Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects

Muhammad Usman Hadi[1,*], Qasem Al-Tashi[2,*], Rizwan Qureshi[2,*], Abbas Shah[3], Amgad Muneer[2], Muhammad Irfan[4], Anas Zafar[5], Muhammad Bilal Shaikh[6], Naveed Akhtar[7], Mohammed Ali Al-Garadi[8], Jia Wu[2], and Seyedali Mirjalili[9,10]

[1]*School of Engineering, Ulster University, Belfast, BT15 1AP, United Kingdom (m.hadi@ulster.ac.uk)*
[2]*Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA (qaal@mdanderson.org; frizwan@mdanderson.org; amabdulraheem@mdanderson.org; JWu11@mdanderson.org)*
[3]*Department of Electronics Engineering, Mehran University of Engineering and Technology, Jamshoro,76062 Pakistan (zaigham.shah@faculty.muet.edu.pk)*
[4]*Faculty of Electrical Engineering, Ghulam Ishaq Khan Institute (GIKI) of Engineering Sciences and Technology, Swabi, 23460 Pakistan (mirfan@giki.edu.pk)*
[5]*Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan (anaszafar98@gmail.com)*
[6]*Center for Artificial Intelligence and Machine Learning (CAIML), Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Perth, Australia (mbshaikh@our.ecu.edu.au)*
[7]*Computing and Information Systems, The University of Melbourne, 700 Swanston Street, Carlton 3010, VIC Australia*
[8]*Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA*
[9]*Centre for Artificial Intelligence Research and Optimization, Torrens University Australia, Fortitude Valley, Brisbane, QLD 4006, Australia (ali.mirjalili@torrens.edu.au)*
[10]*University Research and Innovation Center, Obuda University, 1034 Budapest, Hungary*

Abstract

Within the vast expanse of computerized language processing, a revolutionary entity known as Large Language Models (LLMs) has emerged, wielding immense power in its capacity to comprehend intricate linguistic patterns and conjure coherent and contextually fitting responses. Large language models (LLMs) are a type of artificial intelligence (AI) that have emerged as powerful tools for a wide range of tasks, including natural language processing (NLP), machine translation, vision applications, and question-answering. This survey provides a comprehensive overview of LLMs, including their history, architecture, training methods, applications, and challenges. We begin by discussing the fundamental concepts of generative AI and the architecture of generative pre-trained transformers (GPT). We then provide an overview of the history of LLMs, their evolution over time, and the different training methods that have been used to train them. We then discuss the wide range of applications of LLMs, including medical, education, finance, engineering, media, entertainment politics, and law. It also discusses how LLMs are shaping the future of AI and their increasing role in scientific discovery, and how they can be used to solve real-world problems. Next, we explore the challenges associated with deploying LLMs in real-world scenarios, including ethical considerations, model biases, interpretability, and computational resource requirements. We also highlight techniques for enhancing the robustness and controllability of LLMs and addressing bias, fairness, and quality issues in Generative AI. Finally, we conclude by highlighting the future of LLM research and the challenges that need to be addressed in order to make this technology more reliable and useful. This survey is intended to provide researchers, practitioners, and enthusiasts with a comprehensive understanding of LLMs, their evolution, applications, and challenges. By consolidating the state-of-the-art knowledge in the field, this article is anticipated to serve as a valuable resource for learning the current state-of-the art as well as further advancements in the development and utilization of LLMs for a wide range of real-world applications. The GitHub repo for this project is available at Github-Repo.

Index Terms

Large Language Models, Large Vision Models, Generative AI, Conversational AI, LangChain, Natural language processing, Computer Vision, GPT, ChatGPT, Bard, AI chatbots

Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects

I. Introduction

Language modeling (LM) is a fundamental task in natural language processing (NLP) that aims to predict the next word or a character in a given sequence of text [1], [2]. It involves developing algorithms and models that can understand and generate coherent human language. The primary objective of LM is to capture the probability distribution of words in a language, which allows the model to generate new text [3], complete sentences [4], and predict the likelihood of different word sequences [5], [6]. Early language models, such as n-gram models [7], were based on simple statistical techniques that estimated the probabilities of word sequences using frequency counts [8], [9]. However, with the rise of deep learning in NLP [10], the availability of enormous amounts of public datasets [11], and powerful computing devices [12] to process these big data with complex algorithms, has led to the development of large language models.

