Transformative Trends: A Comprehensive Review of Large Language Models (LLMs) in Healthcare

Chetna Kumari[1], Merin Meleet[2]

[1]Student, Department of ISE, R V College Of Engineering
[2]Assistant Professor, Department of ISE, R V College Of Engineering

---***---

Abstract - Recent years have witnessed notable progress in Language Model (LM) technology, especially with the introduction of Large Language Models (LLMs). With the use of complex neural networks and enormous volumes of data, these models—like the GPT (Generative Pre-trained Transformer) series—have completely changed the field of natural language processing. This study offers a thorough analysis of LLMs, examining their uses as well as the difficulties and moral issues that come with implementing LLMs in practical settings. Moreover, it investigates the wide range of uses of LLMs in many fields, such as question answering, drug discovery, language translation, and text production. By means of this review, we hope to offer ideas on future directions in this quickly developing field as well as insights into the state-of-the-art in LLM research today.

Key Words: Large Language Models, LangChain, Natural language processing, GPT, ChatGPT, AI chatbots, Indirect Prompt Injection

1.INTRODUCTION

The landscape of natural language processing (NLP) has been profoundly transformed by the emergence of Large Language Models (LLMs), characterized by their vast scale, sophisticated neural architectures, and ability to process and generate human-like text [6]. The goal of language modeling (LM), a core task in natural language processing (NLP), is to forecast the word or character that will appear next in a given text sequence[6]. It entails creating models and algorithms that can comprehend human language and produce coherent speech. The main goal of LM is to represent the probability distribution of words in a language. This enables the model to produce new sentences[8],[13], create new text[7], and forecast the likelihood of certain word combinations[9].

Natural language creation and comprehension have been transformed by large language models (LLMs) like ChatGPT and Bard. They are highly helpful in a variety of fields (such as search engines, customer assistance, and translation) because they have strong problem-solving abilities, contextual awareness, deep language comprehension, and human-like text production capabilities[1].

In many real-world tasks, LLMs are becoming an essential component of powerful agents that use LLMs for reasoning, tool utilization, and observational adaptation[17]. LLMs are being included into applications more and more frequently because of their flexible features, which are simple to adjust using natural language cues[3]. LLMs have demonstrated remarkable proficiency in language understanding and generation. With the increase in model size, they are better equipped to handle complex tasks and even exhibit emergent abilities [36]. Unique skills like instruction following, step-by-step reasoning, and zero shot generalization are demonstrated by LLMs in contrast to standard models. Self-supervised learning on a massive corpus (more than 1.4 trillion tokens) combined with human feedback tweaking makes such skills possible [37].

2. ARCHITECTURE OF LLM

LLMs were created expressly to understand and generate human language. Education, communication, content creation, article writing, healthcare, research, entertainment, and information distribution are just a few of the fields in which LLMs have been employed and contributed considerably to the field of AI [34]. The history of language models is seen in Figure 1[35].

Artificial neural networks (ANNs) were first proposed in the 1940s by Warren McCulloch and Walter Pitts [18]. Subsequently, the first language models were developed in the 1950s and 1960s [19]. Rule-based models and early neural networks were among these models. Their application of precisely defined linguistic elements and norms facilitated language processing [20]. In the 1980s and 1990s, language models based on statistics were developed. These models are part of a class of models used in machine learning (ML) and natural language processing (NLP) to identify and measure statistical patterns and correlations in linguistic data [21]. Word embeddings were first introduced to the field of NLP in the mid-2000s. This was considered a significant advancement and garnered a lot of attention [22]. The process of representing words in a continuous vector space is known as word embedding. The method uses a vector space to represent the words in order to capture their semantic links [23]. A major development in LLMs was the introduction of neural language models in the mid-2010s [24]. These models used artificial neural networks to interpret, generate, or predict human language in addition to using deep learning techniques to understand language patterns from large amounts of textual data[25].

