scientific reports

Check for updates

OPEN

# Industrial applications of large language models

Mubashar Raza[1], Zarmina Jahangir[2], Muhammad Bilal Riaz[4,5]✉, Muhammad Jasim Saeed[2] & Muhammad Awais Sattar[3]

**Large language models (LLMs) are artificial intelligence (AI) based computational models designed to understand and generate human like text. With billions of training parameters, LLMs excel in identifying intricate language patterns, enabling remarkable performance across a variety of natural language processing (NLP) tasks. After the introduction of transformer architectures, they are impacting the industry with their text generation capabilities. LLMs play an innovative role across various industries by automating NLP tasks. In healthcare, they assist in diagnosing diseases, personalizing treatment plans, and managing patient data. LLMs provide predictive maintenance in automotive industry. LLMs provide recommendation systems, and consumer behavior analyzers. LLMs facilitates researchers and offer personalized learning experiences in education. In finance and banking, LLMs are used for fraud detection, customer service automation, and risk management. LLMs are driving significant advancements across the industries by automating tasks, improving accuracy, and providing deeper insights. Despite these advancements, LLMs face challenges such as ethical concerns, biases in training data, and significant computational resource requirements, which must be addressed to ensure impartial and sustainable deployment. This study provides a comprehensive analysis of LLMs, their evolution, and their diverse applications across industries, offering researchers valuable insights into their transformative potential and the accompanying limitations.**

**Keywords** Large Language models, LLMs, NLP, Transformers

Communication has vital importance in each and every aspect of human life. There are a number of languages that are used by humans to communicate with each other. In this era of advanced and innovative technologies. We required efficient ways to communicate with machines[1,2]. Natural language processing (NLP) is a fragment of AI that provides mechanism for humans to interact with computers and machines[3–5]. NLP involves enabling computers to understand, interpret, and generate human language in a way that is both meaningful and useful. NLP combines computational linguistics with AI techniques to process and analyze large amounts of natural language data. NLP encompasses a variety of tasks, including text analysis, extraction of meaningful information, machine translation for converting text between languages, speech recognition to convert spoken language into text, and text generation for producing human-like text[6,7]. Machines do not posse the ability to understand and generate content in human languages. AI enables them to understand and generate content in human language[8]. To give the machines with human like ability to read, write and generate the content is a scientific challenge[9].

Advancements in AI, NLP, and availability of immense amount of training data, contributed to the evolution of LLMs. LLMs are fragment of language models (LMs) that uses neural networks, immense number of training parameters, and unlabeled text[10]. They have the self-supervised learning approach which allows them to train on huge amount of unlabeled data[11]. This approach provides them advantages over supervised learning models, as the data does not require to be labelled manually. Most of the LLMs are built on transformer architectures[12]. Transformer is advanced architecture based on neural networks. Due to the presence of self-attention mechanism, the transformer has the ability to better understand the connection between different input variables and parameters[13]. LLMs work on two stage training pipeline which enhances their learning efficiency. The first training stag is pre training and the second training stage of pipeline is fine tuning[14]. In the first stage they are trained on immense amount of unlabeled data using self-supervised training approach. In the second stage they are trained on specific and labelled data. Combination of these two stages enable them to provide high accuracy[15].

[1]Department of Computer Science, COMSATS University, Sahiwal Campus, Islamabad, Pakistan. [2]Department of Computer Science, Riphah International University, Lahore Campus, Lahore, Pakistan. [3]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden. [4]IT4Innovations, VSB – Technical University of Ostrava, Ostrava, Czech Republic. [5]Applied Science Research Center, Applied Science Private University, Amman, Jordan. ✉email: muhammad.bilal.riaz@vsb.cz; bilalsehole@gmail.com

LLMs have their foundations in the early language models and neural networks. The early language models were developed using statistical models and n-gram models[16,17]. The early language models were failed to express the long terms and context in languages[18]. After the availability of large data sets, the researchers started to explore neural networks in more detail in such a way that they could help to improve language models. Finally the researchers achieved their millstone in the form of Recurrent Neural Networks (RNNs) and LSTMs[19]. The RNNs were able to model the sequential data in languages but they also have limitations in term of long-term dependencies. The LLMs were started to emerge after the introduction of transformer architecture[20]. The transformer architecture is a efficient in handling long-term dependencies[21]. Today's advanced language models including Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT), mT5, RoBERTa, XLNet and LLaMA are based on LLM's and transformer architectures[22–24].

The LLMs use pipeline in their architecture[25]. In the first stage of pipeline the training data is preprocessed. The second stage of pipeline consists of model training where data passes through different steps including random parameter initialization, numerical data input, loss function calculation, parameter optimization, and iterative training. After the training stage the model can be tested and deployed. Research shows that LLMs have the to perform potential outstanding in specialized NLP tasks and applications in specific domains. LLMs also performed well in different domains of industry including automotive, e-commerce, education, finance and banking, health care and medicine[26–31]. In the automotive industry, LLMs enhance in-car virtual assistants, enabling voice commands and real-time translation services, improving driver safety and convenience. E-commerce platforms leverage LLMs for personalized shopping experiences, optimizing search results, product recommendations, and customer service interactions using chatbots. In education, LLMs facilitate personalized learning, automated grading, and intelligent tutoring systems, making education more effective. The finance and banking industry benefits from LLMs through advanced fraud detection and risk assessment. Healthcare and medicine utilize LLMs for predictive diagnostics, patient data analysis, and leading to improved patient outcomes and operational efficiencies. Across these domains of industries, LLMs drive innovation by automating tasks, enhancing user experiences, and providing deep insights.

## Background of LLMs

The concept of machine translation and language models emerged after the Alan Turing's 1950 paper[32]. Alan Turing's paper is a foundation of AI and language processing. From 1950 to 1980 statistical methods and rule-based systems were explored for language models and language processing[33]. Those early models and language processing methods had limitations in case of complex language processing tasks. In 1986 RNNs were introduced[34]. RNNs are based on neural networks[35]. RNNs are capable of processing language data but they have limitations when it comes to long range dependencies in languages[36]. In 1997 Long Short-Term Memory (LSTM) were introduced. They addressed the limitations of RNNs[19,37]. LSTMs have limitations related to word embeddings, and representations of words as numerical vectors[38]. In 1997, Bidirectional Long Short-Term Memory (BiLSTM) networks were introduced, extending the functionality of LSTMs by processing input sequences in both forward and backward directions. While traditional LSTMs handle long-term dependencies only in a unidirectional manner, BiLSTMs overcome this limitation by capturing contextual information from both past and future contexts simultaneously[39]. After 2010 deep learning revolutionized NLP, with models like Word2Vec and GloVe creating powerful word embeddings[40,41]. Neural machine translation (NMT) emerged, surpassing traditional statistical methods.

Word2Vec is a word embedding technique in NLP that generates dense vector representations for words, capturing their meanings and relationships based on contextual usage[42]. It was developed in 2013, by a team led by Tomas Mikolov at Google[43]. Word2Vec employs two main architectures: Continuous Bag of Words (CBOW) and Skip-gram[44]. CBOW predicts a target word from surrounding context words, while Skip-gram predicts context words from a target word. By using techniques like negative sampling and hierarchical softmax, Word2Vec efficiently learns these word embeddings from large text corpora[45]. These embeddings place semantically similar words close to each other in the vector space, enabling applications such as word similarity measurement, clustering, analogies, and as features in various machine learning models[46]. Word2Vec has significantly advanced the field of NLP by providing a robust and scalable way to understand and manipulate textual data.

Global Vectors for Word Representation (GloVe), is a powerful word embedding technique used in NLP to capture the semantic relationships between words[47]. It was introduced in 2014. IT was developed by researchers at Stanford University, GloVe differs from Word2Vec by leveraging both local context information and global statistical information from a corpus[36]. It constructs a co-occurrence matrix, where each entry represents how frequently a pair of words appears together within a certain context window[48]. By factorizing this matrix, GloVe generates word vectors that encode meaningful semantic relationships, ensuring that similar words are placed closer together in the vector space[49]. This approach combines the strengths of traditional count-based methods and predictive models, resulting in embeddings that perform well on a variety of NLP tasks, such as machine translation.

In 2015, Google introduced the initial LLM that uses deep learning algorithms[50]. It was referred as Google Neural Machine Translation (GNMT) model. This model uses huge amount of training data during training. As compared to previous language models it was able to handle complex NLP tasks. In 2017, transformer architecture was introduced[51]. Transformer architecture played an vital role in the development of LLMs such as BERT[52]. The main objective behind the development of the transformer models was to overcome the limitations of earlier models such as RNNs and LSTM[53]. Transformer models are able to capture the long-term dependencies in text. In 2018, Google introduced BERT[54]. Introduction of BERT was a vital advancement in NLP. The BERT is a pre-trained model and it can be fine-tunned on specific domain of NLP. OpenAI introduced GPT model in 2018 [55]. It was based on transformer architecture. The first version of GPT is called GPT-1. In 2019, GPT-2 was

introduced[56]. GPT-2 consists of 1.5 billion parameters[57]. In 2020, GPT-3 was introduced and in 2023 GPT-4 was introduced[58].

## Motivation

Although researchers have discussed and covered the applications of LLM's in various industries but the previous studies have limitations. These studies did not cover the many important aspects of LLM's in industries including domain specific applications, modern architectures, security, privacy and ethics. Most of the studies covered one or two domains of industry. They did not cover applications of LLMs in top industries and their comparison. Most of the studies are not peer reviewed research works[8,28,59–62]. Absence of these key points motivated the authors to write this review articles. This review article extensively explores the current review articles to identify their limitations and cover their gaps. The objective of this study is to cover the modern architectures of LLMs, applications of LLMs in top industries and to address the issues of security, privacy, and ethics related to the use of LLMs in industries. In this review article we focused on top industries including finance and banking, healthcare and medicine, education, ecommerce, and automotive. Table 1 summarizes the comparison of this study with previous studies.

## Main contribution

- Providing an extensive overview of LLMs including their background, evaluation, modern architectures, and their applications in top industries.
- Describing the applications of LLMs in top industries including finance and banking, healthcare and medicine, education, ecommerce, and automotive.
- Describing the open issues and challenges of LLMs related to data security, privacy and ethics.
- Describing the important aspects of LLMs.
- Describing contemporary LLMs and their architectures.
- Investigating the evaluation metrics for LLMs in detail.

| Studies | Security and Privacy | Ethics | Scope | Approach |
|---|---|---|---|---|
| Can et al., [27] | No | No | Investigates the Multimodal LLMs in autonomous driving and related systems. | Review of existing literature and tools related to LLMs in autonomous driving systems. |
| Zijian et al., [61] | No | No | Investigates the LLMs in mobility forecasting within transportation systems. | Review of existing literature related to LLMs in transportation systems. |
| Dingkai et al., [63] | No | No | Explores the applications of LLMs in intelligent transportation systems. | Review of existing literature related to LLMs in intelligent transportation systems. |
| Qingyang et al., [28] | No | Yes | Examines the applications of LLMs in e-commerce. | Review of existing literature related to LLMs in e-commerce. |
| Xiaonan et al., [64] | No | No | Investigates the LLMs based recommendations in e-commerce. | Review of existing literature related to LLMs in e-commerce. |
| Jin et al., [65] | No | No | Examines the LLM based personalization systems to improve customer experience. | Review of existing literature related to LLM based personalization systems. |
| Shen et al., [62] | No | No | Investigates the LLM based technologies in educational settings. | Survey of existing literature related to applications of LLMs in education. |
| Hanyi et al., [66] | No | No | Investigates the use of LLMs for smart education. | Survey of existing literature related to applications of LLMs in smart education. |
| Stefan et al., [67] | No | No | Explores the potential of LLMs in education, and how their challenges might be addressed through game-based learning. | Review of existing literature related to LLMs and their challenges in education. |
| Lixiang et al., [68] | No | Yes | Explore the uses of LLMs for automating educational tasks, while addressing the associated challenges. | Review of existing literature related to LLMs in education. |
| Nadia et al., [69] | No | No | Examines the impact of ChatGPT on education by focusing on its influence on the teaching. | Review of existing literature related to impact of ChatGPT on education. |
| Jean et al., [30] | No | No | Investigates the applications of financial LLMs in finance domain. | Review of existing literature related to financial LLMs. |
| Yinheng et al., [60] | No | No | Investigates the applications of LLMs in finance domain. | Review of existing literature related to LLMs in finance. |
| Huaqin et al., [70] | No | No | Investigates the applications of LLMs in finance domain. | Review of existing literature related to LLMs in finance. |
| Godwin et al., [71] | No | No | Investigates the applications of LLMs in banking industry. | Review of existing literature related to LLMs in banking. |
| Yining et al., [73] | No | Yes | Examines the application of LLMs in the various healthcare domains. | Review of existing literature related to LLMs in medical industry. |
| Yanxin et al., [74] | No | No | Investigates the applications of LLMs in the medical industry. | Review of existing literature related to LLMs in medical industry. |
| Ping et al., [75] | No | No | Explore the integration of generative AI and LLMs into healthcare and medical practices. | Review of existing literature related to LLMs in medical industry. |
| This Study | Yes | Yes | Investigates the applications of LLMs in industries including healthcare, medicine, automotive, e-commerce, education, finance and banking. | Detailed review on Industrial Applications of LLMs. |

**Table 1. Comparison with previous studies.**

The remaining sections of this study have been organized as: The section II covers and discusses the literature, section III covers the methodology, section IV discusses the LLMs in detail, section V discusses the domain specific applications, section VI Case studies and empirical evidence, section VII discusses the issues and challenges, section VIII Ethical considerations and responsible deployment, section IX covers the future directions of LLMs, section X discusses the limitations of this study and section XI covers the conclusion.

## Literature review

In last decade LLMs have become evolution in AI. The number of advancements and improvements in LLMs are growing day by day. The applications of LLMs in industries are also increasing in an unpredictable manner. Many research works have been conducted in the past to explore the LLMs and their applications in the industries.

Can et al.,[27] investigate multimodal LLMs for autonomous driving. The authors cover the background of multimodal LLMs and the development of multimodal using LLMs. The study also covers the background of autonomous driving. The study investigates the role of multimodal LLMs in transportation and driving. Zijian et al.,[61] explore the applications and role of LLMs in transportation system. The study examines how LLMs improves the transportation system by forecasting the traffic information. The study examines how LLMs are helpful in handling the demands and limitations of transportation system. The study also explores the utilization of LLMs in predicting the human travel. The study highlights if we use LLMs in transportation system then we can predict the human travel and manage the demands of transport. In this way the LLMs are also helpful to manage the urban planning. Dingkai et al.,[63] investigate the applications of LLMs in autonomous vehicles, traffic management, and transportation safety. The study also discusses the limitations and advantages of LLMs in traffic management and autonomous driving. The study explores some datasets used to train LLMs for traffic management and autonomous driving. The research delves into the development of LLMs for the said domains and fields.

