Large language model to multimodal large language model: A journey to shape the biological macromolecules to biological sciences and medicine

Manojit Bhattacharya,[1,5] Soumen Pal,[2,5] Srijan Chatterjee,[3,5] Sang-Soo Lee,[3] and Chiranjib Chakraborty[4,5]

[1]Department of Zoology, Fakir Mohan University, Vyasa Vihar, Balasore, Odisha 756020, India; [2]School of Mechanical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India; [3]Institute for Skeletal Aging & Orthopedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon, Gangwon-Do 24252, Republic of Korea; [4]Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Kolkata, West Bengal 700126, India

After ChatGPT was released, large language models (LLMs) became more popular. Academicians use ChatGPT or LLM models for different purposes, and the use of ChatGPT or LLM is increasing from medical science to diversified areas. Recently, the multimodal LLM (MLLM) has also become popular. Therefore, we comprehensively illustrate the LLM and MLLM models for a complete understanding. We also aim for simple and extended reviews of LLMs and MLLMs for a broad category of readers, such as researchers, students in diversified fields, and other academicians. The review article illustrates the LLM and MLLM models, their working principles, and their applications in diversified fields. First, we demonstrate the technical concept of LLMs, working principle, Black Box, and the evolution of LLMs. To explain the working principle, we discuss the tokenization process, token representation, and token relationships. We also extensively demonstrate the application of LLMs in biological macromolecules, medical science, biological science, and other areas. We illustrate the multimodal applications of LLMs or MLLMs. Finally, we illustrate the limitations, challenges, and future prospects of LLMs. The review acts as a booster dose for clinicians, a primer for molecular biologists, and a catalyst for scientists, and also benefits diversified academicians.

INTRODUCTION

Since the launch of ChatGPT by OpenAI on November 30, 2022, large language models (LLMs) have become popular quickly. The medical and scientific communities have been thrilled to use LLMs in various biology, medicine, and science areas. During the last phase of 2022, Stokel-Walker reported in *Nature* that ChatGPT can write Smart essays.[1] After that, people found that the written responses of this LLM chatbot were rapid and often invariable from humans. Conversely, researchers found that LLMs can process the text from queries and can respond, revolutionizing the field.

LLMs are one of the significant achievements among the recent noteworthy technologies of artificial intelligence (AI). The rapid development of AI has led to the development of this sophisticated technology of LLMs. As a component of AI, LLMs are trained with billions of words derived from internet-based content, books, articles, and massive texts. During the production of text, the LLM chatbots also extract input from billions of unidentified general web pages. However, using AI and NLP (natural language processing), LLM chatbots can recognize questions and offer automated answers. They are LLM-based dialog agents.[2,3] This model can answer free-text queries without being trained with a specific topic. The model's explicit training objective is constructed on the generative pre-trained transformer (GPT) architecture. It is to predict the next word efficiently and can be prepared in a sentence or paragraph. Their performance has become increasingly human-like through dialog as a dialog agent (Figure 1). The most essential benefit of GPT models is their processing speed. The GPT model can answer complex input queries in just a few seconds. LLMs use a machine learning model based on the architectures of neural networks.[2,4–6] The skill resembles cognitive capability.

The publicly popularized LLM chatbot, ChatGPT was GPT model 3.5. Millions of users started to use it within a few months after it was released. It was noted that about 100 million users used ChatGPT within 2 months after its launch.[7] Subsequently, interest in LLM chatbots has increased very fast across academic domains. Afterward, it has been used across industrial domains.[8,9] In medical science, it has been applied in different areas such as cardiology,[10] orthopedics,[11] radiology,[12] infectious disease,[13] drug resistance,[14] surgery,[15] etc. Other than medical science, ChatGPT has been used in different academic areas such as pharmacology and

Correspondence: Chiranjib Chakraborty, Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Kolkata, West Bengal 700126, India.
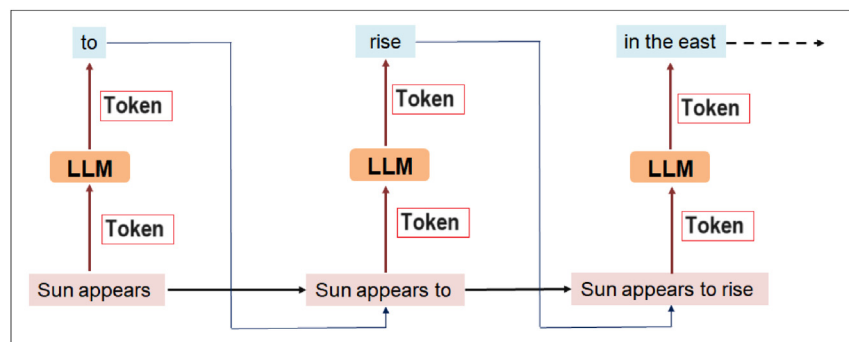**E-mail:** drchiranjib@yahoo.com

Figure 1. The working principle of a LLM in generating a single-token sequence
The model generates tokens, ultimately forming a complete sentence.

Basic technical concept of LLMs

The role of an LLM is to respond to questions. The model is trained with a considerable number of tokens. It is based on transformer architecture and NLP. The transformer architecture is a type of deep learning neural network architecture. The significant questions are: Which token will most probably come next? What will the weightage of a token be? The LLM works through the probability distribution. The probability distribution is as follows:

$$P(\omega_{n+1}|\omega_1 \ldots \ldots \omega_n),$$ where $\omega_1 \ldots \omega_n$ is a sequence of tokens (the context) and ω_{n+1} is the predicted next token.[2]

How the LLM works

Language models (LMs) have been developing for several years to enhance the intellect of machine languages. Technically, LMs function with probability value allocation to word sequence and deliver appropriate text output (Figure 1).[23] LMs utilize tokens as basic units for understanding and generating text output, where tokens refer to parts of text that can be words, characters, or sub-words. With the use of tokens, LMs can grasp the connection and relation among words, enabling them to produce grammatically textual output that is right.[24] LLMs, i.e., pre-trained LMs with millions of parameters, can grasp and produce output strikingly similar to human language. The LLMs operate with the tokenization theory and can apply millions of tokens to produce output.

Tokenization process

In tokenization, the text is split into small sections known as tokens, which are subsequently utilized for analysis in LLMs. Tokenization is crucial in text pre-processing, preparing input tokens for LMs. The commonly employed methods for tokenization are WordPiece and BPE (byte-pair encoding), which are utilized by significant models such as bidirectional encoder representations from transformers (BERT) as well as the GPT. Nevertheless, the tokenization's effect may vary in languages that are rich morphologically, such as the Turkic languages, where the addition of prefixes and suffixes can generate numerous words. In such cases, a newly formed tokenizer operating at the morphological level can challenge the already established tokenizers.[25] Within LLMs, tokenization also holds importance in speech processing. The tokenization process in discrete speech notably contributes to integrating speeches into the LLMs.

drug discovery,[16] law,[17] education,[18] biomedical engineering,[19] finance,[20] etc.

Subsequently, in 2023, OpenAI released GPT model 4, which is a multimodal LLM (MLLM). MLLMs can be trained with video, audio, or image, along with all the training parameters that are used to trained LLM. It can be trained with more comprehensive training parameters. MLLM is a robust model.[21] Therefore, it is considered a more advanced version of LLM. GPT-3.5 and GPT-4 used different parameters and tokens. It was reported that GPT-3 was using over 175 billion machine learning parameters. GPT-3's parameters (175 billion) help the LLM's vast understanding of language and knowledge across various domains. This machine learning model was trained on 300 billion tokens. GPT-3.5 is more fine-tuned than GPT-3's abilities. GPT-3.5 is more refined and exhibits more accuracy compared with GPT-3. Similarly, the GPT-4 model was developed to incorporate about 1.8 trillion parameters. At the same time, this MLLM is trained with 13 trillion tokens.[22]

Here, this comprehensive review article illustrates the basic technical concept of LLM, its working principle, LLM and its Black Box, and the evolution of LLMs. We also illustrate the application of LLMs in medical science, biological science, and other areas. We illustrate the multimodal applications of LLMs or MLLMs. Finally, we illustrate the limitations, challenges, and future prospects of LLMs. The article has been simplified and extended, which will help a broad group of readers to understand the topic better.

Nonetheless, the discrete process gives rise to information loss, which impairs the performance. One unique speech-representing codec called RepCodec has been proposed for semantic speech tokenization to enhance the functioning of these speech tokens (discrete).[26] In the context of text summarization, tokenization is an essential technique for obtaining the needed information precisely.[27] Summaries help lower the time needed to read and facilitate researching documents. However, tokenization can potentially introduce ambiguity, and there exists a lack of clarity on whether the divided tokens attain optimal performance for the intended task.[28] An innovative, straightforward, and effective method named GrowLength has been introduced to expedite the pre-training process of LLMs. This method gradually enhances training length during the pre-training period, thus reducing computing expenses and improving efficiency.[29]

Token representation

LLMs utilize embedding and encoding mechanisms to represent tokens. These mechanisms enable the model to comprehend and generate natural language by capturing the tokens' semantic and syntactic information. Embedding mechanisms map tokens to high-dimensional vectors, thereby facilitating the model's grasp of their semantic meaning. One commonly adopted approach involves utilizing pre-trained embeddings of the words, such as GloVe or Word2Vec, which offer distributed descriptions of words depending on their contextual usage.[30] Within LLMs, token embeddings are acquired through pertaining and fine-tuning, enabling the model to capture intricate linguistic patterns and relationships between tokens.[31] Encoding mechanisms, on the other hand, process token embeddings to capture sequential and contextual information. In LLMs, this is typically accomplished by employing self-attention mechanisms such as the transformer architecture, which allows it to allocate important weights for each token in the overall input sequence.[32] Consequently, the model becomes proficient in capturing long-range dependencies and contextual information, which proves critical for numerous NLP tasks. In addition to self-attention, recent research has explored novel token encoding mechanisms, such as recurrent alignment and contrastive losses, designed to encapsulate nuanced semantic relationships between tokens and optimize the embedding space.[33]

Token context and relationships

LLMs such as GPT-3 and BERT can capture context and establish relationships between tokens by generating token representations that retain crucial contextual knowledge necessary for various tasks. An exemplification of this is observed in the SPAE (Semantic Pyramid AutoEncoder), which empowers LLMs that are frozen to undertake comprehension and generate tasks requiring non-linguistic procedures such as videos or visual representation of objects. SPAE accomplishes this by converting raw pixels into interpretable lexical tokens taken out from the word stock of the LLM, thereby capturing both semantic meaning and intricate details necessary for visual reconstruction.[34] In addition, contextual LMs, such as BERT produce token representations that retain context-specific knowledge essential for tasks at the type level, thereby indicating the context sensitivity of processes

such as similarity estimation and relatedness estimation.[35] The contextual knowledge facilitates LLMs in capturing intricate relationships between tokens, rendering them suitable for various tasks, including image generation and semantic estimation. Attention mechanisms in the models play a pivotal role in contextual understanding by allowing the models to selectively focus on different segments of the input sequence to capture pertinent information. An illustration of this is the tri-attention framework in NLP, which explicitly incorporates the query, key, and context interactions by including context as the third dimension in the computation of relevance scores. Thus, this framework surpasses traditional bi-attention approaches and pre-trained neural LMs in various NLP jobs[36] Furthermore, the Vit-BiGRU-Attention sentiment classification model utilizes attention mechanisms to assign varying weights to individual words, thereby enhancing the comprehension of emotions and determining the polarity of emotions in user comments, ultimately leading to improved accuracy in sentiment classification.[37]

As a machine learning model, LLM architecture primarily consists of numerous layers of neural networks, such as recurrent, feedforward, embedding, and attention layers. It uses a probabilistic model, tokenization process and representation, and neural architecture at a time to generate human-like language text.

