Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 04 April-2025, Page No.- 4578-4593

DOI: 10.47191/etj/v10i04.26, I.F. – 8.482

A Comprehensive Survey on Large Language Models: Architectures, Applications, and Ethical Considerations

Chreesk Sabah M. Ali[1], Dr. Ibrahim Mahmood Ibrahim[2]

[1,2]Akre University for Applied Sciences, Technical College of Informatics, Department of Information Technology, Duhok, Kurdistan Region, Iraq.

ABSTRACT: The rapid progress and revolutionary potential of large language models (LLMs) are examined in this survey, with particular attention paid to transformer-based designs and scaling strategies that power models such as GPT-4, Gemini, and LLaMA. Focusing on advancements in natural language comprehension, translation, text production, and information retrieval, this study examines various applications in the fields of healthcare, finance, education, and public administration. Along with addressing ethical issues such as algorithmic bias, disinformation, privacy threats, and environmental impacts, the study promotes robust auditing frameworks, consistent evaluation measures, and environmentally friendly practices. The study offers a comprehensive framework for future research and the responsible development of large language models (LLMs), outlining the complex supply chain and development processes with a focus on ethical integration, accountability, and transparency.

KEYWORD: Extensive Language Models, Transformer Architectures, Natural Language Processing Applications, Ethical Implications, Algorithmic Bias, Disinformation, Privacy and Security, Environmental Consequences, Assessment Metrics.

1. INTRODUCTION

Recent years have witnessed an explosive surge in large language models (LLMs). LLMs are a category of natural language processing (NLP) models that utilize deep learning techniques to enhance human-computer interaction through speech [1]. The potential of large language models has been apparent from the impressive amount of research and growth of the model, including but not limited to machine learning model scaling techniques; numerous types of models based on text, pictures, audio, textual reviews, social connections, or a combination of these; and contrasting quality measures as per the context of production and training. [2]

Recent results on task-based financial language models (LLMs), multilingual language models, biomedical and clinical language models, vision-language models, and code language models are cutting-edge. Moreover, a comprehensive literature review has been conducted to date, and only a limited number of subcategories have been investigated. An extensive survey has been provided on a significant LLM subcategory, encompassing the processes, characteristics, datasets, transformer versions, and comparison metrics for each type, accompanied by a brief overview of ten different types [3]. Furthermore, it is part of the broader public debate and opposing visions on the transformative power of such technology and its impact on society. The intentionality, characteristics, and cynical use of more research on GPT-3-type design principles have been found. It stresses that LLMs exhibit several remarkable capabilities besides generating human-like language, such as excellent recent notes. [4][5]

In recent years, Large Language Models (LLMs) have indeed gained considerable traction in various fields of natural language processing, text understanding, and many related areas. This advancement has opened up exciting possibilities in a wide range of tasks, including generating, analyzing, and regulating text that was previously considered impractical or overly complicated [6]. However, it is essential to acknowledge that a substantial portion of the work conducted in this domain, including the scaling of LLM technologies, has not adequately addressed ethical questions that span several critical dimensions. Engaging in intelligent conversation about these pressing concerns demands a certain level of understanding of the intricate LLM game: how it functions, what it is not able to do effectively, and what ethical implications it raises are likely, considering the rapid evolution of its capabilities. [7]

2. BACKGROUND AND EVOLUTION OF LARGE LANGUAGE MODELS

The emergence of large language models (LLMs) such as GPT-3 has sparked an immense wave of significant interest in the field of natural language processing due to their remarkably enhanced capabilities to understand and generate human language. These advanced models possess the ability to process vast amounts of data and exhibit a level of comprehension that can closely mimic human-like responses. However, with these advancements come serious concerns

about issues related to model biases, discrimination, and overall reliability, which remain prominent in discussions surrounding the use of LLMs. Reports have indicated the potential of these powerful LLMs to produce convincing forgeries unintentionally and spread misinformation, which could substantially degrade the quality of public discourse and mislead the general public into adopting inaccurate beliefs [8][9]. The multifaceted issues surrounding gray-box LLMs may exacerbate the rise of misinformation and intensify existing societal conflicts. Thorough investigations into the functioning and outputs of current large language models (LLMs) have validated lingering concerns regarding their role in the dissemination of fake news, conspiracy theories, and various forms of medical misinformation, all of which pose significant potential socio-political risks to societies worldwide. Open AI's earlier model, GPT-2, serves as a foundational example of the capabilities exhibited by large language models (LLMs), utilizing extensive datasets such as Common Crawl and WebText for practical training. These sophisticated models, featuring an astonishing number of parameters that can reach up to 1.5 billion, produce high-quality content and can stylistically adapt to various user prompts, often making it increasingly challenging for individuals to detect their artificial nature and origins. [10][11]

2.1. Historical Context

Large Language Models (LLMs) have revolutionized natural language processing, demonstrating their adaptability in a wide range of applications. This review examines their historical background and origins, emphasizing the significance of understanding earlier models. The field of computational linguistics, which began in the 1950s, facilitated significant advancements in natural language processing (NLP). Initial approaches were rule-based, requiring complex instructions from linguists. Over the past two decades, machine learning has become a central focus in natural language processing (NLP). Key developments include a 1990 paper on statistical methods and a pivotal 2001 model that emphasized data-driven techniques. The 2013 introduction of word embeddings revolutionized NLP by creating meaningful representations for predictive models, forming the foundation for today's large pre-trained models. Examining influential projects and literature is crucial for grasping the rise of LLMs. [12][13]

Figure 1: The figure rpresentation about Historical Context

2.2. Key Architectures

Since the introduction of the transformer, neural networks utilizing this framework have dominated the field of natural language processing (NLP). The BERT model popularized transformer-based models, while GPT-3 advanced language models (LMs) with a high number of parameters. The success of LMs depends on more extensive neural networks that improve language representation for various NLP tasks. The transformer architecture simplifies training and scales well, making it ideal for LMs. Its versatility has led to the development of many models, with the attention mechanism enabling the contextual computation of word importance and thereby enhancing text understanding. [14]