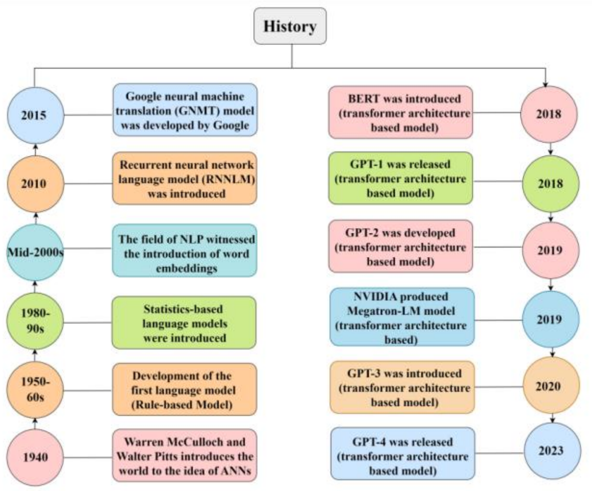
History
2015
Google neural machine translation (GNMT) model was developed by Google
2010
Recurrent neural network language model (RNNLM) was introduced
Mid-2000s
The field of NLP witnessed the introduction of word embeddings
1980-90s
Statistics-based language models were introduced
1950-60s
Development of the first language model (Rule-based Model)
1940
Warren McCulloch and Walter Pitts introduces the world to the idea of ANNs
BERT was introduced (transformer architecture based model)
2018
GPT-1 was released (transformer architecture based model)
2018
GPT-2 was developed (transformer architecture based model)
2019
NVIDIA produced Megatron-LM model (transformer architecture based)
2019
GPT-3 was introduced (transformer architecture based model)
2020
GPT-4 was released (transformer architecture based model)
2023

Fig. 1

Google released the first major neural language model that used deep learning techniques in 2015. The Google Neural Machine Translation (GNMT) model was the name given to the technology [26]. Massive amounts of multilingual textual data were used to train the model. This development represents a significant advancement in the machine translation area [27]. The Transformer model, which debuted in 2017, furthered the progress of language models [28]. The transformer model has greatly influenced natural language processing (NLP) and has been crucial in the creation of language models like Generic Pre-trained Transformers (GPT) and Bidirectional Encoder Representations from Transformers (BERT) [29]. One notable development in the field of NLP is the 2018 launch of BERT by Google AI [30]. The transformer architecture served as the study's fundamental framework. Due to its emphasis on unidirectional language modeling, the previous NLP-based language model had limitations when it came to comprehending contextual information prior to the development of BERT. Google presented BERT as a way to get around this specific restriction [31].

The goals of the GPT-2 series' design were to rectify the shortcomings of the GPT-1 series, which it replaced. GPT-2 was created using the transformer design, same as GPT-1. Alec Radford presented GPT-2, a language model with 1.5 billion parameters that was created on a deep neural network, in 2019 [32]. NVIDIA released the LLM Megatron-LM in 2019 [33]. This concept is based on the transformer architecture, just like GPT. The model has 8.3 billion parameters in total, which is a significantly greater number than the combined parameter count of GPT-1 and GPT-2 [30].

3. APPLICATIONS OF LLM IN HEALTHCARE SECTOR

The pharmaceutical sector, which is fueled by constant innovation and a tenacious quest of therapeutic development, is vital to the worldwide healthcare system. It is a dynamic, multidimensional field with many important facets that all contribute to the finished product[10]. Significant uses of LLM have been discovered in the medical domain, including the creation of clinical notes, the interpretation of test findings, and the anonymization of patient data. Their capacity to produce writing that is human-like suggests that medical documentation will be more accurate and efficient, which might lessen the administrative burden on medical staff and free up more time for patient care. But there are obstacles to overcome before applying LLM in healthcare, most notably the linguistic diversity and medical jargon that are common in multiethnic hospital environments. Due to this intricacy, LLM must be able to read and interpret medical terminology in a variety of languages as well as grasp the subtleties of clinical communication[11].

The application of large language models (LLMs) has revolutionary potential in the dynamic field of healthcare, transforming many aspects of medical documentation, interpretation, and patient care. This study examines how LLMs are being used in the healthcare industry, with a focus on numerous innovative applications that are changing patient care, medical documentation, and interpretation.

A. Disease Diagnosis:

To help medical personnel diagnose diseases effectively and precisely, LLMs can be used to analyze medical data, including patient symptoms, medical histories, and diagnostic test results. Large volumes of healthcare data may be processed by LLMs, who can then use that data to find patterns and trends that could help diagnose diseases[4].

B. Therapeutic Discovery:

By examining biological information, medication interactions, and molecular structures, LLMs are essential in the therapeutic discovery process. They can help scientists find possible treatment targets, forecast the effectiveness of drugs, and create brand-new drug compounds with the right characteristics[4].