LLM and its "Black Box"

Neuroscience models mainly guide artificial neural networks (ANNs), which brain mechanisms encourage and develop. Unfortunately, networks generated by neurons are described as unclear as those generated by the brain. The data was observed to diffused, so it was not straightforward to solve. AI is developed using the ANN model, which cannot be adequately described in the Black Box of AI. Researchers are trying to describe the inner workings of a complex Black Box model of AI.[38] The Black Box can be applicable to LLM or ChatGPT models.[39,40] However, the interpretability of the LLM output cannot be adequately explained. At the same time, tokenization processes cannot be adequately explained. Therefore, the output of LLMs can be described as the Black Box of LLMs and understanding it is a challenge for the researcher.[41–43]

Evolution of LLMs

The evolution of LLMs is evidence of the rapid pace of AI research and innovation. The journey of LLMs started with simpler LMs, and the present journey continues with the development of massive neural networks such as GPT-3.5 and GPT-4 (Figure 2). The present LLM, GPT-4, is a multimodal LLM with immense capabilities across vision, video, audio, language, and 3D. It also claims billions of parameters along with a safety research and monitoring system.[44] The timelines of LLM evolution include multiple stages from the last few years. It evolved significantly over the years, with advancements in model architecture and training methodologies.[45]

Before the era of deep learning, rule based and statistical methods were dominated by NLP. Models such as ALICE (1995) and Eliza (1966) put the foundation for conversational agents. However, they
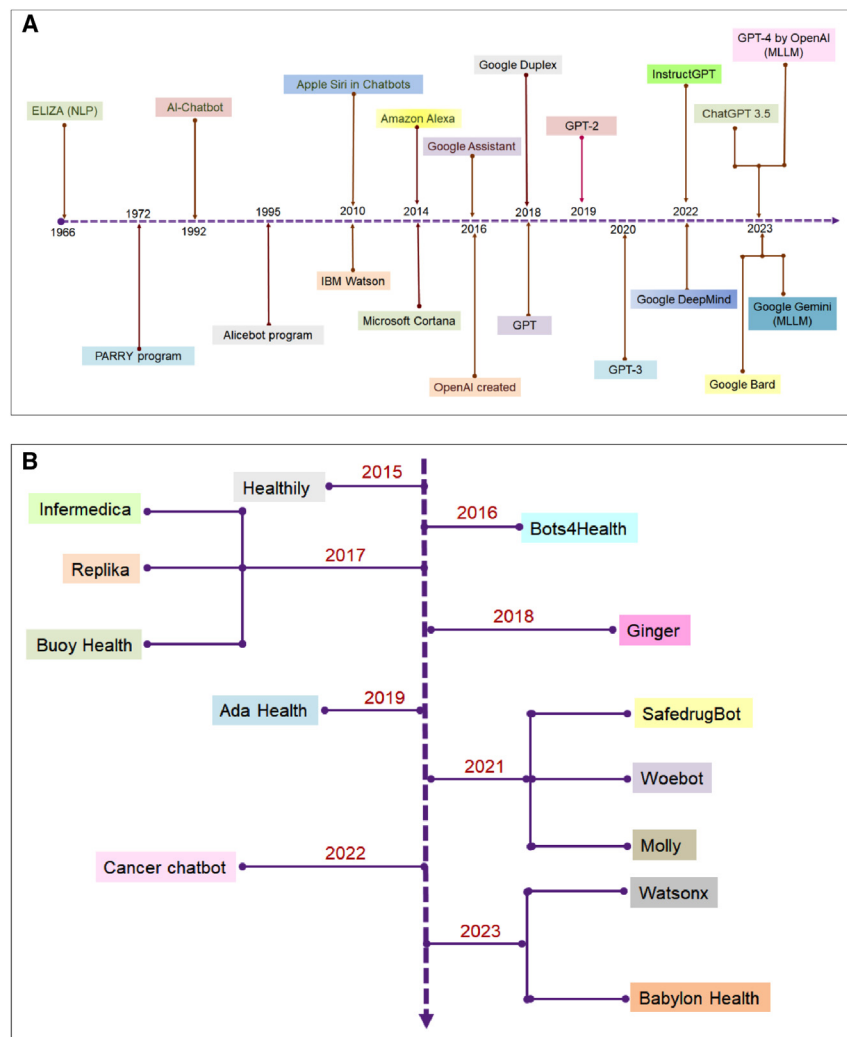
Figure 2. The evolution of a LLM through chatbots
(A) The evolution of LLM chatbots with a timeline that finally developed the recent MLLM. (B) Some significant medical chatbots follow the LLM or MLLM models.

as Word2Vec and GloVe, represented words as continuous vector spaces. These embeddings captured semantic relationships between words, offering better representations than traditional methods.[50] In 2017, statistical language processing came out, and the introduction of deep learning techniques revolutionized NLP. The RNN and LSTM networks showed promise in sequential data tasks, paving the way for more advanced models. The appearance of Google transformer architectures and attention mechanisms marked a turning point in LLM evolution. Transformers use attention mechanisms to process words about all other words in a sentence, significantly improving contextual understanding.[51]

OpenAI's GPT series, starting with GPT-1, then GPT-2, and the massive GPT-3, showcased the power of pre-trained models fine-tuned for various tasks. GPT-3's unprecedented scale generated coherent and contextually relevant text.[52,53] GPT-2 was released in 2018 and can use 1.5 billion parameters. This version of GPT used the extensive collection of free novel books dataset called the BooksCorpus dataset, containing 11,308 novels. It has been noted that it includes about 1×10^9 words or around 74 million sentences.[6] Similarly, GPT-3 was released in 2020, and it can handle 175 billion parameters. GPT-3 is assumed to be 100 times more extensive than the previous GPT (GPT-2). The training dataset comprises 45 terabytes and 5 corpora. It contains Wikipedia, WebText2, Common Crawl (webpages), Books1, and Books2. GPT-3 is one of the most sophisticated LLMs until today.[6] Due to user requirements, GPT-3 was evolved into GPT-4, an MLLM.

Simultaneously, several other pre-trained LMs were developed in recent years (Table 1), which include open pre-trained transformer, pathways language model (PaLM), anthropic-LM, language model for dialog applications, MT-NLG, and LLaMA.[6,54]

LLMs stand as a superior consequence of the remarkable progress in AI research. Their evolution from basic LMs to powerful transformers has redefined the possibilities of NLP. While their applications across domains offer immense benefits, ethical considerations must guide their deployment to ensure a responsible and equitable social integration.[55] As we continue to explore the potential of LLMs, a balanced

were rule based and lacked a proper language understanding.[46] In the 2000s, statistical models such as Hidden Markov Models and n-grams improved language processing by considering probabilities of word sequences. These models were data driven but limited in handling complex language distinctions.[47]

In subsequent consideration of advancement LLMs, a recurrent neural network (RNN) model is an ANN model that processes and converts sequential data inputs into sequential data outputs. RNNs were among the first neural network architectures applied to sequential data, such as text, and were initially created in 1980s. Currently, it is ideally suited for machine learning problems involving sequential data. However, it suffered from vanishing gradient problems, limiting it ability to capture long-range dependencies. The long short-term memory (LSTM) networks addressed the vanishing gradient problem by introducing a memory cell in 1997.[48] It allowed them to capture long-term dependencies in sequential data, improving performance in language-related tasks.[49] Then, in 2013, word embeddings, such

Table 1. Different recently developed LLMs and their developers

Sl. no.	LLM name	Developer	Year of release	Remarks
1.	LLaMA	Meta	2023	It is the collection of foundation language models ranging from 7B to 65B parameters. This is modeled on trillions of tokens, and shows that it has the potential to train state-of-the-art models using publicly obtainable datasets exclusively, deprived of resorting to exclusive and inaccessible datasets.
2.	ChatGPT	OpenAI	2022	It is based on an LLM, and enables users to refine and steer a specific conversation toward a preferred length, style, format, and level of detail, as well as language.
3.	Flamingo	DeepMind	2022	It is a single visual language model (VLM), sets a new state-of-the-art in few-shot learning on an extensive range of open-ended multimodal tasks. It is also able to tackle a number of challenging problems with just a handful of task-specific examples, lacking any additional training prerequisite.
4.	DALL-E	OpenAI	2021	It uses developed text-to-image models by OpenAI by deep learning methods to produce digital images from natural language descriptions, known as prompts.
5.	Anthropic-LM	Anthropic	2023	A combination of AI chatbot and LLM that can help doctors to diagnose diseases accurately and efficiently. An improved version has also been designed to be safer than other models and is sometimes called a potential ChatGPT killer.
6.	Turing-NLG	Microsoft	2020	It is a 17-billion-parameter language model by Microsoft that outperforms the state-of-the-art on many downstream NLP tasks. In addition to completing an unfinished sentence, it can generate direct answers to questions and summaries of input documents.
7.	Minerva	Google	2022	Minerva is based on the pathways language model (PaLM). It is a 540-billion-parameter language model that can generalize across different domains and tasks. It was trained on a dataset of scientific papers and web pages that contain mathematical expressions.
8.	Wu Dao 2.0	Academia	2021	It has the ability to perform natural language processing and image recognition, in addition to generation of text and images. The model cannot only write essays, poems, and couplets in traditional Chinese, it can both generate alt text based on a static image and generate nearly photorealistic images based on natural language descriptions.
9.	Imagen	Google	2023	Imagen is a text-to-image diffusion model with an exceptional degree of photorealism and a deep level of language understanding. It builds on the power of large transformer language models in appreciative text and hinges on the strength of diffusion models in a high-fidelity image generation group.
10.	GPT-3	OpenAI	2020	It is a decoder-only transformer model of a deep neural network, which overtakes recurrence and convolution-based architectures It contains sustainable identified themes, emotions, and sentiment from surveys, reviews, live chat logs, help desk tickets, and more.
11.	Megatron	Nvidia	2021	It is an extremely optimized and well-organized library for training LLMs. With this model parallelism, language models can be trained with billions of weights and then used in NeMo for downstream tasks.

approach that combines technological advancement with ethical mindfulness will shape the future of AI and human-machine interactions.[56] A continuous drive marks the evolution of LLMs for larger models and further evolution into MLLM, which provides better pre-training strategies and increased attention for future considerations to ensure responsible AI development and deployment.

Prompt engineering

LLMs or chatbot technologies use deep learning and NLP to learn the language patterns for conversations with humans from a large amount of text data. In conversations with humans, LLMs depend on proper input or high-quality prompting, which is called "prompt engineering."[57,58] Therefore, asking the right question to an LLM is essential. This is a new area of research that focuses on refining, designing, and implementing instructions or prompts to improve LLMs' output.