Large Language Models (LLMs) [13], sometimes referred to as "transformative [14]" or "next-generation [15]" language models, represent a significant breakthrough in NLP [16]. These models leverage deep learning techniques, particularly transformer architectures [17], to learn and understand the complex patterns and structures present in language data [18]. A key characteristic of LLMs is their ability to process vast amounts of data, including unstructured text, and capture semantic relationships between words and phrases [19]. These models can also process visual, audio [20], audiovisual [21], as well as multi-modal data [22] and learn the semantic relationships between them. These models have significantly enhanced the capabilities of machines to understand and generate human-like language [23].

The history of LLMs can be traced back to the early development of language models and neural networks [24]. The journey begins with the era of statistical language models [25]. In this stage, researchers primarily relied on probabilistic approaches [26] to predict word sequences. Classic examples include n-grams, Hidden Markov Models (HMMs) [27] and Maximum Entropy Models [28]. N-grams, for instance, are sequences of adjacent words or tokens that are used to predict the likelihood of the next word based on the preceding ones. While rudimentary by today's standards, these models marked a crucial starting point in the field of natural language understanding. They allowed for basic text generation and word prediction but were limited in their ability to capture complex contextual relationships [29] [30]. [31]. Then a shift towards more data-driven methodologies has been witnessed [32]. Researchers began to explore machine learning algorithms to improve language understanding [33] These models learned patterns and relationships within large text corpora. Support Vector Machines (SVMs) is a notable example from this [34]. Machine learning models brought a more sophisticated approach to NLP tasks, allowing for the development of applications like spam detection [35] and sentiment analysis [36].

The emergence of deep learning marked a pivotal moment in the development of LLMs [37]. Neural networks, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks gained prominence [38]. These deep learning architectures delved deeper into the data, allowing them to capture more intricate features and long-range dependencies within text. This stage significantly improved the models' ability to understand context, making them suitable for tasks like machine translation and speech recognition [16], [39]. However, deep learning also faced challenges with vanishing gradients [40] and long-term dependencies [41], limiting their effectiveness.

The breakthrough in LLMs came with the introduction of the Transformer architecture in the seminal work "Attention is All You Need" by Vaswani et al. in 2017 [42]. The Transformer model, based on the self-attention mechanism [43], enabled parallelization and efficient handling of long-range dependencies. It laid the foundation for models like OpenAI's GPT (Generative Pre-trained Transformer) series [44] and BERT (Bidirectional Encoder Representations from Transformers) [45] by Google, which achieved groundbreaking results in a wide range of language tasks. These mechanisms enabled models to consider the entire context of a sentence or document, allowing for true contextual understanding [46]. Transformer-based models, often pre-trained on massive text corpora, can generate coherent and contextually relevant text, revolutionizing applications like chatbots [47], text summarization [48], and language translation [49].

Since then, LLMs have undergone several developmental stages, with models increasing in size and complexity. The GPT series, starting with GPT-1 and continuing with GPT-2 and GPT-3 [50], has successively grown in the number of parameters, allowing for more sophisticated language understanding and generation capabilities [51] . Likewise, BERT-inspired models have seen advancements in pre-training strategies, such as ALBERT [52] (A Lite BERT) and RoBERTa [53], which further improved performance and efficiency.

Furthermore, advancements in LLMs have extended to more specific domains, with models designed for specialized tasks like medical language processing [54], scientific research [55], website development [56] and code generation [57]. Moreover,

TABLE I: List of Acronyms and corresponding definitions.

Acronym	Definition
AI	Artificial Intelligence
AGI	Artificial General Intelligence
BBH	Big Bench Hard
BERT	Bidirectional Encoder Representations from Transformers
CV	Computer Vision
CTRL	Conditional Transformer Language Model
FFF	Fused Filament Fabrication
GANs	Generative Adversarial Networks
GNMT	Google Neural Machine Translation
GPT	Generative Pre-Trained transformers
GPT-3	Generative Pre-trained Transformer 3
GPT-4	Generative Pre-trained Transformer 4
GPUs	Graphical Processing Units
GRUs	Gated Recurrent Units
LLaMA	Large Language Model Meta AI
LLM	Large Language Models
LM	Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
MLM	Masked Language Modeling
NSP	Next Sentence Prediction
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
PLMs	Pre-trained Language Models
RLHF	Reinforcement Learning Human Feedback
RNN	Recurrent neural networks
RNNLM	Recurrent neural network language model
SLMs	Statistical Language Models
T2V	Text to video
T5	Text-to-Text Transfer Transformer
TPUs	Tensor Processing Units
USMLE	United States Medical Licensing Exam
VL-PTMs	Vision-Language Pre-trained Models
XLNet	eXtreme Language Understanding Network

efforts have been made to address ethical concerns [58], interpretability [59], and reducing biases in LLMs to ensure responsible and equitable use [60]. The development stages of large models have witnessed a constant quest for larger models, improved pre-training strategies, and specialized domain adaptations [61], [62]. As research continues, the potential applications and impact of LLMs on various fields, including education, healthcare, and human-computer interaction, continue to expand, inspiring further innovations and advancements.

