relevant responses.

Like any machine learning and deep learning algorithm, working with GPT also has a workflow. Figure.3 shows the workflow or the pipeline of utilizing GPT in any application domain. The pipeline starts with data collection and preparation by performing pre-processing operations such as cleaning, tokenizing, and formatting. Once the data is prepared, a suitable pre-trained model should be selected and fine-tuned with the available data. Before inferencing the model, the test data should be encoded in the way that was done during the pre-training phase. Then, the outputs of the model are decoded to represent understandable human texts. If the results of the inference are satisfactory, then the model can be moved to a production environment for deploying by integrating the model into an application or service by using APIs. After the deployment, the model's performance should be continuously monitored, and based on feedback, the model should be continuously re-trained.

GPT models are trained using vast text datasets, enabling them to generate human-like responses. Collecting such datasets can be challenging as they must represent various topics while avoiding biases that could affect model outputs. Training requires substantial computational resources due to their complex architectures and large parameter sizes (e.g., GPT-3 has 175 billion parameters [24]), often involving powerful GPUs or TPUs over weeks or months. Key challenges include obtaining large, diverse, and high-quality datasets, managing the complexity of model architecture, and ensuring efficient use of resources. Evaluating model performance can be complex due to the subjective nature of language tasks. Metrics such as perplexity or BLEU scores may not fully capture the quality of generated text in all contexts, necessitating more comprehensive evaluation strategies. Balancing model performance and avoiding biases or errors in generated content is a constant challenge.

In the realm of medical research, the incorporation of GPTs has emerged as an area of burgeoning interest. The medical

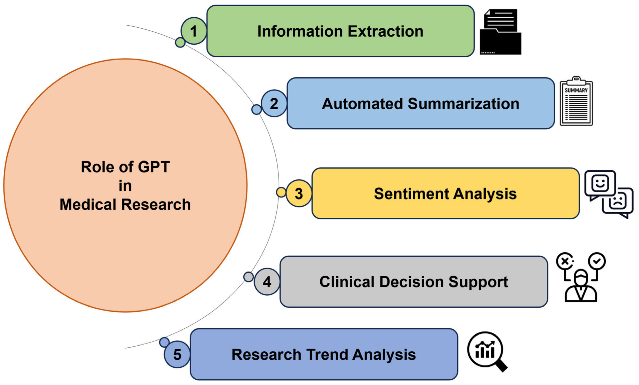


FIGURE 4. Different roles of GPT in medical research.

sector abounds with a plethora of unstructured textual data sources, such as Electronic Health Records (EHRs), clinical narratives, scholarly articles, and feedback from patients. This wealth of data harbors immense potential for extracting salient insights, which, if harnessed correctly, could unlock new vistas in medical understanding and patient care [25]. Figure. 4 shows the summary of different roles of GPT in medical research. The rest of the sections briefly describe these roles.

1) **Information Extraction:** A quintessential utility of GPTs in the medical research paradigm is the extraction of information [26]. This process entails distilling structured data from the vague and large volume of unstructured textual narratives. GPT algorithms, for instance, can discern and isolate particular medical entities such as pathologies, symptomatology, pharmacological treatments, and medical interventions from the labyrinth of clinical documentation [27]. The information thus garnered can be employed in a spectrum of critical tasks, including categorizing patients, monitoring disease patterns, and providing clinical decision support systems.
2) **Automated Summarization:** Another crucial application of GPTs in medical research is the generation of concise summaries [28]. Confronted with the relentless surge of scholarly medical writings, clinicians and researchers find it increasingly arduous to remain abreast of the latest discoveries [29]. Here, GPT algorithms offer a lifeline by condensing extensive research treatizes into digestible synopses, thereby allowing medical professionals to quickly assimilate the crux of research findings without perusing the full text [11].
3) **Sentiment Analysis:** The field of sentiment analysis, or opinion mining, is gaining momentum within medical applications of GPTs [30]. This analytic approach involves parsing texts to deduce the sentiment conveyed by the author. Applied within the medical context, sentiment analysis can scrutinize patient feedback, social media commentary, and other patient-originated texts to glean insights into patient experiences, their viewpoints, feelings about different treatment methods, attitudes towards healthcare providers, and perspectives on various medical conditions [31].
4) **Clinical Decision Support:** GPT models are increasingly pivotal in augmenting clinical decision-making processes [32]. By meticulously parsing and synthesizing data from a patient's electronic health record, GPTs help clinicians render more nuanced and informed medical decisions. For instance, GPT-driven algorithms can analyze patient data and medical histories to recommend potential diagnostic tests or treatments, assisting clinicians in making informed decisions. These systems can evaluate complex clinical scenarios, suggesting the most relevant investigations or interventions based on current medical guidelines and patient-specific factors, thereby enhancing the efficiency and accuracy of clinical decision-making process [33].
5) **Research Trend Analysis:** GPTs also serve as instrumental tools in dissecting and interpreting the trajectory of medical research [34]. By sifting through voluminous quantities of scholarly articles, GPTs can discern and bring to light prevailing trends, predominant themes, and discernible lacunae within the corpus of existing research. This intelligence is invaluable for researchers contemplating investigative avenues and funding bodies tasked with allocating resources to foster scientific inquiry.

GPT models hold immense potential for revolutionizing medical research. They are crucial for parsing and understanding complex medical data, enabling more efficient research processes. The relevance of GPT in this field stems from its ability to process vast amounts of medical literature and patient data, providing insights that can accelerate medical discoveries and improve patient care. The importance of GPT models lies in their capacity to transform large, unstructured medical datasets into actionable knowledge, making them invaluable tools in the ongoing advancement of medical research and healthcare.

IV. THE POWER AND POTENTIAL OF GPTS IN MEDICAL RESEARCH

The power and potential of GPTs in the realm of medical research are vast and multifaceted. This section delves deeper into the capabilities of these advanced AI models and how they can revolutionize the field of medical research. Some of the key tools and libraries used in medical research are summarized in Table. 2.

1) **Literature Review and Data Analysis:** GPTs stand at the vanguard of modernizing the medical literature review and data analysis processes [43], [44]. The large volume of medical literature defies manual review; however, GPTs can be harnessed to parse and synthesize this information, distilling essential findings, discerning patterns, and contrasting disparate

TABLE 2. Key NLP tools and libraries for medical research.

Tool/Library	Description	Applications
MedSpaCy [35]	Customizable NLP pipelines for clinical data	Information extraction, clinical named entity recognition
cTAKES [36]	Clinical NLP system with UIMA framework	Drug information extraction, symptom and disease mention extraction
BioBERT [37]	BERT pre-trained on biomedical corpora	Medical document classification, relation extraction
BioNLP [38]	Suite of NLP tools for biomedical text mining	Gene mention identification, mutation mention extraction
Neji [39]	Modular web service for biomedical NER (Named Entity Recognition)	Drug and disease mention extraction, clinical entity recognition
Med7 [40]	Transformer-based NER model trained on Electronic Health Records (EHR)	For extracting seven clinical entities related to patient health records such as medication name, dosage level, amount of time the drug was prescribed etc.
ClinicalBERT [41]	BERT pre-trained on clinical notes	ClinicalBERT shall be used in predicting 30-days hospital re-admission
Width.ai [42]	Summarization pipeline for summarizing human written medical prescription records	Medical Record Summarization

studies with finesse [45]. This enhances the literature review's efficiency and ensures comprehensive coverage of original and innovative research.

2) **Unprecedented Scale of Data Processing:** The capacity of GPTs to ingest and dissect data on a scale hitherto unattainable presents remarkable advantages [46], [47]. Within medical research, this translates to the analysis of thousands of documents—from scholarly articles to clinical trial data—far surpassing the temporal capabilities of human researchers. Such expeditious information processing can expedite research trajectories, catalyzing swifter breakthroughs and medical progress [48].

3) **Advanced Natural Language Understanding:** Engineered with a sophisticated grasp of linguistic nuances, GPTs can navigate the complex vocabulary of medical science, interpreting intricate research findings and emulating human prose [49]. This capability in language processing offers a pathway for automating tasks that are often repetitive yet essential, such as synthesizing findings from various studies, deciphering complex data sets, and preparing detailed research reports [43].

4) **Predictive Capabilities:** Beyond text generation and comprehension, GPTs possess the predictive ability to make inferences based on processed data [50].

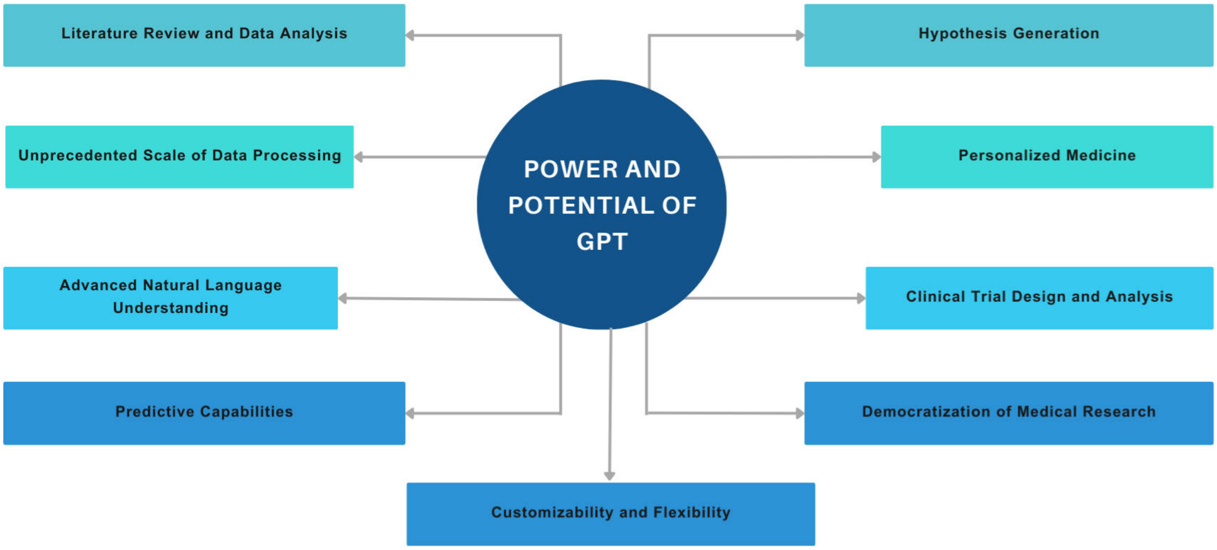

Literature Review and Data Analysis
Hypothesis Generation
Unprecedented Scale of Data Processing
POWER AND POTENTIAL OF GPT
Personalized Medicine
Advanced Natural Language Understanding
Clinical Trial Design and Analysis
Predictive Capabilities
Democratization of Medical Research
Customizability and Flexibility

FIGURE 5. Power and potential of GPTs in medical research.