Different scientists use prompt engineering to yield more valuable outputs for precise instructions from LLMs. Kleinig et al. illustrated how to use prompt engineering in ophthalmology.[57] Venerito et al. explained the use of prompt engineering in rheumatology research

and described how prompts are methodically constructed in this area of research.[59] Polak and Morgan explained how to extract proper data from research papers.[60] Therefore, prompt engineering is valuable in biological macromolecules, biological sciences, and medicine.

Application of LLMs in biological macromolecules

Recently, LLMs have been used in various fields of biological macromolecules (Figure 3). Researchers have been trying to use LLMs to understand the properties and functions of biological macromolecules.

Proteins

LLMs have been used in various fields of protein research. Researchers have been using LLMs to understand the properties and functions of proteins. Therefore, researchers are using LLMs in different areas of protein science. They have also attempted to develop protein-centric LLMs to perform protein-related tasks. Recently, Zhuo et al. have suggested one LLM for protein entitled PROTLLM to perform protein-language-related and protein-associated tasks. It is a versatile crossmodal LLM that can perform dynamic protein-centric assignments and dynamic protein mounting.[61] Likewise, Guo et al. suggested one LLM for the protein named Proteinchat to
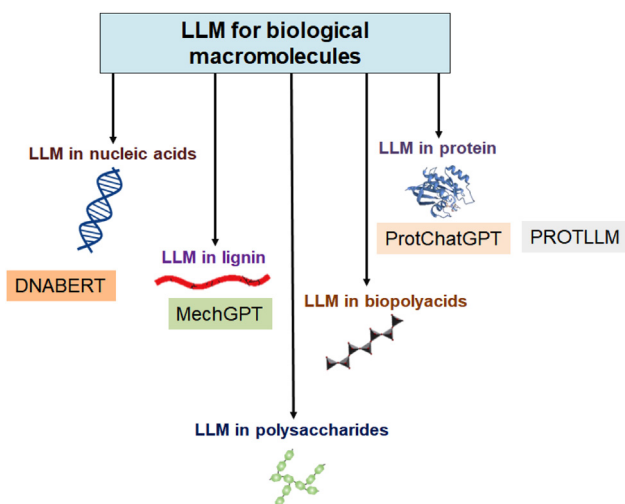
Figure 3. Some significant applications of a LLM in the field of biological macromolecules.

perform chatbot-like functionalities on three-dimensional protein structures.[62] Similarly, Wang et al. suggested another LLM for the protein named ProtChatGPT for comprehending proteins with an LLM.[63] At the same time, Wang et al. developed another LLM model for a protein called InstructProtein. The LLM model helps to comprehend the protein language and align humans through knowledge instruction.[64] However, LLMs have started to be used to explore novel research of different areas of protein structure and function, and other different areas.

Nucleic acid research

LLMs have recently been used in various fields of nucleic acid research. We recently explained LLM's role in the field.[65] Researchers have conducted GeneTuring using six GPT models: new Bing, ChatGPT, BioMedLM, BioGPT, GPT-3, and GPT-2. The open company developed ChatGPT, GPT-3, and GPT-2. The GeneTuring test was conducted with an exhaustive QA database with 600 genomics questions. New Bing's overall performance was the best.[66] Ji et al. use a pre-trained transformer model called DNABERT. Using this pre-trained model, they have indicated transcription factor binding sites, splice sites, and promoters.[67] Recently, a group of researchers used DeepMind to forecast the effective and improved gene expression prediction from DNA sequences.[68]

Polysaccharides and lignin

LLMs have recently started to be used to explore various fields of polysaccharides and lignin. Researchers have studied ChatGPT's performance in responding to glycobiology and carbohydrate chemistry queries. Williams and Fadda explored ChatGPT's answer style to different glycobiology and carbohydrate chemistry questions. They found that the model can answer short and descriptive questions correctly. However, they found that answers contained fabricated text.[69] Researchers developed an LLM for material modeling. Buehler developed MechGPT for materials modeling and mechanics, which

can be used for lignin modeling.[70] David et al. developed an LLM or ANN to assess sugar output from Kraft waste-based lignocellulosic pre-treatments. They tried to understand that domain-specific knowledge can help accelerate the progression of lignocellulosic waste pre-treatment.[71]

Application of LLMs in biological sciences

Drug discovery and development

LLMs play a significant role in drug discovery, offering valuable contributions in various stages of the drug development process. They can compute extensive pieces of scientific literature, extracting relevant information about potential drug targets, biomarkers, and mechanisms of action.[72] Likewise, they aid in identifying potential drug targets by analyzing biological and biomedical texts to understand the relationships between genes, proteins, and diseases. LLMs can predict potential drug interactions, assessing the likelihood of adverse or synergistic therapeutic effects.[9] Moreover, the LLMs can contribute to monitoring adverse events related to drugs by analyzing medical literature, clinical trial reports, and social media. They also help match eligible patients to clinical trials by analyzing electronic health records, medical literature, and patient data. The application of LLMs in drug discovery demonstrates their ability to process and understand large volumes of biomedical information, offering valuable insights that can identify novel therapeutic targets, which initiate faster methods of novel therapeutics development.[16] The application of LLMs in drug discovery demonstrates their ability to process and understand large volumes of biomedical information, offering valuable insights that can trigger the faster process of identification of novel therapeutics and their development.

Molecular biology and computational biology

Chatbot technology is evolving very fast and is appearing as a new AI tool for molecular biologists and computational biologists. Several researchers have checked the potentiality of LLMs in molecular biology. Recently, Ross and Gopinath illustrated the process of learning the structural biophysics of DNA using an LLM.[73] Lubiana et al. give ten tips to assist computational biologists in optimizing the research workflow with an LLM or ChatGPT. The tips include enhancing data clean up, writing code efficiently, improving data visualization, and prompt engineering. They also advise that we should only depend a little on ChatGPT.[74] Tiwari et al. used ChatGPT/GPT-4 or an MLLM to understand pathway enrichment and annotation gaps with comparative analysis using the manual curation process (conventional process). They determined some promising capabilities of this MLLM.[75] Levine et al. developed Cell2sentence, a GPT-2 model based on an LLM. It can be used to teach biological science, especially single-cell transcriptomics.[76] However, the application of LLMs to MLLMs is increasing day by day in molecular biology and computational biology.

Application of LLMs in medical sciences

LLM models have been used from time to time and have been applied in different fields of medical science (Figure 4). LLMs can enhance diagnosis and support clinical judgment in medicine. However, to
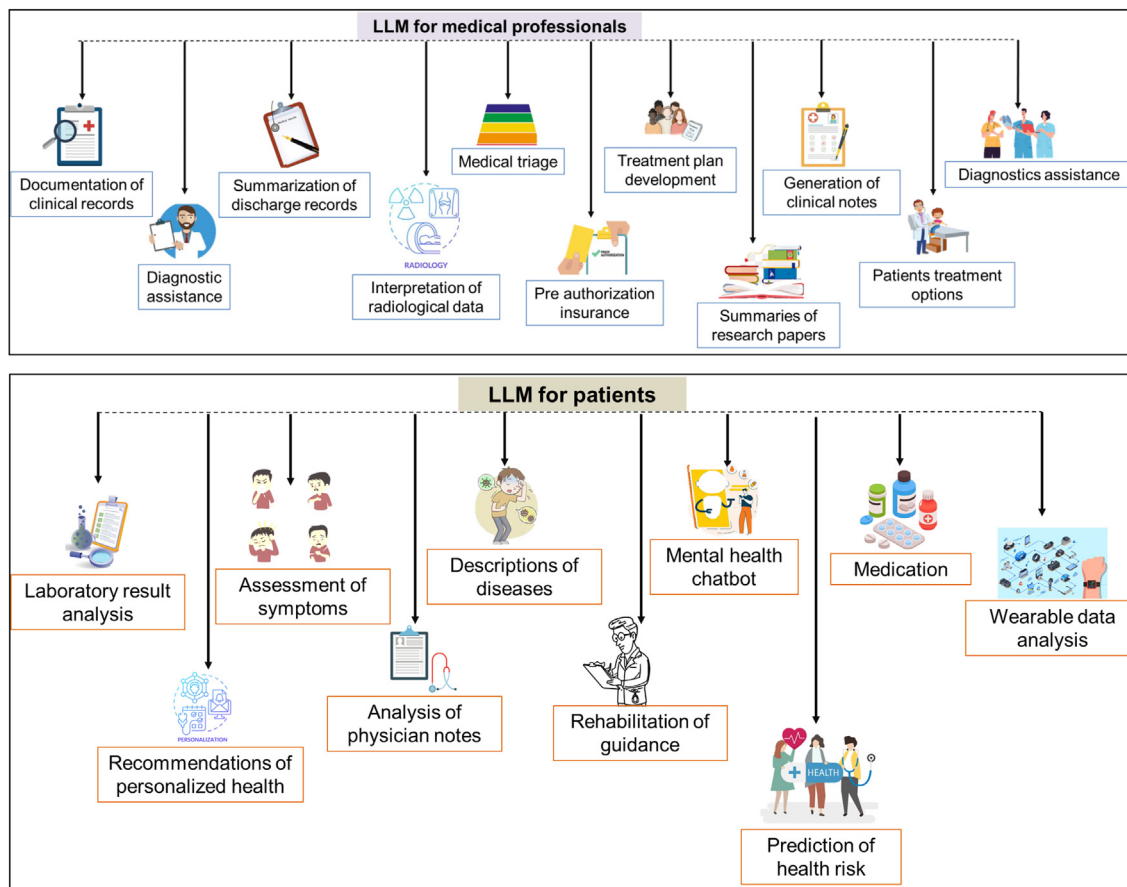
Figure 4. Some significant clinical applications of a LLM in the medical field.

make them function well in the medical field, particular difficulties must be overcome.[77] LLMs may completely transform the healthcare industry by improving diagnosis accuracy, predicting the course of diseases, and supporting physicians in their decision-making.[78,79]

LLMs in medical education

LLMs might be improved by concentrating on specialized medical literature to stay relevant and up to date. They can also be customized for different languages and scenarios, improving global access to medical knowledge and information.[77] Recently, there have been numerous instances where the application of LLM technology, notably ChatGPT, has been documented.[6] After passing the US Medical Licensing Exams, ChatGPT became well-known in the medical community. GPT-4 performs far better than GPT-3.5, its predecessor.[80,81] These medical LMs, however, were exceptionally trained on texts related to medicine or biology. They come in useful for jobs such as question-answering, translating, and summarizing. Examining if smaller models trained on pertinent data can perform as well at a reduced cost is necessary, given the high cost of training and utilizing these models. For example, at the cost of $600, the Center for Research on Foundation Models at Stanford University made a model named Alpaca that matched

the performance of OpenAI's text-davinci-003 with just 4% of its parameters.[82] By strengthening critical medical competencies such as factual knowledge and interpersonal communication, LLMs can elevate the standard of care for patients. For example, ChatGPT has demonstrated success in medical licensure exams demonstrating its substantial medical knowledge and ability to participate in medical reasoning.[80,81,83] LLMs' medical reasoning and concept understanding can be improved even further by providing focused instruction that includes questions akin to those on a medical test and expertly chosen sample answers. GPT-4 presently exhibits the most significant medical domain knowledge among LLMs. Nonetheless, LLMs face a fundamental limitation: they frequently reproduce pre-existing medical biases[84] and sustain inequalities associated with socioeconomic status, gender, ethnicity, and other characteristics.[80,85]