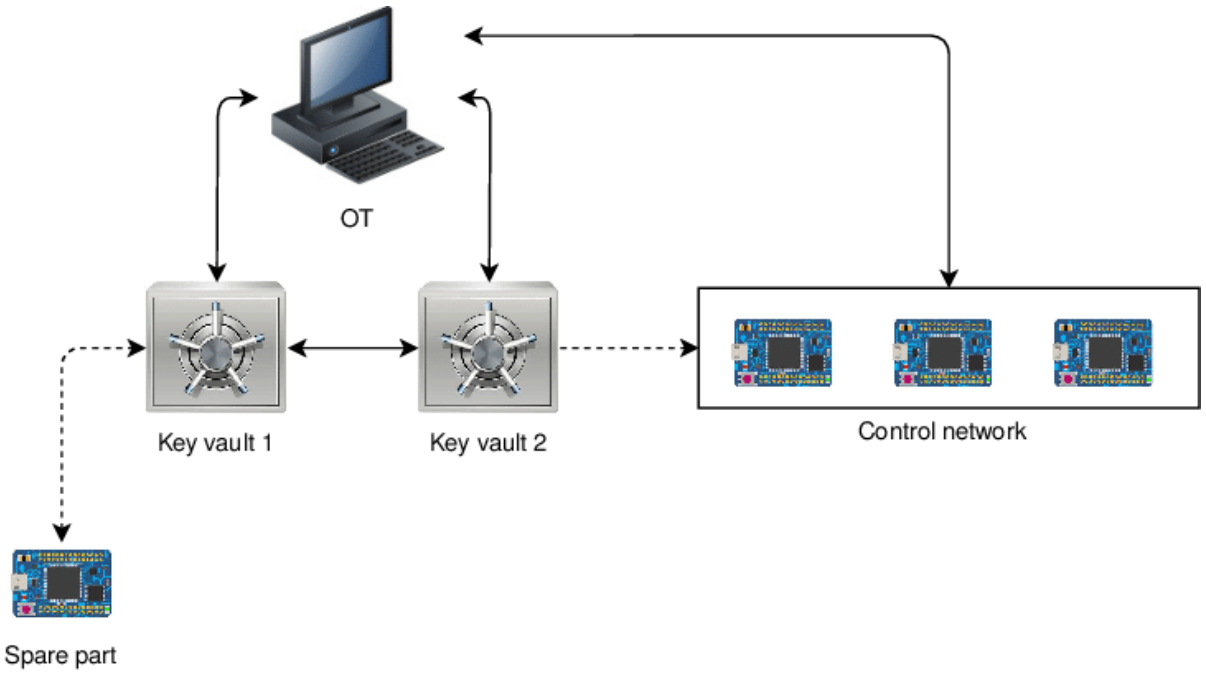

Figure 2: The figure rpresentation about Key Architectures

3. APPLICATIONS OF LARGE LANGUAGE MODELS

Large Language Models (LLMs) have gained popularity in natural language processing, with applications in various sectors, notably in business, enhancing consumer engagement. An examination of past works and a critical evaluation of available resources can support future research. This study reviews the development of LLM research, highlighting the need to gather diverse evidence to understand the architectures and methodologies of LLMs. It presents an expert review of notable models, offering insights into their practical usage. The research examines the current applications and future potential of large language models (LLMs). It begins with an introduction to the significance of LLMs, outlining the methodology for surveying the literature, synthesizing key materials, assessing the evolution, capabilities, advantages, and limitations of LLMs, and identifying research gaps. [15]

3.1. Natural Language Understanding

Natural Language Understanding (NLU) is crucial in large language models (LLMs), enabling machines to accurately interpret human language. LLMs utilize advanced algorithms and extensive training data to enhance comprehension across various Industrial Manufacturing Process (IMP) scenarios, including sentiment analysis and conversational AI. These models can analyze extensive text and infer meanings, understanding context without human input. However, nuances in language remain a challenge. While recommendation models aim to improve clarity, they might miss intended meanings. Machines can struggle to connect dots and extract information, often finding it difficult to handle subtle requests or poorly structured prompts, which complicate the understanding of language nuances. [16]

3.2. Text Generation

Large language models (LLMs) are revolutionizing content creation by enabling users to generate text that resembles human language and explore new applications. Unlike earlier methods, which used rule-based generation, LLMs create context-informed text. They can produce various content types, such as articles, poetry, and change logs, but may show inconsistencies in longer texts. Ethical concerns arise over potential misleading content and deception. Recent advancements focus on fine-tuning for specific domains instead of training from scratch. One notable project used a medical dataset to promote transparency. Innovations in meta-learning and input formats have improved domain alignment. Current research aims to enhance prediction accuracy, develop classifiers for biased content, and ensure thoughtful Romanization for non-English languages. Diverse applications in retail, finance, and digital humanities underscore the support of LLMs for creative fields, such as music curation. [17][18]

3.3. Language Translation

Large language models (LLMs) have significantly improved machine-generated text, particularly in the field of translation. This review focuses on advancements in LLM architectures essential for translation, which is as important as text classification and generation. Accurate translation involves well-formed sentences that preserve the original meaning. The deep learning attention mechanism has significantly enhanced translation quality, as demonstrated by benchmarks such as WMT14 and WMT19, where large models outperform conventional NLP systems [19]. For example, Facebook AI's Efficient Multilingual can handle low-resource language pairs with 15 models. Real-time

translation systems facilitate global communication, enhancing daily life and work. LLMs also expand cross-cultural communication, fostering understanding among diverse populations. Continued improvements are anticipated to strengthen relationships through dialogue, enabling these models to translate various texts idiomatically and enrich our understanding of literature and history. [20]

3.4. Information Retrieval

The exponential growth in textual data has brought large language models (LLMs) into focus, which have become dominant in NLP application scenarios across various industries. Through extensive arrangements, numerous parameters, and pre-training, LLMs have advanced NLP to unprecedented efficiency, with user-friendly APIs enabling widespread use. Specifically, this paper aims to illustrate the recent advancements in LLMs' architectures, applications, and ethical considerations. By covering a plethora of knowledge on the current LLM trend, this can be broadly beneficial for developers, researchers, and users. [21]

3.4.1.GPT-4

The arrangement of transformer architecture has sparked a surge in vitality in NLP. As an imitation of technologies developed by several versions of GPT models, proposals were made. GPT-4 models outperform human evaluations in advancing language understanding. Due to the increasing number of parameters in LLMs, large models far exceed the pre-training range of token numbers. Thus, large-scale language models face considerable challenges in assembling and computing. Meanwhile, multiple successive works have been done to incentivize the empowerment of lightweight LLMs. To rival human performance in language evaluations, LLMs should have at least 20 billion parameters. Consequently, favorable settings for updating need to be provided and essentially enhanced. [22][23]

3.4.2. Applications

LMs have achieved breakthroughs in text-matching tasks, transforming into text search models for practical answers. A search-to-write mechanism enables companies to engage with search engines and address users' inquiries. Deployed conversational bots used by e-commerce platforms must communicate using consistent information. These conversations employ an abductive approach, resulting in the development of a search kiosk for immersive dialogue support. Among various applications, inventive assistants excel by employing the retrieve-and-interact n-gram strategy to enhance answer models and streamline responses to diverse questions. This personal assistant goes beyond experimental use by proposing an expansion with public question banks. [24]

4. ETHICAL CONSIDERATIONS IN THE DEVELOPMENT AND DEPLOYMENT OF LARGE LANGUAGE MODELS

Large language models (LLMs) have excelled in various tasks due to their extensive training on vast amounts of text data. However, safety concerns have arisen, particularly about misinformation and hate speech generated by models like GPT-3. It's essential to detect false content and design deceptive prompts to ensure the ethics and safety of AI. Improving Large Language Model (LLM) defenses against harmful prompts can enhance their robustness. Recent developments in pre-trained models have advanced natural language processing and data analysis tools, but have also raised issues regarding misinformation and harmful behavior. This essay assesses the R&D landscape, examining the risks associated with LLMs and offering strategies for companies, academics, and regulators to mitigate these risks while maximizing technological benefits. [25][26]

4.1. Privacy and Security

The recent surge in LLM research has advanced various natural language processing tasks, but it has also raised significant concerns about user privacy and confidentiality in AI interactions. Discussions focus on the harmful impacts of AI dialog systems on user security. This section highlights recent studies on privacy and security related to cross-attention models, critiquing the risks associated with information leakage and unauthorized data use. Methodologies and adversarial attacks aim to secure training and inference data, utilizing techniques ranging from essential randomization to differential privacy. McKinsey's research has also raised concerns about potential lawsuits and ethical issues, prompting dialogue on data protection and AI fairness, especially within EU frameworks. Emphasis on transparency and user consent is crucial for rebuilding trust, avoiding exploitative agreements, addressing challenges in LLM adoption, and prioritizing user safety and privacy. [27]