In medical research, such capabilities could encompass predicting clinical trial outcomes, projecting epidemiological trends, or unveiling nascent research frontiers [51]. These predictive insights are invaluable, potentially steering research and policy decision-making.

5) **Customizability and Flexibility:** The inherent generalist design of GPT models is a key strength, offering a broad base from which they can be finely tailored to specific research requirements [52]. For example, a GPT model can be specifically trained on a narrow corpus of medical literature, relevant to a particular area of study. This flexibility ensures that GPTs are not only highly adaptable to diverse research needs but can also be precisely customized for targeted applications, significantly enhancing their practicality and effectiveness in a range of research scenarios [53].
6) **Democratization of Medical Research:** GPTs have the potential to democratize medical research by automating complex analyses and rendering research more approachable [54]. This democratization could engender a broader, more extensive participation in medical research, potentially igniting a restoration of innovation and inclusivity in research findings and outcomes. While hypothesis generation focuses on aiding researchers in formulating research questions, democratization extends beyond this by broadening participation and diversity in research. It not only facilitates a wider range of individuals and institutions to engage in medical research but also promises to culturally shift the research landscape towards more inclusivity and innovation, thereby enriching the research outcomes and contributions significantly.
7) **Clinical Trial Design and Analysis:** GPTs are poised to revolutionize the framework of clinical trials such as streamlining participant selection, orchestrating trial protocols, and sift through complex trial data [55], [56]. The integration of GPTs in this arena could greatly enhance the precision and efficiency of trials, thereby expediting the journey of novel therapies from conception to clinical application.
8) **Personalized Medicine:** The capability of GPTs in data analysis is particularly relevant to the rapidly growing field of personalized medicine. By assimilating and interpreting vast datasets, GPTs can pinpoint treatment modalities tailored to the individual nuances of patients' genetic profiles, lifestyles, and other pertinent factors [57]. Such targeted analysis can culminate in highly individualized and potentially more effective treatment regimens.
9) **Hypothesis Generation:** The utility of GPTs extends into the creative realm of hypothesis generation. Through meticulous analysis of existing research and data, GPTs can unveil areas where knowledge is lacking and propose new avenues for investigation [58]. This capacity to generate viable research hypotheses can direct scientific inquiry towards the most fertile and uncharted territories of medical research, thereby fostering a climate of innovation and discovery.

In summary, the power and potential of GPTs in medical research are vast. However, realizing this potential will require careful implementation, ongoing training and education, and robust ethical and regulatory frameworks. Figure. 5 summarizes the overall power and potential of GPTs in medical research.

V. CASE STUDIES OF MEDICAL APPLICATIONS

GPTs are currently used for systematic summarization and various other applications of medical and bio-medical research. The following subsection presents various case studies to showcase the potential usage of GPTs in various application domains of medical research.

A. SUMMARIZATION OF BIOMEDICAL RESEARCH USING GPT-3

The medical research landscape is inundated with vast amounts of information, with new studies being published at an unprecedented rate. These publications carry vital insights that can influence patient care, policy decisions, treatment to a clinical condition, evidence for the treatment and further research. However, the sheer volume of medical research publications poses a challenge for professionals who need to stay updated. This has spurred interest in leveraging advanced language models, like GPT-3, to systematically summarize medical research.

The study titled "Summarizing, Simplifying, and Synthesizing Medical Evidence Using GPT-3 (with Varying Success)" [59] embarked on a mission to evaluate the efficacy of GPT-3, specifically its GPT3-D3 variant, in summarizing biomedical articles. The overarching goal was to determine whether such state-of-the-art models could be trusted to distill complex medical research into accurate, coherent, and useful summaries. Some of the salient aspects of the study are as follows:

1) **Evaluation Framework:**
   a) **Data Source:** The study primarily sourced biomedical abstracts for the summarization tasks. These abstracts spanned various sub-domains within biomedicine, ensuring a comprehensive evaluation.
   b) **Evaluation:** Domain experts with formal medical training were enlisted to critically assess the generated summaries. Their evaluations gauged the summaries on parameters such as faithfulness to the original text, coherence of content, and the utility of the summarized information.
   c) **Tasks:**
      i) Single Document Summarization: GPT3-D3 was tasked with producing summaries for individual biomedical abstracts.
      ii) Multi-document Summarization: The model was challenged to synthesize information from multiple documents into a singular summary.

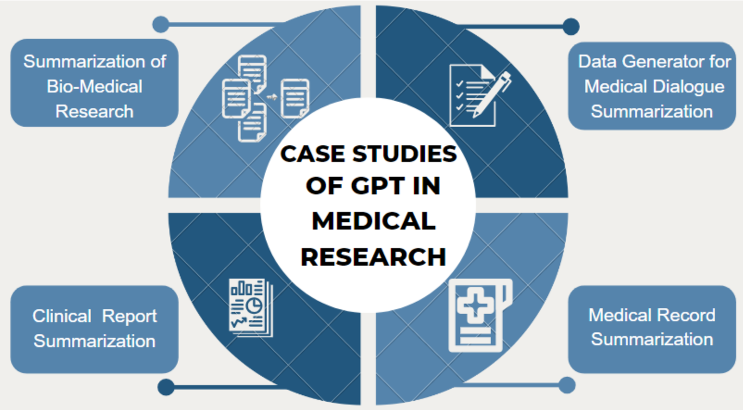

Summarization of Bio-Medical Research
Data Generator for Medical Dialogue Summarization
CASE STUDIES OF GPT IN MEDICAL RESEARCH
Clinical Report Summarization
Medical Record Summarization

FIGURE 6. Usage of GPT for summarization in various domains of medical research.

2) **Key Findings:**

a) **Performance Evaluation:** GPT3-D3 exhibited a remarkable ability to distill individual biomedical abstracts into concise summaries. These metrics included measures of coherence, accuracy, and conciseness. Coherence assessed how logically connected and understandable the summaries are, while accuracy evaluated how faithfully the summaries represent the original text. Conciseness measured the effectiveness of the summaries in distilling the main points without unnecessary verbosity. The majority of these summaries were found to be coherent, capturing the main essence of the research without distorting the original message.

b) **Analysis of Errors:** The study delved deeper into understanding the nature of errors made by GPT3-D3. The study utilized qualitative analysis by domain experts. These experts, with deep understanding of biomedical content, compared the model-generated summaries against original abstracts. Two prominent error patterns emerged:

i) *Minor Inaccuracies:* These were subtle errors where the model might have slightly misrepresented a detail from the original abstract. While not drastically altering the meaning, these inaccuracies could lead to misinterpretations in certain contexts.

ii) *Omissions:* A more concerning pattern was the omission of crucial details. In certain summaries, GPT3-D3 left out key findings or implications, which could impact the utility of the summary for readers wanting a comprehensive overview.

c) **Complexities in Multi-document Tasks:** GPT3-D3's performance exhibited a noticeable decline in the multi-document summarization task. Synthesizing information from multiple sources into a singular cohesive summary proved challenging for the model. The resulting summaries often lacked the depth and breadth required to encompass the key findings from all documents. The struggles faced by GPT3-D3 in this task of summarizing medical research articles highlighted the intricacies involved in multi-document summarization. It underscored the need for models to be better equipped to handle the complexities of synthesizing diverse sources while ensuring that no vital information is lost in the process.

GPT3-D3's involvement in biomedical summarization showcased its significant potential and progress. Excelling in single-document summarization, it highlighted the model's capabilities in distilling complex medical information. While there are areas for improvement in multi-document tasks, these challenges mark important steps towards advancing language models. As the quest to transition from information chaos to clarity in medical research continues, the study provides a roadmap highlighting both the milestones achieved and the terrains yet to be conquered.

B. MEDICALLY AWARE GPT-3 AS A DATA GENERATOR FOR MEDICAL DIALOGUE SUMMARIZATION

The scholarly investigation entitled "Medically Aware GPT-3 as a Data Generator for Medical Dialogue Summarization" [60] embarked upon an explorative endeavor to ascertain the proficiency of GPT-3 in fabricating synthetic training data, with a focus on encapsulating medically salient information. The primary objective was to assess the reliability of such experimental models in condensing intricate medical dialogues into precise and intelligible summaries.

The application of GPT-3 as the foundational mechanism for synthesizing medical data represents a pioneering stride in Large Language Models (LLMs). This distinctive use case unveils a novel vista on integrating language models into the medical sphere. The model could infer the informational content from only 210 human-annotated examples to replicate the usefulness of a much larger collection of 6400 examples. This evidences the effectiveness and creativity of the methodology, providing a solution to the enduring problem of data scarcity.

1) **Components of the Research Study:**
   a) **Data Genesis:** The study aimed to catalyze the transformation of 210 human-labeled instances into a dataset. The major objective of the research methodology is to embody the richness and heterogeneity of 6400 examples into the dataset.
   b) **Utilization of GPT-3:** GPT-3 played a very significant role in the study. It underwent meticulous training to engender synthetic training data that was not only abundant in content but also replete with clinical relevance.