Qingyang et al.,[28] present the fairness, applications, and challenges faced by LLMs in e-commerce industry. The study reveals that LLMs offer innovative solutions in e-commerce industry and they enhance the customer experience. The research work discusses the pretraining, finetuning, and prompting of LLMs. The study covers the role of LLMs in product reviews, customer support and product recommendations. The research explores the broad applications of LLMs in e-commerce. Xiaonan et al.,[64] introduce applications of LLMs in e-commerce in the form of recommendation systems. They review the latest advances in LLM techniques for recommendation systems. Additionally, the authors provide a comprehensive discussion on the future directions of LLM-driven recommendation systems. This study addresses the urgent need for a deeper understanding of LLM-driven recommendation systems due to the rapid development in this research area. Jin et al.,[65] discuss the applications of LLMs in human computer interaction, personalization systems and recommendation system. The study investigates the impact of LLMs for improving customer experience.

Shen et al.,[62] investigate the different technologies and tools of LLMs in education. The paper investigates the tools that are related to students and teachers. The study also discusses the technological advancements in LLMs related to education. The authors discuss the risks associated with the use of LLMs in education. Their research provides comprehensive study of LLMs in education. Hanyi et al.,[66] investigate the role of LLMs in education. The study summarizes the role of LLMs in improving teaching methodology and education models. The study then discusses the integration of LLMs in education. Stefan et al.,[67] investigate the applications of LLMs in education. The research addresses the challenges faced by LLMs in education. The author discusses that playful and game-based learning can solve those challenges of LLMs. The study also discusses the generative AI in education. Lixiang et al.,[68] investigate the practical challenges of LLMs in education. They discuss the ethical concerns of using LLMs for grading, feedback, and question generation. The study addresses that these ethical concerns are obstacles for LLMs in education. Nadia et al.,[69] discusses the impact of ChatGPT in education and teaching process. The study summarizes the research conducted after the introduction of ChatGPT. The study discusses that the impact of ChatGPT is positive but it is critical for teachers and students.

Jean et al.,[30] investigate the applications of LLMs in finance. The study discusses the performance, history, and techniques of LLMs in finance. They discuss the training data, finetuning methods and datasets of LLMs in finance. Yinheng et al.,[60] discusses the current techniques used with LLMs in finance. The study covers the pre training, zero shot learning, custom training, and pre training of LLMs in finance. The research emphasizes the decision framework for professionals of finance to select appropriate LLMs. Huaqin et al.,[70] investigated the applications of LLMs in financial domain. The study discusses that the use of LLMs is increasing gradually in finance. They claim that professionals are using LLMs for financial report generation, analyzing investor sentiments, and forecasting the trends of market. Godwin et al.,[71] discuss the applications of LLMs in banking industry. The study discusses the role of LLMs in banking domain. The study focuses on the text-based communication, and personalized interactions. The study also investigates the role of LLMs in customer support, automation of tasks, and decision-making process. Christian et al.,[72] investigate the LLMs for financial advice. They claim that larger models are better as compare to models trained on average size of datasets. The research delves into usefulness of LLMs for financial advice.

Yining et al.,[73] summarize the applications of LLMs in the medical industry. The cited study covers medical text data processing, public health awareness, and clinical settings of LLMs. The study also covers information extraction, summarization and question answering techniques related to LLMs. Yanxin et al.,[74] investigate development of LLMs for medicine industry. The study explores the techniques used in LLMs for medicine domain. The study also discusses the directions for the integration of LLMs in medicine industry. Ping et al.,[75] cover the role of generative AI and LLMs in healthcare industry. They investigate the integration of generative AI and LLMs in healthcare. The study focuses on the benefits of LLMs in healthcare including decision making process, information retrieval and medical data management. The study also compares the LLMs in healthcare with the typical rule-based AI systems and other machine learning models. Marco et al.,[31]

summarize the applications of LLMs in healthcare including biomedical NLP, literature summarization, and clinical documentation management.

### Limitations and drawbacks of existing studies

This section examines the limitations and gaps in existing research on LLMs. It highlights the areas where previous studies fall short. Some existing research works do not address an evaluation of LLMS, their modern architectures, and their applications in top industries, for example[27,61,63,64]. Whereas this research work provides an extensive overview of LLMs including their background, evaluation, modern architectures, and their applications in top industries including finance and banking, healthcare and medicine, education, ecommerce, and automotive. The research works do not investigate the critical open issues and challenges of LLMs related to data security, privacy and ethics[62,66,67,69]. Whereas this research work thoroughly investigates and highlights the open issues and challenges of LLMs related to data security, privacy and ethics. A large number of studies do not delve into contemporary LLMs and their architectures[30,70,71,73]. Whereas this research analyzes and investigates the contemporary LLMs and their architectures. The studies cited in this review paper are industry-specific and cover only one or two industries[69,72,76–78], whereas this review focuses on the applications, issues, and challenges of LLMs in the top and most important industries.

## Methodology

The research material cited in this study have been acquired from well-known and recognized scientific journals and conferences from 2020 to 2024. The research articles have been searched from well-known research platforms including IEEE Xplore, ACM Digital Library, Google Scholar, ScienceDirect, Springer. Initially more than 300 papers were selected related to keywords. After the initial selection, a comprehensive study of papers has been conducted and we finalized more than 100 research articles. The final research articles have selected based on keywords, topic and industrial domains. To conduct a comprehensive search, the main keywords used are "LLMs", "Natural Language processing", "Deep Learning" and "machine learning". The combinations of these keywords and some other keywords specific to different domains of industries are used to compile material for this study. These keywords helped to find the relevant articles for this study. The extensive search has been conducted to find the relevant and quality articles. Table 2 shows the details of keywords and their combinations which are used to conduct the literature search for this study.

## Large Language models

LLMs models are combination of AI algorithms and NLP techniques. They are powerful tools which are used in various industries to enhance language related applications. They generate human like text based on the context

| Sr #. | Keywords and their Combinations | Domain |
|---|---|---|
| 1 | Large Language Models in Healthcare | Healthcare |
| 2 | LLMs for medical diagnosis | Healthcare |
| 3 | Natural Language Processing in Healthcare | Healthcare |
| 4 | Machine Learning and LLMs in healthcare | Healthcare |
| 5 | Deep Learning and LLMs in healthcare | Healthcare |
| 6 | Large Language Models in Automotive Industry | Automotive |
| 7 | LLMs for autonomous vehicles | Automotive |
| 8 | Natural Language Processing in automotive | Automotive |
| 9 | Machine Learning and LLMs in vehicle | Automotive |
| 10 | Deep Learning and LLMs in vehicle | Automotive |
| 11 | Large Language Models in E-commerce | E-commerce |
| 12 | Natural Language Processing in E-commerce | E-commerce |
| 13 | Machine Learning and LLMs in customer service | E-commerce |
| 14 | Deep Learning and LLMs in customer service | E-commerce |
| 15 | Large Language Models in Education | Education |
| 16 | Machine Learning and LLMs in Education | Education |
| 17 | Natural Language Processing in education | Education |
| 18 | Deep Learning and LLMs in Education | Education |
| 19 | Large Language Models in Finance | Finance |
| 20 | LLMs for financial forecasting | Finance |
| 21 | Machine Learning and LLMs in financial analysis | Finance |
| 22 | Deep Learning and LLMs in financial analysis | Finance |
| 23 | Natural Language Processing in banking | Banking |
| 24 | Machine Learning and LLMs in banking | Banking |
| 25 | Deep Learning and LLMs in banking | Banking |

**Table 2. Keywords and their combinations used to search literature.**

user inputs. They have improved the chatbots, content generation platforms, virtual assistants and customer service automation.

## Important aspects of Llms

LLMs have very complex architecture and they have various aspects depending upon the category of model. Figure 1 organizes the important aspects of LLMs. In the subsection, the important aspects of LLMs have been covered.

*Attention mechanism*

The attention mechanism is a vital part of LLMs[79]. The attention mechanism is used to find the representation of input sequences connecting different tokens[79–81]. The main idea is to calculate a set of attention weights that determine the importance of each part of the input data in relation to each part of the output data. There are various attention mechanisms used in LLMs including self-attention, multi-head attention, and positional encoding[82].

*Training methods*

Different training methods are used for the model training of machine learning models but LLMs are trained using distributed methodologies. The reason is that LLMs require huge amount of data and computational power[83]. Distributed methodologies of LLMs are model parallelism, data parallelism, optimizer parallelism, tensor parallelism, pipeline parallelism, and federated learning[84,85].

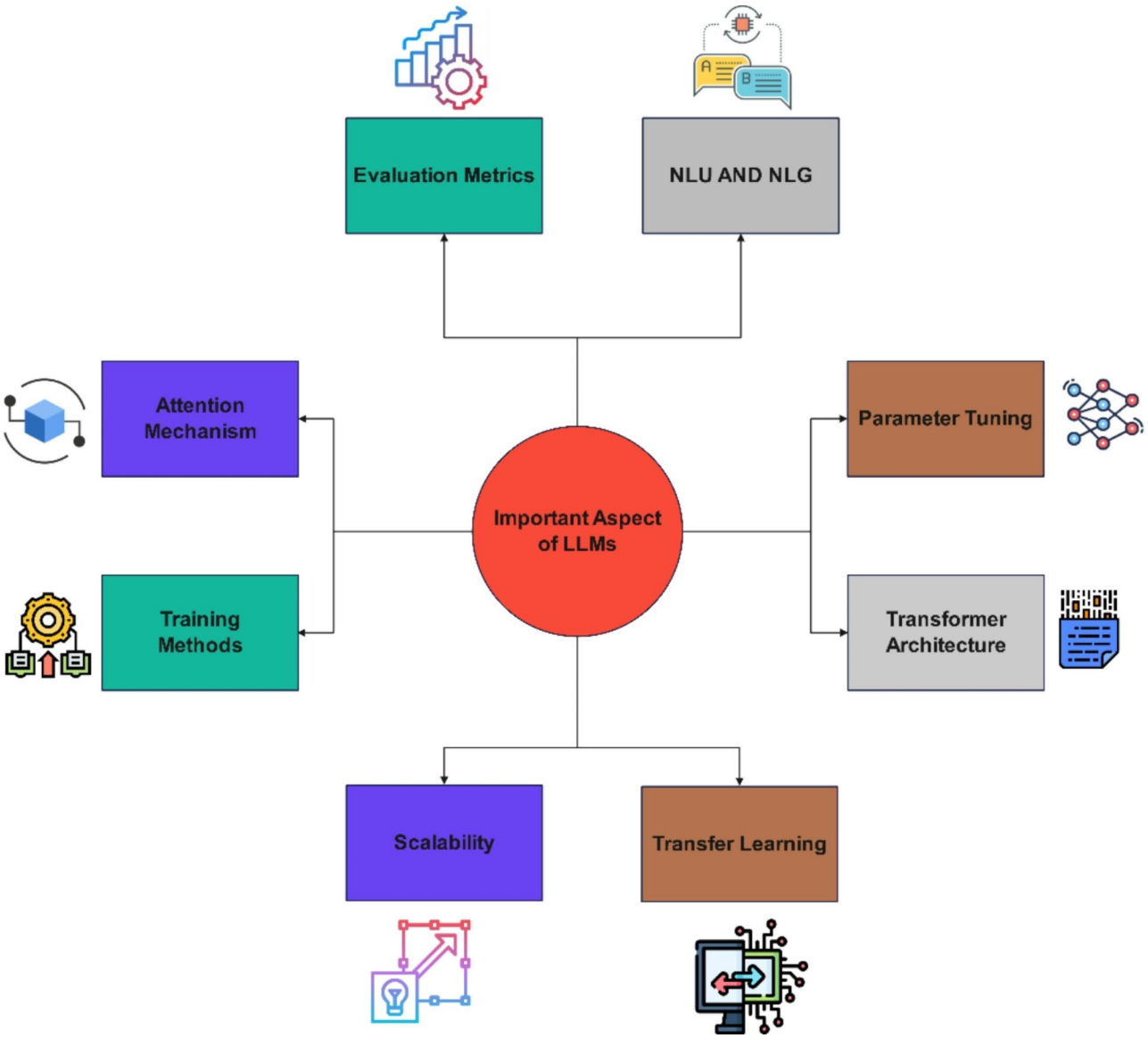


**Fig. 1**. **Important Aspects of LLMs**.

*Parameter tuning*
The LLMs are first pre-trained on huge datasets and then they can be fine-tuned for specific applications[86]. This approach of training provides us customized models. Customized models are highly efficient and accurate[87]. There are various parameter tuning techniques used to fine tune the LLMs including prompt tuning, prefix tuning, adapter tuning, parameter efficient fine tuning, sparse fine tuning, and parameter sharing[88–90]. These kinds of techniques optimize the performance without requiring extensive computational resources and large amount of data[91]. Parameter tuning techniques allow the LLMs to be deployed in resource constrained environment[92].

*Transformer architecture*
Transformer architecture is a foundation of various NLP models. The Transformer architecture is a deep learning model introduced by Vaswani et al.,[93] in 2017. The main purpose of transformer was to handle sequential data in neural networks[94]. The key components of transformer architecture are input embedding, positional encoding, encoder, decoder, and attention mechanism. Transformer architecture uses attention mechanism to find dependencies[95,96]. The architecture can handle inputs of varying lengths. Due to its efficiency and versatility, this architecture has been replaced typical neural networks[97]. Transformer architecture has been specifically used in LLMs.

*Evaluation metrics*
To assess the performance of LLMs, various metrics are used depending on the specific task.

i. **Intrinsic evaluation metrics**: These metrics measure the model's linguistic and semantic capabilities without relying on downstream tasks. It Evaluates the fluency of the language model[98].

- **Perplexity**: Measures how well a language model predicts a test set. A lower perplexity shows better performance. It evaluates the fluency of the language models[99,100].
- **Bleu (bilingual evaluation understudy)**: Compares the overlap between n-grams of the generated and reference texts. It is common in machine translation[101].
- **Meteor (metric for evaluation of translation with explicit ordering)**: It was introduced as an Improvement upon BLEU by considering synonyms, stemming, and word order[102].
- **Bertscore**: Uses contextual embeddings to compute semantic similarity between generated and reference text[103].

ii. **Human evaluation metrics**: Human evaluation metrics are used for assessing and comparing how LLMs perform on evaluation sets. It is a way to assess the performance of LLMs by asking people to judge the model's output. Human evaluations are often combined with automated metrics to provide a more comprehensive view of the model's performance[104].

- **Fluency**: Evaluates how naturally and smoothly the model's output adheres to linguistic norms. It measures the model's ability to produce grammatically correct, coherent, and contextually appropriate sentences, making the output easy to read and understand[105].
- **Coherence**: Refers to a quantitative or qualitative measure of how logically consistent, contextually appropriate, and smoothly connected the output of the model is within a given text. Coherence evaluates whether the generated content aligns with the input prompt and flows logically without contradictions, abrupt topic shifts, or incoherent phrasing[106].
- **Relevance**: A measure of how closely the model's outputs align with the user's query. It evaluates the appropriateness, accuracy, and contextual suitability of the generated response[107].
- **Bias and fairness**: They are used to evaluate and ensure the equitable behavior of LLMs across different demographic groups, contexts, or use cases[108].

iii. **Efficiency and resource utilization metrics**: LLMs requires a comparable large number of computational resources as compared to ordinary machine learning models. So, specialized evaluation metrics are used to evaluate them for resource utilization. These are crucial for deploying LLMs in production environments[109].