The use of LLMs in healthcare is progressing rapidly, driven by the widespread availability of LLMs, including their availability to students and some research-based initiatives. This involvement can be validated by the involvement of ChatGPT in the Epic Systems Corporation's software, as reported in a recent article.[86] The potential applications are diverse, ranging from streamlining administrative tasks

such as assisting on the instructions related to patient discharge, insurance filings, as well as obtaining some prior authorizations for medical services.[87] Moreover, there is a prospect of enhancing the standard-of-care by finding the older medical history from intricate records of patients along with a detailed checking on some of the standardized operating procedures. Among the emerging applications, two stand out: the capability of LLMs to analyze vast amounts of data in an unstructured form in electronic health records and their potential to aid in clinical documentation.[87,88] Incorporating these models into the educational framework can stimulate deep critical thinking, encourage creative work, and provide innovative learning experiences. Furthermore, gaining a profound understanding of these models prepares the students for working in the healthcare industry, which is also closely related to AI. Evaluating the application of ChatGPT in medical science is a crucial stride in harnessing the technological potential to guide forthcoming changes in the new era of medical science. More notably, the next wave of healthcare professionals needs to not only be familiar with these modern technologies but also possess the skills to responsibly and effectively employ them in the delivery of patient care.[89] In addition, the emergence of ChatGPT has generated new insights into AI-powered chatbots and their possible uses, attracting considerable attention worldwide. In recent months, there has been growing interest among scientists and medical professionals in implementing the applications of LLMs in medicine.[90]

LLM-based medical tool or device for medical education and research

LLMs are trained on medical data based on various codes and text. After examining this training set of data for more than 80 medical LMs, Wornow et al. distinguished two significant groups.[91] Firstly, textual resources such as progress notes or PubMed abstracts train specific models. They learn by making predictions about the words that will come next in these papers, just like generic LMs such as GPT-3 do. The effectiveness of utilizing domain adaptation, transfer learning, and alternative methodologies in the medical field is demonstrated by multiple examples of LLMs that have been specifically fine-tuned for medical purposes. BioBERT, a biological LM based on the BERT architecture, was refined by leveraging large biomedical datasets such as PMC full-text articles and the abstracts available in PubMed. As a result, there were significant improvements in several biological NLP tasks, such as problem-solving, relation extraction, question-answering, and named entity recognition.[92] ClinicalBERT, a distinct model, was subjected to fine-tuning using the MIMIC-III dataset, which comprises the electronic health records from patients in critical care units. Fine-tuning exhibited enhanced efficacy in clinical NLP assignments, including diagnosis categorization, patient mortality rate prediction, and de-identification.[93] BlueBERT, a model constructed according to the basis of the BERT architecture, has already been pre-trained on an extensive collection of biomedical texts and has demonstrated exceptional performance in multiple biomedical NLP tasks. These tasks include named relation extraction, biomedical problem-solving, and entity recognition.[94] The cases above highlight the effectiveness of utilizing domain-specific fine-

tuning, transfer learning, domain adaptation, and alternative methods to harness the capabilities of LLMs in various fields of medical science.[77] Recently, a specialized version named Med-PaLM 2 (Google) that was trained on medical data achieved state-of-the-art results similar to the level of proficiency exhibited by human doctors.[95] Recently, specialized LLMs in different fields of medical science have been developed and applied periodically, and some of them are PMC-LLaMA, ClinicalCamel, MedAlpaca, BioGPT, BioMedLM, Med-PaLM2, and ChatDoctor (Table 2). These specialized LLMs have revolutionized the field of medical science.

LLMs in different clinical fields

By examining extensive medical data, the LLMs can quickly develop specialized knowledge in various medical sectors, including radiology, pathology, and oncology.[103–105] Notably, the release of OpenAI's ChatGPT quickly sparked a massive revolution in other clinical fields, such as ophthalmology, nephrology, cardiology, and orthopedics.[11] Some LLMs are also trained using patient record sequences of medical codes. These models pick up new information by anticipating the codes for the next day or comprehending the time intervals between particular codes. They consider the sequence and the chronology of medical occurrences documented in a patient's file. For instance, if trained on some particular codes, these models can predict the chance of a stroke, heart attack, or renal failure. Rather than producing text, these models yield a fixed-length, high-dimensional vector that machines can read as an "embedding" of the patient's medical record. With as little as 100 training data examples, these embeddings can be used to build models predicting 30-day readmissions, prolonged hospital stays, and in-patient death.[106] Domain-specific LLMs tailored to specific domains could offer valuable new features in various clinical domains. For instance, foresight, an LLM that was built on the GPT architecture and trained on unstructured data from over 811,336 electronic health records, showed promise in accurately predicting and forecasting outcomes during validation trials.[107]

However, education and specific training are essential for effectively integrating LLMs into medical practice. Given the growing importance of LLMs in healthcare, medical personnel must fully understand their capabilities and limitations. This knowledge will allow them to utilize these technologies in clinical settings effectively. To fully equip future medical practitioners, medical curricula must incorporate the fundamental principles of utilizing LLMs. It will ensure that students gain the necessary knowledge and abilities to navigate and exploit these technological developments.[77]

Application of LLMs in other areas
Financial modeling and sentiment analysis

LLMs are increasingly used in financial modeling and sentiment analysis, offering advanced NLP capabilities to analyze and interpret financial data and model and measure market sentiment. It can transcribe and analyze earnings calls, extracting critical information about a company's performance, outlook, and management discussions to know the market trends.[108] In addition, LLMs can analyze text data related to companies, industries, and economic conditions to assess

Table 2. Application of specialized LLMs in different fields of medical science

Sl. no.	LLM	Year of release	Remarks	Reference
1.	PMC-LLaMA	2023	PMC-LLaMA, an open-source language model that is developed by refinement of an open-source language model on a total of 4.8 million biomedical academic papers for added injecting medical knowledge, improving its capability in the medical domain.	Wu et al.[96]
2.	ClinicalCamel	2023	It is an open LLM obviously tailored for clinical research. Fine-tuned from LLaMA-2 using QLoRA, Clinical Camel achieves state-of-the-art performance across medical benchmarks among openly available medical LLMs. Leveraging efficient single-GPU training.	Toma et al.[97]
3.	MedAlpaca	2023	MedAlpaca was developed by instruction fine-tuning of the LLaMA 13B and 7B models on Medical Meadow data. It is also the assembly of reformatted instruction-response pairs with datasets for medical NLP tasks and data derived from various internet sources.	Han et al.[98]
4.	BioGPT	2023	BioGPT is a domain-specific GPT language model for biomedical text generation and mining. BioGPT follows the transformer language model backbone, and is pre-trained on 15M PubMed abstracts from scratch.	Luo et al.[99]
5.	BioMedLM	2022	BioMedLM is based on a HuggingFace GPT model (decoder-only transformer) with 2.7B parameters and a maximum context length of 1,024 tokens. It also uses a custom biomedical tokenizer trained on PubMed abstracts with a vocabulary size of 28,896.	Karkera et al.[100]
6.	Med-PaLM2	2022	Med-PaLM is a large language model (LLM) designed to provide high quality answers to medical questions. It is also available to Google Cloud customers, who are able to explore a range of applications, from basic tasks to complex workflows. It has been aligned to the medical domain and evaluated using medical exams, medical research, and consumer queries.	Luo et al.[101]
7.	ChatDoctor	2023	This is a specified language model with improved accuracy in medical advice refining the large language model meta-AI (LLaMA) by a large dataset of patient-doctor dialogs obtained from a widely used online medical consultation platform.	Li et al.[102]

and quantify various financial, operational, and market risks.[109] Subsequently, LLMs process customer reviews, feedback, and comments to measure sentiment about products, services, or brands. This information is valuable for companies to understand customer satisfaction, allowing companies to make data-driven decisions to improve strategies in both financial modeling and sentiment analysis.[110] LLMs influence their ability to understand and generate human-like language to process vast amounts of textual data, providing valuable insights for decision-making in the financial domain.[111]

Legal research and analysis

LLMs are increasingly employed in legal research and analysis, transforming how legal professionals access, process, and understand legal information. They assists legal researchers in analyzing and summarizing case law, providing concise overviews of legal precedents and decisions; more specifically, they automatically summarize lengthy legal documents, including contracts, pleadings, and briefs, facilitating quick review by legal professionals.[112] LLMs contribute to trademark searches and analysis by processing and summarizing relevant information from trademark databases and legal texts. Applying LLMs in legal research and analysis enhances efficiency, accuracy, and the accessibility of legal information, transforming how legal professionals approach various tasks within the legal domain.[113] LLMs can review and analyze legal contracts, helping legal professionals identify vital terms, risks, and obligations.

Customer service by chatbots and email response

LLMs power chatbots for customer support, answering queries, resolving issues, and conversationally providing information. LLMs enhance the capability of chatbots to handle multiturn conversations, maintaining context and providing coherent responses across different user inputs. They engage in natural language conversations, accurately interpret user queries, and provide more relevant responses, providing users with a more human-like and intuitive interaction experience.[113,114] They also can help analyze user sentiment during interactions, allowing chatbots to respond appropriately and adapt their tone based on the user's emotional context.[115] LLMs can help generate responses to customer emails, improving efficiency in handling customer inquiries. This helps to categorize and prioritize incoming emails, directing them to the appropriate department or team for more efficient handling in both chatbot-driven customer service and email communication. LLMs are crucial in automating processes, improving response accuracy, and enhancing the overall customer experience by providing more natural and intelligent interactions.[116]

Education field

LLMs have found diverse applications in the education sector, transforming various aspects of teaching, learning, and administrative processes, supporting the automated grading of assignments and exams, and providing quick and consistent feedback to students.

Furthermore, they can assist in language learning by providing grammar explanations, vocabulary explanations, and conversational practice.[117] Different conditions provide reliable assistance to students with homework, offering explanations and guidance on various subjects. LLMs help generate research proposals and offer guidance on structuring and framing research questions. Presently, the LLMs contribute to language translation, breaking down language barriers and making educational content accessible to a global audience. Finally, they contribute to student performance data analysis, helping educators make data-driven decisions.[118] Therefore, the application of LLMs in education is vast and continually evolving, potentially enhancing learning experiences, streamlining administrative processes, and providing personalized support to students and educators.[119]

Marketing field

LLMs play a crucial role in marketing across various aspects, applying NLP capabilities to enhance communication, analyze data, and optimize strategies. It is capable in creating compelling copy for digital advertising campaigns, ensuring messages resonate with the target audience, blog posts, articles, and other content for marketing purposes, and maintaining consistency and quality.[120,121] Subsequently, the LLMs can contribute in analyzing competitor strategies, monitoring industry trends, and identifying areas for differentiation. Likewise, creating personalized marketing messages based on user data improves customer engagement and conversion rates.[122] Besides, the LLMs power chatbots for customer support, providing instant responses to queries and guiding users through the customer journey.[123] In short, integrating LLMs in marketing enhances efficiency, personalization, and data-driven decision-making, making them valuable tools in the dynamic and competitive marketing landscape.