2) **Principal Discoveries:**
   a) **Models Subjected to Evaluation:**
      i) The authors of [60] assessed the efficacy of GPT-3-ENS (their proposed model) as a generator of labeled data by comparing it with models architected for both abstractive and hybrid (abstractive and extractive) summarization.
      ii) PEGASUS [61] was employed for abstractive summarization tasks. PEGASUS, which stands for **P**re-training with **E**xtracted **G**ap-sentences for **A**bstractive **SU**mmarization **S**equence-to-sequence models, is a deep learning model developed by Google Research. It is designed specifically for the task of abstractive text summarization.
      iii) Dr. Summarize (DRSUM) [62], along with its high-performing variant (2M-PGEN), was utilized for extractive summarization. DRSUM is a model or system designed for extractive summarization. The 2M-PGEN variant is an enhanced version of the original model with improved features like phrase generation capabilities or modifications that allow it to handle larger datasets or produce more accurate summaries.
   b) **Specifics of Implementation:**
      i) When interfacing with GPT-3 via the OpenAI API, various parameters can be meticulously tuned for optimal performance. These include selecting the model type (e.g., davinci), defining the input prompt, setting the maximum response length (Max Tokens), and adjusting 'Temperature' for creativity in responses. Other parameters like 'Top P' control response diversity, while 'Frequency Penalty' and 'Presence Penalty' manage repetitiveness and topic variety, respectively. Additional settings like 'Stop Sequences' dictate where the model stops generating text, and 'Echo' determines if the prompt is included in the output. Advanced options like 'Best Of' and 'Logprobs' enabled finer control over output quality and analytical insights, respectively, while 'User Data' provides context for more relevant responses.
      ii) GPT-ENS is a medically aware GPT-3 ensemble algorithm used to generate the best summary from the medical dialogues. The algorithm takes dialogue snippets and labeled examples as input. The authors have also introduced an in-house extractor - MEDICALENTITYRECOGNIZER, to extract the medical concept from the given text. The extractor has access to the Unified Medical Language System (UMLS) [63]. GPT3 was used to generate a summary for the given dialogue snippet sample. Then, the generator summary is fed as an input to the extractor for extracting medical context. This process is repeated for the specified ensembling trails, and the best summary (the summary that has the highest recall value) is returned.
      iii) The optimal number of instances for priming GPT-3 was discerned to be 21, bounded by the constraints of the context window length. The context window length refers to the maximum amount of text (in terms of tokens, which can be words or parts of words) that the model can consider at one time when generating a response. For GPT-3, this limit is around 2048 tokens.
   c) **Training of Summarization Models with GPT-3-ENS Data:**
      i) Models such as PEGASUS and DRSUM were trained on both human-annotated data (H6400) and data synthesized by GPT-3-ENS.
      ii) Remarkably, with only 210 human-annotated examples, GPT-3-ENS was adept at generating a enourmous volume of training data for PEGASUS and DRSUM models, yielding performance on par with or surpassing the reliance on 6400 human-annotated examples.
      iii) The performance of PEGASUS was notably enhanced when trained on data generated by GPT-3-ENS, intimating that such data constitutes high-quality training material for abstractive summarization models like PEGASUS.
      iv) Conversely, DRSUM's performance was consistent when utilizing GPT-3-ENS synthesized data, hinting that such data may be optimally suited for purely abstractive models.
   d) **Synergy of Human-Annotated and GPT-3-ENS Synthesized Data:**
      i) Given the limitations of GPT-3's priming context, it was postulated that the most effective summaries would emanate from a model trained on a dataset amalgamating human and GPT-3-ENS labeled examples.
      ii) The introduction of a mixing parameter α allowed for the calibration of the ratio of GPT-3-ENS labeled examples to human-labeled data.

iii) A composite dataset, integrating human-annotated and GPT-3-ENS synthesized data, invariably bolstered nearly all automated metrics for PEGASUS and DRSUM across all values of α.

iv) Clinicians exhibited a preference for summaries derived from models trained on the mixed dataset, affirming its superiority in capturing medical information and the overall essence of the summary.

These findings underscore the potential of GPT-3-ENS in generating synthetic training data for medical dialogue summarization and demonstrate its effectiveness when compared with traditional human-labeled data. The study also highlights the nuanced differences in performance when the synthesized data is used for training different types of summarization models.

C. CLINICAL REPORT SUMMARIZATION

The potential of GPT models has evolved tremendously. Another area where GPTs can be applied in medical research is clinical report summarization. The clinical reports, written by radiologists, describe the findings of medical imaging for diagnosis, treatment, and follow-up of diseases. The clinical reports are crucial to review and understand the current state of the patients, including their disease condition and the effect of the treatment. It is crucial to collect and compile information promptly and with precision. Chien et al. [64] experimented with using GPTs to summarize radiologic reports. The study was done on longitudinal aneurysm reports. The study was conducted with the motive to harness the benefits of GPTs and to understand their strengths and weaknesses in medical research. A brief description of the work is described as below:

1) **Summarization Models Used:** The study compared the performance of five advanced summarization models, namely BARTcnn [65], LongT5booksum [66], LED-booksum [67], LEDlegal [68], and LEDclinical [69], with expert-generated summaries. In addition, publicly available case reports of brain aneurysms from PubMed were used to evaluate the open access and online model GPT3davinci [70]. All of these are pre-trained models, trained on large datasets and using billions of parameters to generate summaries.
2) **Clinical Data Collection:** As part of this experimental analysis, clinical imaging reports of 64 aneurysms from 52 patients were collected. These reports were obtained between 2005 and 2022. Out of these 52 patients, 8 were males and 44 were females. The study used a total of 137 clinical imaging reports, which included three different modalities for aneurysm imaging: MR angiography, CT angiography, and DSA. Also, 100 clinical reports on brain aneurysms were collected from PubMed for experimental analysis.
3) **Evaluation and Results:** The summaries generated by the GPT models are compared with the ground truth summaries (which are expert-generated). ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [71] and BERTscore [72] were used as evaluation metrics. The summaries were also evaluated by four experts based on different aspects such as information accuracy, redundancy, comprehensiveness, and readability. Although the generative models examined were not specifically designed for clinical imaging reports, they could summarize most of the crucial information accurately. However, these models still have some limitations and room for improvement. Among the five state-of-the-art models tested, BARTcnn performed the best in generating patient clinical reports. The comparative analysis report indicated that GPT3davinci performed superior in summarizing case reports, while BARTcnn's performance was second.

D. MEDICAL RECORD SUMMARIZATION

The task of summarizing medical records is another area where GPTs are being applied. A state-of-the-art SOTA GPT-4 [42] medical record summarization pipeline has been developed by Width.ai, which utilizes advanced AI techniques for efficient summarization of human written medical prescription records. The article discusses the development of a medical record summarization pipeline using GPT-4. The pipeline is designed to handle the challenges of medical record summarization, which include the complexity of medical language, the need for high accuracy, and the importance of maintaining patient privacy.

The pipeline is built in three stages:

1) **Preprocessing:** In this stage, the medical records are cleaned and prepared for the model. Custom OCR (Optical Character Recognition) was used to extract text from the medical records. This is an essential step as the accuracy of the extracted text improves the efficiency of subsequent stages. This involves removing irrelevant information, standardizing the format of the data, and anonymizing patient information to maintain privacy.
2) **Summarization:** The cleaned data is then passed through the GPT-4 model, which generates a summary of the medical record. The model is trained to understand medical language and to extract the most important information from the record.
3) **Postprocessing:** The generated summaries are then post-processed to ensure they are accurate and coherent. This involves checking the summary against the original record and making any necessary corrections.

The article also discusses the benefits of using GPT-4 for this task. GPT-4's large model size and ability to understand context make it well-suited to the complexity of medical language. Furthermore, the model's ability to generate coherent and concise summaries can help healthcare professionals quickly understand a patient's medical history, leading to more efficient and effective care. The article concludes by noting that while the pipeline is promising, it is still a work in progress.

TABLE 3. Summary of various use-cases of GPT model used in medical domain for summarization.

Work	Task	GPT Model Used	Results	Limitations
Summarizing, Simplifying, and Synthesizing Medical Evidence Using GPT-3 (with Varying Success) [59]	Single and Multi-document summarization	GPT3-D3	GPT3-D3 outperforms the task of single document summarization	• Minor inaccuracies and omissions are observed. • Synthesizing information from multiple sources into a singular cohesive summary proved challenging for the model.
Medically Aware GPT-3 as a Data Generator for Medical Dialogue Summarization [60]	Dataset synthesis using GPT and summarizing medical dialogue	• GPT-3-ENS • PEGASUS • DRSUM	The effectiveness of summarization improved with the model trained with the data synthesised by GPT3-ENS	• There is room for improvement of coherency. • The relevance or importance of the information in the dialog needs to be considered as a metric for evaluation.
AI-Assisted Summarization of Radiologic Reports: Evaluating GPT3davinci, BARTcnn, LongT5booksum, LEDbooksum, LEDlegal, and LEDclinica [64]	Summarize radiologist report summarization using GPT	• BARTcnn • LongT5booksum • LEDbooksum • LEDlegal • LEDclinical • GPT3davinci	• The summaries generated were evaluated using automated metrics and expert evaluation. • BARTcnn was better at summarizing clinical reports than the other models, and GPT3davinci was better at summarizing case reports more concisely.	• Though the models considered for the experimentation were not specifically tailored for summarization application, the models still performed well in summarizing. • However, there is still room for improvement.
GPT-4 Medical Record Summarization Pipeline [42]	Summarizing human written medical prescription records	GPT-4	The trained large language model was able to produce coherent and concise summaries	The pipeline proposed is a work in progress, and there is a lot of scope for improving the generated summary.

A summary of the case studies discussed above is presented in Table 3. Overall, GPTs have shown promise in systematic summarization of medical research. They have the ability to analyze large amounts of data, understand complex medical information, and generate concise summaries. Ongoing research and advancements in GPT technology aim to address these challenges and further improve the effectiveness of GPTs in solving the problem of systematic summarization in medical research.

VI. CHALLENGES AND LIMITATIONS

In the realm of medical research summarization, the advent and application of GPTs have brought forth a revolutionary potential [73]. However, as with any technological advancement, particularly in the sensitive sphere of medical science, there exists a suite of challenges and limitations that require meticulous scrutiny [46]. From ensuring unwavering accuracy in the distilled information to addressing the opaque nature of these advanced models, the path to seamless integration of GPTs into medical summarization is fraught with complexities [74]. Additionally, the ethical quandaries of handling patient data and the computational demands of such models further complicate their ubiquitous adoption [75], [76]. This section delves into these challenges, offering a comprehensive analysis of the obstacles and considerations that stakeholders must grapple with in harnessing the capabilities of GPTs for medical research summarization. Figure 7 provides the summary of various challenges and limitations a GPT model has on medical research, which is explained in the subsections below:

A. FAITHFULNESS AND ACCURACY

The application of GPTs in the systematic summarization of condition-specific medical research requires a high degree of accuracy and faithfulness to the original content [77]. Any deviation from factual accuracy can lead to misunderstandings, misinterpretations, and potentially harmful consequences.