- **Inference time**: The amount of time it takes for a trained ML model, such as LLMs, to process input data and produce an output or response[110].
- **Memory and compute requirements**: Refers to the computational resources necessary to train, fine-tune, or deploy the LLMs[111].
- **Energy efficiency**: A metric that measures how effectively an LLM utilizes computational resources (electricity, memory, and processing power) to perform tasks such as generating text, answering queries, or processing data. It evaluates the trade-off between energy consumption and performance, aiming to optimize the balance between environmental sustainability and computational output[112].

iv. **Novel metrics**: The ordinary evaluation metrics are not enough to judge and evaluate the developing LLMs so with the growing capabilities of LLMs, newer metrics are being developed.

- **Holistic evaluation of language model (helm)**: HELM was introduced in 2021 by a team of researchers from the University of California, Berkeley, who wanted to provide a more holistic and actionable evaluation of LLMs in real-world applications[113]. The goal was to help developers and researchers better understand the potential and limitations of these models across a wide range of use cases. It was introduced to evaluate LLMs

in a comprehensive and multifaceted manner, going beyond traditional benchmarks like accuracy or perplexity. Instead of focusing only on isolated metrics, HELM aims to assess models in the context of multiple tasks and domains to understand their overall performance, fairness, robustness, and potential societal impacts[114]. This evaluation metric considers various aspects of LLM behavior, such as bias, ethical considerations, interpretability, and performance across diverse use cases[115].

- **Winograd schema challenge**: It was introduced by Terry Winograd in 2011. It refers to a metric used to evaluate the performance of LLMs in terms of their ability to understand and resolve ambiguities in natural language[116]. It specifically tests a model's ability to resolve pronouns in sentences, where the correct answer depends on world knowledge and context, rather than simple syntactic rules. A Winograd Schema consists of a pair of sentences that differ by only one or two words, typically involving a pronoun. The challenge is that, in each pair, the pronoun's reference is ambiguous, and the correct reference can only be determined by reasoning about the context[117].
- **Knowledge F1**: It was introduced in 2020 as part of the "Fact-based Evaluation of Language Models" approach, mainly to evaluate models like T5 and BERT on tasks that require factual knowledge retrieval[118]. It is a performance measure used in evaluating LLMs based on their ability to recall factual knowledge. It combines precision and recall into a single metric, where precision measures how many of the facts retrieved by the model are correct, and recall assesses how many of the correct facts are retrieved by the model. This metric is particularly useful for tasks that require factual accuracy, such as question answering and knowledge retrieval[119].

## Natural Language Understanding (nlu) and generation Language generation (nlg)

NLU and NLG are two critical components of LLMs that enable machines to process and produce human language. NLU refers to a model's ability to comprehend and interpret input text, extracting meaning, context, and intent from natural language. It involves tasks such as sentiment analysis, entity recognition, and syntactic parsing[120]. On the other hand, NLG focuses on generating human-like, coherent, and contextually appropriate text based on input data. Together, NLU and NLG allow LLMs to understand complex queries, engage in meaningful conversations, and produce relevant responses, making them fundamental in applications like chatbots, translation, summarization, and content creation[121].

## Scalability

Scalability refers to the ability of the model to handle increasing amounts of data, users, or tasks without a significant decrease in performance[122]. It involves the model's capacity to expand in both computational power and complexity, enabling it to process larger datasets, generate more sophisticated outputs, and support a growing number of simultaneous requests. Scalability in LLMs also relates to their ability to be deployed across different environments, from local machines to cloud infrastructures, ensuring that as demand increases, the system can adapt and maintain its effectiveness[123].

## Transfer learning

Transfer learning is a technique where a model pre-trained on a large dataset is fine-tuned on a specific, often smaller, domain-specific dataset. This approach leverages the knowledge acquired during pre-training to improve performance on specialized tasks without requiring the model to be trained from scratch[124]. By transferring the general understanding developed during pre-training, LLMs can quickly adapt to new tasks, making them more efficient and effective in scenarios with limited task-specific data. This is particularly useful in industries or fields where labeled data is rare but general language knowledge is applicable[125].

## Contemporary Llms

Contemporary LLMs represent a significant advancement in the field of artificial intelligence and NLP. These models are built upon deep learning architectures with billions of parameters, enabling them to understand and generate human-like text with outstanding fluency. Figure 2 organizes the contemporary LLMs. In the subsections of this section the contemporary LLMs have been covered in detail.

### *Gpt series*

GPT series have 4 LLMs including GPT-1, GPT-2, GPT-3 and GPT-4. GPT-4 is latest and advanced model[58,126,127]. All of these models are built on transformer architecture[128]. There models are trained on huge datasets and later on fine-tuned for specific applications. These models are used in generating coherent, contextually relevant text, making them useful for applications in content creation, language translation, and conversational agents[129].

### *Bert*

BERT was developed by google in 2018 to address the limitations of earlier language models[130,131]. It was specifically developed to handle the intricacies of languages[132]. BERT is a remarkable advancement to handle the context of words in the sentences. Earlier language models process the text sequentially whereas BERT processes the text in both directions, right to left and left to right[133,134]. This approach allows the BERT to capture the full context of words in sentences. This bidirectional approach improves the performance of language models to deeply understand the context of words in text.

### *Albert*

A lite BERT or ALBERT was developed by google in the late 2019 [135]. It is a lighter version of BERT. It uses optimization techniques such as factorized embedding parameterization and cross-layer parameter sharing to reduce the number of parameters[136]. These techniques make it memory efficient and suitable for resource

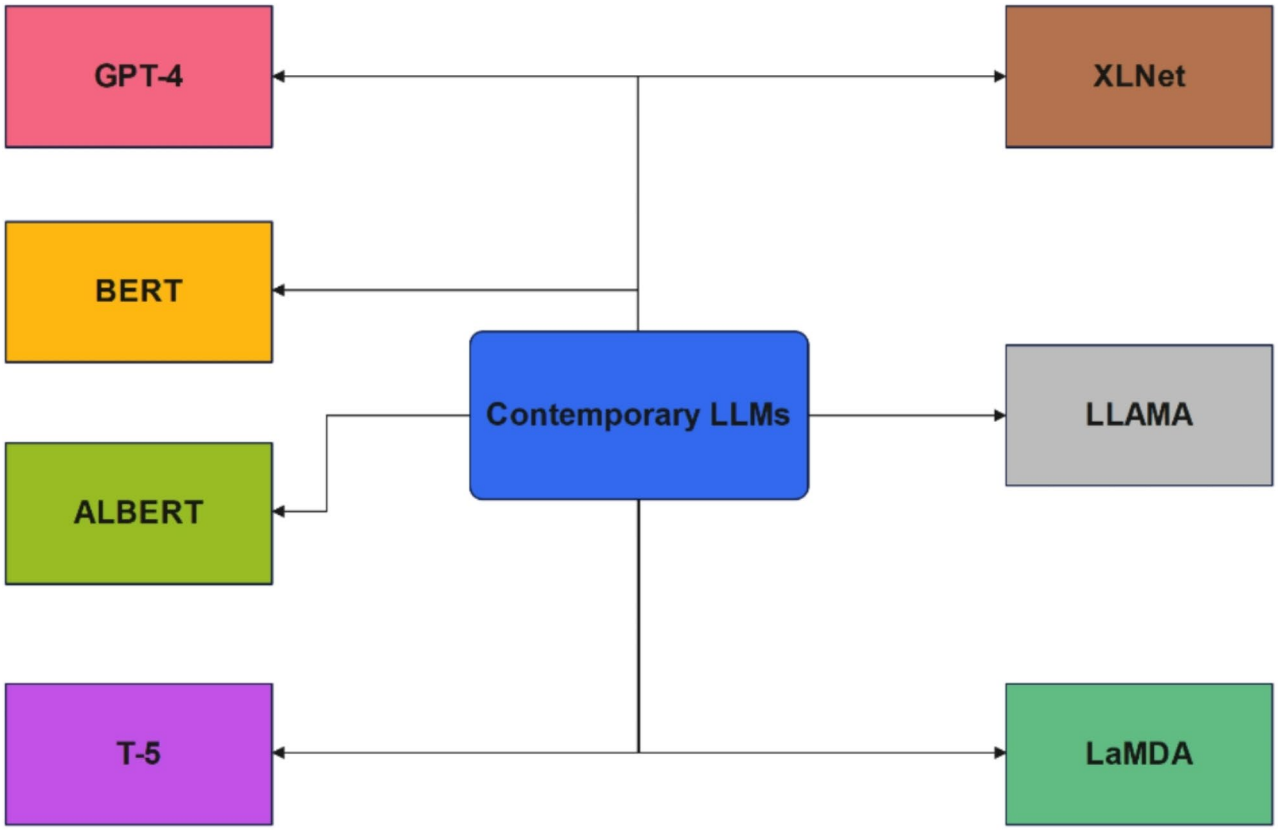


**Fig. 2. Contemporary LLMs.**

constrained environment[137]. The main purpose behind the development of this model was to design a lightweight LLM without compromising the performance.

*T5*

The Text-to-text transfer transformer or T5 was developed by google in 2019 [138]. This model uses the approach of text-to-text task. Text to text task means the input and output of T5 are always in the form for strings[139]. T5 uses transformer architecture. It was pre trained on a huge and multipurpose dataset which is called C4 or Colossal Clean Crawled Corpus[140]. The main purpose behind the development of this model was to simplify the approach to NLP tasks. This single model can be fine-tuned for various applications and NLP tasks.

*XLNet*

XLNet was designed by researchers at Carnegie Mellon University, Pittsburgh and Google in 2019 [141]. It is a combination of autoregressive LLMs and bidirectional LLMs[142]. During training it use permutations of the input sequences. The use of permutation approach assists the model to achieve remarkable performance on a variety of NLP tasks[143]. The main purpose behind the development of XLNet was to overcome the limitations of bidirectional models.

*LaMDA*

LaMDA was developed by researchers at Google in 2021 [144]. LaMDA is a conversational model. The main purpose behind the development of LaMDA was conversations[145]. LaMDA was trained on dialogue datasets so that it can be engaged in human like conversations[146]. LaMDA is suitable for virtual assistants and customer support applications. LaMDA was developed to improve interactions between AI and humans, improving the user experience.

*LLAMA*

The Large Language Model Meta AI (LLAMA) was developed by AI researchers at Facebook in 2023 [147]. This model was developed to enhance the capabilities of LLMs in NLP tasks. The main purpose behind the development of this model was to improve the performance of LLMs in research and practical applications[148]. This model uses transformer architecture and advanced training techniques to get the outstanding performance on various NLP tasks.

## Domain specific applications

LLMs are transforming various industries by providing tailored solutions that leverage specialized knowledge. By focusing on specific domains, LLMs can deliver highly accurate, context-aware insights that drive innovation and efficiency across various fields. In this study we discuss the applications of LLMs in top industries including automotive, e-commerce, education, finance and banking, health care and medicine. Figure 3 summarizes the applications of LLMs.

## Health care and medicine

LLMs have diverse applications in medical field. LLMs have been warmly welcomed in many medical domains. LLMs helps the medical professionals in problem solving and learning[149]. According to the studies specialized AI chatbots will be further improved in near future[150]. LLMs are helpful in clinical decision making[151].

LLMs are used in question answering related to medical fields[152]. LLMs are helpful in for the investigation of patient data, and medical surveys[153]. LLMs are used in biotechnology field to address the challenges bio[154]. LLMs provide information to patient related to healthcare and treatment[155]. LLMs are used for X-ray analysis[156]. Medical text analysis is a challenge for medical professionals[153]. Medical text consists of the aberrations and technical terms related to specific fields[157]. LLMs help the medical professionals to structure the raw medical

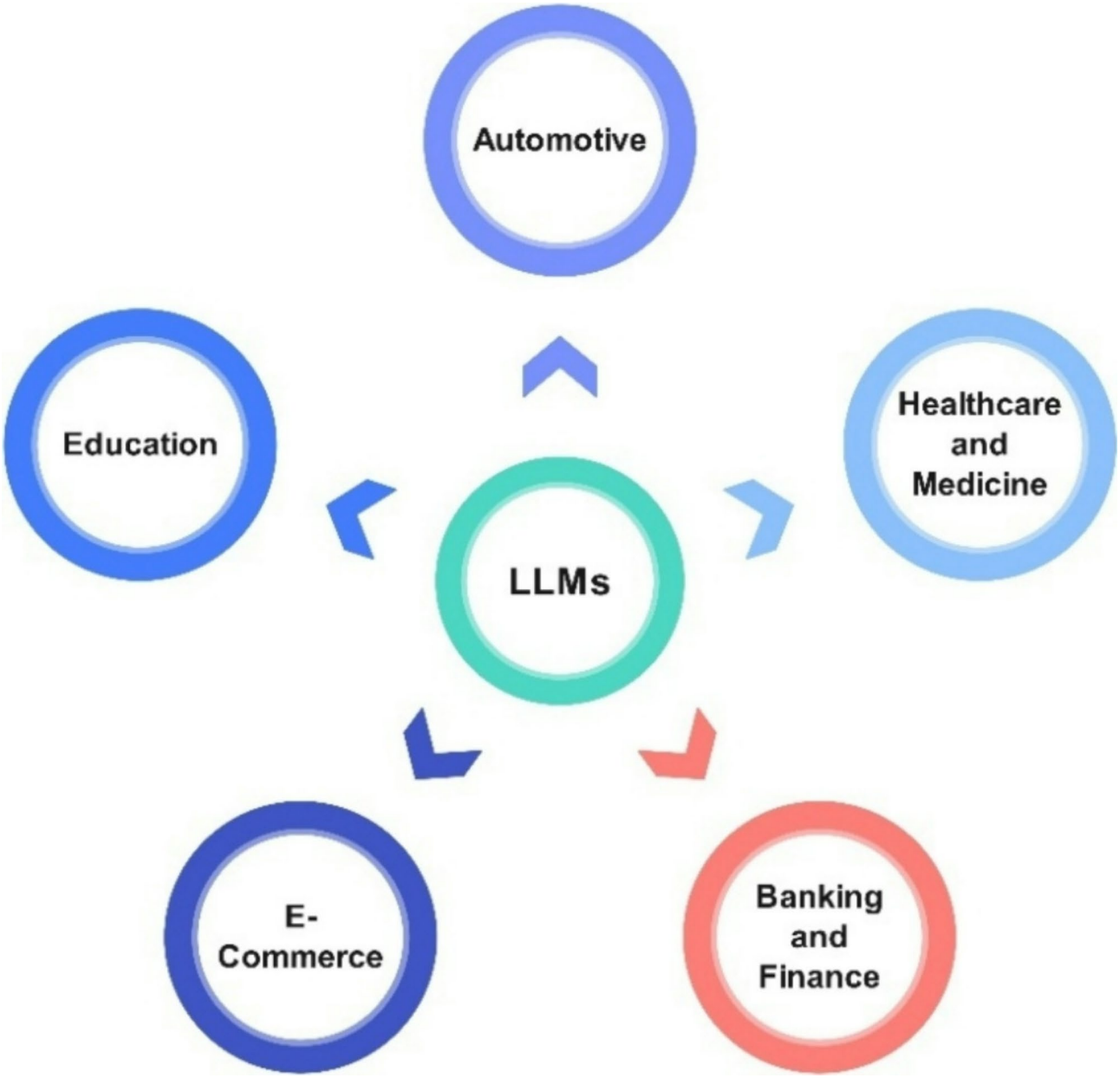


**Fig. 3**. Domain Specific Applications of LLMs.

data and retrieve specific information from the medical text. LLMs help to extract information from medical notes[158].