Human resources

LLMs are being increasingly applied in various human resources (HR) aspects, helping to streamline processes, improve communication, and enhance decision-making and, similarly, to help optimize job descriptions to attract a diverse pool of candidates and ensure clarity in expectations.[124] They support automating the interview scheduling, saving time, and improving the candidate experience. Specifically, helping to answer common queries from new employees during the on-boarding process, providing information about company policies, benefits, and procedures. They finely contribute to analytics by analyzing employee data, generating reports, and providing insights into workforce trends. They also classify the hiring data to identify areas for improvement in diversity and recommend strategies to enhance diversity in the workforce.[125] Incorporating LLMs in HR enhances efficiency, personalization, and data-driven decision-making, making them valuable tools for HR professionals in managing various aspects of the employee life cycle.[126]

E-commerce

LLMs play a significant role in the e-commerce sector, contributing to various aspects of online retail. It can generate compelling product de-

scriptions, improving the quality and consistency of e-commerce content.[127] LLMs analyze customer reviews to provide insights into product feedback content based on user preferences and behavior, enhancing the shopping experience and sentiment.[54] Subsequently, it enables personalized shopping experiences by tailoring website content, offers, and promotions to individual user profiles, predictive analytics models, forecasting future sales trends, and customer behavior based on historical data. The integration of LLMs in e-commerce contributes to enhanced customer experiences, improved content creation, and management efficiency, and more data-driven decision-making for businesses operating in the online retail space.[128]

Research and academia

LLMs have significantly impacted research and academia, transforming various aspects of the scholarly landscape. LLMs have performed a satisfactory role in automated literature reviews by summarizing and extracting relevant information from a vast corpus of academic papers.[129] LLMs contribute to the automatic generation of concise and accurate abstracts for research papers, aiding in understanding complex topics, and writing sections of research papers, providing language suggestions, and aiding in the overall structure of academic writing.[130] LLMs are employed in plagiarism detection tools to identify and highlight potential instances of plagiarism in academic writing. Currently, LLMs are used for automated grading of assignments and exams, providing quick and consistent feedback to students. Besides these, LLMs contribute to building and maintaining institutional knowledge bases, making information easily accessible to researchers and academics.[131] Applying LLMs in research and academia accelerates processes, improves writing and communication, and enhances overall efficiency in various scholarly activities.

Multimodal applications of LLMs or MLLMs

MLLMs can be trained with video, audio, image, and text (Figure 5). Researchers and entrepreneurs have noted the diverse applications of MLLMs and their immense possibilities. MLLMs are increasingly combined with computer vision models for tasks involving text and images, such as images, videos, or audio, to provide a richer and more comprehensive user experience and visual question-answering systems.[132] The versatility of MLLMs allows them to be applied to many tasks across diverse industries, demonstrating their potential to improve efficiency and provide intelligent solutions. More specifically, LLMs assist in summarizing content within images, providing brief textual descriptions for visually impaired users or those who prefer text-based information.[133] They contribute to understanding and interpreting visual cues within chatbot conversations, providing more context-aware responses. For assistance, in integrating textual and visual content for presentations, ensuring coherence and relevance between the spoken or written text and visual aids, MLLMs enhance search engines' capabilities for more accurate results.[134] MLLMs contribute to creating immersive experiences in virtual or augmented reality by providing natural language understanding alongside visual and auditory elements. Such an essential combination of LLMs in multimodal applications enhances the ability to process, understand, and generate content that spans multiple modalities, providing users
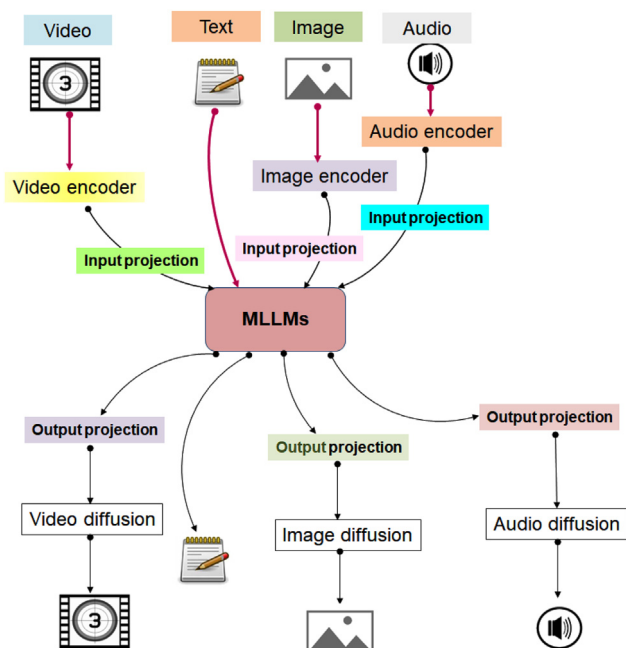
Figure 5. The MLLM's working principle.

with more immersive and contextually rich experiences.[135] Recently, researchers have developed several MLLM models. One such interesting MLLM is Next-GPT, which can be used as an any-to-any MLLM model.[136] Another MLLM model is Bliva, which is a very simplistic MLLM. It can handle text-rich visual questions.[137] Han et al. developed an MLLM-based Chartllama for charts.[138] Several other MLLMs have been developed for different applications, such as mplug-owl,[139] Palm-e,[140] Mm-llms,[141] M3exam,[142] Kosmos-g,[143] etc.

MLLMs in biological macromolecules to biological sciences

MLLMs have been applied in biological science to explore different possibilities. Significant models have been developed to explore the next-generation possibilities. Researchers have developed a GIT-Mol, an MLLM model to explore the possibilities of complex molecular science that integrates text information, images, and graphs. It can predict chemical reactions and compound name recognition.[144] Recently, Lin et al. developed a multimodal deep learning model for multiclass glaucoma surgery and its outcome prediction. The multimodal neural network improves clinical decision-making for postoperative management.[145] Xu et al. developed Protst, a next-generation MLLM model, to explore biomedical texts and protein sequences.[146]

Similarly, an MLLM model called MuSe-GNN han developed, which performs gene presentation from multimodal biological graph data.[147] Huang et al. illustrated the application and prospects of an MLLM in dentistry. The researchers explain how MLLMs can shape the future landscape of dentistry.[21] However, there are many possibilities for the MLLM model to solve complex problems in biological science.

Limitations of LLMs

LLMs, such as GPT models (GPT-3.5 and GPT-4), have achieved remarkable success in various NLP tasks. However, they also come with certain significant limitations (Table 3). LLMs can generate coherent and contextually relevant text, but they often lack deep understanding of the world and common sense reasoning. They can generate nonsensical or incorrect responses in specific contexts.[148] LLMs can inadvertently preserve biases present in the training data, which can result in biased or unfair outputs, especially when dealing with sensitive topics such as gender, race, or religion. Mitigating biases in LLMs remains a significant challenge.

Moreover, while LLMs excel in understanding and generating text based on context, they may struggle with long-term dependencies or maintaining coherence over extended passages.[6] This can lead to inconsistencies or inaccuracies in generated text, especially in complex or significant scenarios. The LLMs typically require vast amounts of data for pre-training, which can be expensive and resource-intensive. Furthermore, they may stumble with generalizing to out-of-domain or low-resource domains where training data are limited. While fine-tuning LLMs on specific tasks can improve performance, it often requires careful selection of hyperparameters, task-specific data, and fine-tuning strategies.[149]

In addition, fine-tuning may only sometimes lead to optimal performance, especially for tasks with unique requirements or constraints. In the case of safety and ethical concerns, the LLMs have the potential to generate harmful or malicious content, including misinformation, hate speech, or inappropriate material.[150] Ensuring LLMs' safe and ethical use poses significant challenges for researchers and practitioners. Training and deploying LLMs can be computationally expensive and resource-intensive, requiring powerful hardware and substantial infrastructure. It can limit accessibility to LLMs for researchers and organizations with limited resources.[151] Addressing these limitations requires ongoing research and development efforts in bias mitigation, robustness testing, model interpretability, and ethical AI frameworks. In addition, interdisciplinary collaboration involving experts from diverse fields, such as linguistics, psychology, and ethics, is essential to foster the responsible development and deployment of LLMs.

CHALLENGES AND FUTURE PROSPECTS

Tokenization in LLMs, although an essential aspect, has its share of challenges. One notable obstacle is the management of out-of-vocabulary tokens. LLMs face difficulties when they encounter words or phrases not present in their training data, affecting their ability to comprehend and generate relevant output. To mitigate this challenge, continuous refinement and adaptation of the model is required to keep up with the ever-evolving nature of language. Furthermore, tokenization may encounter difficulties in capturing nuanced semantic meanings. Ambiguities and polysemy in language can lead to multiple interpretations of a single token, posing challenges for models to discern the intended meanings accurately. Addressing such ambiguities necessitates advancements in contextual understanding and

Table 3. Different limitations of LLMs with their mitigating strategies

Sl. no.	Types of LLM limitations	Description/remarks	Mitigating strategies
1.	Ethical concerns	The responses may be risky, biased, or offensive nature. Having threat of privacy and other security breaches. No established accountability was present for the consequences of model outputs. No consensus on what roles AI should and should not play in medicine.	Refinement to decrease the incidence of undesirable outputs. Formation of governance systems and managing the experts. Fixing of a reporting system for end users to flag the dangerous responses. Consensus-building creativities connecting the patients and medical practitioners.
2.	Coherence	Model outputs are centered on learned associations between the words in spite of considerate input information (queries) used in outputs. Falsified facts are offered as if they were true.	The regenerating model architecture and training tactics used to develop true semantic knowledge. Fine-tuning to exclude presentation of wrong information.
3.	Accuracy	GPT-3 is restricted to data of 570 GB. Models are limited to learning probabilistic associations between words, these are not trained to understand. Training data are obtained from unverified and invalidated resources, websites, books, etc.	Training data validation for insecurity indicators. Fine-tuning used to enhance medical accuracy. Self-improvement over intelligent prompts (like, chain-of-thought).
4.	Recency	The training datasets of GPT do not comprise content created after September, 2021. All concerning datasets necessarily "cut off" at a random time or date.	Assembly of the training data from more current sources. The real-time internet access (e.g., Bing AI, BlenderBot 3, Sparrow).
5.	Transparency and interpretability	It is not clear that how models generate answers from input queries, architectural data, and algorithms (Black Box problem). It is uncertain which parts of the training dataset are gearing, in created responses.	Prerequisite for outputs to cite which portions of the dataset added to the model's answers section. It has explicable AI research and development.

disambiguation techniques. In addition, tokenization can be resource-intensive, particularly in models with extensive vocabularies. Processing vast amounts of data for tokenization may result in increased computational demands, thereby limiting the scalability of models for real-time applications or resource-constrained environments.