C. Drug Target Discovery:

By mapping possible drug targets, LLMs like ChatGPT help pharmaceutical researchers find new targets for drug development. Through the examination of intricate biological data and interactions, LLMs can expedite target selection, resulting in a quicker time to drug discovery[28].

D. Toxicity Prediction:

Throughout the drug discovery and development process, LLMs are able to forecast the toxicity characteristics of compounds that resemble drugs. LLMs can offer insights into the possible toxicity of novel drug candidates by examining biological data and molecular structures. This aids in the process of prioritizing safe and efficient compounds for additional research and development[4].

E. Drug-Drug Interaction Analysis:

LLMs, like ChatGPT, can assist researchers in comprehending what is referred to as drug-drug interactions (DDIs), which are interactions between various medications. LLMs can help medical practitioners prevent potentially dangerous medication combinations by offering insights into how various pharmaceuticals interact with one another through the analysis of pharmacological data and molecular structures.[4]

F. Research and Development: The pharmaceutical industry's research and development procedures are undergoing a radical transformation thanks in large part to the introduction of LLMs. LLMs are used in drug discovery to forecast complex drug-target and drug drug interactions, offering insightful information for creating novel therapeutic molecules. Furthermore, LLMs play a crucial role in target identification, allowing scientists to anticipate the protein targets of natural substances that are employed to treat particular illnesses. Pharmaceutical businesses are able to accelerate the drug development pipeline and improve the efficacy of research activities by utilizing the power of LLMs[10].

G. Regulatory Affairs: LLMs are proving to be essential instruments in the field of regulatory affairs, helping to guarantee adherence to strict requirements and standards. Pharmaceutical firms can simplify regulatory procedures, evaluate complicated regulatory papers, and promote efficient communication within the regulatory environment by utilizing the skills of LLMs. Regulatory compliance processes like document analysis and information retrieval can be automated with the use of LLMs, which improves regulatory affairs management's correctness and efficiency. The pharmaceutical industry's dedication to maintaining standards for pharmaceutical product safety, efficacy, and quality is demonstrated by the use of LLM technologies into regulatory procedures[10].

H. Quality Control: LLMs are being used more frequently to improve quality control methods, which are essential to upholding the highest standards throughout the pharmaceutical production process. Pharmaceutical businesses can guarantee the safety, effectiveness, and dependability of their goods by using LLMs in data processing and interpretation. In order to analyze product quality data, spot anomalies, and spot any problems before they become serious throughout the production process, LLMs are essential. LLMs optimize quality control processes by using their sophisticated data processing and pattern recognition skills, which eventually improves product quality and complies with regulatory standards[10].

I. Clinical Trials: With their increased skills in data processing and interpretation, LLMs are revolutionizing the clinical trial scene in the pharmaceutical sector. LLMs are used in clinical trial settings to evaluate enormous volumes of patient data and medical literature, giving researchers and medical practitioners insightful information. Pharmaceutical businesses can improve patient care throughout clinical trial processes, support complicated medical queries, and increase contact with patients by integrating LLMs into chatbot systems. The utilization of chatbots driven by LLM in clinical trials is a prime example of how the industry is dedicated to utilizing cutting-edge technologies to improve patient involvement and optimize trial procedures[4], [10].

J. Supply Chain Optimization: LLMs are being used more frequently in the pharmaceutical business to optimize the supply chain operations that guarantee patients receive pharmaceutical medicines on time. Pharmaceutical businesses can improve inventory control, distribution network optimization, and supply chain management by utilizing LLMs' analytical capabilities. LLMs make it possible to analyze supply chain data, spot patterns and trends, and forecast changes in demand, which improves operational effectiveness and reduces costs. Supply chain optimization using LLM technology demonstrates the industry's dedication to improving logistics and guaranteeing smooth product delivery to satisfy patient needs[10].

K. LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Through evaluation studies with real psychiatrists and patients, the LLM-empowered Chatbots for Psychiatrist and Patient Simulation proved successful in psychiatric circumstances. The study demonstrated the viability of simulating psychiatrists and patients with mental problems with ChatGPTpowered chatbots. The chatbots were able to perform professional diagnostic processes, provide correct diagnoses, and provide a better overall healthcare experience. The significance of customized prompts was highlighted by an investigation on the effects of prompt designs on chatbot behavior and user experience. Overall, the study's findings point to the potential for LLM-powered chatbots to improve user experiences and diagnostic dialogues in mental health settings, highlighting their usefulness in situations involving psychiatric outpatients.[5]

L. Automation of discharge note generation using LLM: The automation of discharge note generation using LLM benefits healthcare professionals and patients by enhancing efficiency, accuracy, continuity of care, communication, and patient outcomes. LLM automates the creation of comprehensive discharge notes, saving time for healthcare professionals and reducing administrative burdens. The accuracy and completeness of discharge notes generated by LLM ensure that healthcare providers have standardized and detailed summaries of patient information, facilitating seamless continuity of care and improving communication among providers. Ultimately, the use of LLM in discharge note generation contributes to better patient care outcomes through improved coordination, decision-making, and overall quality of care[11].

M. LLM in Chatbot-based systems in Healthcare: Natural language generation (NLG) and natural language understanding (NLU) capabilities of LLMs like GPT-4 greatly improve chatbot-based systems in the healthcare industry[15]. With the use of these models, chatbots can more precisely and contextually understand and react to patient inquiries, resulting in more efficient communication and individualized support for the management of chronic illnesses. LLMs such as GPT-4 can produce replies that resemble those of a human, which enhances the trustworthiness, empathy, and engagement of chatbot engagements for patients. This human-like conversational skill is critical in healthcare settings, where patient involvement and trust are essential for fostering adherence to medical prescriptions and self-management techniques. The employment of LLMs improves the efficiency of chatbot engagements because they can respond to patient inquiries promptly and accurately, potentially lessening the workload for healthcare personnel and enhancing patient outcomes overall[12].

There are more chances for innovation in healthcare applications because of the ongoing breakthroughs in the sector and the continual evolution of LLMs with newer versions as GPT-4. With these developments in LLM technology, chatbot based solutions may be able to be further enhanced, giving patients managing chronic diseases more individualized and efficient support. Healthcare chatbots can utilize language learning models (LLMs) to provide customized recommendations, motivational messages, and compassionate responses. This allows patients to take charge of their health and follow treatment plans. By facilitating effective, compassionate, and patient-centered communication, the incorporation of LLMs into chatbot systems not only improves the caliber of patient encounters but also has the potential to revolutionize the way healthcare services are delivered[12].

4. SECURITY THREATS RELATED WITH LLM IN HEALTHCARE SECTOR

The use of LLMs in security-related tasks carries some possible hazards and threats. Among these issues is the vulnerability of LLMs to user-level attacks, whereby adversaries can manipulate and social engineer them due to their humanlike reasoning capabilities. LLMs can be used maliciously to create phony material, misleading messages, and phishing emails, among other things. This presents serious security risks for both individuals and companies. The confidentiality, integrity, and security of LLMs are further threatened by model extraction and parameter extraction assaults, for which there is a dearth of research focused on finding effective solutions. Strict control of LLM outputs also makes it more difficult to identify and stop the creation of malicious content, which increases the dangers of disseminating false or damaging information[1].

The concerns that have been uncovered emphasize how important it is to solve the security flaws in LLMs and put strong defenses in place to lessen the possibility of detrimental effects in security-related applications. In order to successfully handle the changing landscape of data protection, privacy, and cybersecurity practices, the incorporation of LLMs in security duties calls for a reevaluation of cybersecurity laws, regulations, and compliance standards. Therefore, taking proactive steps is crucial to enhancing LLMs' security posture and resilience against new threats as well as guaranteeing their safe and responsible deployment in situations when security is crucial[1].

The paper highlights a potential weakness in ChatGPT's ability to handle misspellings, as it fails to identify mistakes and may overlook certain symptoms during conversations. This limitation suggests that LLMs like ChatGPT[16] may struggle with accurately interpreting and responding to text inputs that contain errors or misspellings, which could impact the overall performance and reliability of the chatbots in psychiatric scenarios[5].

Large Language Models (LLMs) have a number of drawbacks, one of which is their tendency to produce ”hallucination,” or erroneous or illogical outputs, which can impair the quality and dependability of generated text. The intricacy and thorough data processing involved in LLMs might lead to the production of false or erroneous information, thus adding flaws in the final product. Biases in training data and the societal ramifications of text generated by AI give rise to ethical considerations, underscoring the significance of responsible deployment. There is a chance that activities requiring human judgment will be overly delegated to LLMs, which could reduce the role of humans in decision-making. It is imperative to tackle these adverse effects in order to guarantee the moral and efficient deployment of LLMs in diverse contexts[2].