1) **Factual Inconsistency:** The issue of factual inconsistency in medical summaries arises when the

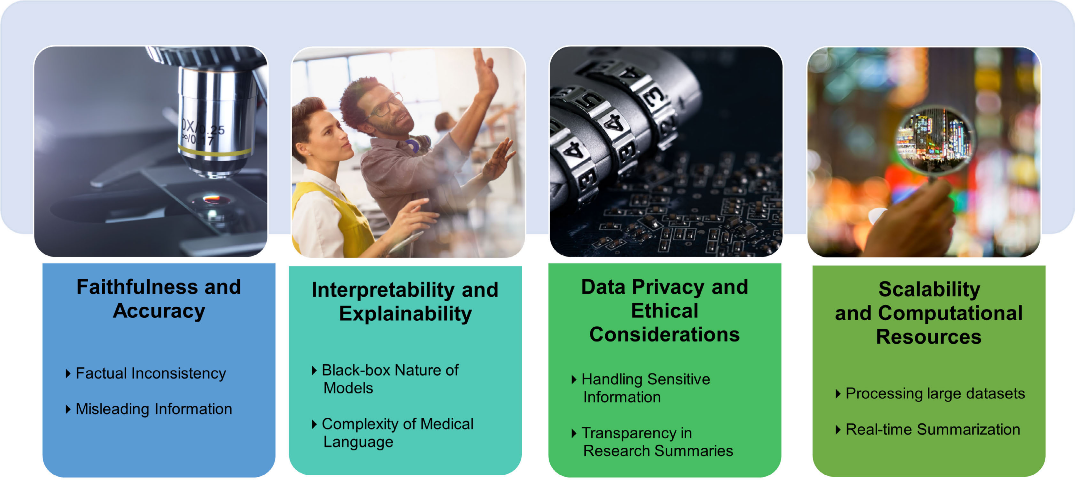

Faithfulness and Accuracy
Factual Inconsistency
Misleading Information
Interpretability and Explainability
Black-box Nature of Models
Complexity of Medical Language
Data Privacy and Ethical Considerations
Handling Sensitive Information
Transparency in Research Summaries
Scalability and Computational Resources
Processing large datasets
Real-time Summarization

FIGURE 7. Challenges and limitation of GPT in medical research.

content produced does not align with established facts or evidence-based medical practices [78], [79]. The implications of such discrepancies are profound, potentially affecting clinical outcomes and patient care. The primary factors contributing to these inconsistencies include:

a) ***Training Data Quality:*** The accuracy of a machine learning model is highly contingent on the quality of its training data. In the context of medical summaries, if the data includes inaccuracies or biases, these can propagated through the model's outputs [80], [81]. Therefore, ensuring data integrity by incorporating peer-reviewed and clinically verified sources is essential to minimize such errors.

b) ***Model Interpretation:*** The complex nature of medical terminology and concepts poses a significant challenge to GPT models, which may result in misinterpretations [82]. Enhanced training approaches like Unified Medical Language Systems (UMLS), Medical Ontologies, Clinical Decision Support Systems (CDSS), etc., that focus on the medical domain can improve the model's understanding, reducing the risk of factual discrepancies.

c) ***Lack of Expert Supervision:*** The absence of medical expert involvement during model training can lead to gaps in the model's knowledge, particularly regarding the nuances of medical accuracy [83], [84]. Collaborative efforts between data scientists and medical professionals can help bridge these gaps, leading to more reliable summaries.

2) **Misleading Information:** In the domain of medical research summarization, the propagation of misleading information is a significant issue that undermines the validity of conveyed findings [85]. Such misinformation can stem from multiple factors, including but not limited to:

a) ***Over-Simplification:*** The process of distilling complex medical research into a summary can sometimes result in the over-simplification of intricate concepts. This over-simplification by GPT models may lead to the omission of vital nuances and caveats that are essential for a complete understanding of the medical artifact.

b) ***Loss of Context:*** Summarization inherently involves compressing extensive information into a more digestible form. During this process, GPTs may inadvertently omit crucial context, yielding summaries that could misrepresent the original research's intentions or conclusions [86], [87]. This context loss can change the perceived meaning of the research findings and potentially guide readers to incorrect interpretations.

c) ***Training Data Bias:*** Bias in the data used to train GPTs can lead to skewed summarizations that reflect these biases [88], [89], [90]. Such replication of bias can have serious implications, especially in the sensitive context of medical information.

d) ***Algorithmic Bias:*** The design of GPT algorithms themselves may unintentionally introduce bias, potentially affecting certain groups or medical conditions disproportionately [91].

Mitigating the risk of misleading information requires careful design of the summarization process, including clear guidelines on what constitutes an acceptable level of simplification, rigorous testing against biases, and validation by medical experts.

Ensuring faithfulness and accuracy in the application of GPTs for systematic summarization of medical artifacts is a complex and multifaceted challenge [92]. It requires a careful balance between the need for concise summarization and the imperative to maintain factual accuracy and avoid misleading information [93]. Collaborative efforts between

AI researchers, medical experts, and regulatory bodies, along with the implementation of robust validation processes, are essential to address these challenges effectively.

B. INTERPRETABILITY AND EXPLAINABILITY

Interpretability and explainability are crucial in the context of medical research, where understanding the reasoning behind a decision or summary is often as important as the decision or summary itself [94], [95]. These requirements pose a unique challenges in the application of GPTs for systematic summarization of medical research. These challenges are as follows:

1) **Black-Box Nature of Models:** The utilization of GPTs in medical research summarization introduces several challenges due to their "black-box" nature, where the internal mechanisms are not readily interpretable [96]. These challenges include:
   a) ***Understanding Model Decisions:*** The decision-making process within GPTs involves a complex interplay of neural network weights and biases. This complexity obscures the reasoning behind how specific summaries are generated, making it challenging to dissect and understand the model's rationale for its outputs [97]. The ability to interpret these decisions is crucial, especially when summaries influence medical decisions.
   b) ***Trust and Acceptance:*** Medical professionals and researchers often require clear justification for the information they use, which is hindered by the opaque nature of GPTs [98]. The absence of a transparent explanation for how summaries are derived can lead to skepticism and hinder the acceptance of such advanced technologies in clinical practice.
   c) ***Regulatory Compliance:*** The medical field is subject to stringent regulations, including the need for transparency in decision-support systems [99]. The black-box nature of GPTs poses a significant challenge in meeting these legal mandates, as it can be difficult to demonstrate the model's decision-making pathway in a way that satisfies regulatory standards.

   Efforts to increase transparency include the development of tools that visualize the decision-making process within the model or the creation of simpler, more interpretable models.
2) **Complexity of Medical Language:** Medical language's specialized and intricate nature poses significant interpretability challenges for summarization models like GPTs. GPT generated summaries often exhibit complexities due to the specialized nature of medical terminology and concepts. This necessitates a high degree of domain-specific expertise for accurate interpretation, potentially constraining their accessibility to non-specialists [100]. Additionally, the inherent ambiguity and nuanced language of medical texts pose challenges for GPTs in capturing the full semantic depth in summaries. Despite achieving technical correctness, these summaries may lack the contextual nuances and detailed comprehension apparent to human experts, leading to potential oversimplifications or missed subtleties in representing the original research content [101].

C. DATA PRIVACY AND ETHICAL CONSIDERATIONS

Data privacy and ethical considerations hold exceptional significance in the medical sphere, particularly when handling sensitive patient information and confidential research findings. The application of GPTs for the systematic summarization of medical research accentuates these considerations, introducing several challenges:

1) **Handling Sensitive Information:** The deployment of GPTs in the medical summarization process requires the handling of delicate data, which, if not managed correctly, could lead to breaches of confidentiality and privacy [102], [103]. The following measures can be adopted to mitigate potential breaches of confidentiality and privacy:
   a) ***Consent and Authorization:*** Acquiring proper consent and ensuring necessary authorizations for the use of sensitive medical data is a legal and ethical imperative [104], [105]. This process is compounded by varying regulations across jurisdictions, necessitating a comprehensive approach to compliance.
   b) ***Data Anonymization:*** The anonymization of personal data is critical to safeguard the privacy of individuals [106], [107]. It is crucial that the techniques used for anonymization are robust enough to prevent re-identification. Re-identification can be particularly challenging with rich medical datasets where multiple data points can lead to patient identification.
   c) ***Security Measures:*** Protecting sensitive medical data demands stringent security protocols [108], [109]. This includes not only safeguarding the data during the summarization process but also ensuring the secure storage and transmission of information to prevent data breaches.
2) **Transparency in Research Summaries:** Transparency within the process of generating medical research summaries is a critical component of ethical standards. Ensuring that the information conveyed is both reliable and verifiable requires the following:
   a) ***Accuracy and Integrity:*** It is crucial that summaries generated by GPTs remain true to the original research content, preserving the scientific accuracy and the integrity of the data [86]. Any alteration or misinterpretation introduced in the summarization process could have significant repercussions for clinical practice and patient outcomes [110].
   b) ***Disclosure of Methods:*** A transparent summarization process involves clear communication about the algorithms and techniques employed [111]. This not

only facilitates reproducibility and peer review but also builds trust among medical professionals who rely on these summaries for critical insights.

Data privacy and ethical considerations in leveraging GPTs for systematic summarization of medical artifacts are multifaceted and require careful attention to legal, regulatory, and moral standards. Addressing these requirements is imperative to maintain the integrity and trust in the use of GPTs for medical research summarization. This involves not only technical solutions but also a framework of policies and procedures that prioritize the ethical handling of data. As GPT technology continues to advance, it is essential that these systems are designed and implemented with a strong emphasis on privacy, security, and ethical considerations, aligning with the highest standards of medical data governance.

D. SCALABILITY AND COMPUTATIONAL RESOURCES

The application of GPTs in the systematic summarization of medical artifacts requires significant computational resources. Scalability, or the ability to efficiently process large volumes of data and adapt to growing demands, is a critical consideration. The challenges associated with this requirement is as follows:

1) **Processing Large Datasets:** Medical research often involves vast amounts of data, including clinical records, research papers, and experimental results [112]. The computational demand for processing and summarizing vast datasets is significant [113]. Crafting algorithms that are both fast and precise in processing large amounts of data is a sophisticated endeavor [22]. Efficiency is key to ensuring that the summarization process is practical and scalable.
2) **Real-time Summarization:** In some medical applications, real-time summarization of research findings may be required [114]. This presents the following additional challenges:
   a) ***Latency:*** Achieving minimal delay in the summarization process is essential, as any significant latency could impede timely decision-making in clinical environments [115].
   b) ***Concurrency:*** The ability to process several summarization tasks concurrently puts a strain on computational resources. Optimizing these resources to handle multiple requests without performance degradation is critical [116].
3) **Cost Considerations:** The deployment and scaling of GPTs for medical summarization necessitate considerable computational resources, which carry the following associated costs:
   a) ***Hardware Costs:*** The initial investment and maintenance of specialized hardware like GPUs, which are required for their processing capabilities, can represent a significant financial burden [117].
   b) ***Cloud Computing Costs:*** Leveraging cloud services offers scalability and flexibility but also incurs ongoing costs that can accumulate, impacting the financial viability of projects [118].