Despite the challenges mentioned above, the future of LLM tokenization holds promising prospects. The current research and development activities focus on mitigating the existing constraints and unlock new capabilities. One avenue of exploration is the enhancement of handling out-of-vocabulary tokens. Future LLMs may incorporate more effective mechanisms to adapt to novel language elements, thus reducing the impact of encountering unfamiliar tokens and improving overall language coverage. Another exciting frontier involves advancements in contextual understanding. Research into more sophisticated attention mechanisms and context aggregation techniques can lead to models better equipped to capture subtle nuances in language, enabling more accurate and context-aware tokenization. Improving the interpretability and explainability of tokenization processes is an ongoing focal point. As LLMs become increasingly integrated into various applications, understanding the rationale behind token-level decisions becomes crucial for building trust and ensuring ethical use. Efforts to optimize computational efficiency are also underway. Future LLMs are expected to leverage innovative architectures or strategies to streamline tokenization processes, making them more accessible for various applications and devices. Therefore, while challenges certainly exist, the continuous evolution of LLM tokenization holds immense potential. LLMs can achieve greater accuracy, adaptability, and applicability in diverse linguistic contexts by addressing current limitations and embracing future possibilities. As the field of NLP advances, tokenization within LLMs is poised to play a pivotal role in shaping the next generation of intelligent language understanding systems.

CONCLUSION
LLMs have revolutionized biological sciences and medicine, resulting in transformative and fundamental applications and faster progress in medicine and different areas of biological sciences. LLMs are helping to generate new hypotheses in these areas. The LLM model also helps clinical decision-making and understanding of possible future outcomes. MLLMs make it faster and provide a broader range of opportunities. There are ample opportunities to research those areas using LLMs or MLLMs. Researchers have explained that the range of possibilities is vast.

However, possible risks generate substantial concerns among researchers, experts, and users. Successful validation of the LLM and MLLM technologies will benefit human society at large. At the same time, ethics, safety, and potential human replacement are the most significant concerns. However, we are hopeful that future researchers will use the technologies in a way that will do justice to society by properly utilizing them.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Validation, M.B., S.P., S.C., and S.-S.L.; writing – original draft, M.B., S.P., S.C., S.-S.L., and C.C.; writing – review & editing, C.C.; figure and table development, M.B.; formal analysis, S.C. and S.-S.L.; conceptualization, C.C.; investigation, C.C.; supervision, C.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays - should professors worry? Nature. Online ahead of print. https://doi.org/10.1038/d41586-022-04397-7.

2. Shanahan, M., McDonell, K., and Reynolds, L. (2023). Role play with large language models. Nature 623, 493–498.

3. Chakraborty, C., Bhattacharya, M., Pal, S., and Lee, S.-S. (2023). From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology 7, 100164.

4. Blank, I.A. (2023). What are large language models supposed to model? Trends Cognit. Sci. 27, 987–989.

5. Meskó, B., and Topol, E.J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ digital medicine 6, 120.

6. Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F., and Ting, D.S.W. (2023). Large language models in medicine. Nat. Med. 29, 1930–1940.

7. Meyer, J.G., Urbanowicz, R.J., Martin, P.C.N., O'Connor, K., Li, R., Peng, P.C., Bright, T.J., Tatonetti, N., Won, K.J., Gonzalez-Hernandez, G., and Moore, J.H. (2023). ChatGPT and large language models in academia: opportunities and challenges. BioData Min. 16, 20.

8. Wang, X., Anwer, N., Dai, Y., and Liu, A. (2023). ChatGPT for design, manufacturing, and education. Procedia CIRP 119, 7–14.

9. Pal, S., Bhattacharya, M., Islam, M.A., and Chakraborty, C. (2023). ChatGPT or LLM in next-generation drug discovery and development: pharmaceutical and biotechnology companies can make use of the artificial intelligence-based device for a faster way of drug discovery and development. Int. J. Surg. 109, 4382–4384.

10. Lautrup, A.D., Hyrup, T., Schneider-Kamp, A., Dahl, M., Lindholt, J.S., and Schneider-Kamp, P. (2023). Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. Open heart 10, e002455.

11. Chatterjee, S., Bhattacharya, M., Pal, S., Lee, S.S., and Chakraborty, C. (2023). ChatGPT and large language models in orthopedics: from education and surgery to research. J. Exp. Orthop. 10, 128.

12. Bajaj, S., Gandhi, D., and Nayar, D. (2024). Potential Applications and Impact of ChatGPT in Radiology. Acad. Radiol. 31, 1256–1261.

13. Cheng, K., Li, Z., He, Y., Guo, Q., Lu, Y., Gu, S., and Wu, H. (2023). Potential Use of Artificial Intelligence in Infectious Disease: Take ChatGPT as an Example. Ann. Biomed. Eng. 51, 1130–1135.

14. Chakraborty, C., Pal, S., Bhattacharya, M., and Islam, M.A. (2024). ChatGPT or LLMs can provide treatment suggestions for critical patients with antibiotic-resistant infections: A next-generation revolution for medical science? Int. J. Surg. 110, 1829–1831.

15. Cheng, K., Li, Z., Guo, Q., Sun, Z., Wu, H., and Li, C. (2023). Emergency surgery in the era of artificial intelligence: ChatGPT could be the doctor's right-hand man. Int. J. Surg. 109, 1816–1818.

16. Chakraborty, C., Bhattacharya, M., and Lee, S.S. (2023). Artificial intelligence enabled ChatGPT and large language models in drug target discovery, drug discovery, and development. Mol. Ther. Nucleic Acids 33, 866–868.

17. Choi, J.H., Hickman, K.E., Monahan, A.B., and Schwarcz, D. (2021). ChatGPT goes to law school. J. Leg. Educ. 71, 387.

18. Adeshola, I., and Adepoju, A.P. (2023). The opportunities and challenges of ChatGPT in education. Interact. Learn. Environ. 2023, 1–14.

19. Pal, S., Bhattacharya, M., Lee, S.S., and Chakraborty, C. (2024). A Domain-Specific Next-Generation Large Language Model (LLM) or ChatGPT is Required for Biomedical Engineering and Research. Ann. Biomed. Eng. 52, 451–454.

20. Dowling, M.M., and Lucey, B.M. (2023). ChatGPT for (finance) research: The Bananarama conjecture. SSRN Journal 53, 103662.

21. Huang, H., Zheng, O., Wang, D., Yin, J., Wang, Z., Ding, S., Yin, H., Xu, C., Yang, R., Zheng, Q., and Shi, B. (2023). ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. Int. J. Oral Sci. 15, 29.

22. Brynjolfsson, E., Li, D., and Raymond, L.R. (2023). Generative AI at Work. NBER Working Paper 31161, 1–67. National Bureau of Economic Research. https://doi.org/10.3386/w31161.

23. Borders, T.L., and Volkova, S. (2021). An Introduction to Word Embeddings and Language Models.

24. Vu, T.-T., Phung, D., and Haffari, G. (2020). Effective unsupervised domain adaptation with adversarially trained language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.01739.

25. Toraman, C., Yilmaz, E.H., Şahinuç, F., and Ozcelik, O. (2023). Impact of tokenization on language models: An analysis for turkish. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22, 1–21.

26. Huang, Z., Meng, C., and Ko, T. (2023). Repcodec: A speech representation codec for speech tokenization. Preprint at arXiv. https://doi.org/10.48550/arXiv.2309.00169.

27. Islam, T., Hossain, M., and Arefin, M.F. (2021). Comparative analysis of different text summarization techniques using enhanced tokenization. In 2021 3rd International Conference on Sustainable Technologies for Industry 40 (STI) (IEEE), pp. 1–6.

28. Hiraoka, T., Shindo, H., and Matsumoto, Y. (2019). Stochastic tokenization with a language model for neural text classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1620–1629.

29. Jin, H., Han, X., Yang, J., Jiang, Z., Chang, C.-Y., and Hu, X. (2023). Growlength: Accelerating llms pretraining by progressively growing training length. Preprint at arXiv. https://doi.org/10.48550/arXiv.2310.00576.

30. Bałazy, K., Banaei, M., Lebret, R., Tabor, J., and Aberer, K. (2021). Direction is what you need: Improving word embedding compression in large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2106.08181.

31. Fu, C.-L., Chen, Z.-C., Lee, Y.-R., and Lee, H-y (2022). Adapterbias: Parameter-efficient token-dependent representation shift for adapters in nlp tasks. Preprint at arXiv. https://doi.org/10.48550/arXiv.2205.00305.

32. Xu, C., Zhai, B., Wu, B., Li, T., Zhan, W., Vajda, P., Keutzer, K., and Tomizuka, M. (2021). You only group once: Efficient point-cloud processing with token representation and relation inference module. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), pp. 4589–4596.

33. Zhang, Z., Shu, C., Chen, Y., Xiao, J., Zhang, Q., and Zheng, L. (2022). Icaf: Iterative contrastive alignment framework for multimodal abstractive summarization. In 2022 International Joint Conference on Neural Networks (IJCNN) (IEEE), pp. 1–8.

34. Yu, L., Cheng, Y., Wang, Z., Kumar, V., Macherey, W., Huang, Y., Ross, D., Essa, I., Bisk, Y., Yang, M.H., et al. (2024). Spae: Semantic pyramid autoencoder for multimodal generation with frozen llms. Adv. Neural Inf. Process. Syst. 36, 1–13.

35. Chronis, G., and Erk, K. (2020). When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships. In Proceedings of the 24th Conference on Computational Natural Language Learning, pp. 227–244.

36. Yu, R., Li, Y., Lu, W., and Cao, L. (2022). Tri-Attention: Explicit Context-Aware Attention Mechanism for Natural Language Processing. Preprint at arXiv. https://doi.org/10.48550/arXiv:221102899.

37. Wang, X., and Wang, R. (2022). Text sentiment classification based on Vit-BiGRU-attention model. In International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022), 12456 (SPIE), pp. 338–343.

38. Castelvecchi, D. (2016). Can we open the black box of AI? Nature 538, 20–23.

39. Schwartz, I.S., Link, K.E., Daneshjou, R., and Cortés-Penfield, N. (2024). Black Box Warning: Large Language Models and the Future of Infectious Diseases Consultation. Clin. Infect. Dis. 78, 860–866.

40. Chakraborty, C., Bhattacharya, M., Islam, M.A., and Agoramoorthy, G. (2023). ChatGPT indicates the path and initiates the research to open up the black box of artificial intelligence. Int. J. Surg. *109*, 4367–4368.

41. Demszky, D., Yang, D., Yeager, D.S., Bryan, C.J., Clapper, M., Chandhok, S., Eichstaedt, J.C., Hecht, C., Jamieson, J., Johnson, M., et al. (2023). Using large language models in psychology. Nat. Rev. Psychol. *2*, 688–701.

42. Editorials (2023). ChatGPT is a black box: how AI research can break it open. Nature *619*, 671–672.

43. Ullah, E., Parwani, A., Baig, M.M., and Singh, R. (2024). Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology - a recent scoping review. Diagn. Pathol. *19*, 43.

44. Rice, S., Crouse, S.R., Winter, S.R., and Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. Technol. Soc. *76*, 102426.

45. Lehman, J., Gordon, J., Jain, S., Ndousse, K., Yeh, C., and Stanley, K.O. (2023). Evolution through large models. In Handbook of Evolutionary Machine Learning (Springer), pp. 331–366.