Zhou et al.,[159] investigated the traditional Chinese medicine for epidemic prevention and treatment. They proposed a LLM for question answering for epidemic prevention and treatment using traditional Chinese medicine. They said traditional Chinese medicine has rich and practical literature related to epidemic prevention and treatment. The literature is complex and huge which is a challenge for medical professionals to extract the required information. The literature consists of numerous books. They said their proposed LLM can solve this problem and it is better than traditional models. They said the proposed model can efficiently handle the huge and complex literature related to traditional Chinese medicine. Another study conducted by Zhe et al.,[160] related traditional Chines medicine. They fine-tuned different LLMs for traditional Chinese medicine formula classification. They claimed that their fine-tuned models achieved better accuracy and performance as compared to previous models.

Yutao et al.,[161] studied the medication guidance and drug reaction prediction and proposed a model for medication guidance and drug reaction forecasting. Their model uses two stage training, in first part the model is trained on drug datasets.

and in second part the model is trained on real datasets of patients. Their model is helpful for medical professionals to improve the healthcare services. Abdulkader et al.,[162] fine tunned the T5 model for medical report summarization. Medical reports are not easy to understand for the public so they fine tunned the model so that could help the public. Senay et al.,[78] proposed a model to automate the administrative tasks in healthcare. Their model is useful for healthcare professionals. Their proposed model can perform appointment scheduling, documentation, and retrieval of medical records. Dimitrios et al., proposed a model to provide diagnostic suggestions to the patients. They said the diagnosis process is a complex one and their proposed model can be helpful for the patients. Luis et al.,[163] proposed an Internet of Medical Things (IoMT) system integrated with LLM. Their proposed model can monitor Parkinson's disease (PD), which is a neurodegenerative disorder.

*Impacts of Llms on health care and medicine*

LLMs are reducing time for clinical documentation and reporting by automating the tasks for example medical transcription and summarization. Nuance Dragon Medical One software uses LLMs to transcribe doctor-patient conversations with up to 98% accuracy, decreasing documentation time by 30–40% [164,165]. LLMs accelerated processing of massive amounts of medical literature to identify relevant studies, trends, and insights. IBM Watson for Health analyzes big datasets to find patterns, and insights saving researchers 20–30% of their time in literature reviews[166,167]. LLMs provide personalized treatment plans based on genetic, environmental, and lifestyle factors[168]. LLMs have improved insurance claim processing time by 30%, reducing errors[169].

*Limitations of Llms in health care and medicine*

LLMs often face difficulty with domain-specific accuracy in medical contexts. For example, general-purpose GPT-4 may generate responses that lack the depth or precision required for complex medical cases[170]. LLMs require enormous, high-quality datasets to perform efficiently. In health care and medicine, data availability is often restricted by privacy laws (for example, HIPAA) and ethical considerations[171]. Deploying LLMs in health care systems requires substantial infrastructure investment and integration with existing workflows[172]. Regulatory laws for LLMs in health care are still evolving, making it difficult to attain compliance[173].

*Proven vs. emerging applications of Llms in health care and medicine*

There are many applications of LLMs that can be found in health care and medicine. Some of them are proven while some of them are still under development. Proven applications include clinical documentation, diagnostic decision support, and chatbots for patient engagement. Whereas emerging applications include AI-Driven surgery assistance, and autonomous health monitoring systems. Researchers are still working on LLMs integration with robots and it still require much improvement and validation[174]. Autonomous health monitoring systems are still mainly in prototype stages and have yet to be adopted on a broad scale due to concerns over accuracy[170].

*Automotive*

LLMs have countless applications in the automotive industry. LLMs have applications in automotive industry including chain management, predicting shortage, and improving production schedules. In case of autonomous vehicles, applications of LLMs are NLP assistant, voice-activated controls, and real-time navigation and predictive maintenance[77]. Bhavin et al.,[175] proposed a multi model vehicle system based on LLM and cloud. The proposed system is a real time automotive system which can assist the driver in driving and provide navigation support and object detection. The proposed model also provides the data privacy and security in system. Zhi-Qi et al.,[176] proposed a model based on LLMs for electric vehicle battery supply chain. Their proposed model can predict the disruption in the supply chain of batteries for electric vehicles. Zeba et al.,[177] proposed a model using fusion approach based on LLM and computer vision. The proposed model detects and differentiates the objects on road. The model also detects the lines on road. After the detection the proposed model translate the detected objects into text form for driver. Mobina et al.,[178] proposed a model in the form of digital voice assistant. Their proposed model is based on LLMs and SVM classifier. The model translates the voice commands for vehicle.

*Impacts of Llms on automotive*

LLMs assist in design and engineering to streamline the process. BMW uses LLMs to create design process for vehicle parts and reduce the time required[179]. Google uses LLMs for its self-driving cars[180]. Tesla uses LLMs in its autopilot system. Audi uses LLMs in its chatbot to provide information related to vehicle features, and sales[181].

Toyota and Honda use LLMs to optimize inventory levels and delivery schedules, leading to a 15–20% reduction in supply chain costs[182]. Automotive companies use LLMs to perform analysis on legal documents. Mercedes-Benz uses LLMs to analyze regulatory changes related to vehicle safety and emissions, streamlining compliance reporting and reducing compliance-related costs by 10–15% [183,184].

*Limitations of Llms in automotive*

Environmental factors and road conditions vary across different areas that may lead to consistency issues with LLMs integrated in autonomous driving systems[185]. Autonomous driving system driven vehicle diagnostics require diverse, and high-quality datasets to perform efficiently[186]. However, these datasets may not fully capture the range of real-world conditions and environmental factors. Training, implementing and deploying LLMs in automotive industry require computational and financial resources[77].

*Proven vs. emerging applications of Llms in automotive*

Proven applications of automotive industry include autonomous driving and driver assistance, predictive maintenance and diagnostics, manufacturing process optimization, voice assistants and in-car interaction[187,188]. Whereas full autonomous driving (Level 5), AI-driven design and customization, real-time traffic and route optimization, and personalized vehicle experience are still emerging and researchers are working on them for improvements[189–191].

*E-commerce*

LLMs are widely adopted in the e-commerce industry. These days e-commerce has become the vital part of global economy. After the advancements in the internet, the traditional shopping approaches have been replaced by e-commerce. E-commerce provides convenience to the consumers. Although e-commerce has many benefits but it has challenges as well, for example language barrier for consumers. To overcome the challenges of e-commerce, LLMs play an important role. Dehong et al.,[192] proposed an LLM based e-commerce machine translation model. They claimed that their proposed model can handle domain specific terms and keywords, providing better performance and accuracy in translation. They fine tunned their model on Chinese and English bilingual terms. Kaidi Chen et al.,[193] proposed an LLM based machine translation model for e-commerce. They claimed that their proposed model can handle the domain related words and special formulas of e-commerce domain. They claimed that their proposed model provides more robustness as compared to previous machine translation models. Chenhao et al.,[194] proposed an LLM- ensemble based model for product attribute value extraction. They claimed that their proposed model improves the recommendation system for consumers by improving the process of product value extraction. Ben et al.,[195] proposed an LLM based relevance modeling for e-commerce search engines. They claimed that their proposed model ensures that products selected based on consumer query is aligned with the intent of consumer.

*Impacts of Llms on e-commerce*

Amazon uses LLM powered chatbots to handle and process customer inquiries[196]. These chatbots can resolve 70–80% of customer service queries without human intervention[197]. This reduces the need for human customer service agents lessens the cost. Amazon uses LLMs for product recommendations and personalization[198]. Walmart LLMs to analyze customer profile data and behavior across different areas, permitting for personalized email marketing[199]. Zara uses LLMs to predict demand for different clothing styles across various regions, adjusting production schedules and inventory levels accordingly[182]. Other different brands use LLMs for content creation and social media[200].

*Limitations of Llms in e-commerce*

In case of recommendation systems, LLMs may face challenges in understanding vague queries or generating accurate recommendations when customer input is imprecise[201]. The performance of LLMs totally relies on the quality and range of training data, which may not always represent the full range of customer behaviors[202]. It can be computationally intensive for LLMs to Handle large-scale, real-time operations in e-commerce, such as dynamic pricing updates[195]. LLMs operate as black-box systems, making it problematic to explain why a particular recommendation or search result was made[203].

*Proven vs. emerging applications of Llms in e-commerce*

Although LLMs have significant impact on e-commerce but some of its applications are proven and some are still emerging. Proven applications include customer service and support, product recommendations, marketing and advertising, fraud detection and prevention, search optimization, pricing optimization, content creation and copywriting[204,205]. Whereas emerging applications include inventory management and demand forecasting, regional and cultural adaptation, advanced behavioral analytics, and ethical recommendations[206,207].

*Education*

Large Language Models (LLMs) are transforming the education industry by enhancing personalized learning and administrative efficiency. They can assist in crafting customized learning experiences by analyzing student performance and tailoring content to individual needs. They also support teachers by automating administrative tasks, generating lesson plans, and grading assignments. Additionally, LLMs facilitate language translation for diverse learners, fostering a more inclusive educational environment. Ehsan et al.,[208] proposed a method for distilling fine-tuned LLMs into smaller, efficient neural networks for deployment on resource-constrained devices. The authors trained a student model using the prediction probabilities of a teacher model achieving comparable accuracy with state-of-the-art models while being significantly smaller and faster. Zheyuan et

al.,[209] proposed an LLM based framework. Their proposed model is a multi-agent classroom simulation model. Their proposed model is a classroom mechanism for automatic classroom teaching. Liuqing et al.,[210] proposed education approach to learn bio-inspired design. They claimed that bio-inspired designs are difficult to understand and learn. They also said that learning bio-inspire designs depends upon on teacher. The authors said that their proposed model is helpful to learn bio-inspired designs. Victor et al.,[211] proposed a web application based on LLMs for education. The authors said their proposed application is subscription free. The teachers can upload their own datasets to fine tune the model.

*Impacts of Llms on education*
LLMs are providing personalized learning to students. Duolingo which is an education based mobile app, uses LLMs to enhance learning and engagements by 20% [212]. LLMs are also used to create educational content including quizzes, assignments, and lecture plans. Khan Academy uses LLMs to create practice problems for students[213]. LLMs identify gaps in student understanding through data analysis and offer targeted interventions[214]. LLMs provide assistance to researchers by summarizing the research articles and identifying the trends. LMMs are used to provide personalized feedback to students[215].

*Limitations of Llms in education*
LLMs may misinterpret user queries when applied to multiple topics. LLMs may produce incorrect information for complex and diverse topics[216]. LLMs often lack the ability to provide effectively to various learning preferences, such as visual, and auditory. Training data may not represent all cultural, linguistic, or regional contexts resulting in biased outputs[217]. Finetuning LLMs with existing syllabus and regulatory standards can be a time consuming and complex task[218].

*Proven vs. emerging applications of Llms in education*
Proven applications of LLMs include personalized learning, automated content generation, intelligent tutoring systems, language translation, enhanced assessment and feedback[219]. Emerging applications include advanced behavioral analytics, adaptive simulations for teacher training, regional and cultural adaptation, and gamified learning experiences[220].

*Finance and banking*
LLMs are gradually being used in finance and banking to enhance customer service, and improve decision-making. They can analyze massive amounts of data to provide insights, automate tasks such as customer queries and fraud detection, and assist in risk management. Shijie et al.,[221] proposed a model specialized for finance. They said that they trained their model on massive financial dataset. Zarza et al.,[222] proposed a model for financial planning and budgeting based on LLM. They claimed that their proposed model is effective for both houses and corporates. Their proposed model suggests solutions to manage budget plans. Boyu et al.,[223] proposed a model for financial sentiment analysis and investment decision making. They claimed that their proposed model improves the sentiment analysis and investment decision making process. George et al.,[76] proposed a model which is a combination of LLM and cloud environment. The authors said the proposed model improves the operations and compliance in banking system. The authors said their proposed models overcomes the traditional challenges of banking. They said their model also enhances the customer experience. Daniel et al.,[224] proposed a model for automatic topic modeling and their categorization for tagging retail banking transactions. They used LLM and zero shot prompting.

*Impacts of Llms on finance and banking*
LLMs can enhance fraud detection systems by analyzing the transactional data. JPMorgan Chase uses LLMs to keep an eye on transactions, which reduces the fraud chances by 40% [225]. Bank of America obtained an LLMs based chatbot named Erica which handles customer inquiries and bill payments[226]. Citi bank uses AI models to reduce loan approval time by 50% [227]. Compliance team of Standard Chartered Bank report a 40% reduction in manual reviews and an 85% increase in detection precision[228]. BloombergGPT provides insights on market conditions by analyzing and processing textual data from numerous sources. BloombergGPT is trained on 50-billion parameters and it was built for finance[229].

*Limitations of Llms in finance and banking*
Due to biased training data LLMs can predict inaccurate market trends resulting in financial losses[70]. Banking and finance produce huge amount of live data which can overwhelm the LLMs in real time applications[230]. Attackers may exploit vulnerabilities of LLMs to manipulate the outputs[231]. Ensuring customer data privacy during training and implementation of LLMs is a challenge[232].

*Proven vs. emerging applications of Llms in finance and banking*
In finance and banking, proven applications of LLMs are fraud detection, customer support automation, document processing, trading and market predictions, and personalized banking[70]. Emerging applications are advanced behavioral analytics, emotional feedback analysis, and blockchain integration[233].

*Case studies and empirical evidence*
To provide the concrete evidence of integration of LLMs in industry, this section investigates some practical applications. IBM Watson uses LLMs to analyze extensive medical literature and patient data, providing evidence-based diagnoses and treatment recommendations. This application assists the healthcare professionals in making informed decisions, thereby improving patient outcomes and streamlining the diagnostic process[234–236]. Alexa

is cloud-based voice assistant that was developed by Amazon in 2014. Now Amazon Alexa is powered by LLMs. Amazone has implemented a custom LLM to make Alexa more efficient in its conversations. Alexa depends upon LLM to understand, process and response user queries[237]. Coca-Cola is a famous international brand. It always uses innovative marketing strategies since its birth. It is one of the most iconic brands. Coca-Cola has integrated an advanced language model GPT-4 in its marketing operations. Coca-Cola is using GPT-4 for content generation, information retrieval, social media, and generating comprehensive reports[234,238,239].

The famous OTT platform Netflix is using LLMs for personalized recommendation system. The main purpose of the Netflix is to innovate its personalized recommendation system. Netflix uses LLMs to analyzes extensive data of users, find valuable insights and provide the users with their desired content[234,240]. Spotify is a famous audio streaming service. Like Netflix Spotify also uses LLMs to improve its music recommendation system. Spotify uses LLMs to analyze user listening habits, playlists, and interactions with the platform. LLMs provide Spotify a way to understand user preferences and recommend them with their desired music[234,241]. The New York Times is famous leading global media. It is using LLMs to improve advertising strategies. LLMs enables advertisers to maximize their influence by suggesting the best places for ad campaigns based on the advertisement's messaging. By refining strategies LLMs enhance campaign performance[234,242].