   These cost factors must be carefully considered in the context of medical summarization projects using GPTs to ensure sustainable and cost-effective operation.

Scalability and computational resource considerations are central to the successful application of GPTs in the systematic summarization of medical artifacts.

VII. FUTURE DIRECTIONS

The rapid evolution of artificial intelligence, particularly GPTs, has opened doors to unprecedented possibilities in various sectors. GPTs have shown a tremendous performance in medical research especially in the application of summarization. Some of the existing works which uses the GPT for medical research summarization is presented in Table 4. In the realm of medical research, where the sheer volume of data often overwhelms professionals, GPTs stand as a beacon of hope [119]. Their capability to distill vast amounts of information into concise, coherent summaries promises to revolutionize the way researchers and medical professionals interact with data [120]. As we delve into the future prospects of GPTs in medical research summarization, it becomes paramount to understand both the potential integrations that can enhance research methods and the technological advancements that will steer this transformation. This section provides a holistic view of these dimensions, painting a picture of the imminent future where GPTs and medical research symbiotically evolve.

1) **Enhanced Integration of GPTs for Medical Research Summarization:**
   a) ***Research Database Query Systems:*** GPTs could be harnessed as sophisticated query processors within voluminous databases of medical research papers. They are poised to interpret intricate queries propounded by scholars and return pertinent literature or even precise excerpts that resonate with the inquiry [121].
   b) ***Automated Annotation and Emphasis Mechanisms:*** By integrating GPTs with annotation systems for scholarly papers, could autonomously accentuate pivotal discoveries, experimental approaches, or outcomes, thereby streamlining the review process [122].
   c) ***Sophisticated Cross-referencing Tools:*** As medical professionals and scholars peruse an article, GPTs could be employed to instantaneously collate cognate studies or furnish insights from ancillary literature, an asset of incalculable worth for contextualizing the expanse of a specific medical inquiry [123].
   d) ***Automated Metadata Synthesis:*** GPTs can be programmed to synthesize metadata for fresh academic treatises when such metadata is unavailable. This synthesis encompasses descriptors, abstracts,

and classifications, thereby enhancing the efficiency of indexing and retrieval from extensive data sources [124].

e) ***Synergy with Scholarly Review Platforms:*** Within forums where medical experts evaluate or deliberate over scholarly papers, GPTs could suggest instantaneous synopses, facilitating a swift comprehension of the literature during colloquies [125].

f) ***Tools for Trend Examination:*** While processing and summarizing a large volume of academic papers over time, GPTs could be integrated with tools that identify and clarify emerging trends, prevalent methodologies, and key discoveries within specific areas of medical research. This capability would enable researchers to quickly grasp the current state of research in particular medical conditions, observing how themes and treatment strategies have evolved.

g) ***Template-guided Reporting Constructs:*** GPTs might be attuned to fabricate reports based on predefined templates for scholars chronicling specific medical conditions, thereby assuring uniformity across diverse research papers.

h) ***Fusion with Visualisation Instruments:*** The summation of data-intensive research is often expedited by graphical elucidation. GPTs could merge with visualization apparatus to yield not merely textual condensations but also diagrammatic portrayals contingent on the content of medical research papers [128].

2) **Technological Advancements of GPTs for Medical Research Summarization:**

a) ***Domain-Specific Refinement:*** A key future development for GPT models is their refinement and calibration to specific medical domains. While these models are generally trained on vast and varied datasets, the ability to fine-tune them for particular medical specialties or conditions is crucial. This ensures that the models develop a deep and nuanced understanding of the unique complexities and terminologies inherent in different medical fields [129].

b) ***Enhanced Contextual Comprehension:*** Future iterations of GPT models are expected to exhibit significantly improved contextual comprehension. This means that they will be better equipped to grasp the underlying context and intended meaning of medical research findings [23].

c) ***Adaptive Feedback Mechanisms:*** Another progressive aspect is the incorporation of adaptive feedback mechanisms in GPT models. This involves the ability of these models to learn iteratively from feedback, thereby continuously improving the precision and relevance of their generated summaries. Such a feature allows for ongoing refinement of the summarization process, making it more aligned with user needs and expectations, and thereby improving the overall quality and utility of the summaries over time [130].

d) ***Multimodal Functionalities:*** Future developments in GPT models are also expected to include improved multimodal functionalities. This refers to the ability of these models to process and interpret data from various modalities, including visual (images, charts) and auditory (speech, audio recordings) inputs. [131].

e) ***Integrated Data Synthesis:*** Data synthesis is essential when the available data is insufficient or to enhance data privacy. A significant future direction for GPT models is their enhanced ability to integrate and synthesize different types of data. This involves not just processing textual content but also incorporating supplementary data formats such as graphs, charts, and tables. The aim is to produce comprehensive and holistic summaries that encompass all aspects of research articles, including their various data representations [132].

f) ***Automated Scholarly Referencing:*** GPTs may be instructed to autonomously identify and reference pertinent studies, thus upholding the accuracy and integrity that is essential in the medical discipline [133].

g) ***Instantaneous Abstracting:*** Technological progress has expedited the operational tempo of GPTs, enabling the generation of immediate summaries, pivotal for real-time deliberations or pressing research appraisals [134]. The future aspect here involves further enhancing the speed and efficiency of GPTs to support immediate, on-the-spot summarization and analysis.

h) ***Seamless System Integration:*** Future advancements should focus on integrating GPT models more smoothly with various research tools, databases, and analytical platforms. This integration is aimed at making GPTs a ubiquitous and versatile component in the medical research infrastructure, enhancing their accessibility and utility across different platforms and applications [135].

i) ***Augmented System Resilience:*** The future direction is to improve the resilience of GPT models. This involves using techniques like adversarial training to make these models more robust and reliable, especially when dealing with complex or ambiguous medical texts [136]. The goal is to ensure consistent quality in the generated summaries, regardless of the complexity of the source material.

j) ***Progressive Knowledge Acquisition:*** An important future aspect is the continuous learning capability of GPT models. Instead of requiring complete retraining to update their knowledge base, these models are evolving to incrementally assimilate new information. This aspect is particularly critical in the medical field, where staying abreast of the latest research and

TABLE 4. Applications of GPT models in medical research paper summarization.

Application	Summary
Advanced Summarization with GPT-4 [126]	• Utilizes the advanced capabilities of GPT-4 to conduct both section-wise and comprehensive full-note summarizations. • The efficiency of extracting relevant information from medical dialogues is improved.
Summarization using Pure GPT-3 [59]	• Leveraging the general capabilities of GPT-3. • This application focuses on summarizing medical dialogues to provide concise insights. • It is beneficial for clinicians and researchers with time constraints.
COVID-19 Research Summarization with GPT-2 [120]	• In response to the global pandemic, GPT-2 was employed to summarize a vast number of research articles related to COVID-19. • Aiding in the swift dissemination of crucial findings to the global medical community.

findings is essential for maintaining the relevance and accuracy of the models [137].

VIII. CONCLUSION

In conclusion, the application of GPTs in the systematic summarization of medical research has shown significant promise. This survey has provided a comprehensive overview of the current state-of-the-art technologies and methods, highlighting the power and potential of GPTs in transforming the way medical research data is processed, understood, and utilized. We have discussed the various applications of GPTs in the medical field, from diagnosing and treating patients to summarizing systematic reviews and medical records. We have also highlighted the challenges and limitations of GPTs, emphasizing the need for further research and development to overcome these hurdles. The case studies presented in this paper demonstrate the effectiveness of GPTs in solving the problem of information overload in medical research. As the field of artificial intelligence continues to evolve, we anticipate that GPTs will play an increasingly critical role in the systematic summarization of medical research, leading to more efficient and effective healthcare delivery. Further research should concentrate on overcoming the existing constraints of GPT models in medical document summarization. The future improvement should focus more on enhancing the accuracy and reliability of summaries for complex medical texts, ensuring they capture critical information without misinterpretation. Additionally, exploring the integration of GPT models with domain-specific medical knowledge bases could enrich the contextuality and relevance of the summaries. The development of tailored GPT variants, trained on diverse medical literature and patient records, may also be beneficial. These efforts will not only improve the efficiency of processing medical documents but also potentially aid in decision-making and providing insights in healthcare settings.

REFERENCES

[1] G. Han, X. Liu, F. Han, I. N. T. Santika, Y. Zhao, X. Zhao, and C. Zhou, "The LISS—A public database of common imaging signs of lung diseases for computer-aided detection and diagnosis research and medical education," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 648–656, Feb. 2015.

[2] J. Liu, J. Ma, J. Li, M. Huang, N. Sadiq, and Y. Ai, "Robust watermarking algorithm for medical volume data in Internet of Medical Things," *IEEE Access*, vol. 8, pp. 93939–93961, 2020.

[3] S. S. Furuie, M. S. Rebelo, R. A. Moreno, M. Santos, N. Bertozzo, G. H. M. B. Motta, F. A. Pires, and M. A. Gutierrez, "Managing medical images and clinical information: InCor's experience," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 1, pp. 17–24, Jan. 2007.

[4] M. S. Sharif, M. Abbod, A. Al-Bayatti, A. Amira, A. S. Alfakeeh, and B. Sanghera, "An accurate ensemble classifier for medical volume analysis: Phantom and clinical PET study," *IEEE Access*, vol. 8, pp. 37482–37494, 2020.

[5] A. Gkoulalas-Divanis, G. Loukides, and J. Sun, "Toward smarter healthcare: Anonymizing medical data to support research studies," *IBM J. Res. Develop.*, vol. 58, no. 1, pp. 9:1–9:11, Jan. 2014.

[6] X. Zhang, P. Geng, T. Zhang, Q. Lu, P. Gao, and J. Mei, "Aceso: PICO-guided evidence summarization on medical literature," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2663–2670, Sep. 2020.

[7] M. Yang, C. Li, Y. Shen, Q. Wu, Z. Zhao, and X. Chen, "Hierarchical human-like deep neural networks for abstractive text summarization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2744–2757, Jun. 2021.