46. Rajaraman, V. (2023). From ELIZA to ChatGPT: History of Human-Computer Conversation. Reson. *28*, 889–905.

47. Room, C. (2023). N-Gram Model. Algorithms *17*, 1–30.

48. Valdenegro, D. A LLM (2023). Digest for Social Scientist. Preprint at arXiv *3*, 1–11.

49. Egan, S., Fedorko, W., Lister, A., Pearkes, J., and Gay, C. (2017). Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC. Preprint at arXiv. https://doi.org/10.48550/arXiv.171109059.

50. Dyde, T. (2023). Documentation on the emergence, current iterations, and possible future of Artificial Intelligence with a focus on Large Language Models. In Bachelor's Thesis (Turku University of Applied Sciences), pp. 1–53.

51. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Adv. Neural Inf. Process. Syst. *30*.

52. Kalyan, K.S. (2024). A survey of GPT-3 family large language models including ChatGPT and GPT-4. Natural Language Processing Journal 6, 100048.

53. Kublik, S., and Saboo, S. (2022). GPT-3 (O'Reilly Media, Incorporated), pp. 1–150.

54. Roumeliotis, K.I., Tselikas, N.D., and Nasiopoulos, D.K. (2024). LLMs in e-commerce: A comparative analysis of GPT and LLaMA models in product review evaluation. Natural Language Processing Journal 6, 100056.

55. Moore, S., Tong, R., Singh, A., Liu, Z., Hu, X., Lu, Y., Liang, J., Cao, C., Khosravi, H., Denny, P., et al. (2023). Check for updates Empowering Education with LLMS-The Next-Gen Interface and Content Generation. In Artificial Intelligence in Education Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky: 24th International Conference, AIED 2023, Tokyo, Japan, July 3–7, 2023, Proceedings, *32* (Springer Nature).

56. Moore, S., Tong, R., Singh, A., Liu, Z., Hu, X., Lu, Y., Liang, J., Cao, C., Khosravi, H., Denny, P., et al. (2023). Empowering education with llms-the next-gen interface and content generation. In International Conference on Artificial Intelligence in Education (Springer), pp. 32–37.

57. Kleinig, O., Gao, C., Kovoor, J.G., Gupta, A.K., Bacchi, S., and Chan, W.O. (2024). How to use large language models in ophthalmology: from prompt engineering to protecting confidentiality. Eye *38*, 649–653.

58. Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., and Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. NPJ Digit. Med. *7*, 41.

59. Venerito, V., Lalwani, D., Del Vescovo, S., Iannone, F., and Gupta, L. (2024). Prompt engineering: The next big skill in rheumatology research. Int. J. Rheum. Dis. *27*, e15157.

60. Polak, M.P., and Morgan, D. (2024). Extracting accurate materials data from research papers with conversational language models and prompt engineering. Nat. Commun. *15*, 1569. https://doi.org/10.1038/s41467-024-45914-8.

61. Zhuo, L., Chi, Z., Xu, M., Huang, H., Zheng, H., He, C., Mao, X.L., and Zhang, W. (2024). ProtLLM: An Interleaved Protein-Language LLM with Protein-as-Word Pre-Training. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.07920.

62. Guo, H., Huo, M., Zhang, R., and Xie, P. (2023). Proteinchat: Towards achieving chatgpt-like functionalities on protein 3d structures. Preprint at TechRxiv. https://doi.org/10.36227/techrxiv.23120606.v1.

63. Wang, C., Fan, H., Quan, R., and Yang, Y. (2024). ProtChatGPT: Towards Understanding Proteins with Large Language Models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2402.09649.

64. Wang, Z., Zhang, Q., Ding, K., Qin, M., Zhuang, X., Li, X., and Chen, H. (2023). Instructprotein: Aligning human and protein language via knowledge instruction. Preprint at arXiv. https://doi.org/10.48550/arXiv.231003269.

65. Chatterjee, S., Bhattacharya, M., Lee, S.S., and Chakraborty, C. (2023). Can artificial intelligence-strengthened ChatGPT or other large language models transform nucleic acid research? Mol. Ther. Nucleic Acids *33*, 205–207.

66. Hou, W., and Ji, Z. (2023). GeneTuring tests GPT models in genomics. Preprint at bioRxiv. Cold Spring Harbor Laboratory Preprints. https://doi.org/10.1101/2023.03.11.532238.

67. Ji, Y., Zhou, Z., Liu, H., and Davuluri, R.V. (2021). DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Bioinformatics *37*, 2112–2120.

68. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P., and Kelley, D.R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nat. Methods *18*, 1196–1203.

69. Williams, D.O., and Fadda, E. (2023). Can ChatGPT pass Glycobiology? Glycobiology *33*, 606–614.

70. Buehler, M.J. (2024). MechGPT, a Language-Based Strategy for Mechanics and Materials Modeling That Connects Knowledge Across Scales, Disciplines, and Modalities. Appl. Mech. Rev. *76*, 021001.

71. David, A.N., Sewsynker-Sukai, Y., Meyer, E., and Kana, E.G. (2023). Harnessing Artificial Neural Networks and large language models for bioprocess optimization: Predicting sugar output from Kraft waste-based lignocellulosic pretreatments. Ind. Crop. Prod. *206*, 117686.

72. Vert, J.-P. (2023). How will generative AI disrupt data science in drug discovery? Nat. Biotechnol. *41*, 750–751.

73. Ross, T.D., and Gopinath, A. (2024). Chaining thoughts and LLMs to learn DNA structural biophysics. Preprint at arXiv. https://doi.org/10.48550/arXiv.240301332.

74. Lubiana, T., Lopes, R., Medeiros, P., Silva, J.C., Goncalves, A.N.A., Maracaja-Coutinho, V., and Nakaya, H.I. (2023). Ten quick tips for harnessing the power of ChatGPT in computational biology. PLoS Comput. Biol. *19*, e1011319.

75. Tiwari, K., Matthews, L., May, B., Shamovsky, V., Orlic-Milacic, M., Rothfels, K., Ragueneau, E., Gong, C., Stephan, R., Li, N., et al. (2023). ChatGPT usage in the Reactome curation process. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.08.566195.

76. Levine, D., Lévy, S., Rizvi, S.A., Pallikkavaliyaveetil, N., Chen, X., Zhang, D., Vrkic, I., and Zhong, A. (2023). Cell2sentence: Teaching large language models the language of biology. Preprint at bioRxiv, 557287. https://doi.org/10.1101/2023.09.11.557287.

77. Karabacak, M., and Margetis, K. (2023). Embracing Large Language Models for Medical Applications: Opportunities and Challenges. Cureus *15*, e39305.

78. Wang, S., Zhao, Z., Ouyang, X., Wang, Q., and Shen, D. (2023). Chatcad: Interactive computer-aided diagnosis on medical image using large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2302.07257.

79. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., and Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit. Med. *4*, 86.

80. Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., and Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit. Health *2*, e0000198.

81. Nori, H., King, N., McKinney, S.M., Carignan, D., and Horvitz, E. (2023). Capabilities of gpt-4 on medical challenge problems. Preprint at arXiv. https://doi.org/10.48550/arXiv.230313375.

82. Taori, R., GI, Zhang, T., Dubois, Y., and Li, X. (2023). Stanford Alpaca: code and documentation to train Stanford's Alpaca models and generate the data. https://github.com/tatsu-lab/stanford_alpaca.

83. Gilson, A., Safranek, C.W., Huang, T., Socrates, V., Chi, L., Taylor, R.A., and Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med. Educ. 9, e45312.

84. Agniel, D., Kohane, I.S., and Weber, G.M. (2018). Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ 361, k1479.

85. Shaikh, O., Zhang, H., Held, W., Bernstein, M., and Yang, D. (2022). On second thought, let's not think step by step! Bias and toxicity in zero-shot reasoning. Preprint at arXiv. https://doi.org/10.48550/arXiv.221208061.

86. Asmas, K. (2023). Epic to Integrate GPT-4 into its EHR through Expanded Microsoft Partnership. Medcity News. https://medcitynews.com/2023/04/epic-to-integrate-gpt-4-into-its-ehr-through-expanded-microsoft-partnership/.

87. Landi, H. (2023). Doximity rolls out beta version of ChatGPT tool for docs aiming to streamline administrative paperwork. Fierce Healthcare. https://www.fiercehealthcare.com/health-tech/doximity-rolls-out-beta-version-chatgpt-tool-docs-aiming-streamline-administrative.

88. Lee, P., Bubeck, S., and Petro, J. (2023). Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N. Engl. J. Med. 388, 1233–1239.

89. Safranek, C.W., Sidamon-Eristoff, A.E., Gilson, A., and Chartash, D. (2023). The Role of Large Language Models in Medical Education: Applications and Implications. JMIR Med. Educ. 9, e50945.

90. Chakraborty, C., Pal, S., Bhattacharya, M., Dash, S., and Lee, S.S. (2023). Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front. Artif. Intell. 6, 1237704.

91. Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., Pfeffer, M.A., Fries, J., and Shah, N.H. (2023). The shaky foundations of large language models and foundation models for electronic health records. NPJ Digit. Med. 6, 135.

92. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., and Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36, 1234–1240.

93. Huang, K., Altosaar, J., and Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. Preprint at arXiv. https://doi.org/10.48550/arXiv.1904.05342.

94. Peng, Y., Yan, S., and Lu, Z. (2019). Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. Preprint at arXiv. https://doi.org/10.48550/arXiv.190605474.

95. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S., Cole-Lewis, H., Neal, D., et al. (2023). Towards expert-level medical question answering with large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2305.09617.

96. Wu, C., Zhang, X., Zhang, Y., Wang, Y., and Xie, W. (2023). Pmc-llama: Further finetuning llama on medical papers. Preprint at arXiv. https://doi.org/10.48550/arXiv.230414454.

97. Toma, A., Lawler, P.R., Ba, J., Krishnan, R.G., Rubin, B.B., and Wang, B. (2023). Clinical camel: An open-source expert-level medical language model with dialogue-based knowledge encoding. Preprint at arXiv. https://doi.org/10.48550/arXiv.230512031.

98. Han, T., Adams, L.C., Papaioannou, J.-M., Grundmann, P., Oberhauser, T., Löser, A., Truhn, D., and Bressem, K.K. (2023). MedAlpaca–an open-source collection of medical conversational AI models and training data. Preprint at arXiv. https://doi.org/10.48550/arXiv.230408247.

99. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., and Liu, T.Y. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. Briefings Bioinf. 23, bbac409.

100. Karkera, N., Acharya, S., and Palaniappan, S.K. (2023). Leveraging pre-trained language models for mining microbiome-disease relationships. BMC Bioinf. 24, 290.

101. Luo, L., Ning, J., Zhao, Y., Wang, Z., Ding, Z., Chen, P., Fu, W., Han, Q., Xu, G., Qiu, Y., et al. (2024). Taiyi: a bilingual fine-tuned large language model for diverse biomedical tasks. J. Am. Med. Inf. Assoc, ocae037. https://doi.org/10.1093/jamia/ocae037.

102. Li, Y., Li, Z., Zhang, K., Dan, R., Jiang, S., and Zhang, Y. (2023). Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. Cureus 15, e40895. https://doi.org/10.7759/cureus.40895.