4.2. Misinformation and Disinformation

Artificial intelligence research in natural language processing has led to the development of large language models (LLMs) that generate coherent text, supporting applications such as customer service, content generation, and translation. However, these models can also produce misleading information, which poses risks to public safety and impacts areas such as health, stock markets, and elections. It's essential to understand LLM-generated misinformation to prevent its misuse. Users, developers, and organizations should enhance datasets to identify misinformation, manage model outputs, evaluate source credibility, monitor accountability, and improve search capabilities. The deterioration of institutional trust and global information disorder complicates the recovery from COVID-19 and scientific progress. Incorporating suggested ethical considerations can help mitigate the risks associated with

LLMs, underscoring the importance of responsible use and research in information integrity technologies. [28][29]

4.3. Environmental Impact

As computational electricity use in industry and services rises, so does energy consumption in AI development. Concerns focus on the environmental impact of training large language models (LLMs). This project aims to raise awareness, emphasizing the importance of green considerations as AI technologies evolve. Training a single model can require multiple GPUs for days or weeks, creating a significant carbon footprint. Studies on carbon footprint during training and emissions disclosures by AI labs have emerged. Proposals include energy-efficient practices and architectures inspired by biology. Community initiatives, like greenconference.ai, emphasize sustainability through carbon reporting and tree planting. However, progress toward reductions has been slow, indicating a need for better coordination, knowledge sharing, and accountability. [30]

5. EVALUATION METRICS AND BENCHMARKS

Recently, Large Language Models (LLMs) have drawn much attention due to their promising performance, surpassing other machine learning models across various natural language processing tasks. The successful development of various large language models (LLMs) has motivated a comprehensive survey of the most recent LLM models, examining not only model architectures, pre-training strategies, and underlying datasets but also noteworthy applications and domains in which they have been employed. As LLMs have elicited growing interest in terms of ethical, legal, and societal considerations, an extended overview is also presented concerning potential risks, misuse, and worthy open research challenges to facilitate the future development and deployment of LLM technologies. [31]

The rise to prominence of LLMs over the past three years, in part due to the breakthrough of pre-trained transformer model families, has generated curiosity to evaluate, compare, and contrast different LLMs. The performance of an LLM should ideally be measured using robust, generative evaluation frameworks rather than during its pre-training. Since each publication to date proposes evaluation based on fixed metrics and datasets, comparing the wide-ranging discoveries is complicated. It is desirable to unify and standardize the evaluation methodology as proposals are submitted. Several models have been devised recently, necessitating a thorough evaluation approach. [32]

6. LITERATURE REVIEW

By breaking down the entire ecosystem into discrete parts, **Wang et al. (2024)** [33] offer a thorough investigation of the LLM supply chain. The supply chain is divided into layers by its analysis, including computing resources, development toolchains, datasets, models, and applications. They also examine the interdependencies that make creating and implementing LLMs particularly challenging. By focusing on both software engineering and security and privacy, the authors highlight existing constraints, such as model versioning, reproducibility, and dependency vulnerabilities, while also providing a research agenda for the future that aims to address these issues. This study calls for a concerted effort to ensure scalable, safe, and morally sound LLM development, establishing a crucial foundation for further research. LM developments necessitate a concerted effort to establish a critical basis for further research.

[34] **In 2024, Javaid and Saeed** investigate how large language models (LLMs) can enhance unmanned aerial vehicle (UAV) technologies in this research. Their research demonstrates how incorporating LLMs can enhance real-time data processing, facilitate more independent decision-making, and facilitate more effective mission planning by utilizing adaptive algorithms and improved natural language interpretation. The research thoroughly analyzes several LLM designs designed specifically for UAV applications. It assesses how well they can get over conventional restrictions in obstacle detection, navigation, and operational coordination. The authors lay the groundwork for the next generation of intelligent, responsive, and effective UAV systems by describing both the benefits and the technical challenges.

Mökander & Floridi (2024) [35] put out a new paradigm for auditing LLMs. It is divided into three interconnected layers: governance audits, model audits, and application audits. Their strategy aims to address the complex moral and societal issues raised by LLMs by providing methodical supervision throughout the entire lifecycle of an LLM. The governance audit assesses technology providers' policies and procedures to ensure that moral and legal requirements are met. The model audit verifies the LLMs for biases, safety, and technical soundness. The application audit closely examines how these models are utilized in practical situations. This multi-layered approach offers a workable plan for achieving openness and accountability in the rapidly evolving field of LLM development and deployment.

Papageorgiou & Tjortjis [36] investigate how LLMs can be integrated into e-government systems, emphasizing the development of a repeatable and flexible architectural framework. Their research utilizes Retrieval-Augmented Generation (RAG) to develop sophisticated virtual assistants that enhance the delivery of public services by improving user engagement and data processing. The study demonstrates how a well-designed LLM-based architecture can revolutionize digital government services by addressing key concerns, including scalability, cost-effectiveness, and transparency. By offering case studies and real-world examples, the writers provide insightful information on how practitioners and legislators can utilize LLMs to foster creativity and enhance operational effectiveness in public administration.

Kong & Zohren (2024) [37] investigate how large language models (LLMs) can revolutionize financial and investment management. They highlight how LLMs can process intricate datasets, enhance sentiment analysis, and optimize decision-making through innovative applications such as agent-based modeling and asset-liability management. The paper presents comprehensive benchmarks and carefully selected datasets, providing practitioners with a robust framework for integrating LLMs into conventional financial processes. It divides financial applications into three categories: language tasks, forecasting, and financial reasoning. In addition to showcasing the technological advancements in LLM structures, this study explores the opportunities and practical challenges of utilizing these models in high-stakes financial settings.

A thorough analysis of the new applications of LLMs in the medical industry is provided by **Omiye & Daneshjou (2022)** [38], who describe how these models are being utilized to augment jobs, including writing clinical notes, responding to patient inquiries, and preparing for medical exams. Both the potential for LLMs to enhance clinical decision support and the inherent risks, such as the possibility of producing medically incorrect outputs and exacerbating biases, that may impact patient care are rigorously evaluated in this research. This fair assessment highlights the importance of strict ethical standards and validation as large language models (LLMs) are increasingly integrated into healthcare systems.

Large language models (LLMs) transform passive recommendation engines into active, interactive platforms that can comprehend and anticipate user wants**. Chen et al. (2024)** [39] examine this convergence and emphasize a paradigm change. According to the study, LLMs offer dynamic personalization through natural language interfaces, enhancing the user experience by providing more specialized content and facilitating seamless integration with other tools, such as search engines and APIs. Meanwhile, the document addresses issues such as protecting user privacy, ensuring transparency, and outlining a plan for future AI-powered personalization advancements.

A thorough analysis of the state of LLMs in medicine is provided by **Thirunavukarasu and Ting (2024)** [40], who emphasize how these tools can potentially transform clinical practice by improving natural language comprehension and decision support. From early versions to sophisticated systems like ChatGPT, this essay recounts the evolution of large language models (LLMs). It explores their uses in clinical and educational contexts, including the creation of clinical summaries and the promotion of patient interaction. It also highlights the necessity of continuous model improvement, robust validation, and cautious handling of intrinsic constraints, such as bias and reliability issues, to fully realize their potential in healthcare.

Yan and Liu (2024) [41] provide a comprehensive ethical framework specifically designed for integrating large language models into learning environments. Their analysis presents a systematic and multi-layered approach to addressing these challenges, highlighting key ethical concerns, including algorithmic bias, data privacy, and the potential for disseminating misleading information. The authors emphasize the importance of open and responsible procedures, stressing the significance of regulatory oversight to ensure that AI-powered teaching resources promote equity and respect moral principles by comparing AI development with human education. This paradigm establishes the foundation for upcoming studies in ethical AI education and serves as a guide for legislators.