## Open issues and challenges

LLMs provide a massive number benefits but they also have issues and challenges. Over the time many issues and challenges have been overcome by researchers but many of them are still open for research and debate[243]. In this section we highlight the open issues and challenges of LLMs.

### Open issues

This section covers the open issues of LLMs in industry. Figure 4 summarizes the open issues of LLMs.

- **Ethical Issues**: The LLMs are trained on massive datasets. The questions that arise here are: Who can use the dataset? How can the dataset be used? And when can the dataset be used? The ethical issue related to the use of dataset are still open to discuss. The datasets can consist of biased data leading to the biased outputs from LLMs[244]. The LLMs can also provide hate speech and misinformation.
- **Data Privacy Issues**: The training datasets of LLMs can consist of personal data which is an open issue for LLMs. Data privacy preserving techniques are required to train the models without compromising user privacy. As the use of data is increasing in LLM models, the privacy concerns are also increasing[59].
- **Adversarial and Cyber-attacks**: LLMs are vulnerable to cyber-attacks. Security of LLMs is an open issue. Improving the security of LLMs against cyber-attacks is a big concern[245]. LLMs can be vulnerable to adversarial inputs that manipulate their outputs in harmful ways. Understanding how to strengthen models against such attacks is a critical area of research.

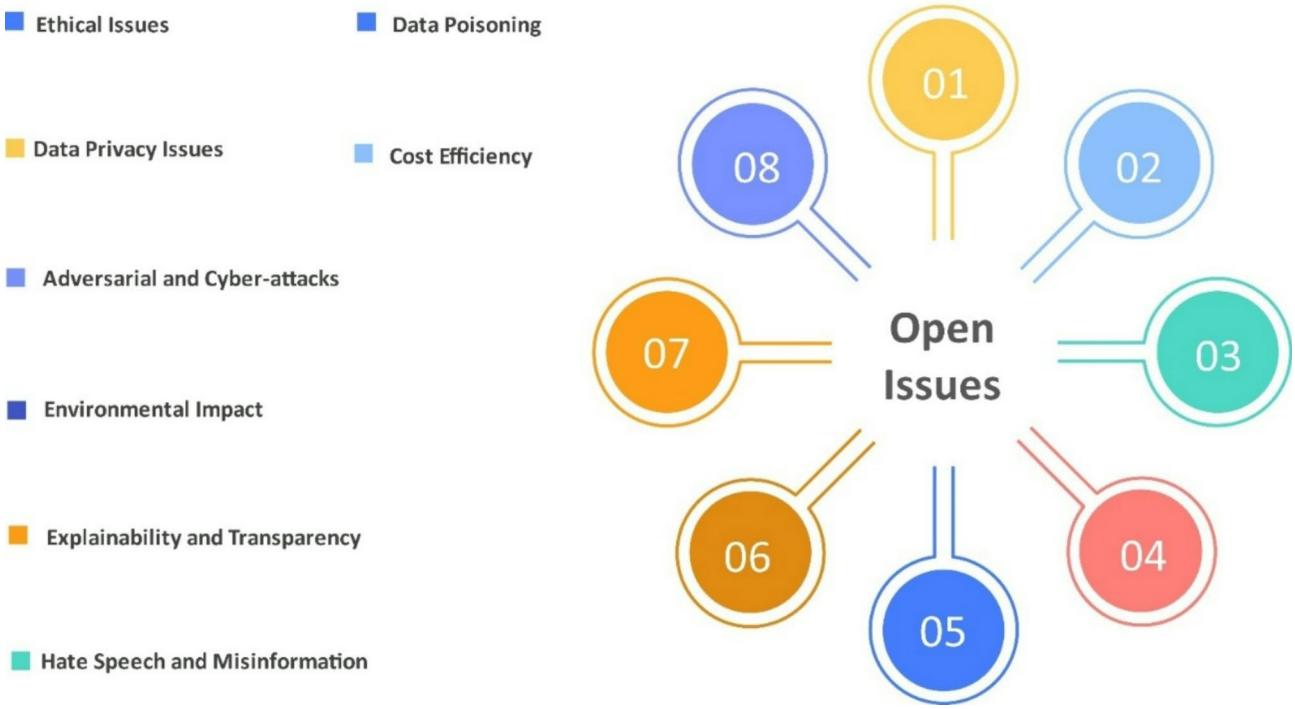


**Fig. 4**. Open Issues of LLMs.

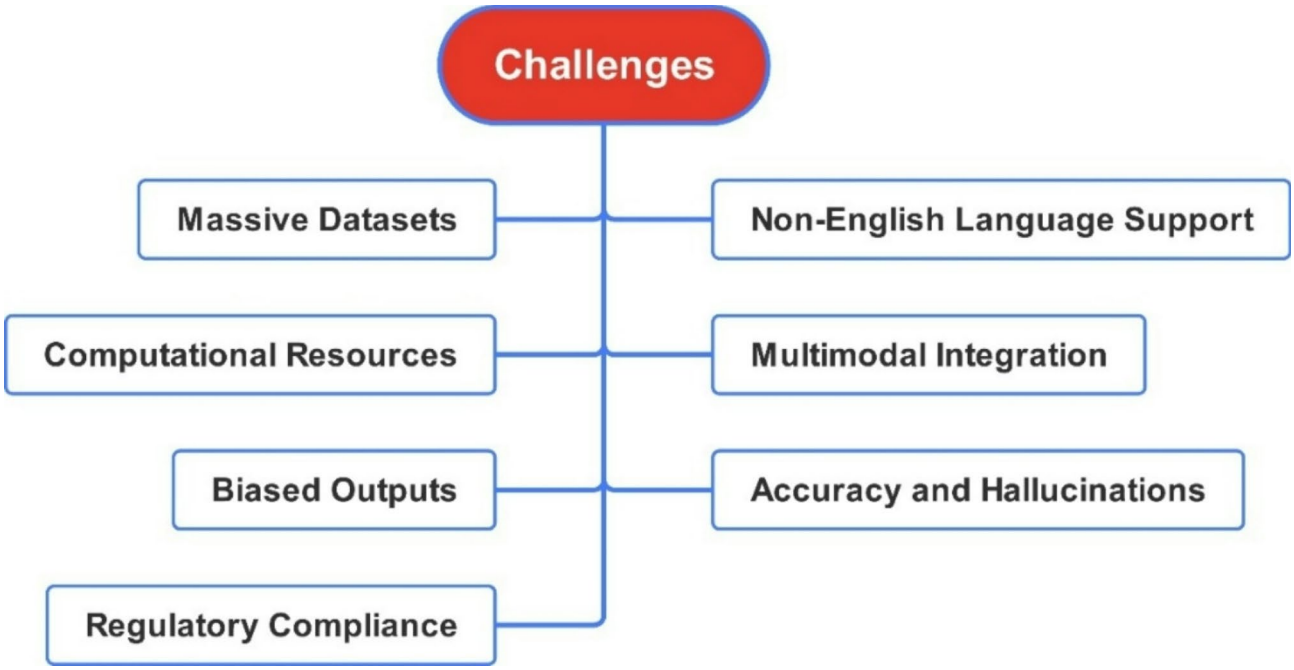


**Fig. 5**. Challenges of LLMs.

- **Environmental Impact**: Training and deploying LLMs require considerable computational resources, leading to significant energy consumption and carbon emission. The environmental footprint of these models is an open issue that calls for the development of more energy-efficient algorithms[246].
- **Explainability and Transparency**: LLMs operate as black-box models, making it difficult to understand how they generate specific outputs. This lack of explainability raises concerns in critical domains like healthcare and finance, where understanding the rationale behind decisions is essential[247].
- **Hate Speech and Misinformation**: LLMs can unintentionally generate harmful content, including hate speech or misinformation, which can have real-world consequences. The responsibility of developers to mitigate these risks is a critical area for further exploration[248].
- **Data Poisoning**: Attackers may introduce malicious data into training sets, leading to compromised model integrity[249].
- **Cost Efficiency**: The financial burden associated with developing and maintaining LLMs remains a significant barrier for many organizations. High costs related to data acquisition, processing power, and ongoing model training can deter smaller enterprises from leveraging these technologies[250].

*Challenges*

The use of LLMs is increasing gradually in the industry leading to challenges. This section covers the open challenges of LLMs in industry. Figure 5 Summarizes the challenges of LLMs.

- **Massive Datasets**: LLMs are trained on massive and complex datasets. The source of datasets is internet. Due to their size and complexity, it is a challenge to maintain the security and privacy and quality of datasets[251]. Handling and processing a massive data are a challenge itself.
- **Computational Resources**: Due to the huge amount of training datasets LLMs require huge set of computational resources[252]. Some models require special hardware for their training. Energy consumption is high for LLMs training. These are the open challenges for LLMs.
- **Biased Outputs**: Biased outputs from LLMs presents a significant challenge, as LLMs can unintentionally reflect and amplify biases present in their training data[253]. This can lead to unfair results, particularly in sensitive areas such as hiring, law enforcement, or healthcare, where impartiality is critical.
- **Regulatory Compliance**: With the emergence of regulations such as GDPR, ensuring compliance while utilizing large datasets poses a challenge. Organizations must navigate these legal frameworks while balancing innovation with privacy rights[254].
- **Non-English Language Support**: A significant gap exists in the performance of LLMs across different languages, particularly non-English languages. This limitation restricts access to advanced AI capabilities for non-English speaking populations. Efforts must be directed towards developing robust models that can understand and generate content in a variety of languages without compromising quality[255].
- **Multimodal Integration**: The integration of multiple data modalities (text, images, audio) into LLMs presents an open challenge that could expand their capabilities significantly. Current models primarily focus on text-based inputs, which limits their applicability in diverse fields such as healthcare diagnostics or customer service where multimodal understanding is crucial[256].

- **Accuracy and Hallucinations**: Ensuring the accuracy of outputs generated by LLMs is paramount. The phenomenon of "hallucinations," where models produce reasonable but incorrect information, poses risks in applications that rely on factual accuracy. Addressing this challenge requires improved training methodologies and validation processes to enhance the reliability of generated content[257].

## Ethical considerations and responsible deployment

The deployment of LLMs in industry raises significant ethical concerns that must be carefully addressed to ensure responsible use. Key ethical considerations include the potential for bias in model predictions, privacy issues regarding sensitive data, and the opacity of decision-making processes within these systems[258,259]. To mitigate these risks, it is essential to implement frameworks that prioritize transparency, fairness, and accountability. For instance, organizations should establish rigorous bias detection mechanisms, maintain clear data governance policies, and ensure that LLMs are understandable and auditable. Additionally, it is important to involve interdisciplinary teams comprising ethicists, domain experts, and technologists, during the design and deployment phases to continuously evaluate the social impact LLMs. Practical recommendations for responsible deployment include adopting established ethical guidelines, such as the "Ethics Guidelines for Trustworthy AI" from the EU, and implementing regular audits to assess the ethical performance of LLMs[260,261]. By encouraging a culture of accountability and ongoing scrutiny, we can ensure that LLMs are deployed in ways that align with social values and mitigate the risks of harm.

### Future directions of Llms

LLMs are prominent emerging technologies and continue to be a dynamic area of AI research. Efforts are focused on advancing techniques to tackle data privacy concerns, enhance security, and address ethical considerations such as bias and fairness. Additionally, the integration of LLMs with domain-specific knowledge through specialized fine-tuning will enable more accurate and context-aware applications. As these advancements are made, LLMs will be deployed in ways that maximize their benefits while ensuring ethical use, scalability, and broader societal advantages, steering in a new era of AI innovation.

*Limitations*

One of the important limitations of this study is availability of limited review literature related to LLMs and their applications in the industry. Best efforts are made to find the related literature. The available material is covered thoroughly to address and cover all the related aspects. Due to the restriction of resources this study covered applications of LLMs in prominent industries, although LLMs have applications in some other industries as well. We covered and discussed the details of modern LLM models and architectures but an in-depth analysis can be done on each of them.

## Conclusion

Based on neural networks and transformer architecture, LLMs have evolved in a remarkable way. LLMs have changed the field of NLP. LLMs resulted in extraordinary expansion in NLP. LLMs have revolutionized the text generation and processing. The use of LLMs is increasing and expending in the industry. Applications of LLMs can be found in almost each and every domain of industry. Although LLMs are considered evolutionary and powerful across various fields, they also have limitations and challenges. This study has provided an insightful and meaningful review of LLMs and their applications in the industry. This research work covered the important aspects of LLMs, and contemporary LLMs. The study has also examined industrial domain specific applications of LLMs, including healthcare and medicine, automotive, e-commerce, education finance and banking. The research work covered the open issues of LLMs including ethical issues, data privacy issues, security issues, environmental impacts, explainability and transparency. The study also covered the open challenges of LLMs including massive datasets, computational resources, biased outputs and regulatory compliance. As the field of LLMs research and development is expanding swiftly, this review would be a valuable literature for the researchers looking for literature related to the applications of LLMs in the industry. The study focused on the significance of LLMs in the industry. LLMs represent advancements in NLP and AI, revolutionizing the domain of problem-solving in the industry; however, they are still under development and require many improvements.

## Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 17 October 2024; Accepted: 11 April 2025
Published online: 21 April 2025

## References

1. Khurana, D., Koli, A., Khatter, K. & Singh, S. Natural Language processing: state of the art, current trends and challenges. *Multimed Tools Appl.* **82**, 3713–3744 (2023).
2. Kosch, T. et al. A survey on measuring cognitive workload in human-computer interaction. *ACM Comput. Surv.* **55**, 1–39 (2023).
3. Chowdhary, K. & Chowdhary, K. R. Natural Language processing. *Fundam Artif. Intell.* **2020**, 603–649 (2020).
4. Fanni, S. C., Febi, M., Aghakhanyan, G. & Neri, E. Natural language processing. in *Introduction to Artificial Intelligence* 87–99 (Springer, 2023).
5. Eisenstein, J. *Introduction To Natural Language Processing* (MIT Press, 2019).
6. Bayer, M. et al. Data augmentation in natural Language processing: a novel text generation approach for long and short text classifiers. *Int. J. Mach. Learn. Cybern.* **14**, 135–150 (2023).