In summary and as can be seen from Fig 1; LM research has received widespread attention and has undergone four significant development stages including: statistical language models, machine learning models, deep learning models and transformer-based models. In this research, we mainly focus on LLMs and foundation AI models for language and vision tasks. A list of commonly used acronyms in this article with definitions is given in Table I.

Modern language model called ChatGPT [63] was developed by OpenAI [64]. It is based on the GPT-3.5 architecture [65] and was trained using a sizable amount of internet-sourced text data, including books, articles, wikis and websites (Table II) [66]. ChatGPT is exceptional at producing human-like responses and having conversations with users. In computer vision (CV), researchers are actively engaged in the development of vision-language models inspired by the capabilities of ChatGPT. These models are specifically designed to enhance multimodal dialogues, where both visual and textual information are important [67]. Moreover, the advancements in the field have led to the introduction of GPT-4 [65], which has further expanded the capabilities of language models by seamlessly integrating visual information as part of the input. This integration of visual data empowers the model to effectively understand and generate responses that incorporate both textual and visual cues, enabling more contextually rich and nuanced conversations in multimodal settings.

A. Survey Motivation

The revolutionary ChatGPT has captivated the attention of the community, sparking a wealth of fascinating reviews and discussions on the advancements of LLMs and artificial intelligence [68], [69], [70], [71], [72], [73]. For example, the role of ChatGPT in education is evaluated in [74], healthcare and medicine in [75], [54], protein sequence modeling in [76] and protein generation in [77], a survey on generative AI in [78], and scientific text modeling in [79] and text generation in [80]. The use of LLM in finance is evaluated in [81], impact on labor market in [82], on code writing capabilities in [83], deep fakes in [84], legal aspects in [85], AI for drug discovery in [86], and ML for cancer biomarkers in [87]. The advancements in pre-training, fine-tuning, utilization and capability evaluation of LLMs is presented in [68] and a survey on autonomous agents in [88]. The recent progress in visio-language pre-trained models is discussed in [69] and the knowledge graphs construction and reasoning are explained in [89], selection inference is in [59], and quantum-inspired machine learning in [90]. The survey vision language mpre-trained models [69] presents an overview of various techniques for encoding raw images and texts into single-modal embeddings as a fundamental aspect, also discusses prevalent architectures of Vision-Language Pre-trained Models (VL-PTMs), focusing on their ability to effectively model the interaction between text and image representations.

Despite the growing number of studies on LLMs, there remains a scarcity of research focusing on their technical intricacies and effective utilization. Also, the field is progressing at a very fast pace, so a review article will contribute a lot to the field. Therefore we write this paper in the form of a tutorial. Our primary objective is to explore, learn, and evaluate language models across various domains. We delve into the working principles of language models, analyze different architectures of the GPT family, and discuss strategies for their optimal utilization. Furthermore, we provide detailed insights into writing prompts, and visual prompting techniques, leveraging GPT-plug-ins, and harnessing other AI/LLM tools. These aspects are generally not covered by the existing related articles. Our comprehensive examination also encompasses a discussion on the limitations associated with LLMs, including considerations related to security, ethics, economy, and the environment. In addition, we present a set of guidelines to steer future research and development in the effective use of LLMs. We hope that this paper will contribute to a better understanding and utilization