A thorough analysis exploring the complex nature of bias in large language models (LLMs) is given by **Guo & Liu (2024)** [42]. Extrinsic biases that emerge when the model is applied in downstream tasks are distinguished from intrinsic biases, which are ingrained during the training phase due to the use of unbalanced or unrepresentative data. The authors rigorously assess techniques at the data, model, and output levels to identify and reduce these biases by making analogies with traditional statistical biases. Their research offers crucial insights for the just and responsible development of AI systems, providing a comprehensive toolkit for bias detection and reduction. It also highlights the moral and practical implications of using biased large language models (LLMs) in sensitive domains, such as healthcare and legal decision-making.

By examining and classifying a large amount of literature covering a variety of activities such as code generation, summarization, translation, and vulnerability identification, **Zheng & Wang (2024)** [43] look into the integration of LLMs inside software engineering. Based on a comprehensive collection of 123 papers from key academic sources, their systematic evaluation highlights the transformational potential of LLMs and their existing limits in automating complicated software development processes. The paper provides a helpful roadmap for future research and tool optimization in this rapidly evolving field by critically evaluating the efficacy of LLMs in addressing challenges such as semantic understanding and vulnerability analysis, and offering a comprehensive taxonomy of software engineering tasks enhanced by LLMs.

10. By following the development of AI from early symbolic systems to the cutting-edge transformer architectures that support contemporary models like GPT-4, **Desai & Mehta (2023)** [44] provide a thorough examination of LLMs. Self-attention mechanisms and scaling rules are two important developments that have elevated LLMs to the forefront of natural language processing, and the paper highlights both. The authors critically examine enduring issues, such as innate biases, moral quandaries, and the significant computational resources required for training, in addition to discussing their impressive ability to produce logical and contextually relevant text. By providing a fair assessment of the

revolutionary possibilities of LLMs while warning against their drawbacks, this comprehensive synthesis offers insightful viewpoints on the broader discussion of general artificial intelligence.

To investigate the cultural aspects of LLM outcomes, **Yuan and Zhang (2024)** [45] compare how well they align with mainstream Chinese cultural values in both Chinese and English-speaking contexts. According to the study, notable differences exist between the two languages in how these models represent principles such as collectivism, Daoist balance, and Confucian harmony. The authors demonstrate the crucial influence of the linguistic environment on innate cultural biases in AI systems through the use of quantitative measurements and thematic analysis. Their results underscore the need for culturally sensitive AI that can accommodate a diverse range of user backgrounds and promote inclusivity, providing crucial guidance for the future development of language models that are both culturally sensitive and bias-mitigated.

In **2024, Langston and Ashford** [46] present a new method that utilizes massive language models to automatically generate concise and logical summaries from the abstracts and titles of multiple documents. Their approach, which combines abstractive and extracted summarization methods, responds to the increasing demand for effectively condensing large volumes of textual material. The research excels in technical fields such as science and healthcare, but it also highlights issues with readability and coherence. The discipline of document summarization is advanced by these findings, opening the door for further advancements in hybrid summarization techniques.

Specifically, concentrating on cultural hallucination, **Zhao, Huang, and Li (2024)** [47] propose a comparative research study that assesses how large language models address ethical questions unique to a particular culture. Their research uses automated evaluation measures to evaluate semantic similarity, cultural relevance, and ethical consistency across various models. The results show that although the performance of current models is encouraging, there are still notable differences in their capacity to produce culturally appropriate answers. A deeper cultural understanding must be incorporated into model training to reduce biases and ensure the ethical application of AI, as emphasized in this work.

Kasneci et al. (2024) [48] examine the transformative potential of large language models in educational settings, highlighting the opportunities and challenges of their integration. The authors emphasize the need for educators to acquire crucial digital literacy skills while also discussing how these models can enhance student engagement, tailor learning experiences, and improve content development. They warn that problems such as model brittleness, interpretability issues, and ethical concerns, including bias and misuse, must be properly addressed through rigorous monitoring and informed instructional approaches.

Wong, Leung, and Wong (2024) [49] systematically evaluate four open-source big language models (GPT-Neo, Bloom, FLAN-T5, and Mistral-7B**)** against predetermined benchmarks. Their research focuses on computing efficiency, scalability, and adaptability, and compares the accuracy of the models in language processing and generation tasks. By identifying crucial architectural elements and training approaches that impact model performance, the study offers practical recommendations for enhancing LLMs for practical uses while balancing output quality and resource requirements.

By comparing performance in English and Chinese contexts using the Prompt Bench benchmark, **Wang, Ouyang, and Wang (2024)** [50] provide a thorough examination of commercial large language models. Significant performance discrepancies are observed, particularly when processing Chinese text, which they attribute to variations in the diversity of training materials and linguistic complexity. This cross-linguistic assessment underscores the need for substantial enhancements in model designs and training datasets to equip LLMs with more robust and equitable multilingual capabilities.

Ong et al. (2024) [51] discuss the moral and legal implications of using big language models in the medical industry. Their point of view highlights essential challenges, including concerns about data privacy, intellectual property issues, and the opacity of LLM training and adaptation processes. The writers make the case for creating thorough frameworks and flexible governance techniques that can keep pace with the rapid advancements in technology while ensuring that these models are effectively incorporated into clinical practice, as illustrated by the Colling Ridge conundrum.

The transformational potential of big language models in medicine is thoroughly reviewed by **Karabacak and Margetis (2023)** [52], emphasizing applications ranging from clinical decision-making to diagnostic support. The authors emphasize transfer learning, domain-specific fine-tuning, and reinforcement learning using expert input to maximize LLM performance for medical tasks. Additionally, they address critical issues, including model bias, data privacy, and the necessity of interdisciplinary cooperation to ensure that these technologies are used in healthcare settings ethically and effectively.

7. DISCUSSION AND COMPARISON

Table 1: Comparison of all the studies that were reviewed.

Ref	Models and Technologies	Focuses on	Benefit	Challenges	LLM Supply Chain	Outcomes
[33]	LLMs and MLLMs that have already received training (e.g., GPT-4, Gemini, LLaMA)	Outlining the LLM supply chain for downstream applications, foundation models, and infrastructure	gives a thorough study plan and a well-organized definition for the next studies	Reproducibility, security flaws, dependency management, and complex system design	Main goal: a thorough analysis of the interdependencies and layers of the supply chain	identifies important research gaps and suggests a forward-looking research plan
[34]	Integrating LLMs (transformer-based models) with UAV platforms	Using LLM integration to improve UAV autonomy and decision-making	Better autonomous UAV operations and real-time data analysis	Real-time responsiveness, communication security, and integration with sensor data	Indirect: Pay more attention to an application case than the entire supply chain.	Lays forth a plan for upcoming LLM-enabled UAV uses
[35]	Foundation models, such as the GPT series	Putting in place a three-tiered auditing system for LLMs (application, model, and governance)	offers a guide for methodical audits to reduce technical and ethical hazards	Responsibility distribution, moral and legal dilemmas, and emerging behaviors	Not directly: Pay more attention to auditing practices than supply chain management	offers a methodical methodology for auditing to control LLM risks
[36]	LLMs with modular architectures and retrieval-augmented generation (RAG)	Including LLMs in e-government services helps create systems that are repeatable and scalable	Improves the delivery of public services by using automation and better data processing	Ensuring ethical use, scalability, and repeatability in public systems	Not straightforward: prioritizes supply chain specifics before architectural integration	provides a framework for e-government applications built on modular LLM
[37]	Feinberg, the GPT series, and LLM variations unique to finance	Financial reasoning, sentiment analysis, and textual analysis are among the ways that LLMs are used in finance	improves decision-making, tailored recommendations, and financial analysis	Accurate sentiment analysis, real-time processing, and data complexity	Not directly: Showcases real-world financial applications	examines benchmark applications and finds opportunities in
[38]	GPT-3.5/4, ChatGPT, and more	Using LLMs in research, medical education,	Simplifies clinical procedures and increases effectiveness in	Risk of bias, inaccuracy, and privacy issues	Not straightforward: Instead of the supply	gives a summary of the risks and hazards,