103. Yan, A., McAuley, J., Lu, X., Du, J., Chang, E.Y., Gentili, A., and Hsu, C.N. (2022). RadBERT: Adapting Transformer-based Language Models to Radiology. Radiol. Artif. Intell. 4, e210258.

104. Santos, T., Tariq, A., Das, S., Vayalpati, K., Smith, G.H., Trivedi, H., and Banerjee, I. (2022). PathologyBERT - Pre-trained Vs. A New Transformer Language Model for Pathology Domain. In AMIA Annual Symposium Proceedings AMIA Symposium, 2022 (American Medical Informatics Association), pp. 962–971.

105. Kather, J.N. (2023). Artificial intelligence in oncology: chances and pitfalls. J. Cancer Res. Clin. Oncol. 149, 7995–7996.

106. Steinberg, E., Jung, K., Fries, J.A., Corbin, C.K., Pfohl, S.R., and Shah, N.H. (2021). Language models are an effective representation learning technique for electronic health record data. J. Biomed. Inf. 113, 103637.

107. Kraljevic, Z., Bean, D., Shek, A., Bendayan, R., Hemingway, H., Yeung, J.A., Deng, A., Baston, A., Ross, J., Idowu, E., et al. (2022). Foresight–generative pretrained transformer (GPT) for modelling of patient timelines using Ehrs. Preprint at arXiv. https://doi.org/10.48550/arXiv.2212.08072.

108. Deng, X., Bashlovkina, V., Han, F., Baumgartner, S., and Bendersky, M. (2023). What do llms know about financial markets? a case study on reddit market sentiment analysis. In Companion Proceedings of the ACM Web Conference 2023, pp. 107–110.

109. de Zarzà, I., de Curtò, J., Roig, G., and Calafate, C.T. (2023). Optimized Financial Planning: Integrating Individual and Cooperative Budgeting Models with LLM Recommendations. AI 5, 91–114.

110. Xing, F. (2024). Designing Heterogeneous LLM Agents for Financial Sentiment Analysis. Preprint at arXiv. https://doi.org/10.48550/arXiv.240105799.

111. Li, Y., Wang, S., Ding, H., and Chen, H. (2023). Large Language Models in Finance: A Survey. In Proceedings of the Fourth ACM International Conference on AI in Finance, pp. 374–382.

112. Cui, J., Li, Z., Yan, Y., Chen, B., and Yuan, L. (2023). Chatlaw: Open-source legal large language model with integrated external knowledge bases. Preprint at arXiv. https://doi.org/10.48550/arXiv.230616092.

113. Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. (2020). Scaling laws for neural language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.200108361.

114. Nicolescu, L., and Tudorache, M.T. (2022). Human-computer interaction in customer service: the experience with AI chatbots—a systematic literature review. Electronics 11, 1579.

115. Stoilova, E. (2021). AI chatbots as a customer service and support tool. ROBONOMICS: The Journal of the Automated Economy 2, 21.

116. Soni, V. (2023). Large language models for enhancing customer lifecycle management. J. Empir. Soc. Sci. Stud. 7, 67–89.

117. Tayan, O., Hassan, A., Khankan, K., and Askool, S. (2024). Considerations for adapting higher education technology courses for AI large language models: A critical review of the impact of ChatGPT. Machine Learning with Applications 15, 100513.

118. Gan, W., Qi, Z., Wu, J., and Lin, J.C.-W. (2023). Large language models in education: Vision and opportunities. Preprint at arXiv. https://doi.org/10.48550/arXiv.231113160.

119. Hosseini, M., Gao, C.A., Liebovitz, D.M., Carvalho, A.M., Ahmad, F.S., Luo, Y., MacDonald, N., Holmes, K.L., and Kho, A. (2023). An exploratory survey about using ChatGPT in education, healthcare, and research. PLoS One 18, e0292216.

113. Brand, J., Israeli, A., and Ngwe, D. (2023). Using gpt for market research, pp. 1–35. Available at SSRN 4395751. https://doi.org/10.2139/ssrn.4395751.

121. Arsenijevic, U., and Jovic, M. (2019). Artificial intelligence marketing: chatbots. In 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI) (IEEE), pp. 19–193.

122. Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.230310130.

123. Kaczorowska-Spychalska, D. (2019). How chatbots influence marketing. Management 23, 251–270.

124. Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G.J., Beltran, J.R., Boselie, P., Lee Cooke, F., Decker, S., DeNisi, A., et al. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. Human Res. Mgmt. Journal 33, 606–659.

125. Agossah, A., Krupa, F., Perreira Da Silva, M., and Le Callet, P. (2023). LLM-based Interaction for Content Generation: A Case Study on the Perception of Employees in an IT department. In Proceedings of the 2023 ACM International Conference on Interactive Media Experiences, pp. 237–241.

126. Gan, C., Zhang, Q., and Mori, T. (2024). Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening. Preprint at arXiv. https://doi.org/10.48550/arXiv.240108315.

127. Wang, H., and Na, T. (2024). Rethinking E-Commerce Search. ACM SIGIR Forum 57, 1–19.

128. Gao, D., Chen, K., Chen, B., Dai, H., Jin, L., Jiang, W., Ning, W., Yu, S., Xuan, Q., Cai, X., and Yang, L. (2024). Llms-Based Machine Translation for E-Commerce, pp. 1–12. Available at SSRN 4682559. https://doi.org/10.2139/ssrn.4682559.

129. Antu, S.A., Chen, H., and Richards, C.K. (2023). Using LLM (Large Language Model) to Improve Efficiency in Literature Review for Undergraduate Research. In Proceedings of the Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation 2023 co-located with 24th International Conference on Artificial Intelligence in Education (AIED 2023), Tokyo, Japan, July 7, 2023, 3487 (CEUR-WS.org), pp. 1–9.

130. Bom, H.-S.H. (2023). Exploring the Opportunities and Challenges of ChatGPT in Academic Writing: a Roundtable Discussion. Nucl. Med. Mol. Imaging 57, 165–167.

131. Ülkü, A. (2023). Artificial intelligence-based large language models and integrity of exams and assignments in higher education: the case of tourism courses. Tourism & Management Studies 19, 21–34.

132. Qin, J., Wu, J., Chen, W., Ren, Y., Li, H., Wu, H., Xiao, X., Wang, R., and Wen, S. (2024). DiffusionGPT: LLM-Driven Text-to-Image Generation System. Preprint at arXiv. https://doi.org/10.48550/arXiv.240110061.

133. Han, J., Zhang, R., Shao, W., Gao, P., Xu, P., Xiao, H., Zhang, K., Liu, C., Wen, S., Guo, Z., et al. (2023). Imagebind-llm: Multi-modality instruction tuning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2309.03905.

134. Estecha-Garitagoitia, M., Rodríguez-Cantelar, M., Ruiz, A.G., García, C.G.F., Romero, S.E., Conforto, C., Fernández, A.S., Salvador, L.F., and D'Haro, L.F. (2023). THAURUS: An Innovative Multimodal Chatbot Based on the Next Generation of Conversational AI. Alexa Prize SocialBot Grand Challenge 5.

135. Meskó, B. (2023). The impact of multimodal large language models on health care's future. J. Med. Internet Res. 25, e52865.

136. Wu, S., Fei, H., Qu, L., Ji, W., and Chua, T.-S. (2023). Next-gpt: Any-to-any multimodal llm. Preprint at arXiv. https://doi.org/10.48550/arXiv.230905519.

137. Hu, W., Xu, Y., Li, Y., Li, W., Chen, Z., and Tu, Z. (2023). Bliva: A simple multimodal llm for better handling of text-rich visual questions. Preprint at arXiv. https://doi.org/10.48550/arXiv.230809936.

138. Han, Y., Zhang, C., Chen, X., Yang, X., Wang, Z., Yu, G., Fu, B., and Zhang, H. (2023). Chartllama: A multimodal llm for chart understanding and generation. Preprint at arXiv. https://doi.org/10.48550/arXiv.231116483.

139. Ye, Q., Xu, H., Xu, G., Ye, J., Yan, M., Zhou, Y., Wang, J., Hu, A., Shi, P., Shi, Y., et al. (2023). mplug-owl: Modularization empowers large language models with multimodality. Preprint at arXiv. https://doi.org/10.48550/arXiv.230414178.

140. Driess, D., Xia, F., Sajjadi, M.S., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A., Tompson, J., Vuong, Q., Yu, T., et al. (2023). Palm-e: An embodied multimodal language model. Preprint at arXiv. https://doi.org/10.48550/arXiv.230303378.

141. Zhang, D., Yu, Y., Li, C., Dong, J., Su, D., Chu, C., and Yu, D. (2024). Mm-llms: Recent advances in multimodal large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.240113601.

142. Zhang, W., Aljunied, M., Gao, C., Chia, Y.K., and Bing, L. (2024). M3exam: A multilingual, multimodal, multilevel benchmark for examining large language models. Adv. Neural Inf. Process. Syst. 36, 5484–5505.

143. Pan, X., Dong, L., Huang, S., Peng, Z., Chen, W., and Wei, F. (2023). Kosmos-g: Generating images in context with multimodal large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.231002992.

144. Liu, P., Ren, Y., Tao, J., and Ren, Z. (2024). GIT-Mol: A multi-modal large language model for molecular science with graph, image, and text. Comput. Biol. Med. 171, 108073.

145. Lin, W.C., Chen, A., Song, X., Weiskopf, N.G., Chiang, M.F., and Hribar, M.R. (2024). Prediction of multiclass surgical outcomes in glaucoma using multimodal deep learning based on free-text operative notes and structured EHR data. J. Am. Med. Inf. Assoc. 31, 456–464.

146. Xu, M., Yuan, X., Miret, S., and Tang, J. (2023). Protst: Multi-modality learning of protein sequences and biomedical texts. In International Conference on Machine Learning (PMLR), pp. 38749–38767.

147. Liu, T., Wang, Y., Ying, R., and Zhao, H. (2024). MuSe-GNN: Learning Unified Gene Representation From Multimodal Biological Graph Data. Adv. Neural Inf. Process. Syst. 36, 1–17.

148. Hou, Y., Yeung, J., Xu, H., Su, C., Wang, F., and Zhang, R. (2023). From Answers to Insights: Unveiling the Strengths and Limitations of ChatGPT and Biomedical Knowledge Graphs. Res. Sq. https://doi.org/10.21203/rs.3.rs-3185632/v1.

149. Hadi, M.U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M.B., Akhtar, N., Wu, J., and Mirjalili, S. (2023). A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage. Preprint at TechRxiv. https://doi.org/10.36227/techrxiv.23589741.v1.

150. Qi, S., Cao, Z., Rao, J., Wang, L., Xiao, J., and Wang, X. (2023). What is the limitation of multimodal llms? a deeper look into multimodal llms through prompt probing. Inf. Process. Manag. 60, 103510.

151. Clusmann, J., Kolbinger, F.R., Muti, H.S., Carrero, Z.I., Eckardt, J.N., Laleh, N.G., Löffler, C.M.L., Schwarzkopf, S.C., Unger, M., Veldhuizen, G.P., et al. (2023). The future landscape of large language models in medicine. Commun. Med. 3, 141.