LLMs have a lot of drawbacks and security problems when they are integrated into applications[13]. Because LLMs are weak intermediaries, users can be manipulated by adversarial prompts or false document summaries, and adversaries can compromise models and cause other security breaches by using Indirect Prompt Injection (IPI) attacks to take advantage of LLMs[14]. LLMs' flexibility and self-governing behavior open doors for user manipulation and the spread of false information, escalating security risks related to data privacy and illegal access. Because LLM-integrated applications are inherently vulnerable to a variety of attacks, including fraud, infiltration, and biased output, it is critical to address security vulnerabilities in order to protect users from exploitation and maintain system integrity[3].

5. CONCLUSION

Our paper offers a thorough analysis of the uses of LLMs, in the healthcare industry, highlighting the revolutionary potential of these models in the areas of patient care, medical documentation, and interpretation. We have highlighted the important contributions of LLM technology to improving healthcare practices by exploring the use of LLMs in disease diagnosis, therapeutic discovery, drug target discovery, toxicity prediction, drug-drug interaction analysis, chatbot-based systems, research and development, regulatory affairs, quality control, clinical trials, and supply chain optimization. Specifically, the use of LLMs to automate the creation of discharge notes has shown promise as a useful application that can improve patient outcomes, efficiency, accuracy, communication, and continuity of care. Our study has also shown security risks related to LLM integration in apps, such as censorship issues, malicious exploitation, model extraction, parameter extraction, and userlevel attack vulnerabilities. Proactive steps are necessary to reduce risks, strengthen defenses, and coordinate with changing cybersecurity regulations and compliance requirements in order to guarantee the safe and responsible deployment of LLMs in security-related duties and integrated applications. Future developments in LLM technology and its application to healthcare applications have the potential to improve patient care even more, expedite medical procedures, and accelerate the delivery of healthcare.

REFERENCES

[1] Y. Yao et al., ”A Survey on Large Language Model (LLM) Security and Privacy: The Good, The Bad, and The Ugly,” High-Confidence Computing, vol. 4, no. 2, pp. 100211, 2024.

[2] O. Topsakal and T. C. Akinci, ”Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast,” Florida Polytechnic University, FL, USA and University of California at Riverside, CA, USA.

[3] K. Greshake et al., ”Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection,” 2023.

[4] S. Pal et al., ”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,” 2023.

[5] S. Chen et al., ”LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation,” 2023.

[6] M. U. Hadi, Q. A. Tashi, R. Qureshi, et al., ”Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects,” TechRxiv, Nov. 16, 2023.

[7] T. Iqbal and S. Qureshi, ”The survey: Text generation models in deep learning,” Journal Name, vol. 34, no. 6, pt. A, pp. 2515-2528, Jun. 2022.

[8] ”HONEST: Measuring Hurtful Sentence Completion in Language Models,” NAACL, 2021. [Online]. Available: https://aclanthology.org/2021.naacl-main.191

[9] B. Min et al., ”Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey,” Nov. 1, 2021.

[10] Y. Han and J. Tao, ”Revolutionizing Pharma: Unveiling the AI and LLM Trends in the Pharmaceutical Industry,” last revised Jan. 22, 2024.

[11] H. Jung et al., ”Enhancing Clinical Efficiency through LLM: Discharge Note Generation for Cardiac Patients,” Apr. 8, 2024.

[12] S. Montagna et al., ”Data Decentralisation of LLM-Based Chatbot Systems in Chronic Disease Self-Management,” Sep. 2023.

[13] R. Tang, Y.-N. Chuang, and X. Hu, ”The Science of Detecting LLMGenerated Text,” 2024.

[14] R. Pedro, D. Castro, P. Carreira, and N. Santos, ”From Prompt Injections to SQL Injection Attacks: How Protected is Your LLM-Integrated Web Application?", 2023.

[15] M. Abbasian, I. Azimi, A. M. Rahmani, and R. Jain, ”Conversational Health Agents: A Personalized LLM-Powered Agent Framework,” 2024. [16] Y. Wu, F. Roesner, T. Kohno, N. Zhang, and U. Iqbal, ”SecGPT: An Execution Isolation Architecture for LLM-Based Systems,” 2024.