	medical LLMs	and clinical decision support	medical environments	with data in clinical settings	chain, the emphasis is on medicinal applications	along with careful advice
[39]	Emerging customized LLM systems and the GPT series	Using LLMs to engage users and personalize recommender systems actively	provides more individualized user interactions and effective information dissemination	Complex personalization, interaction with external technologies, and data privacy management	Not straightforward: emphasizes personalization strategies over supply chain management	explains the opportunities and difficulties of customizing LLM outputs
[40]	GPT-3.5/4, ChatGPT, and more medical LLM versions	Examining LLM applications in medicine with an emphasis on advancements in clinical and research	enhances clinical judgment and expedites research procedures	Limitations of technology, more validation required, and possible clinical hazards	Not directly: Focuses more on medical applications than the supply chain	outlines the possibilities and potential for future advancement in the healthcare industry
[41]	Using LLMs in pedagogical and ethical contexts	Creating review procedures and ethical standards for AI education	encourages equity, openness, and moral principles in AI teaching	Data privacy concerns, algorithmic bias, and inconsistent ethical standards	Not straightforward: Emphasize moral leadership in schooling	offers a multi-tiered ethical framework for using AI in the classroom
[42]	Different LLMs that emphasize bias evaluation, such as GPT and BERT versions	Analyzing the causes, assessment, and reduction of biases in LLMs	gives a thorough analysis of the causes of bias and suggests ways to reduce it to promote justice	Statistical biases, ethical hazards, and inherent biases in training data	Reviewing bias factors across the lifespan instead of just the supply chain is not direct	combining mitigation strategies with assessment procedures to minimize bias
[43]	Advanced LLMs (such as variants of LaMDA and GPT-4)	Examining how well LLMs integrate and perform in software engineering activities	improves code production and review by increasing software engineering automation.	Performance fluctuations, integration difficulties, and limitations in code comprehension	Not straightforward: emphasizes software engineering applications along the whole supply chain	Sorts tasks and provides information about the efficacy and efficiency of LLM in SE
[44]	Constructed on transformer designs, modern LLMs (such as GPT-4)	A thorough examination of the development, possibilities, and drawbacks of	broad knowledge of the development of AI, new developments in technology, and possible uses	High computational demands, bias, ethical considerations, and	Not straightforward: it provides a theoretical summary rather than specifics about	provides a thorough analysis of the potential, constraints, and upcoming

		LLM in contemporary AI		interpretability problems	the supply chain.	difficulties facing LLMs
[45]	Bilingual LLMs (such as ChatGPT and Gemini)	Comparative evaluation of LLMs' cultural alignment in bilingual (Chinese and English) settings	provides information about cultural prejudices and conformity to traditional Chinese norms	Language, cultural differences, moral dilemmas, and situational issues	Not straightforward: Pay more attention to cultural factors than supply chain management	offers guidance and empirical support for creating AI systems that are sensitive to cultural differences
[46]	Document summarization using LLMs (e.g., GPT-4-based architectures)	Automated summary of abstracts and titles of several documents	Improves the effectiveness of knowledge distribution and information retrieval	Issues with readability and consistency of summaries, as well as domain-specific modifications	Not straightforward: emphasizes applications of summarization over the supply chain	Introduces a paradigm for hybrid summarization that combines abstractive and extractive techniques
[47]	The Gemini 1.5 Flash and ChatGPT 4	Comparative analysis of cultural hallucinations in relation to ethical issues that are culturally distinctive	gives information on how LLMs respond to ethical dilemmas and culturally sensitive content.	Cultural hallucinations and the reduction of inappropriate or erroneous reactions	Not direct: Considers cultural and ethical performance rather than the supply chain.	Demonstrates significant performance discrepancies and suggests evaluation metrics for cultural correctness
[48]	ChatGPT and related LLMs for education	Possibilities and difficulties of using LLMs in teaching	Enhances educational content production, individualized learning, and student involvement	demands ongoing human supervision, digital literacy, and bias-reduction techniques	Not directly: Pay more attention to the effects and applications of education than to the supply chain problem	Focuses on fact-checking and critical thinking while offering suggestions for the responsible integration of LLMs in the classroom
[49]	Open-source LLMs (the Mistral-7B, Bloom, FLAN-T5, GPT-Neo, etc.)	Assessing effectiveness in tasks involving language generation and comprehension	finds the advantages and disadvantages of scalable and reasonably priced NLP applications	Scalability, flexibility, and computational efficiency issues across models	Not direct: Focuses more on supply chain elements than performance evaluation	provides suggestions and optimization techniques to increase model efficiency

[50]	Commercial LLMs (such as Claude from Anthropic, Google's Gemini, ChatGPT-4, and ChatGPT-3.5)	Prompt Bench comparison analysis focusing on Chinese and English performance	draws attention to gaps in multilingual performance and guides linguistic inclusion improvements	Significant differences in performance between Chinese and English, as well as language and cultural obstacles	Not straightforward: Multilingual performance evaluation is prioritized over supply chain management	offers guidance for creating more comprehensive and culturally sensitive LLMs
[51]	LLMs used in medical settings (such as ChatGPT variations)	Regulatory and ethical issues facing LLMs in medicine	provides guidelines for responsible adoption in healthcare contexts and creates a framework for ethical integration	Challenges with data privacy, intellectual property, responsibility, and transparency	Not straightforward: emphasizes medical ethics and regulations over supply chains	offers suggestions and methods for resolving moral and legal issues in medical applications.
[52]	Medical LLMs (such as domain-specific modifications like Bio BERT and Clinical BERT, or variations of ChatGPT)	Accepting LLMs for use in medicine with a focus on clinical decision assistance, diagnostic precision, and healthcare integration	enhances access to current medical knowledge, facilitates clinical decision-making, and increases diagnostic accuracy through domain adaptation and fine-tuning	The intricacy of medical terminology, integration difficulties, guaranteeing objective results, data protection, and ethical and regulatory barriers	Not straightforward: prioritizes healthcare adaptation of LLMs above a comprehensive supply chain perspective	offers a thorough analysis along with 10 recommendations for the ethical, responsible, and successful integration of LLMs in medicine

A comprehensive summary of numerous studies examining the creation and application of large language models (LLMs) across various industries is presented in the table. It classifies studies according to the kind of model or technology, from well-known models like GPT-4 and Gemini to specific systems in education, healthcare, and finance, and describes how each work focuses on different facets of the LLM supply chain, downstream applications, and infrastructure. In addition to addressing issues such as technical integration, security, ethical quandaries, and reproducibility, the research highlights several advantages, including improved decision-making, increased scalability, and tailored interactions. Overall, the table not only presents the current state of LLM deployment in both general and specialized contexts, but it also highlights important areas for future research and offers suggestions for enhancing the LLM ecosystem as a whole.