[8] E. T. R. Schneider, J. V. A. de Souza, Y. B. Gumiel, C. Moro, and E. C. Paraiso, "A GPT-2 language model for biomedical texts in Portuguese," in *Proc. IEEE 34th Int. Symp. Computer-Based Med. Syst. (CBMS)*, Jun. 2021, pp. 474–479.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.

[10] R. Alec, N. Karthik, S. Tim, and S. Ilya, "Improving language understanding with unsupervised learning," *Citado*, vol. 17, pp. 1–12, Oct. 2018.

[11] H. Batra, A. Jain, G. Bisht, K. Srivastava, M. Bharadwaj, D. Bajaj, and U. Bharti, "CoVShorts: News summarization application based on deep NLP transformers for SARS-CoV-2," in *Proc. 9th Int. Conf. Rel., INFOCOM Technol. Optim. (Trends Future Directions) (ICRITO)*, Sep. 2021, pp. 1–6.

[12] H. Zhuang and W. Zhang, "Generating semantically similar and human-readable summaries with generative adversarial networks," *IEEE Access*, vol. 7, pp. 169426–169433, 2019.

[13] J. Sheela and B. Janet, "Caviar-sunflower optimization algorithm-based deep learning classifier for multi-document summarization," *Comput. J.*, vol. 66, no. 3, pp. 727–742, Oct. 2021.

[14] Y. Zhu, X. Yang, Y. Wu, and W. Zhang, "Leveraging summary guidance on medical report summarization," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 10, pp. 5066–5075, Oct. 2023.

[15] X. Cai, S. Liu, J. Han, L. Yang, Z. Liu, and T. Liu, "ChestXRayBERT: A pretrained language model for chest radiology report summarization," *IEEE Trans. Multimedia*, vol. 25, pp. 845–855, 2023.

[16] R. K. Garg, V. L. Urs, A. A. Agrawal, S. K. Chaudhary, V. K. Paliwal, and S. K. Kar, "Exploring the role of chat GPT in patient care (diagnosis and treatment) and medical research: A systematic review," *Health Promotion Perspect.*, vol. 13, no. 3, p. 183, Jun. 2023.

[17] A. Johnson, P. Tom, and R. Mark. (2020). *MIMIC-III Clinical Database (version 1.4)*. [Online]. Available: https://physionet.org/content/mimiciii/1.4/

[18] *Open Access Biomedical Image Search Engine*. Accessed: Jun. 3, 2024. [Online]. Available: https://openi.nlm.nih.gov/

[19] *Medical Information Mart for Intensive Care*. Accessed: Jun. 3, 2024. [Online]. Available: https://mimic-cxr.mit.edu/

[20] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence RNNs and beyond," in *Proc. 20th SIGNLL Conf. Comput. Natural Lang. Learn.*, Berlin, Germany, S. Riezler and Y. Goldberg, Eds., Aug. 2016, pp. 280–290.

[21] L. Shamseer, D. Moher, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation," *Systematic Rev.*, vol. 4, pp. 1–9, 2015.

[22] Y.-Y. Chuang, H.-M. Hsu, K. Lin, R.-I. Chang, and H.-Y. Lee, "MetaEx-GAN: Meta exploration to improve natural language generation via generative adversarial networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 3968–3980, 2023.

[23] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of ChatGPT: The history, status quo and potential future development," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 5, pp. 1122–1136, May 2023.

[24] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 1877–1901. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[25] S. Deepika, S. Shridevi, and N. L. Krishna, "Extractive text summarization for COVID-19 medical records," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, Nov. 2021, pp. 1–5.

[26] C.-Y. Wang, P.-C. Chang, J.-J. Ding, T.-C. Tai, A. Santoso, Y.-T. Liu, and J.-C. Wang, "Spectral–temporal receptive field-based descriptors and hierarchical cascade deep belief network for guitar playing technique classification," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3684–3695, May 2022.

[27] V. Moscato, M. Postiglione, C. Sansone, and G. Sperlí, "TaughtNet: Learning multi-task biomedical named entity recognition from single-task teachers," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2512–2523, May 2023.

[28] J. Choi, J. Park, K. Kyung, N. S. Kim, and J. H. Ahn, "Unleashing the potential of PIM: Accelerating large batched inference of transformer-based generative models," *IEEE Comput. Archit. Lett.*, vol. 22, no. 2, pp. 113–116, Jul. 2023.

[29] Q. Xie, P. Tiwari, and S. Ananiadou, "Knowledge-enhanced graph topic transformer for explainable biomedical text summarization," *IEEE J. Biomed. Health Informat.*, vol. 28, no. 4, pp. 1836–1847, Apr. 2024.

[30] M. Aljebreen, B. Alabduallah, M. M. Asiri, A. S. Salama, M. Assiri, and S. S. Ibrahim, "Moth flame optimization with hybrid deep learning based sentiment classification toward ChatGPT on Twitter," *IEEE Access*, vol. 11, pp. 104984–104991, 2023.

[31] A. E. Saddik and S. Ghaboura, "The integration of ChatGPT with the metaverse for medical consultations," *IEEE Consum. Electron. Mag.*, vol. 13, no. 3, pp. 6–15, May 2024.

[32] O. P. Babalola, O. O. Ogundile, and D. J. J. Versfeld, "A generalized parity-check transformation for iterative soft-decision decoding of binary cyclic codes," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 316–320, Feb. 2020.

[33] H. Mazumdar, C. Chakraborty, M. Sathvik, S. Mukhopadhyay, and P. K. Panigrahi, "GPTFX: A novel GPT-3 based framework for mental health detection and explanations," *IEEE J. Biomed. Health Informat.*, early access, Oct. 30, 2024, doi: 10.1109/JBHI.2023.3328350.

[34] S. Ruksakulpiwat, A. Kumar, and A. Ajibade, "Using ChatGPT in medical research: Current status and future directions," *J. multidisciplinary Healthcare*, vol. 16, pp. 1513–1520, May 2023.

[35] H. Eyre, A. B. Chapman, K. S. Peterson, J. Shi, P. R. Alba, M. M. Jones, T. L. Box, S. L. DuVall, and O. V. Patterson, "Launching into clinical space with medspaCy: A new clinical text processing toolkit in Python," in *AMIA Annu. Symp. Proc.*, vol. 2021, 2021, p. 438.

[36] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 5, pp. 507–513, Sep. 2010.

[37] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Sep. 2019.

[38] W. Gu, X. Yang, M. Yang, K. Han, W. Pan, and Z. Zhu, "MarkerGenie: An NLP-enabled text-mining system for biomedical entity relation extraction," *Bioinf. Adv.*, vol. 2, no. 1, Jan. 2022, Art. no. vbac035.

[39] M. Pérez-Pérez, G. Pérez-Rodríguez, F. Fdez-Riverola, and A. Lourenço. (2017). *Neji: DIY web Services for Biomedical Concept Recognition*. [Online]. Available: https://biocreative.bioinformatics.udel.edu/media/store/files/2017/BioCreative_V5_paper26.pdf

[40] A. Kormilitzin, N. Vaci, Q. Liu, and A. Nevado-Holgado, "Med7: A transferable clinical natural language processing model for electronic health records," *Artif. Intell. Med.*, vol. 118, Aug. 2021, Art. no. 102086.

[41] K. Huang, J. Altosaar, and R. Ranganath, "ClinicalBERT: Modeling clinical notes and predicting hospital readmission," 2019, *arXiv:1904.05342*.

[42] Width.ai. (2023). *GPT-4 Medical Record Summarization Pipeline*. [Online]. Available: https://www.width.ai/post/gpt-4-medical-record-summarization-pipeline

[43] P. Maddigan and T. Susnjak, "Chat2VIS: Generating data visualizations via natural language using ChatGPT, codex and GPT-3 large language models," *IEEE Access*, vol. 11, pp. 45181–45193, 2023.

[44] N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota, and A. Soomro, "A systematic literature review on text generation using deep neural network models," *IEEE Access*, vol. 10, pp. 53490–53503, 2022.

[45] S. Värtinen, P. Hämäläinen, and C. Guckelsberger, "Generating role-playing game quests with GPT language models," *IEEE Trans. Games*, vol. 16, no. 1, pp. 127–139, Dec. 2022.

[46] H. Liu, Y. Cai, Z. Lin, Z. Ou, Y. Huang, and J. Feng, "Variational latent-state GPT for semi-supervised task-oriented dialog systems," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 970–984, 2023.

[47] R. L. de Queiroz and P. A. Chou, "Transform coding for point clouds using a Gaussian process model," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3507–3517, Jul. 2017.

[48] J. Liu, Y. Zhang, B. Gong, W. Lu, and Y. Xiao, "Research on text-based Q&A system technology for medical aesthetic field," in *Proc. IEEE/ACIS 22nd Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2022, pp. 14–19.

[49] S. Singh and A. Mahmood, "The NLP cookbook: Modern recipes for transformer based deep learning architectures," *IEEE Access*, vol. 9, pp. 68675–68702, 2021.

[50] J. Mao, J. Han, and T. Cui, "Development and assessment of improved global pressure and temperature series models," *IEEE Access*, vol. 9, pp. 104429–104447, 2021.

[51] Q. Lu, D. Dou, and T. H. Nguyen, "Textual data augmentation for patient outcomes prediction," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 2817–2821.

[52] K. Shailaja, B. Seetharamulu, and M. A. Jabbar, "Machine learning in healthcare: A review," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 910–914.

[53] M. Research. (2023). *Capabilities of GPT-4 on Medical Challenge Problems*. [Online]. Available: https://www.microsoft.com/en-us/research/publication/2023/03/GPT-4_medical_benchmarks-641a308e45ba9.pdf

[54] Y. Li, Y. Liu, Z. Wang, X. Liang, L. Liu, L. Wang, L. Cui, Z. Tu, L. Wang, and L. Zhou, "A comprehensive study of GPT-4V's multimodal capabilities in medical imaging," *medRxiv*, 2023.

[55] D. Freeman, B. S. Loe, D. Kingdon, H. Startup, A. Molodynski, L. Rosebrock, P. Brown, B. Sheaves, F. Waite, and J. C. Bird, "The revised green et al., paranoid thoughts scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut-offs," *Psychol. Med.*, vol. 51, no. 2, pp. 244–253, Jan. 2021.