[17] Q. Wu et al., ”AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation,” 2023.

[18] M. Kardum, ”Rudolf Carnap—The grandfather of artificial neural networks: The influence of Carnap’s philosophy on walter pitts,” in Guide to Deep Learning Basics. Cham, Switzerland: Springer, 2020, pp. 55–66.

[19] G. Leech, ”Corpora and theories of linguistic performance,” Svartvik, J. Directions Corpus Linguistics, vol. 10, pp. 22–105, Jun. 1992.

[20] J. Hirschberg, B. W. Ballard, and D. Hindle, ”Natural language processing,” ATT Tech. J., vol. 67, no. 1, pp. 41–57, Jan. 1988.

[21] B. Cronin, ”Annual review of information science and technology,” Inf. Today, Medford, OR, USA, 2004, vol. 39.

[22] D. S. Hain, R. Jurowetzki, T. Buchmann, and P. Wolf, ”A textembedding-based approach to measuring patent-to-patent technological similarity,” Technol. Forecasting Social Change, vol. 177, Apr. 2022, Art. no. 121559.

[23] G. Curto, M. F. Jojoa Acosta, F. Comim, and B. Garcia-Zapirain, ”Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings,” AI Soc., vol. 2022, pp. 1–16, Jun. 2022.

[24] P. Azunre, Transfer Learning for Natural Language Processing. New York, NY, USA: Simon and Schuster, 2021.

[25] Y. Shi, M. Larson, and C. M. Jonker, ”Recurrent neural network language model adaptation with curriculum learning,” Comput. Speech Lang., vol. 33, no. 1, pp. 136–154, Sep. 2015.

[26] R. K. Yadav, S. Harwani, S. K. Maurya, and S. Kumar, ”Intelligent chatbot using GNMT, SEQ-2-SEQ techniques,” in Proc. Int. Conf. Intell. Technol. (CONIT), Jun. 2021, pp. 1–5.

[27] D. Luitse and W. Denkena, ”The great transformer: Examining the role of large language models in the political economy of AI,” Big Data Soc., vol. 8, no. 2, Jul. 2021, Art. no. 205395172110477.

[28] H. Chopra, S. Akash, S. Chakraborty, et al., ”Artificial intelligence (AI) paving critical role in drug discovery, drug designing and studying drug-drug interactions - Correspondence,” Int. J. Surg., 2023, doi:10.1097/JS9.0000000000000564.

[29] M. Onat Topal, A. Bas, and I. van Heerden, ”Exploring transformers in natural language generation: GPT, BERT, and XLNet,” 2021, arXiv:2102.08036.

[30] L. Fan, L. Li, Z. Ma, S. Lee, H. Yu, and L. Hemphill, ”A bibliometric review of large language models research from 2017 to 2023,” 2023, arXiv:2304.02020.

[31] C. Sur, ”RBN: Enhancement in language attribute prediction using global representation of natural language transfer learning technology like Google BERT,” Social Netw. Appl. Sci., vol. 2, no. 1, p. 22, Jan. 2020.

[32] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, ”Language models are unsupervised multitask learners,” OpenAI Blog, vol. 1, no. 8, p. 9, 2019.

[33] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, ”Megatron-LM: Training multi-billion parameter language models using model parallelism,” 2019, arXiv:1909.08053.

[34] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. Shaikh, N. Akhtar, J. Wu, and S. Mirjalili, ”A survey on large language models: Applications, challenges, limitations, and practical usage,” TechRxiv, Nov. 16, 2023.

[35] M. A. K. Raiaan, M. S. H. Mukta, K. Fatema, N. M. Fahad, S. Sakib, M. M. J. Mim, J. Ahmad, M. E. Ali, and S. Azam, ”A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges,” 2024.

[36] X. Ma, G. Fang, and X. Wang, ”LLM-Pruner: On the Structural Pruning of Large Language Models,” 2023.

[37] H. Wen, Y. Li, G. Liu, S. Zhao, T. Yu, T. J.-J. Li, S. Jiang, Y. Liu, Y. Zhang, and Y. Liu, ”AutoDroid: LLM-powered Task Automation in Android,” 2023.