1. Extract statistic

The frequency count provides a comprehensive analysis of the various models and technologies mentioned in the given list. To ensure accuracy, each line was carefully examined to identify specific names and then grouped according to similarities. For example, entries like "Google's Gemini" were combined with "Gemini," and split entries like "GPT-3.5/4" were painstakingly separated to count GPT-3.5 and GPT-4 as distinct entities. Furthermore, generic comments were grouped under broad categories to ensure that every distinct reference was recorded without repetition, for example, when referring to the "GPT series" or just "GPT" in bias evaluation situations. This rigorous approach not only clarifies how often each model or technology is referenced but also highlights the varying degrees of attention given to distinct language model improvements across areas such as education, economics, and medical applications, as illustrated in Figure 1.

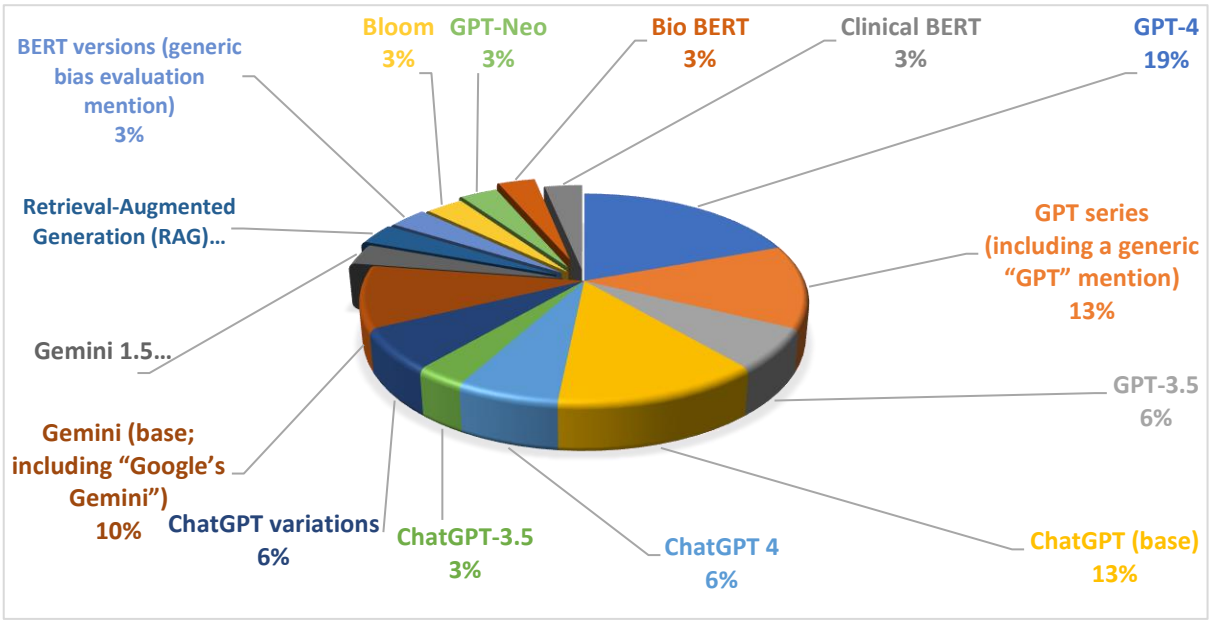

Figure 3: Statical representation about Models and Technologies

The frequency analysis reveals a definite preference for subtle language over direct assertions. The phrase "Not straightforward" occurs eleven times, indicating that the book continuously welcomes intricacy and oblique approaches to its topic. This profusion of qualifiers not only broadens the scope of the discussion but also subtly shifts the emphasis away from a single or overly straightforward understanding of the supply chain. Conversely, the terms "not directly" and "not direct," which appear four and three times, respectively, suggest a moderate propensity to indicate that although the supply chain is an important component, it is not the exclusive topic of conversation. Furthermore, "Main goal" and "Indirect" are mentioned only once, emphasising the overall goals without overburdening the reader with explicit instructions. The careful layering of modifiers highlights a deliberate strategy: by using these diverse phrases, the text encourages a deeper examination of more general themes, such as ethical, cultural, and application-specific nuances, thus enhancing the conversation beyond the specific boundaries of supply chain management, as illustrated in Figure 2.

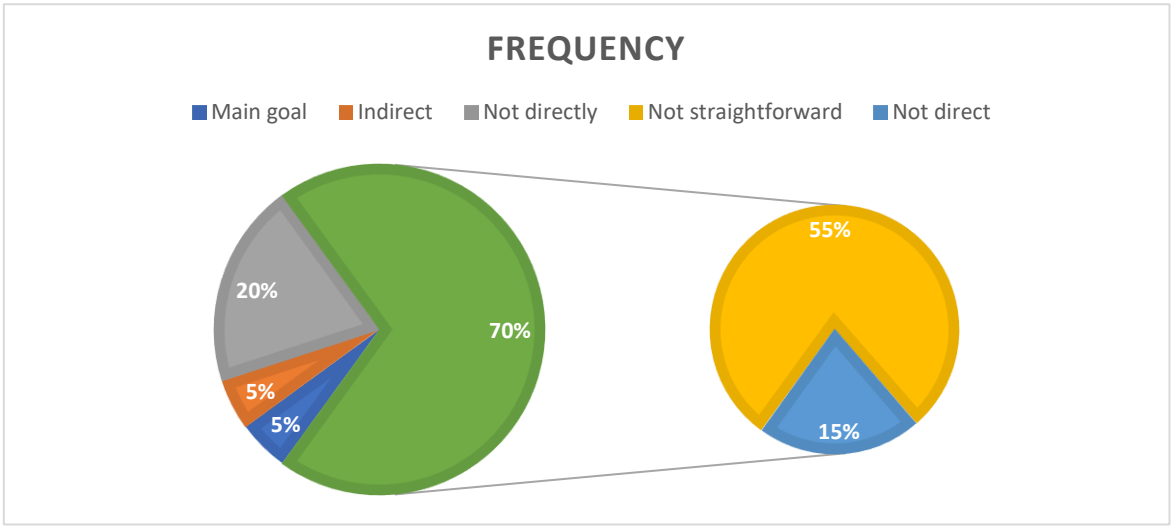

Figure 4: Statical representation about LLM Supply Chain

The distribution chart shows that the verb "offers" occurs most frequently, with six occurrences, highlighting the importance of providing thorough instructions and workable solutions. This recurrence suggests that the overall strategy places equal emphasis on providing practical techniques and advice across various application areas, including evaluating LLM hazards and developing ethical frameworks. Three instances of the verb "provide" follow, demonstrating a definite emphasis on providing thorough analyses, frameworks, and empirical support. The fact that other verbs

like "identifies," "Lays," "examines," "gives," "explains," "outlines," "Sorts," "Introduces," "Demonstrates," and "Focusses" only occur once each demonstrates a well-rounded range of strategies, from identifying research gaps to describing future opportunities and addressing ethical, legal, and cultural considerations. Overall, as shown in Figure 3, the variety of action verbs demonstrates a complex strategy designed to provide both precise insights and rigorous methodologies, ensuring that many facets of LLM integration and the associated challenges are thoroughly addressed.