[56] Certara. (2023). *Can't ChatGPT Do That? Practical Applications for AI in Drug Discovery & Development*. [Online]. Available: https://www.certara.com/blog/cant-chatgpt-do-that-practical-applications-for-ai-in-drug-discovery-development/

[57] S. Zhang, S. M. H. Bamakan, Q. Qu, and S. Li, "Learning for personalized medicine: A comprehensive review from a deep learning perspective," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 194–208, 2019.

[58] Y. Zhou, H. Liu, T. Srivastava, H. Mei, and C. Tan, "Hypothesis generation with large language models," in *Proc. 1st Workshop NLP Sci. (NLP4Science)*. Miami, FL, USA: Association for Computational Linguistics, Nov. 2024, pp. 117–139.

[59] C. Shaib, M. Li, S. Joseph, I. Marshall, J. J. Li, and B. Wallace, "Summarizing, simplifying, and synthesizing medical evidence using GPT-3 (with varying success)," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, Toronto, ON, Canada, 2023, pp. 1387–1407.

[60] B. Chintagunta, N. Katariya, X. Amatriain, and A. Kannan, "Medically aware GPT-3 as a data generator for medical dialogue summarization," in *Proc. 2nd Workshop Natural Lang. Process. Med. Conversations*, 2021, pp. 66–76.

[61] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2019, pp. 11328–11339.

[62] A. Joshi, N. Katariya, X. Amatriain, and A. Kannan, "Dr. Summarize: Global summarization of medical dialogue by exploiting local structures," 2020, *arXiv:2009.08666*.

[63] *Unified Medical Language System*. Accessed: Jun. 3, 2024. [Online]. Available: https://www.nlm.nih.gov/research/umls/index.html

[64] A. Chien, H. Tang, B. Jagessar, K.-W. Chang, N. Peng, K. Nael, and N. Salamon, "AI-assisted summarization of radiologic reports: Evaluating GPT3davinci, BARTcnn, LongT5booksum, LEDbooksum, LEDlegal, and LEDclinical," *Amer. J. Neuroradiol.*, vol. 45, no. 2, pp. 244–248, Feb. 2024.

[65] (2024). *BART (large-sized Model), Fine-tuned on CNN Daily Mail*. Accessed: Jun. 3, 2024. [Online]. Available: https://huggingface.co/facebook/bart-large-cnn

[66] *LongT5booksum*. Accessed: Jun. 3, 2024. [Online]. Available: https://huggingface.co/pszemraj/long-t5-tglobal-base-16384-book-summary

[67] *LED Large Book Summary*. Accessed: Jun. 3, 2024. [Online]. Available: https://huggingface.co/pszemraj/led-large-book-summary

[68] *LED for Legal Summarization of Documents*. Accessed: Jun. 3, 2024. [Online]. Available: https://huggingface.co/nsi319/legal-led-base-16384

[69] *Clinical LED Summarizer*. Accessed: Jun. 3, 2024. [Online]. Available: https://huggingface.co/griffin/clinical-led-summarizer

[70] *GPT Models Documentation*. Accessed: Jun. 3, 2024. [Online]. Available: https://platform.openai.com/docs/models/gpt-base

[71] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Proc. Text Summarization Branches Out*, Barcelona, Spain, Jul. 2004, pp. 74–81. [Online]. Available: https://aclanthology.org/W04-1013

[72] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating text generation with BERT," 2019, *arXiv:1904.09675*.

[73] M. S. Ansary, "A hybrid approach for extractive summarization of medical documents," in *Proc. IEEE Int. Conf. Biomed. Eng., Comput. Inf. Technol. Health (BECITHCON)*, Dec. 2021, pp. 1–4.

[74] J. J. Bird, M. Pritchard, A. Fratini, A. Ekárt, and D. R. Faria, "Synthetic biological signals machine-generated by GPT-2 improve the classification of EEG and EMG through data augmentation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3498–3504, Apr. 2021.

[75] B. Murdoch, "Privacy and artificial intelligence: Challenges for protecting health information in a new era," *BMC Med. Ethics*, vol. 22, pp. 1–5, Dec. 2021.

[76] T. Johnson, K. Kollnig, and P. Dewitte, "Towards responsible, lawful and ethical data processing: Patient data in the U.K.," *Internet Policy Rev.*, vol. 11, no. 1, pp. 1–24, Mar. 2022.

[77] Y. Wang, Y. Qin, D. Deng, J. Wei, Y. Zhou, Y. Fan, T. Chen, H. Sun, L. Liu, S. Wei, and S. Yin, "An energy-efficient transformer processor exploiting dynamic weak relevances in global attention," *IEEE J. Solid-State Circuits*, vol. 58, no. 1, pp. 227–242, Jan. 2023.

[78] S. Li and J. Xu, "A novel clinical trial prediction-based factual inconsistency detection approach for medical text summarization," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2023, pp. 1–8.

[79] H. Rong, G. Chen, T. Ma, V. S. Sheng, and E. Bertino, "FuFaction: Fuzzy factual inconsistency correction on crowdsourced documents with hybrid-mask at the hidden-state level," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 1, pp. 167–183, Jan. 2023.

[80] N. Tran, H. Chen, J. Bhuyan, and J. Ding, "Data curation and quality evaluation for machine learning-based cyber intrusion detection," *IEEE Access*, vol. 10, pp. 121900–121923, 2022.

[81] Y. Zhu, X. Shen, and P. Du, "Denoising-based decoupling-contrastive learning for ubiquitous synthetic face images," *IEEE Access*, vol. 11, pp. 104946–104954, 2023.

[82] S. N. Aakur and S. Sarkar, "Leveraging symbolic knowledge bases for commonsense natural language inference using pattern theory," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13185–13202, Nov. 2023.

[83] expert.ai. (2023). *Adding GPT and LLMs To Your Enterprise Hybrid Approach*. [Online]. Available: https://www.expert.ai/resource/adding-gpt-and-llms-to-your-enterprise-hybrid-approach/

[84] X. Pei, Y. Li, and C. Xu, "GPT self-supervision for a better data annotator," 2023, *arXiv:2306.04349*.

[85] M. Wang, H. Lu, S. Liu, and Z. Zhu, "How to mislead AI-assisted network automation in SD-IPoEONs: A comparison study of DRL- and GAN-based approaches," *J. Lightw. Technol.*, vol. 38, no. 20, pp. 5574–5585, Oct. 15, 2020.

[86] Z. Q. Wang and A. El Saddik, "DTITD: An intelligent insider threat detection framework based on digital twin and self-attention based deep learning models," *IEEE Access*, vol. 11, pp. 114013–114030, 2023.

[87] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the middle: How language models use long contexts," *Trans. Assoc. Comput. Linguistics*, vol. 12, pp. 157–173, Feb. 2024.

[88] G. Lan, S. Xiao, J. Yang, J. Wen, and M. Xi, "Generative AI-based data completeness augmentation algorithm for data-driven smart healthcare," *IEEE J. Biomed. Health Informat.*, early access, Oct. 30, 2023, doi: 10.1109/JBHI.2023.3327485.

[89] B. Ghai and K. Mueller, "D-BIAS: A causality-based human-in-the-loop system for tackling algorithmic bias," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 1, pp. 473–482, Jan. 2023.

[90] E. Ferrara, "Should ChatGPT be biased? Challenges and risks of bias in large language models," 2023, *arXiv:2304.03738*.

[91] J. Alasadi, A. AlHilli, P. K. Atrey, and V. K. Singh, "A generative approach to mitigate bias in face matching using learned latent structure," in *Proc. IEEE 8th Int. Conf. Multimedia Big Data (BigMM)*, Dec. 2022, pp. 150–157.

[92] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, "On faithfulness and factuality in abstractive summarization," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, Jan. 2020, pp. 1906–1919.

[93] V. Adlakha, P. BehnamGhader, X. Han Lu, N. Meade, and S. Reddy, "Evaluating correctness and faithfulness of instruction-following models for question answering," 2023, *arXiv:2307.16877*.

[94] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, "CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 699–711, Feb. 2021.

[95] K. A. Hasenstab, J. Huynh, S. Masoudi, G. M. Cunha, M. Pazzani, and A. Hsiao, "Feature interpretation using generative adversarial networks (FIGAN): A framework for visualizing a CNN's learned features," *IEEE Access*, vol. 11, pp. 5144–5160, 2023.

[96] F. Xing, B. Schuller, I. Chaturvedi, E. Cambria, and A. Hussain, "Guest editorial neurosymbolic AI for sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 1711–1715, Jul. 2023.

[97] A. J. Mahdi, T. Tettamanti, and D. Esztergár-Kiss, "Modeling the time spent at points of interest based on Google popular times," *IEEE Access*, vol. 11, pp. 88946–88959, 2023.

[98] B. Wang, W. Chen, H. Pei, C. Xie, M. Kang, C. Zhang, C. Xu, Z. Xiong, R. Dutta, R. Schaeffer, S. T. Truong, S. Arora, M. Mazeika, D. Hendrycks, Z. Lin, Y. Cheng, S. Koyejo, D. Song, and B. Li, "DecodingTrust: A comprehensive assessment of trustworthiness in GPT models," 2023, *arXiv:2306.11698*.

[99] M. Kuzlu, Z. Xiao, S. Sarp, F. O. Catak, N. Gurler, and O. Guler, "The rise of generative artificial intelligence in healthcare," in *Proc. 12th Medit. Conf. Embedded Comput. (MECO)*, Jun. 2023, pp. 1–4.

[100] V. Kumar, D. Reforgiato Recupero, R. Helaoui, and D. Riboni, "K-LM: Knowledge augmenting in language models within the scholarly domain," *IEEE Access*, vol. 10, pp. 91802–91815, 2022.

[101] Z. Huang, P. K. Damalapati, and E. Wu, "Data ambiguity strikes back: How documentation improves GPT's text-to-SQL," 2023, *arXiv:2310.18742*.

[102] G. Sebastian, "Privacy and data protection in ChatGPT and other AI chatbots: Strategies for securing user information," *SSRN Electron. J.*, vol. 15, no. 1, pp. 1–14, 2023.

[103] O. Aouedi, A. Sacco, K. Piamrat, and G. Marchetto, "Handling privacy-sensitive medical data with federated learning: Challenges and future directions," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 2, pp. 790–803, Feb. 2023.