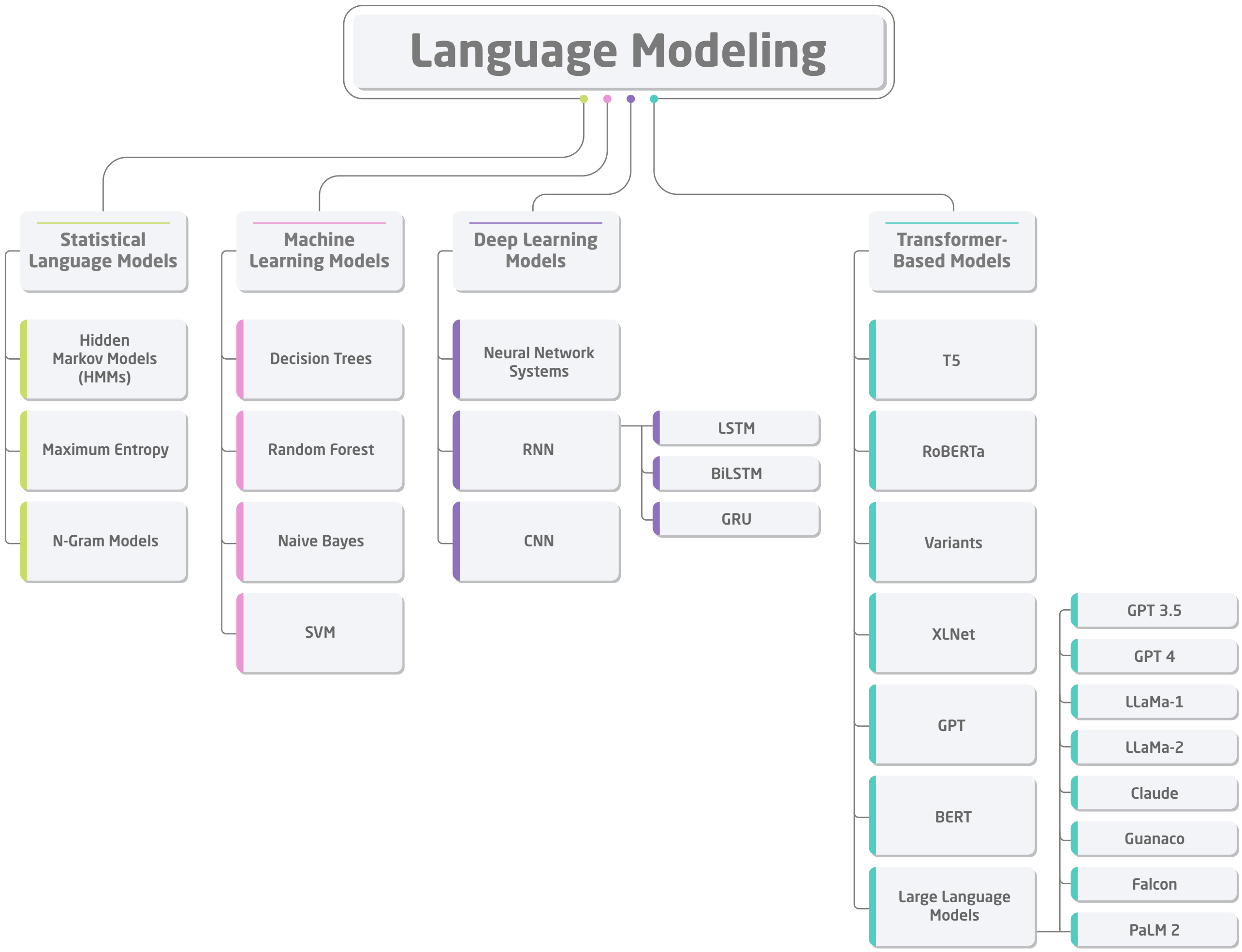


Fig. 1: Types of language modeling. The division of LLMs is categorized into four major blocks: Statistical language models, Machine learning models, Deep learning models and Transformer-based models.

of LLMs.

B. Contributions

The main contributions of this article are as follows:

1) Providing a comprehensive overview of LLMs, including their technical details, advancements, challenges, capabilities and limitations.
2) Presenting a state-of-the-art analysis and comparison of different LLMs.
3) Addressing ethical concerns about LLMs, including their computational requirements and potential for perpetuating biases. We also discuss the limitations of LLMs; including, limited understanding of the physical world, tokenization problems, infomration hallucination, fine-tuning and risk of foundation models.
4) Offering insights into the future potential of LLMs and their impact on society and demonstrating the applications of LLM through four practical use cases in the fields of medicine, education, finance, law, politics, media, entertainment, engineering, and others.
5) This article is uniquely presented in a manner to promote practical usage of LLMs, showcasing the actual LLM outputs to corroborate the discussions.