7. Li, J., Tang, T., Zhao, W. X., Nie, J. Y. & Wen, J. R. Pre-trained Language models for text generation: A survey. *ACM Comput. Surv.* **56**, 1–39 (2024).
8. Zhao, W. X. et al. A survey of large language models. *arXiv Prepr. arXiv2303.18223* (2023).
9. Riedl, M. O. Human-centered artificial intelligence and machine learning. *Hum. Behav. Emerg. Technol.* **1**, 33–36 (2019).
10. Jiang, Z., Xu, F. F., Araki, J. & Neubig, G. How can we know what Language models know? *Trans. Assoc. Comput. Linguist*. **8**, 423–438 (2020).
11. Shen, Y. et al. ChatGPT and other large language models are double-edged swords. *Radiology* vol. 307 e230163 at (2023).
12. Myagmar, B., Li, J. & Kimura, S. Cross-Domain sentiment classification with bidirectional contextualized transformer Language models. *IEEE Access.* **7**, 163219–163230 (2019).
13. Singh, S. & Mahmood, A. The NLP cookbook: modern recipes for transformer based deep learning architectures. *IEEE Access.* **9**, 68675–68702 (2021).
14. Yang, J. et al. Harnessing the power of Llms in practice: A survey on Chatgpt and beyond. *ACM Trans. Knowl. Discov Data*. **18**, 1–32 (2024).
15. Huang, Y. et al. Advancing transformer architecture in long-context large language models: A comprehensive survey. *arXiv Prepr. arXiv2311.12351* (2023).
16. Melis, G., Dyer, C. & Blunsom, P. On the state of the art of evaluation in neural language models. *arXiv Prepr. arXiv1707.05589* (2017).
17. Mikolov, T. & others. Statistical language models based on neural networks. (2012).
18. Naseem, U., Razzak, I., Khan, S. K. & Prasad, M. A comprehensive survey on word representation models: from classical to state-of-the-art word representation Language models. *Trans. Asian Low-Resource Lang. Inf. Process.* **20**, 1–35 (2021).
19. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **404**, 132306 (2020).
20. Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M. & Zhong, J. Attention is all you need in speech separation. in *ICASSP –2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 21–25 (2021). (2021).
21. Xu, M. et al. A survey of resource-efficient llm and multimodal foundation models. *arXiv Prepr. arXiv2401.08092* (2024).
22. Jwa, H., Oh, D., Park, K., Kang, J. M. & Lim, H. Exbake: automatic fake news detection model based on bidirectional encoder representations from Transformers (bert). *Appl. Sci.* **9**, 4062 (2019).
23. Yenduri, G. et al. Gpt (generative pre-trained transformer)--a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access.* **12**, 54608–54649 (2024).
24. Xue, L. et al. mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv Prepr. arXiv11934* (2020). (2010).
25. Zhou, X., Zhao, X. & Li, G. LLM-Enhanced Data Management. *arXiv Prepr. arXiv2402.02643* (2024).
26. Khare, Y. et al. Mmbert: Multimodal bert pretraining for improved medical vqa. in. *IEEE 18th International Symposium on Biomedical Imaging (ISBI)* 1033–1036 (2021). (2021).
27. Cui, C. et al. A survey on multimodal large language models for autonomous driving. in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 958–979 (2024).
28. Ren, Q. et al. A survey on fairness of large language models in e-commerce: progress, application, and challenge. *arXiv Prepr. arXiv2405.13025* (2024).
29. Parker, M. J., Anderson, C., Stone, C. & Oh, Y. A large Language model approach to educational survey feedback analysis. *Int. J. Artif. Intell. Educ.* 1–38 (2024).
30. Lee, J., Stevens, N., Han, S. C. & Song, M. A survey of large language models in finance (finllms). *arXiv Prepr. arXiv2402.02315* (2024).
31. Cascella, M. et al. The breakthrough of large Language models release for medical applications: 1-year timeline and perspectives. *J. Med. Syst.* **48**, 22 (2024).
32. Turing, A. M. Computing machinery and intelligence. *Creat Comput.* **6**, 44–53 (1980).
33. Masri, N. et al. Survey of rule-based systems. *Int. J. Acad. Inf. Syst. Res.* **3**, 1–23 (2019).
34. Grossberg, S. Recurrent neural networks. *Scholarpedia* **8**, 1888 (2013).
35. Salehinejad, H., Sankar, S., Barfett, J., Colak, E. & Valaee, S. Recent advances in recurrent neural networks. *arXiv Prepr. arXiv1801.01078* (2017).
36. Johnson, S. J., Murty, M. R. & Navakanth, I. A detailed review on word embedding techniques with emphasis on word2vec. *Multimed Tools Appl.* **83**, 37979–38007 (2024).
37. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**, 1235–1270 (2019).
38. Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y. & Liu, J. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **11**, 68–75 (2017).
39. Kowsher, M. et al. LSTM-ANN \& BiLSTM-ANN: hybrid deep learning models for enhanced classification accuracy. *Procedia Comput. Sci.* **193**, 131–140 (2021).
40. Church, K. W. Word2Vec. *Nat. Lang. Eng.* **23**, 155–162 (2017).
41. Pennington, J., Socher, R. & Manning, C. D. Glove: Global vectors for word representation. in *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* 1532–1543 (2014). (2014).
42. Di Gennaro, G., Buonanno, A. & Palmieri, F. A. N. Considerations about learning Word2Vec. *J. Supercomput.* **77**, 1–16 (2021).
43. Ma, L. & Zhang, Y. Using Word2Vec to process big text data. in *IEEE International Conference on Big Data (Big Data)* 2895–2897 (2015). (2015).
44. Abubakar, H. D., Umar, M. & Bakale, M. A. Sentiment classification: review of text vectorization methods: bag of words, Tf-Idf, Word2vec and Doc2vec. *SLU J. Sci. Technol.* **4**, 27–33 (2022).
45. Sivakumar, S. et al. Review on word2vec word embedding neural net. in. *international conference on smart electronics and communication (ICOSEC)* 282–290 (2020). (2020).
46. Curto, G., Jojoa Acosta, M. F., Comim, F. & Garcia-Zapirain, B. Are AI systems biased against the poor? A machine learning analysis using Word2Vec and glove embeddings. *AI \& Soc.* **39**, 617–632 (2024).
47. Singgalen, Y. A. Implementation of global vectors for word representation (GloVe) model and social network analysis through wonderland Indonesia content reviews. *J. Sist Komput Dan. Inf.* **5**, 559–569 (2024).
48. Sitender, S., Sushma, N. S. & Sharma, S. K. Effect of GloVe, Word2Vec and fastText embedding on english and hindi neural machine translation systems. in *Proceedings of Data Analytics and Management: ICDAM 2022* 433–447Springer, (2023).
49. Kang, S., Kong, L., Luo, B., Zheng, C. & Wu, J. Principle research of word vector representation in natural language processing. in *International Conference on Electronic Information Engineering and Computer Science (EIECS* vol. 12602 54–60 (2023). (2022).
50. Adawiyah, A. R., Baharuddin, B., Wardana, L. A. & Farmasari, S. Comparing post-editing translations by Google NMT and Yandex NMT. *TEKNOSASTIK* **21**, 23–34 (2023).
51. Mo, Y., Qin, H., Dong, Y., Zhu, Z. & Li, Z. Large language model (llm) ai text generation detection based on transformer deep learning algorithm. *arXiv Prepr. arXiv2405.06652* (2024).
52. Oliaee, A. H., Das, S., Liu, J. & Rahman, M. A. Using bidirectional encoder representations from Transformers (BERT) to classify traffic crash severity types. *Nat. Lang. Process. J.* **3**, 100007 (2023).

53. Wibawa, A. P., Cahyani, D. E., Prasetya, D. D., Gumilar, L. & Nafalski, A. Detecting emotions using a combination of bidirectional encoder representations from Transformers embedding and bidirectional long short-term memory. *Int. J. Electr. \& Comput. Eng.* **13**, 2088–8708 (2023).
54. Areshey, A. & Mathkour, H. Transfer learning for sentiment classification using bidirectional encoder representations from Transformers (BERT) model. *Sensors* **23**, 5232 (2023).
55. Hendy, A. et al. How good are gpt models at machine translation? a comprehensive evaluation. *arXiv Prepr. arXiv2302.09210* (2023).
56. Hanna, M., Liu, O. & Variengien, A. How does GPT-2 compute greater-than? Interpreting mathematical abilities in a pre-trained Language model. *Adv. Neural Inf. Process. Syst.* **36**, 76033–76060 (2024).
57. Bharathi Mohan, G. et al. Text summarization for big data analytics: a comprehensive review of GPT 2 and BERT approaches. *Data Anal. Internet Things Infrastruct.* 247–264 (2023).
58. Kalyan, K. S. A survey of GPT-3 family large Language models including ChatGPT and GPT-4. *Nat. Lang. Process. J.* **6**, 100048 (2023).
59. Yan, B. et al. On protecting the data privacy of large language models (llms): A survey. *arXiv Prepr. arXiv2403.05156* (2024).
60. Li, Y., Wang, S., Ding, H. & Chen, H. Large language models in finance: A survey. in *Proceedings of the fourth ACM international conference on AI in finance* 374–382 (2023).
61. Zhang, Z. et al. Large language models for mobility in transportation systems: A survey on forecasting tasks. *arXiv Prepr. arXiv2405.02357* (2024).
62. Wang, S. et al. Large language models for education: A survey and outlook. *arXiv Prepr. arXiv2403.18105* (2024).
63. Zhang, D., Zheng, H., Yue, W. & Wang, X. Advancing ITS Applications with LLMs: A Survey on Traffic Management, Transportation Safety, and Autonomous Driving. in *International Joint Conference on Rough Sets* 295–309 (2024).
64. Xu, X., Xu, Z., Ling, Z., Jin, Z. & Du, S. Emerging Synergies Between Large Language Models and Machine Learning in Ecommerce Recommendations. *arXiv Prepr. arXiv2403.02760* (2024).
65. Chen, J. et al. When large Language models Meet personalization: perspectives of challenges and opportunities. *World Wide Web.* **27**, 42 (2024).
66. Xu, H., Gan, W., Qi, Z., Wu, J. & Yu, P. S. Large Language Models for Education: A Survey. *arXiv Prepr. arXiv2405.13001* (2024).
67. Huber, S. E. et al. Leveraging the potential of large Language models in education through playful and game-based learning. *Educ. Psychol. Rev.* **36**, 25 (2024).
68. Yan, L. et al. Practical and ethical challenges of large Language models in education: A systematic scoping review. *Br. J. Educ. Technol.* **55**, 90–112 (2024).
69. Yahyazadeh, N. The Influence of ChatGPT in Education: A Comprehensive Review. (2023).
70. Zhao, H. et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv Prepr. arXiv2401.11641* (2024).
71. Godwin Olaoye, H. J. The Evolving Role of Large Language Models (LLMs) in Banking. (2024).
72. Fieberg, C., Hornuf, L. & Streich, D. Using large Language models for financial advice. *Available SSRN* **4850039**, 92 (2024).
73. Huang, Y., Tang, K. & Chen, M. A. Comprehensive Survey on Evaluating Large Language Model Applications in the Medical Industry. *arXiv Prepr. arXiv2404.15777* (2024).
74. Zheng, Y. et al. Large Language Models for Medicine: A Survey. *arXiv Prepr. arXiv2405.13055* (2024).
75. Yu, P., Xu, H., Hu, X. & Deng, C. Leveraging generative AI and large Language models: a Comprehensive Roadmap for Healthcare Integration. in *Healthcare* vol. 11 2776 (2023).
76. George, J. G. Transforming Banking in the Digital Age: The Strategic Integration of Large Language Models and Multi-Cloud Environments.
77. Dhillon, A. S. & Torresin, A. Advancing Vehicle Diagnostic: Exploring the Application of Large Language Models in the Automotive Industry. (2024).
78. Gebreab, S. A., Salah, K., Jayaraman, R., ur Rehman, M. & Ellaham, S. LLM-Based Framework for Administrative Task Automation in Healthcare. in *12th International Symposium on Digital Forensics and Security (ISDFS)* 1–7 (2024). (2024). https://doi.org/10.1109/ISDFS60797.2024.10527275
79. Jin, H. et al. Llm maybe longlm: Self-extend llm context window without tuning. *arXiv Prepr. arXiv2401.01325* (2024).
80. Zhang, T., Yi, J. W., Yao, B., Xu, Z. & Shrivastava, A. Nomad-attention: Efficient llm inference on cpus through multiply-add-free attention. *arXiv Prepr. arXiv2403.01273* (2024).
81. Lin, X. et al. Efficient LLM Training and Serving with Heterogeneous Context Sharding among Attention Heads. *arXiv Prepr. arXiv2407.17678* (2024).
82. Lu, Y. et al. LongHeads: Multi-Head Attention is Secretly a Long Context Processor. *arXiv Prepr. arXiv2402.10685* (2024).
83. Liu, Y. et al. Understanding llms: A comprehensive overview from training to inference. *arXiv Prepr. arXiv2401.02038* (2024).
84. Raje, A. & Communication-Efficient, L. L. M. Training for Federated LearningPh. D. thesis, Carnegie Mellon University Pittsburgh, PA,. (2024).
85. Zeng, F., Gan, W., Wang, Y. & Philip, S. Y. Distributed training of large language models. in *IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)* 840–847 (2023). (2023).
86. McKinzie, B. et al. Mm1: Methods, analysis \& insights from multimodal llm pre-training. *arXiv Prepr. arXiv2403.09611* (2024).
87. Abbasiantaeb, Z., Yuan, Y., Kanoulas, E. & Aliannejadi, M. Let the llms talk: Simulating human-to-human conversational qa via zero-shot llm-to-llm interactions. in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* 8–17 (2024).
88. Lin, X. et al. Data-efficient Fine-tuning for LLM-based Recommendation. in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* 365–374 (2024).
89. Liu, Q. et al. When MOE Meets LLMs: Parameter Efficient Fine-tuning for Multi-task Medical Applications. in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* 1104–1114 (2024).
90. Christophe, C. et al. Med42–Evaluating Fine-Tuning Strategies for Medical LLMs: Full-Parameter vs. Parameter-Efficient Approaches. *arXiv Prepr. arXiv2404.14779* (2024).
91. Xue, T., Wang, Z. & Ji, H. Parameter-efficient tuning helps language model alignment. *arXiv Prepr. arXiv2310.00819* (2023).
92. Han, Z., Gao, C., Liu, J. & Zhang, S. Q. & others. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv Prepr. arXiv2403.14608* (2024).
93. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 5998–6008 (2017).
94. Bhattamishra, S., Patel, A., Blunsom, P. & Kanade, V. Understanding in-context learning in transformers and llms by learning to learn discrete functions. *arXiv Prepr. arXiv2310.03016* (2023).
95. Yousri, R. & Safwat, S. How Big Can It Get? A comparative analysis of LLMs in architecture and scaling. in *International Conference on Computer and Applications (ICCA)* 1–5 (2023). (2023).
96. Peng, B., Narayanan, S. & Papadimitriou, C. On limitations of the transformer architecture. *arXiv Prepr. arXiv2402.08164* (2024).
97. Du, W. et al. Stacking Your Transformers: A Closer Look at Model Growth for Efficient LLM Pre-Training. *arXiv Prepr. arXiv2405.15319* (2024).
98. Cao, Y. T. et al. On the intrinsic and extrinsic fairness evaluation metrics for contextualized language representations. *arXiv Prepr. arXiv2203.13928* (2022).
99. Meister, C. & Cotterell, R. Language model evaluation beyond perplexity. *arXiv Prepr. arXiv2106.00085* (2021).