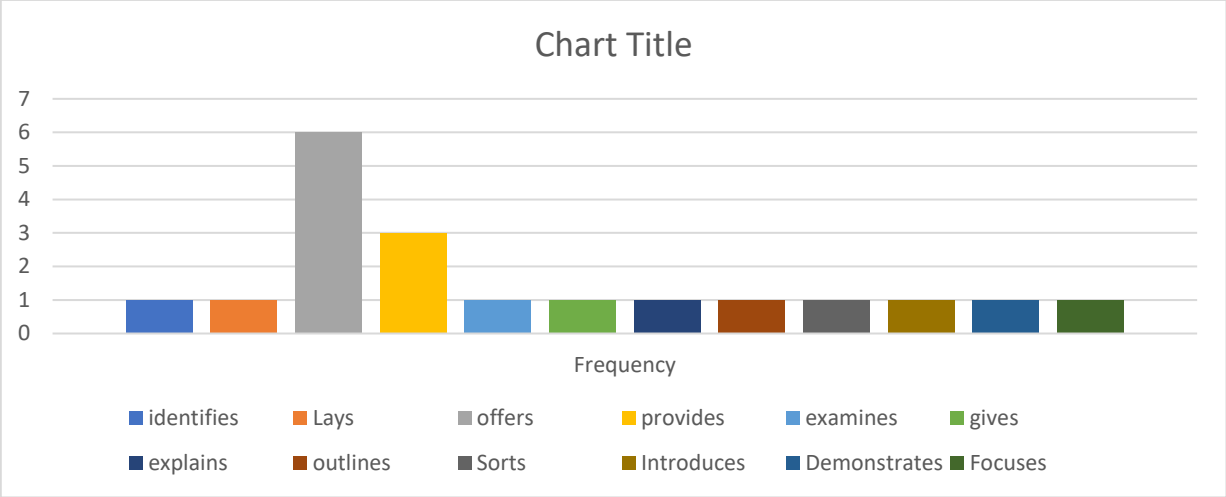

Figure 5: Statical representation about outcome

2. Recommendation

Preserve Formatting and Consistency: PDFs are the best way to make sure that the fonts, graphics, layout, and general appearance of your document are all consistent regardless of the location or browser type. When distributing formal reports, contracts, or marketing materials, this is particularly crucial. You may support the upkeep of professionalism and brand coherence across all platforms by keeping the original design.

Guarantee Accessibility: Enhance the accessibility of your PDFs for diverse users by integrating features like as searchable text, alternative text for images, appropriate headings, and navigational tags. In addition to improving the document's usability for individuals with impairments, this also makes it easier to search inside it. The incorporation of metadata, including document titles, keywords, and authors, enhances categorisation and retrieval.

Boost Security: Make sure that your PDFs are secure by implementing strong security measures at all times. This covers encryption, digital signatures, and password protection. Digital signatures ensure that the content of the document is legitimate and unmodified, while encryption guards against unwanted access. When working with financial records, court documents, or private data, these procedures are essential.

Reduce File Size: Excessive file sizes may cause issues with online sharing, storage, or distribution. Utilise effective compression methods to minimise file size without compromising quality. In addition to ensuring that your document can be shared via email or cloud storage, this optimisation speeds up downloads and uploads, which is especially useful when bandwidth is at a premium.

Facilitate Interactivity: Augment user involvement by integrating interactive components into your PDFs. Features that turn a static document into a dynamic resource include fillable forms, clickable hyperlinks, embedded multimedia, and annotation features. For instructional materials, user manuals, and interactive reports where users gain by interacting directly with the content, interactive PDFs are very helpful.

Promote Archiving and Compliance: When archiving documents for long-term storage, adhere to PDF standards such as PDF/A. The purpose of PDF/A is to guarantee that documents are always available and unchanged, which is crucial for historical, legal, or regulatory records. Following these guidelines will help you stay in compliance with industry rules and protect your papers against potential incompatibilities.

CONCLUSION

In summary, the paper offers a comprehensive overview of large language models (LLMs), tracing their quick development and variety of technological advancements while examining their numerous uses in sectors like public administration, healthcare, education, and finance. It carefully investigates the inner workings of these models, from the broad training datasets and transformer-based structures to sophisticated fine-tuning techniques, and compares these technological developments with the moral, legal, and sociological issues they bring up the study explores issues like algorithmic bias, false information, privacy threats, and environmental effects in great detail. It underscores that, notwithstanding the transformative potential of LLMs in producing human-like language, facilitating nuanced

translations, and improving information retrieval, their implementation necessitates stringent controls. Standardized evaluation metrics, thorough auditing systems, and sustainable processes are recommended in the text to guarantee that LLM technology innovation is in line with the values of openness, accountability, and moral responsibility. Overall, this study provides a clear roadmap for further research and growth, making it a priceless tool for practitioners and scholars. The promise of state-of-the-art language models is balanced with a crucial emphasis on developing safe, effective, and socially conscious applications.

REFERENCES

1. J. A. Goldstein, G. Sastry, M. Musser, R. DiResta, "Generative language models and automated influence operations: Emerging threats and potential mitigations, 2023. [PDF]
2. T. Teubner, C. M. Flath, C. Weinhardt, et al., "Welcome to the era of ChatGPT et al. the prospects of large language models," Business & Information Systems Engineering, vol. 2023, Springer. springer.com
3. C. Ling, X. Zhao, J. Lu, C. Deng, C. Zheng, J. Wang, "Domain specialization as the key to make large language models disruptive: A comprehensive survey,", 2023. [PDF]
4. X. Zhao, J. Lu, C. Deng, C. Zheng, J. Wang, "Beyond One-Model-Fits-All: A Survey of Domain Specialization for Large Language Models," arXiv preprint, 2023. academia.edu.
5. Hasan, D. A., Hussan, B. K., Zeebaree, S. R. M., Ahmed, D. M., Kareem, O. S., & Sadeeq, M. A. M. (2021). The impact of test case generation methods on the software performance: A review. International Journal of Science and Business, 5(6), 33–44. https://doi.org/10.5281/zenodo.4623940
6. S. B. Shah, S. Thapa, A. Acharya, K. Rauniyar, "Navigating the web of disinformation and misinformation: Large language models as double-edged swords," IEEE, 2024. ieee.org
7. D. Barman, Z. Guo, and O. Conlan, "The dark side of language models: Exploring the potential of llms in multimedia disinformation generation and dissemination," Machine Learning with Applications, 2024. sciencedirect.com
8. I. Augenstein, T. Baldwin, M. Cha, et al., "Factuality challenges in the era of large language models and opportunities for fact-checking," *Nature Machine Intelligence*, 2024. google.com
9. Mohammed, S. M., Jacksi, K., & Zeebaree, S. R. M. (2021). A state-of-the-art survey on semantic similarity for document clustering using GloVe and density-based algorithms. Indonesian Journal of Electrical Engineering and Computer Science, 22(1), 552–562. https://doi.org/10.11591/ijeecs.v22.i1.pp552-562.
10. Raaian, M. A. K., Mukta, M. S. H., Fatema, K., and Fahad, N. M., "A review on large language models: Architectures, applications, taxonomies, open issues and challenges," IEEE, 2024. ieee.org.
11. Abdulazeez, A. M., Zeebaree, S. R. M., & Sadeeq, M. A. M. (2018). Design and implementation of electronic student affairs system. Academic Journal of Nawroz University, 7(3). https://doi.org/10.25007/ajnu.v7n3a201.
12. M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects," Authorea, 2023. researchgate.net.
13. Zebari, R. R., Zeebaree, S. R. M., Jacksi, K., & Shukur, H. M. (2019). E-business requirements for flexibility and implementation enterprise system: A review. International Journal of Scientific & Technology Research, 8(11), 655. Retrieved from http://www.ijstr.org.
14. D. Rothman, "Transformers for Natural Language Processing: Build, train, and fine-tune deep neural network architectures for NLP with Python, Hugging Face, and …," 2022. 5.202.73.55.
15. N. Patwardhan, S. Marrone, and C. Sansone, "Transformers in the real world: A survey on nlp applications," Information, 2023. mdpi.com.
16. H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, "A comprehensive overview of large language models,", 2023. [PDF]
17. M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, "A survey on large language models: Applications, challenges, limitations, and practical usage," Authorea, 2023. figshare.com
18. Y. Xie, C. Yu, T. Zhu, J. Bai et al., "Translating natural language to planning goals with large-language models," arXiv preprint arXiv:2302.05128, 2023. [PDF]
19. Y. Wang, Y. S. Lin, R. Huang, J. Wang et al., "Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse …," arXiv preprint arXiv:2405.11505, 2024. [PDF]
20. Z. Tan and M. Jiang, "User modeling in the era of large language models: Current research and future directions," arXiv preprint arXiv:2312.11518, 2023. [PDF]
21. R. M. Samant, M. R. Bachute, S. Gite, and K. Kotecha, "Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature

review and future directions," IEEE Access, 2022. ieee.org
22. P. Liu, Y. Ren, J. Tao, and Z. Ren, "Git-mol: A multi-modal large language model for molecular science with graph, image, and text," Computers in biology and medicine, 2024. [PDF].
23. Ibrahim, R. K., Zeebaree, S. R. M., & Jacksi, K. F. S. (2019). Survey on semantic similarity based on document clustering. Advances in Science Technology and Engineering Systems Journal, 115. https://doi.org/10.25046/aj040515
24. D. M. Petroșanu, A. Pîrjan, and A. Tăbușcă, "Tracing the influence of large language models across the most impactful scientific works," Electronics, 2023. mdpi.com
25. S. Ateia and U. Kruschwitz, "Can open-source LLMs compete with commercial models? Exploring the few-shot performance of current GPT models in biomedical tasks," arXiv preprint arXiv:2407.13511, 2024. [PDF]
26. Z. Liu, A. Zhang, H. Fei, E. Zhang, and X. Wang, "Prott3: Protein-to-text generation for text-based protein understanding,", 2024. [PDF]
27. E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 2023, Elsevier. osf.io
28. X. Wu, R. Duan, and J. Ni, "Unveiling security, privacy, and ethical concerns of ChatGPT," Journal of Information and Intelligence, 2024. sciencedirect.com.
29. Jacksi, K., Dimililer, N., & Zeebaree, S. R. M. (2015). A survey of exploratory search systems based on LOD resources. In Proceedings of the 5th International Conference on Computing and Informatics (ICOCI 2015) (Paper No. 112). Istanbul, Turkey: Universiti Utara Malaysia.
30. O. Oviedo-Trespalacios, A. E. Peden, T. Cole-Hunter, et al., "The risks of using ChatGPT to obtain common safety-related information and advice," Safety Science, vol. 2023, Elsevier. sciencedirect.com
31. S. Tan, B. Knott, Y. Tian, and D. J. Wu, "CryptGPU: Fast privacy-preserving machine learning on the GPU," in *2021 IEEE Symposium on …*, 2021. [PDF]
32. Y. Shen, L. Heacock, J. Elias, K. D. Hentel, B. Reig, and G. Shih, "ChatGPT and other large language models are double-edged swords," Radiology, 2023. rsna.org.
33. Wang, S., Zhao, Y., Hou, X., & Wang, H. (2024). Large Language Model Supply Chain: A Research Agenda. ACM Transactions on Software Engineering and Methodology. https://doi.org/10.1145/3708531.
34. Javaid, S., Fahim, H., He, B., & Saeed, N. (2024). Large Language Models for UAVs: Current State and Pathways to the Future. IEEE Open Journal of Vehicular Technology. https://doi.org/10.1109/OJVT.2024.3446799.
35. Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing Large Language Models: A Three-Layered Approach. AI and Ethics, 4, 1085–1115. https://doi.org/10.1007/s43681-023-00289-2.
36. Papageorgiou, G., Sarlis, V., Maragoudakis, M., & Tjortjis, C. (2024). Enhancing E-Government Services through State-of-the-Art, Modular, and Reproducible Architecture over Large Language Models. Applied Sciences, 14(8259). https://doi.org/10.3390/app14188259.
37. Kong, Y., Nie, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. (2024). Large Language Models for Financial and Investment Management: Applications and Benchmarks. The Journal of Portfolio Management.
38. Omiye, J. A., Gui, H., Rezaei, S. J., Zou, J., & Daneshjou, R. (2022). Large Language Models in Medicine: The Potentials and Pitfalls.\
39. Chen, J., Liu, Z., Huang, X., Wu, C., Liu, Q., Jiang, G., Pu, Y., Lei, Y., Chen, X., Wang, X., Zheng, K., Lian, D., & Chen, E. (2024). When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities. World Wide Web, 27(42).https://doi.org/10.1007/s11280-024-01276-1.
40. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2024). Large Language Models in Medicine: Current Potential and Opportunities for Development.
41. Yan, Y., & Liu, H. (2024). Ethical Framework for AI Education Based on Large Language Models. Education and Information Technologies. https://doi.org/10.1007/s10639-024-13241-6.
42. Guo, Y., Guo, M., Su, J., Yang, Z., Zhu, M., Li, H., Qiu, M., & Liu, S. (2024). Bias in large language models: Origin, evaluation, and mitigation. arXiv. http://arxiv.org/abs/2411.10915v1.
43. Zheng, Z., Ning, K., Zhong, Q., Chen, J., Chen, W., Guo, L., Wang, W., & Wang, Y. (2024). Towards an understanding of large language models in software engineering tasks [Preprint]. arXiv. http://arxiv.org/abs/2308.11396.
44. Desai, B., Patil, K., Patil, A., & Mehta, I. (2023). Large language models: A comprehensive exploration of modern AI's potential and pitfalls. Journal of Innovative Technologies, 6. Retrieved from https://academicpinnacle.com.

45. Yuan, X., Hu, J., & Zhang, Q. (2023). A comparative analysis of cultural alignment in large language models in bilingual contexts.
46. Langston, O., & Ashford, B. (2024, August 02). Automated summarization of multiple document abstracts and contents using large language models. https://doi.org/10.36227/techrxiv.17262754.45577350/v1.
47. Zhao, J., Huang, C., & Li, X. (2024). A comparative study of cultural hallucination in large language models on culturally specific ethical questions [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-4566507/v1.
48. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2024). ChatGPT for good? On opportunities and challenges of large language models for education.
49. Wong, S. M., Leung, H., & Wong, K. Y. (2024). Efficiency in language understanding and generation: An evaluation of four open-source large language models [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-4063228/v1.
50. Wang, S., Ouyang, Q., & Wang, B. (2024). Comparative evaluation of commercial large language models on PromptBench: An English and Chinese perspective [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-3987793/v1.
51. Ong, J. C. L., Chang, S. Y.-H., Liu, N., Ting, D. S. W., Chew, L. S. T., et al. (2024). Ethical and Regulatory Challenges of Large Language Models in Medicine. Lancet Digital Health, 6, e428–e432. https://doi.org/10.1016/S2589-7500(24)00061-X.
52. Karabacak, M., & Margetis, K. (2023). Embracing large language models for medical applications: Opportunities and challenges. Cureus, 15(5), e39305. https://doi.org/10.7759/cureus.39305.