[104] A. N. Ufairah and Y. Widyani, "Experiment on utilizing GPT-3.5 language model to generate DSL of data management for web application development using qore-base platform," in *Proc. IEEE Int. Conf. Data Softw. Eng. (ICoDSE)*, Sep. 2023, pp. 108–113.

[105] M. L. Jones, E. Kaufman, and E. Edenberg, "AI and the ethics of automating consent," *IEEE Secur. Privacy*, vol. 16, no. 3, pp. 64–72, May 2018.

[106] C. Patsakis and N. Lykousas, "Man vs the machine in the struggle for effective text anonymisation in the age of large language models," *Sci. Rep.*, vol. 13, no. 1, p. 16026, Sep. 2023.

[107] A. Tamersoy, G. Loukides, M. E. Nergiz, Y. Saygin, and B. Malin, "Anonymization of longitudinal electronic medical records," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 413–423, May 2012.

[108] Y. Wang, Y. Pan, M. Yan, Z. Su, and T. H. Luan, "A survey on ChatGPT: AI–generated contents, challenges, and solutions," *IEEE Open J. Comput. Soc.*, vol. 4, pp. 280–302, 2023.

[109] J. Ragsdale and R. V. Boppana, "On designing low-risk honeypots using generative pre-trained transformer models with curated inputs," *IEEE Access*, vol. 11, pp. 117528–117545, 2023.

[110] D. B. Johnson et al., "Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the chat-GPT model," *Res. square*, Feb. 2023.

[111] X. Pan, M. Zhang, S. Ji, and M. Yang, "Privacy risks of general-purpose language models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 1314–1331.

[112] H. Müller and D. Unay, "Retrieval from and understanding of large-scale multi-modal medical datasets: A review," *IEEE Trans. Multimedia*, vol. 19, no. 9, pp. 2093–2104, Sep. 2017.

[113] T. Yang, F. Ma, X. Li, F. Liu, Y. Zhao, Z. He, and L. Jiang, "DTATrans: Leveraging dynamic token-based quantization with accuracy compensation mechanism for efficient transformer architecture," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 42, no. 2, pp. 509–520, Feb. 2023.

[114] V. L. Mane, S. S. Panicker, and V. B. Patil, "Summarization and sentiment analysis from user health posts," in *Proc. Int. Conf. Pervasive Comput. (ICPC)*, Jan. 2015, pp. 1–4.

[115] S. Hong, S. Moon, J. Kim, S. Lee, M. Kim, D. Lee, and J.-Y. Kim, "DFX: A low-latency multi-FPGA appliance for accelerating transformer-based text generation," in *Proc. 55th IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Oct. 2022, pp. 616–630.

[116] N. Muennighoff, "SGPT: GPT sentence embeddings for semantic search," 2022, *arXiv:2202.08904*.

[117] S. Liu, P. Li, J. Zhang, Y. Wang, H. Zhu, W. Jiang, S. Tang, C. Chen, Q. Liu, and M. Liu, "16.2 A 28 nm 53.8TOPS/W 8b sparse transformer accelerator with in-memory butterfly zero skipper for unstructured-pruned NN and CIM-based local-attention-reusable engine," in *Proc. IEEE Int. Solid State Circuits Conf. (ISSCC)*, Feb. 2023, pp. 250–252.

[118] Y. Dang, M. Xu, and K. Ye. (2023). *Resource Management for GPT-based Model Deployed on Clouds: Challenges, Solutions, and Future Directions*. Papers with Code. [Online]. Available: https://cs.paperswithcode.com/paper/resource-management-for-gpt-based-model

[119] J. Park, J. Nam, J. Choi, Y.-G. Shin, and S. Park, "Structured medical dataset analysis tool based on ChatGPT," in *Proc. 14th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2023, pp. 837–842.

[120] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and M. M. Kabir, "A survey of automatic text summarization: Progress, process and challenges," *IEEE Access*, vol. 9, pp. 156043–156070, 2021.

[121] Y.-M. Nam, D. Han, and M.-S. Kim, "A parallel query processing system based on graph-based database partitioning," *Inf. Sci.*, vol. 480, pp. 237–260, Apr. 2019.

[122] B. Burger, D. K. Kanbach, S. Kraus, M. Breier, and V. Corvello, "On the use of AI-based tools like ChatGPT to support management research," *Eur. J. Innov. Manage.*, vol. 26, no. 7, pp. 233–241, Dec. 2023.

[123] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, "AI-based conversational agents: A scoping review from technologies to future directions," *IEEE Access*, vol. 10, pp. 92337–92356, 2022.

[124] M. Hu, S. Pan, Y. Li, and X. Yang, "Advancing medical imaging with language models: A journey from N-grams to ChatGPT," 2023, *arXiv:2304.04920*.

[125] M. Javaid, A. Haleem, and R. P. Singh, "ChatGPT for healthcare services: An emerging stage for an innovative perspective," *BenchCouncil Trans. Benchmarks, Standards Eval.*, vol. 3, no. 1, Feb. 2023, Art. no. 100105.

[126] Z. Ahmed. (2023). *Leveraging GPT-4 for Increasingly Dense Summarization*. [Online]. Available: https://medium.com/@zahmed333/leveraging-gpt-4-for-increasingly-dense-summarization-55f01efa0308

[127] S. Wang, J. Li, X. Li, Y. Liu, and X. Liu. (2020). *Automatic Text Summarization of COVID-19 Medical Research Articles Using BERT and GPT-2*. Papers with Code. [Online]. Available: https://paperswithcode.com/paper/automatic-text-summarization-of-covid-19/review/

[128] Z. Chen, C. Zhang, Q. Wang, J. Troidl, S. Warchol, J. Beyer, N. Gehlenborg, and H. Pfister, "Beyond generating code: Evaluating GPT on a data visualization course," 2023, *arXiv:2306.02914*.

[129] Y. Jeong and E. Kim, "SciDeBERTa: Learning DeBERTa for science technology documents and fine-tuning information extraction tasks," *IEEE Access*, vol. 10, pp. 60805–60813, 2022.

[130] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," *Future Gener. Comput. Syst.*, vol. 135, pp. 364–381, Oct. 2022.

[131] M. S. Rahaman, M. M. T. Ahsan, N. Anjum, H. J. R. Terano, and M. M. Rahman, "From ChatGPT-3 to GPT-4: A significant advancement in AI-driven NLP tools," *J. Eng. Emerg. Technol.*, vol. 1, no. 1, pp. 50–60, May 2023.

[132] P. Hämäläinen, M. Tavast, and A. Kunnari, "Evaluating large language models in generating synthetic HCI research data: A case study," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2023, pp. 1–19.

[133] Y. K. Dwivedi et al., "Opinion Paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, Aug. 2023, Art. no. 102642.

[134] C. Ma, W. E. Zhang, M. Guo, H. Wang, and Q. Z. Sheng, "Multi-document summarization via deep learning techniques: A survey," *ACM Comput. Surveys*, vol. 55, no. 5, pp. 1–37, May 2023.

[135] D. Carlander-Reuterfelt, Á. Carrera, C. A. Iglesias, Ó. Araque, J. F. S. Rada, and S. Muñoz, "JAICOB: A data science chatbot," *IEEE Access*, vol. 8, pp. 180672–180680, 2020.

[136] M. Abdullah, A. Madain, and Y. Jararweh, "ChatGPT: Fundamentals, applications and social impacts," in *Proc. 9th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*, Nov. 2022, pp. 1–8.

[137] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, "Class-incremental learning: Survey and performance evaluation on image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5513–5533, May 2023.

BALAMURUGAN PALANISAMY is currently pursuing the Ph.D. degree with the Department of EEE, Birla Institute of Technology and Science, Pilani, Rajasthan, India. His research interests include natural language processing, deep learning, and generative models.

ARJAB CHAKRABARTI is currently pursuing the B.Tech. degree with the Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar. He has completed a few projects in the field of machine learning and web designing. Since Summer 2023, he doing his research internship with the Birla Institute of Technology and Science (BITS), Pilani, under Dr. Vikas Hassija. His research interests include machine learning, reinforcement learning, quantum computing, and deep learning.

ANUSHKA SINGH is currently pursuing the B.Tech. degree with the Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar. Since Summer 2023, she doing her research internship with the Birla Institute of Technology and Science (BITS), Pilani, under Dr. Vikas Hassija. Her research interests include machine learning, reinforcement learning, quantum computing, and deep learning.

VIKAS HASSIJA received the B.Tech. degree from M. D. U. University, Rohtak, India, in 2010, the M.S. degree in telecommunication and software engineering from the Birla Institute of Technology and Science (BITS), Pilani, India, in 2014, and the Ph.D. degree in IoT security and blockchain from the Jaypee Institute of Information Technology (JIIT), Noida. He has done his Postdoctoral Research with the National University of Singapore, Singapore. He is currently working as an Associate Professor at KIIT, Bhubaneswar. He has also worked as an Assistant Professor with JIIT for four years. He has eight years of industry experience and has worked with various telecommunication companies, such as Tech Mahindra and Accenture. His research interests include IoT security, network security, blockchain, and distributed computing.

G. S. S. CHALAPATHI (Senior Member, IEEE) received the B.E. degree (Hons.) in electrical and electronics engineering from the Birla Institute of Technology and Science (BITS) Pilani, in 2009, and the M.E. degree in embedded systems and the Ph.D. degree from BITS Pilani, in 2011 and 2019, respectively. He carried out his postdoctoral research with The University of Melbourne, Australia, under the supervision of Prof. Rajkumar Buyya, and a Distinguished Professor with The University of Melbourne. During his doctoral studies, he has been a Visiting Researcher with the National University of Singapore and Johannes Kepler University, Austria. He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, BITS-Pilani. He has published in reputed journals, such as IEEE Wireless Communication Letters, IEEE Sensors Journal, and *Future Generation Computing Systems*. His research interests include UAVs, precision agriculture, and embedded systems. He is a member of ACM. He is a Reviewer of IEEE Internet of Things Journal and IEEE Access.

AMIT SINGH received the B.E. and M.E. degrees in mechanical engineering from BITS Pilani, in 2009 and 2013, respectively, and the Ph.D. degree in mechanical engineering from the University of California at Los Angeles, with solid and structural mechanics as his major and fluid mechanics as his minor. He was a Postdoctoral Researcher with Johns Hopkins University, working in computational biophysics before joining the Department of Mechanical Engineering, BITS Pilani, in December 2019. Broadly, his research interests include computational mechanics and machine learning, particularly mechanics problems arising from biological phenomena.