100. Colla, D., Delsanto, M., Agosto, M., Vitiello, B. & Radicioni, D. P. Semantic coherence markers: the contribution of perplexity metrics. *Artif. Intell. Med.* **134**, 102393 (2022).
101. Soni, A. Enhancing Multilingual Table-to-Text Generation with QA Blueprints: Overcoming Challenges in Low-Resource Languages. in *International Conference on Signal Processing and Advance Research in Computing (SPARC)* vol. 1 1–7 (2024). (2024).
102. Chauhan, S. et al. Semantic-syntactic similarity based automatic machine translation evaluation metric. *IETE J. Res.* **70**, 3823–3834 (2024).
103. Mander, S., Phillips, J. & LiSAScore Exploring Linear Sum Assignment on BertScore. in *International Conference on Applications of Natural Language to Information Systems* 249–257 (2024).
104. Shankar, S., Zamfirescu-Pereira, J. D., Hartmann, B., Parameswaran, A. & Arawjo, I. Who validates the validators? aligning llm-assisted evaluation of llm outputs with human preferences. in *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* 1–14 (2024).
105. Wang, Y. et al. The fluency-based semantic network of LLMs differs from humans. *Comput. Hum. Behav. Artif. Hum.* **3**, 100103 (2025).
106. Anderson, C., Vandenberg, B., Hauser, C., Johansson, A. & Galloway, N. Semantic coherence dynamics in large language models through layered syntax-aware memory retention mechanism. (2024).
107. Meng, C., Arabzadeh, N., Askari, A., Aliannejadi, M. & de Rijke, M. Query performance prediction using relevance judgments generated by large language models. *arXiv Prepr. arXiv2404.01012* (2024).
108. Chu, Z., Wang, Z. & Zhang, W. Fairness in large Language models: A taxonomic survey. *ACM SIGKDD Explor. Newsl.* **26**, 34–48 (2024).
109. Bai, G. et al. Beyond efficiency: A systematic survey of resource-efficient large language models. *arXiv Prepr. arXiv2401.00625* (2024).
110. Lukasik, M., Narasimhan, H., Menon, A. K., Yu, F. & Kumar, S. Metric-aware LLM inference. *arXiv Prepr. arXiv2403.04182* (2024).
111. Wolters, C., Yang, X., Schlichtmann, U. & Suzumura, T. Memory Is All You Need: An Overview of Compute-in-Memory Architectures for Accelerating Large Language Model Inference. *arXiv Prepr. arXiv2406.08413* (2024).
112. Stojkovic, J., Zhang, C., Goiri, Í., Torrellas, J. & Choukse, E. Dynamollm: Designing llm inference clusters for performance and energy efficiency. *arXiv Prepr. arXiv2408.00741* (2024).
113. Kenthapadi, K., Sameki, M. & Taly, A. Grounding and evaluation for large language models: Practical challenges and lessons learned (survey). in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 6523–6533 (2024).
114. Laskar, M. T. R. et al. A systematic survey and critical review on evaluating large language models: Challenges, limitations, and recommendations. in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* 13785–13816 (2024). (2024).
115. Chang, Y. et al. A survey on evaluation of large Language models. *ACM Trans. Intell. Syst. Technol.* **15**, 1–45 (2024).
116. Reese, M. L. & Smirnova, A. Comparing ChatGPT and Humans on World Knowledge and Common-sense Reasoning Tasks: A case study of the Japanese Winograd Schema Challenge. in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* 1–9 (2024).
117. Zahraei, P. S., Emami, A. & WSC+ Enhancing The Winograd Schema Challenge Using Tree-of-Experts. *arXiv Prepr. arXiv2401.17703* (2024).
118. Christen, P., Hand, D. J. & Kirielle, N. A. Review of the F-Measure: its history, properties, criticism, and alternatives. *ACM Comput. Surv.* **56**, 1–24 (2023).
119. Jiang, Z., Anastasopoulos, A., Araki, J., Ding, H. & Neubig, G. X-FACTR: Multilingual factual knowledge retrieval from pretrained language models. in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* 5943–5959 (2020). (2020).
120. Shaik, R. & Kishore, K. S. Enhancing Text Generation in Joint NLG/NLU Learning Through Curriculum Learning, Semi-Supervised Training, and Advanced Optimization Techniques. *arXiv Prepr. arXiv2410.13498* (2024).
121. Dong, C. et al. A survey of natural Language generation. *ACM Comput. Surv.* **55**, 1–38 (2022).
122. Kenton, Z. et al. On scalable oversight with weak llms judging strong llms. *arXiv Prepr. arXiv2407.04622* (2024).
123. Huang, Y. et al. New solutions on LLM acceleration, optimization, and application. in *Proceedings of the 61st ACM/IEEE Design Automation Conference* 1–4 (2024).
124. Cassano, F. et al. Knowledge transfer from high-resource to low-resource programming languages for code llms. *Proc. ACM Program. Lang.* 8, 677–708 (2024).
125. Kazi, N. & Kahanda, I. Enhancing Transfer Learning of LLMs through Fine-Tuning on Task-Related Corpora for Automated Short-Answer Grading. in *International Conference on Machine Learning and Applications (ICMLA)* 1687–1691 (2023). (2023).
126. Waisberg, E. et al. GPT-4: a new era of artificial intelligence in medicine. *Ir. J. Med. Sci.* **192**, 3197–3200 (2023).
127. Liu, X. et al. GPT understands, too. *AI Open* (2023).
128. Chitty-Venkata, K. T., Emani, M., Vishwanath, V. & Somani, A. K. Neural architecture search for Transformers: A survey. *IEEE Access.* **10**, 108374–108412 (2022).
129. Wang, B. et al. DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models. in *NeurIPS* (2023).
130. Devlin, J., Chang, M. W., Lee, K., Toutanova, K. & Bert Pre-training of deep bidirectional transformers for language understanding. *arXiv Prepr. arXiv1810.04805* (2018).
131. Tan, Y., Jiang, L., Chen, P., Tong, C. & DQMix-BERT Distillation-aware Quantization with Mixed Precision for BERT Compression. in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 311–316 (2023). (2023). https://doi.org/10.1109/SMC53992.2023.10394642
132. Riaz, M. T., Shah Jahan, M., Khawaja, S. G., Shaukat, A. & Zeb, J. TM-BERT: A Twitter Modified BERT for Sentiment Analysis on Covid-19 Vaccination Tweets. in *2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)* 1–6 (2022). (2022). https://doi.org/10.1109/ICoDT255437.2022.9787395
133. Anggrainingsih, R., Hassan, G. M. & Datta, A. C. E. B. E. R. T. Concise and efficient BERT-Based model for detecting rumors on Twitter. *IEEE Access.* **11**, 80207–80217 (2023).
134. Sohrab, M. G., Asada, M., Rikters, M., Miwa, M. & BERT-NAR-BERT A Non-Autoregressive Pre-Trained Sequence-to-Sequence model leveraging BERT checkpoints. *IEEE Access.* **12**, 23–33 (2024).
135. Lan, Z. et al. Albert: A lite bert for self-supervised learning of language representations. *arXiv Prepr. arXiv11942* (2019). (1909).
136. Tripathy, J. K., Chakkaravarthy, S. S., Satapathy, S. C. & Sahoo, M. Vaidehi, V. ALBERT-based fine-tuning model for cyberbullying analysis. *Multimed Syst.* **28**, 1941–1949 (2022).
137. Chiang, C. H., Huang, S. F. & Lee, H. Pretrained language model embryology: The birth of ALBERT. *arXiv Prepr. arXiv02480* (2020). (2010).
138. Mastropaolo, A. et al. Studying the usage of text-to-text transfer transformer to support code-related tasks. in. *IEEE/ACM 43rd International Conference on Software Engineering (ICSE)* 336–347 (2021). (2021).
139. Ni, J. et al. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv Prepr. arXiv2108.08877* (2021).
140. Phan, L. N. et al. Scifive: a text-to-text transformer model for biomedical literature. *arXiv Prepr. arXiv2106.03598* (2021).

141. Yang, Z. et al. Xlnet: generalized autoregressive pretraining for Language Understanding. *Adv. Neural Inf. Process. Syst.* **32**, 5753–5763 (2019).
142. Topal, M. O., Bas, A. & van Heerden, I. Exploring transformers in natural language generation: Gpt, bert, and xlnet. *arXiv Prepr. arXiv2102.08036* (2021).
143. Adoma, A. F., Henry, N. M. & Chen, W. Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. in *17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* 117–121 (2020). (2020).
144. Thoppilan, R. et al. Lamda: Language models for dialog applications. *arXiv Prepr. arXiv2201.08239* (2022).
145. Morales, L., Herrera, M., Camacho, O., Leica, P. & Aguilar, J. LAMDA control approaches applied to trajectory tracking for mobile robots. *IEEE Access.* **9**, 37179–37195 (2021).
146. Ruiz, F. A., Isaza, C. V., Agudelo, A. F. & Agudelo, J. R. A new criterion to validate and improve the classification process of LAMDA algorithm applied to diesel engines. *Eng. Appl. Artif. Intell.* **60**, 117–127 (2017).
147. Touvron, H. et al. Llama: Open and efficient foundation language models. *arXiv Prepr. arXiv2302.13971* (2023).
148. Zhang, R. et al. Llama-adapter: Efficient fine-tuning of language models with zero-init attention. *arXiv Prepr. arXiv2303.16199* (2023).
149. Sayin, B., Minervini, P., Staiano, J. & Passerini, A. Can LLMs Correct Physicians, Yet? Investigating Effective Interaction Methods in the Medical Domain. *arXiv Prepr. arXiv2403.20288* (2024).
150. Wang, S., Liu, T., Kinoshita, S. & Yokoyama, H. M. LLMs May improve medical communication: social science perspective. *Postgrad. Med. J.* **101**, qgae101 (2024).
151. Kim, Y. et al. Adaptive Collaboration Strategy for LLMs in Medical Decision Making. *arXiv Prepr. arXiv2404.15155* (2024).
152. Wang, Y., Ma, X. & Chen, W. Augmenting black-box llms with medical textbooks for clinical question answering. *arXiv Prepr. arXiv2309.02233* (2023).
153. Goel, A. et al. Llms accelerate annotation for medical information extraction. in *Machine Learning for Health (ML4H)* 82–100 (2023).
154. Aparicio, V., Gordon, D., Huayamares, S. G., Luo, Y. & BioFinBERT Finetuning Large Language Models (LLMs) to Analyze Sentiment of Press Releases and Financial Text Around Inflection Points of Biotech Stocks. *arXiv Prepr. arXiv2401.11011* (2024).
155. Kumar, R., Gattani, D. R. K. & Singh, K. Enhancing Medical History Collection using LLMs. in *Proceedings of the Australasian Computer Science Week* 140–143 (2024). (2024).
156. Wang, Z., Luo, X., Jiang, X., Li, D. & Qiu, L. LLM-RadJudge: Achieving Radiologist-Level Evaluation for X-Ray Report Generation. *arXiv Prepr. arXiv2404.00998* (2024).
157. Garc\'\ia-Ferrero, I. et al. MedMT5: An Open-Source Multilingual Text-to-Text LLM for the Medical Domain. in *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* 11165–11177 (2024). (2024).
158. Yang, R. et al. Large Language models in health care: development, applications, and challenges. *Heal Care Sci.* **2**, 255–263 (2023).
159. Zhou, Z., Yang, T. & Hu, K. Traditional Chinese Medicine Epidemic Prevention and Treatment Question-Answering Model Based on LLMs. in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 4755–4760 (2023). (2023). https://doi.org/10.1109/BIBM58861.2023.10385748
160. Wang, Z., Li, K., Ren, Q., Yao, K. & Zhu, Y. Traditional Chinese Medicine Formula Classification Using Large Language Models. in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 4647–4654 (2023). (2023). https://doi.org/10.1109/BIBM58861.2023.10385776
161. Dou, Y. et al. ShennongGPT: A Tuning Chinese LLM for Medication Guidance. in. *IEEE International Conference on Medical Artificial Intelligence (MedAI)* 67–72 (2023). (2023). https://doi.org/10.1109/MedAI59581.2023.00017
162. Helwan, A., Azar, D. & Ozsahin, D. U. Medical Reports Summarization Using Text-To-Text Transformer. in *Advances in Science and Engineering Technology International Conferences (ASET)* 1–4 (2023). (2023). https://doi.org/10.1109/ASET56582.2023.10180671
163. Cardenas, L., Parajes, K., Zhu, M., Zhai, S. & AutoHealth Advanced LLM-Empowered Wearable Personalized Medical Butler for Parkinson's Disease Management. in *IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)* 375–379 (2024). (2024). https://doi.org/10.1109/CCWC60891.2024.10427622
164. Karttunen, P., Vavekanand, R., Xu, Y., Milani, S. & Li, H. Large Language models in healthcare decision support: A review. *Available SSRN* **4892593** (2023).
165. Krishnan, G. et al. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. *Front. Artif. Intell.* **6**, 1227091 (2023).
166. Kumar, A. et al. A survey on IBM watson and its services. in *Journal of Physics: Conference Series* vol. 2273 12022 (2022).
167. Piotrkowicz, A., Johnson, O. & Hall, G. Finding relevant free-text radiology reports at scale with IBM Watson content analytics: a feasibility study in the UK NHS. *J. Biomed. Semant.* **10**, 21 (2019).
168. Lyu, Y. et al. Gp-gpt: Large language model for gene-phenotype mapping. *arXiv Prepr. arXiv2409.09825* (2024).
169. Kang, I., Van Woensel, W. & Seneviratne, O. Using large Language models for generating smart contracts for health insurance from textual policies. In (Eds. Shaban-Nejad, M. Michalowski, & S. Bianco) *AI for Health Equity and Fairness: Leveraging AI To Address Social Determinants of Health* 129–146 (Springer, 2024).
170. Nazi, Z., Al & Peng, W. Large language models in healthcare and medical domain: A review. in *Informatics* vol. 11 57 (2024).
171. Nankya, M., Mugisa, A., Usman, Y., Upadhyay, A. & Chataut, R. Security and privacy in E-Health systems: A review of AI and machine learning techniques. *IEEE Access.* **12** (2024).
172. Reddy, S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement. Sci.* **19**, 27 (2024).
173. Mennella, C., Maniscalco, U., De Pietro, G. & Esposito, M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon* (2024).
174. Bedi, S. et al. Testing and evaluation of health care applications of large Language models: a systematic review. *JAMA* **10** (2024).
175. Desai, B. & Patil, K. Secure and scalable Multi-Modal vehicle systems: A Cloud-Based framework for Real-Time LLM-Driven interactions. *Innov. Comput. Sci. J.* **9**, 1–11 (2023).
176. Cheng, Z. Q. et al. SHIELD: LLM-Driven Schema Induction for Predictive Analytics in EV Battery Supply Chain Disruptions. *arXiv Prepr. arXiv2408.05357* (2024).
177. Wase, Z. M., Madisetti, V. K. & Bahga, A. Object detection Meets LLMs: model fusion for safety and security. *J. Softw. Eng. Appl.* **16**, 672–684 (2023).
178. Moeini, M., Ahmadian, R. & Ghatee, M. Calibrated SVM for Probabilistic Classification of In-Vehicle Voices into Vehicle Commands via Voice-to-Text LLM Transformation. in *8th International Conference on Smart Cities, Internet of Things and Applications (SCIoT)* 180–188 (2024). (2024).
179. Bilgram, V. & Laarmann, F. Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Eng. Manag Rev.* **51**, 18–25 (2023).
180. Osten, W., Bett, C. & Situ, G. The challenge of making self-driving cars: may AI help to overcome the risks, or should we focus on reliable sensor technologies? in *Interferometry and Structured Light 2024* vol. 13135 8–21 (2024).
181. Li, L. et al. Data-centric evolution in autonomous driving: A comprehensive survey of big data system, data mining, and closed-loop technologies. *arXiv Prepr. arXiv2401.12888* (2024).
182. Sanders, N. R. *Supply Chain Management: A Global Perspective* (John Wiley \& Sons, 2025).

183. Mueller-Saegebrecht, S. & Lippert, I. In Tandem with ChatGPT-4: How LLM Enhance Entrepreneurship Education and Business Model Innovation. in *Academy of Management Proceedings* vol. 2024 15473 (2024).
184. Borah, A. & Rutz, O. Enhanced sales forecasting model using textual search data: fusing dynamics with big data. *Int. J. Res. Mark.* **41** (2024).
185. Yang, Z., Jia, X., Li, H. & Yan, J. Llm4drive: A survey of large language models for autonomous driving. in *NeurIPS 2024 Workshop on Open-World Agents* (2023).
186. Baccari, S., Hadded, M., Ghazzai, H., Touati, H. & Elhadef, M. Anomaly detection in connected and autonomous vehicles: A survey, analysis, and research challenges. *IEEE Access.* **12** (2024).
187. Mart\'\inez, I. *The Future of the Automotive Industry* (Springer, 2021).
188. Abdelati, M. H., Mokbel, E. F. F., Abdelwali, H. A., Matar, A. H. & Rabie, M. Revolutionizing automotive engineering with artificial neural networks: applications, challenges, and future directions. *J. Sci. Insights*. **1**, 155–169 (2024).
189. Garikapati, D. & Shetiya, S. S. Autonomous vehicles: evolution of artificial intelligence and the current industry landscape. *Big Data Cogn. Comput.* **8**, 42 (2024).
190. Muzahid, A. J. M., Zhao, X. & Wang, Z. Survey on Human-Vehicle Interactions and AI Collaboration for Optimal Decision-Making in Automated Driving. *arXiv Prepr. arXiv2412.08005* (2024).
191. Tyagi, A. K., Mishra, A. K. & Kukreja, S. Role of Artificial Intelligence Enabled Internet of Things (IoT) in the Automobile Industry: Opportunities and Challenges for Society. in *International Conference on Cognitive Computing and Cyber Physical Systems* 379–397 (2023).
192. Gao, D. et al. LLMs-based machine translation for E-commerce. *Expert Syst. Appl.* **258**, 125087 (2024).
193. Chen, K. et al. General2Specialized LLMs Translation for E-commerce. in *Companion Proceedings of the ACM on Web Conference 2024* 670–673 (2024).
194. Fang, C. et al. Llm-ensemble: Optimal large language model ensemble method for e-commerce product attribute value extraction. in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* 2910–2914 (2024).
195. Chen, B., Dai, H., Ma, X., Jiang, W. & Ning, W. Robust Interaction-based Relevance Modeling for Online E-Commerce and LLM-based Retrieval. *arXiv Prepr. arXiv2406.02135* (2024).
196. Dam, S. K., Hong, C. S., Qiao, Y. & Zhang, C. A complete survey on llm-based ai chatbots. *arXiv Prepr. arXiv2406.16937* (2024).
197. Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S. & Srinivasan, K. A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: applications, open challenges and future research directions. *Comput. Sci. Rev.* **52**, 100632 (2024).
198. Roumeliotis, K. I., Tselikas, N. D. & Nasiopoulos, D. K. Precision-Driven product recommendation software: unsupervised models, evaluated by GPT-4 LLM for enhanced recommender systems. *Software* **3**, 62–80 (2024).
199. Guyt, J. Y., Datta, H. & Boegershausen, J. Unlocking the potential of web data for retailing research. *J. Retail.* **100**, 130–147 (2024).
200. Soni, V. Large Language models for enhancing customer lifecycle management. *J. Empir. Soc. Sci. Stud.* **7**, 67–89 (2023).
201. Lin, J. et al. How can recommender systems benefit from large language models: A survey. *arXiv Prepr. arXiv2306.05817* (2023).
202. Roumeliotis, K. I., Tselikas, N. D. & Nasiopoulos, D. K. LLMs in e-commerce: a comparative analysis of GPT and LLaMA models in product review evaluation. *Nat. Lang. Process. J.* **6**, 100056 (2024).
203. Wu, L. et al. A survey on large Language models for recommendation. *World Wide Web.* **27**, 60 (2024).
204. Provasi, V. The AI revolution: evaluating impact and consequences in copywriting. (2023).
205. Johnsen, M. *AI in Digital Marketing* (Walter de Gruyter GmbH \& Co KG, 2024).
206. Richey, R. G. Jr, Chowdhury, S., Davis-Sramek, B., Giannakis, M. & Dwivedi, Y. K. Artificial intelligence in logistics and supply chain management: A primer and roadmap for research. *Journal of Business Logistics* vol. 44 532–549 at (2023).
207. Li, Y. et al. Large language models for manufacturing. *arXiv Prepr. arXiv2410.21418* (2024).
208. Latif, E., Fang, L., Ma, P. & Zhai, X. Knowledge distillation of llm for education. *arXiv Prepr. arXiv2312.15842* (2023).
209. Zhang, Z. et al. Simulating Classroom Education with LLM-Empowered Agents. *arXiv Prepr. arXiv2406.19226* (2024).
210. Chen, L. et al. BIDTrainer: An LLMs-driven Education Tool for Enhancing the Understanding and Reasoning in Bio-inspired Design. in *Proceedings of the CHI Conference on Human Factors in Computing Systems* 1–20 (2024).
211. Diez-Rozas, V., Estevez-Ayres, I., Alario-Hoyos, C. & Callejo, P. & Delgado Kloos, C. A Web Application for a Cost-Effective Fine-Tuning of Open-Source LLMs in Education. in *International Conference on Artificial Intelligence in Education* 267–274 (2024).
212. Ouyang, Z., Jiang, Y. & Liu, H. The effects of Duolingo, an AI-Integrated technology, on EFL learners' willingness to communicate and engagement in online classes. *Int. Rev. Res. Open. Distrib. Learn.* **25**, 97–115 (2024).
213. Shahzad, T. et al. A comprehensive review of large Language models: issues and solutions in learning environments. *Discov Sustain.* **6**, 27 (2025).
214. Upadhyay, A., Farahmand, E., Muñoz, I., Akber Khan, M. & Witte, N. Influence of LLMs on learning and teaching in higher education. *Available SSRN* **4716855** (2024).
215. Chen, Z. et al. Evolution and prospects of foundation models: from large Language models to large multimodal models. *Comput. Mater. \& Contin* **80**, 1753 (2024).
216. Xu, R. et al. Knowledge conflicts for llms: A survey. *arXiv Prepr. arXiv2403.08319* (2024).
217. Dai, S. et al. Bias and unfairness in information retrieval systems: New challenges in the llm era. in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 6437–6447 (2024).
218. Razafinirina, M. A., Dimbisoa, W. G. & Mahatody, T. Pedagogical alignment of large Language models (LLM) for personalized learning: A survey, trends and challenges. *J. Intell. Learn. Syst. Appl.* **16**, 448–480 (2024).
219. Liu, S., Guo, X., Hu, X. & Zhao, X. Advancing generative intelligent tutoring systems with GPT-4: design, evaluation, and a modular framework for future learning platforms. *Electronics* **13**, 4876 (2024).
220. Yousefi, M., Mullick, J. & Wang, Q. Preparing teachers and students for the challenges of society 5.0: integration of cognitive computing, learning analytics, and gamification of learning. *Preconceptions Policies Strateg Challenges Educ.* **5.0**, 75–99 (2024).
221. Wu, S. et al. Bloomberggpt: A large language model for finance. *arXiv Prepr. arXiv2303.17564* (2023).
222. de Zarzà, I., de Curtò, J., Roig, G. & Calafate, C. T. Optimized financial planning: integrating individual and cooperative budgeting models with LLM recommendations. *AI* **5**, 91–114 (2023).
223. Zhang, B., Yang, H., Zhou, T., Babar, A. & Liu, X. Y. M. Enhancing financial sentiment analysis via retrieval augmented large language models. in *Proceedings of the fourth ACM international conference on AI in finance* 349–356 (2023).
224. de Moraes, D. et al. S. Using Zero-shot Prompting in the Automatic Creation and Expansion of Topic Taxonomies for Tagging Retail Banking Transactions. *arXiv Prepr. arXiv2401.06790* (2024).
225. Policepatil, S. et al. IGI Global,. Financial Sector Hyper-Automation: Transforming Banking and Investing Procedures. in *Examining Global Regulations During the Rise of Fintech* 299–318 (2025).
226. Reda, M. A. Intelligent Assistant Agents: Comparative Analysis of Chatbots through Diverse Methodologies. *GSJ* 12, (2024).
227. Alagic, A. et al. Machine learning for an enhanced credit risk analysis: A comparative study of loan approval prediction models integrating mental health data. *Mach. Learn. Knowl. Extr.* **6**, 53–77 (2024).
228. Saxena, A., Verma, S. & Mahajan, J. Transforming banking: the next frontier. In (eds. Saxena, A., Verma, S. & Mahajan, J.) *Generative AI in Banking Financial Services and Insurance: A Guide To Use Cases, Approaches, and Insights* 85–121 (Springer, 2024).
229. Quinonez, C. & Meij, E. A new era of AI-assisted journalism at Bloomberg. *AI Mag* **45** (2024).

230. Johnsen, M. *Developing AI Applications With Large Language Models*Maria Johnsen,. (2025).
231. Abdali, S., He, J., Barberan, C. J. & Anarfi, R. Can llms be fooled? investigating vulnerabilities in llms. *arXiv Prepr. arXiv2407.20529* (2024).
232. Das, B. C., Amini, M. H. & Wu, Y. Security and privacy challenges of large Language models: A survey. *ACM Comput. Surv.* **57** (2024).
233. Nie, Y. et al. A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges. *arXiv Prepr. arXiv2406.11903* (2024).
234. MindySupport. 9 Cool Case Studies of Global Brands Using LLMs and & Generative, A. I. at (2024). https://hackernoon.com/9-cool-case-studies-of-global-brands-using-llms-and-generative-ai
235. Carneros-Prado, D. et al. Comparative study of large language models as emotion and sentiment analysis systems: A case-specific analysis of GPT vs. IBM Watson. in *International Conference on Ubiquitous Computing and Ambient Intelligence* 229–239 (2023).
236. Chow, J. C. L., Wong, V., Sanders, L. & Li, K. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson assistant platform. in *Healthcare* vol. 11 2417 (2023).
237. Dong, X. L., Moon, S., Xu, Y. E., Malik, K. & Yu, Z. Towards next-generation intelligent assistants leveraging llm techniques. in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 5792–5793 (2023).
238. Martin, A. Artificial Intelligence Transformations in Digital Advertising: Historical Progression, Emerging Trends, and Strategic Outlook. (2024).
239. Zhao, H. et al. A Comprehensive Survey of Large Language Models in Management: Applications, Challenges, and Opportunities. *Challenges, Oppor. (August 14*, (2024). (2024).
240. Agrawal, S., Trenkle, J. & Kawale, J. Beyond Labels: Leveraging Deep Learning and LLMs for Content Metadata. in *Proceedings of the 17th ACM Conference on Recommender Systems* 1 (2023).
241. Jeffri, J. A. C. & Tamizhselvi, A. Enhancing Music Discovery: A Real-Time Recommendation System using Sentiment Analysis and Emotional Matching with Spotify Integration. in *8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* 1365–1373 (2024). (2024).
242. Pearson, S. Computational advertising for meaningful brands, the public purpose, and a sustainable ecology: A call for research into a systems approach and modeling applications of LLMs in marketing and advertising. *J. Curr. Issues \& Res. Advert.* **45**, 357–367 (2024).
243. Pesl, R. D., Stötzner, M., Georgievski, I. & Aiello, M. Uncovering LLMs for Service-Composition: Challenges and Opportunities. in *International Conference on Service-Oriented Computing* 39–48 (2023).
244. Jiao, J., Afroogh, S., Xu, Y., Phillips, C. & Navigating, L. L. M. Ethics: Advancements, Challenges, and Future Directions. *arXiv Prepr. arXiv2406.18841* (2024).
245. Wu, F., Zhang, N., Jha, S., McDaniel, P. & Xiao, C. A new era in llm security: Exploring security concerns in real-world llm-based systems. *arXiv Prepr. arXiv2402.18649* (2024).
246. Rojas, S. Evaluating the environmental impact of large Language models: sustainable approaches and practices. *Innov. Comput. Sci. J.* **10**, 1–6 (2024).
247. Ivanov, Y. Understanding the inner workings of large Language models: interpretability and explainability. *MZ J. Artif. Intell.* **1**, 1–5 (2024).
248. Chen, C. & Shu, K. Combating misinformation in the age of Llms: opportunities and challenges. *AI Mag.* **45**, 354–368 (2024).
249. Bowen, D. et al. Data Poisoning in LLMs: Jailbreak-Tuning and Scaling Laws. *arXiv Prepr. arXiv2408.02946* (2024).
250. Musser, M. A cost analysis of generative language models and influence operations. *arXiv Prepr. arXiv2308.03740* (2023).
251. Liu, Y., Cao, J., Liu, C., Ding, K. & Jin, L. Datasets for large language models: A comprehensive survey. *arXiv Prepr. arXiv2402.18041* (2024).
252. Kausik, B. N. Scaling Efficient LLMs. *arXiv Prepr. arXiv2402.14746* (2024).
253. Long, D. X. et al. LLMs Are Biased Towards Output Formats! Systematically Evaluating and Mitigating Output Format Bias of LLMs. *arXiv Prepr. arXiv2408.08656* (2024).
254. Hassani, S. Enhancing Legal Compliance and Regulation Analysis with Large Language Models. *arXiv Prepr. arXiv2404.17522* (2024).
255. Zhang, X., Li, S., Hauer, B., Shi, N. & Kondrak, G. Don't Trust ChatGPT when Your Question is not in English: A Study of Multilingual Abilities and Types of LLMs. *arXiv Prepr. arXiv2305.16339* (2023).
256. Hamzah, F. & Sulaiman, N. Multimodal integration in large language models: A case study with mistral llm. (2024).
257. Perković, G., Drobnjak, A. & Botički, I. Hallucinations in llms: Understanding and addressing challenges. in *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* 2084–2088 (2024).
258. Pressman, S. M. et al. AI and ethics: a systematic review of the ethical considerations of large language model use in surgery research. in *Healthcare* vol. 12 825 (2024).
259. Stamboliev, E. & Christiaens, T. How empty is trustworthy AI? A discourse analysis of the ethics guidelines of trustworthy AI. *Crit. Policy Stud.* **19**, 1–18 (2024).
260. D\'\iaz-Rodr\'\iguez, N. et al. Connecting the Dots in trustworthy artificial intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation. *Inf. Fusion*. **99**, 101896 (2023).
261. Hickman, E., Petrin, M. & Trustworthy AI and corporate governance: the EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. *Eur. Bus. Organ. Law Rev.* **22**, 593–625 (2021).


## Acknowledgements

The Authors are thankful to their respective universities.

## Author contributions

M.R. and Z.J.; Methodology, Resources, Visualization, Validation, writing original draft, writing review and editing, Formal analysis, M.A.S. and M.B.R.; Investigation, Supervision, Methodology, Resources, Visualization, Validation, writing original draft, Writing review and editing, Formal analysis, Conceptualization, Funding acquisition and administration. M.J.S.; Methodology, Resources, Visualization, Writing – review and editing, validation.

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## Additional information

**Correspondence** and requests for materials should be addressed to M.B.R.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025