The paper is organized as the following sections. Section II provides an introduction to the role of AI in creativity, specifically focusing on generative pre-trained transformers and their significance. Section III presents an overview of LLMs, summarizing a brief history of LLMs and discussing their training and functionality. Taxonomy of LLMs is presented in Section IV and major applications of LLM through different use cases is discussed in Section V. Section VI explores AI-enabled tools that are expected to shape the future. Section VII discusses the practical use cases of GPT plugins and their potential to enhance user productivity and efficiency. Section VIII presents guidelines and working examples using prompting techniques. Section IX proposes the limitations and drawbacks of the current state-of-the-art LLM. Section X presents the impact of LLM on humans and society. Section XI-D presents expert opinions on the subject matter and the authors' perspective on open unanswered avenues. Section XII concludes the survey paper. The overall structure

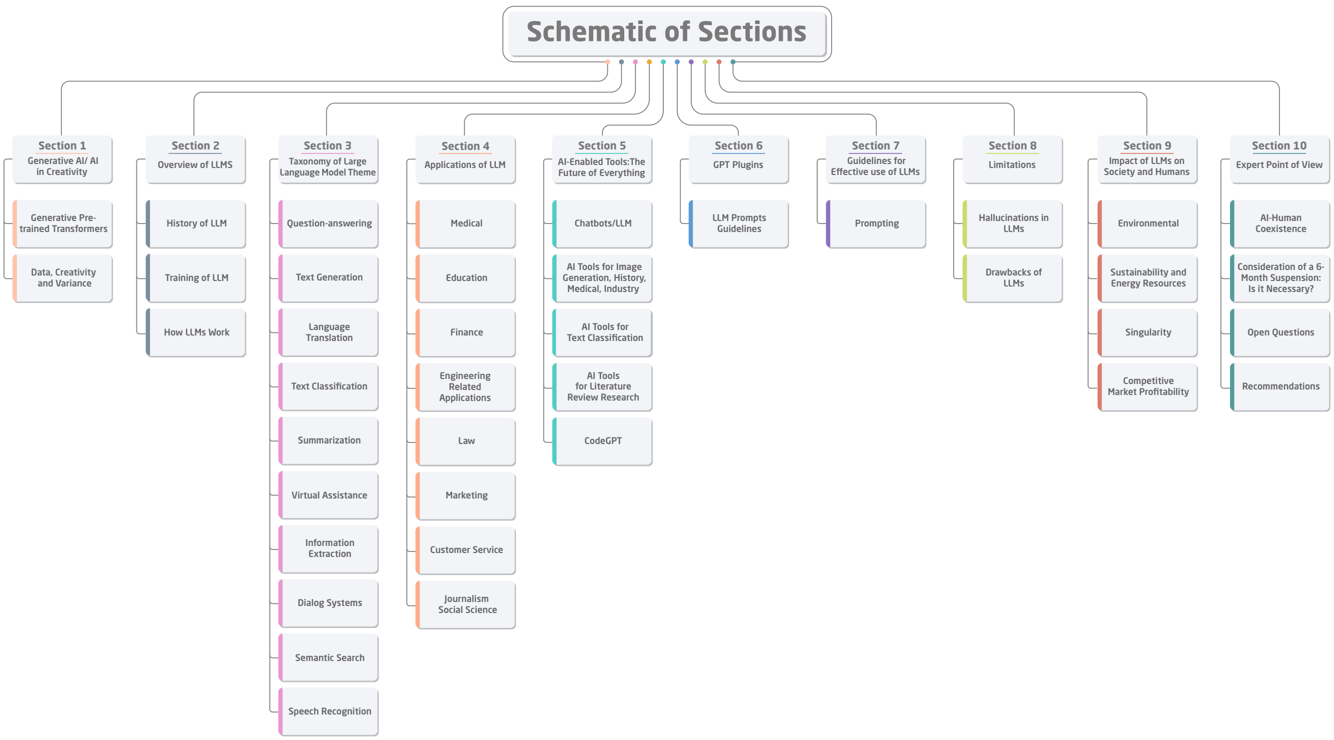


Fig. 2: Schematic Representation of Article Sections - A visual overview of the key sections comprising the structure of the article, providing readers with a roadmap for navigating the content effectively.

TABLE II: **Pre-training data**. Mixtures of data used for pre-training LLaMA [15].

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3TB
C4	15.0%	1.06	783GB
Github	4.5%	0.64	328GB
Wikipedia	4.5%	2.45	83GB
Books	4.5%	2.23	85GB
ArXiv	2.5%	1.06	92GB
Stock Exchange	2.0%	1.03	78GB

of the article is presented in Fig. 2 for a quick reference at a glance.

II. Generative AI

Generative AI (GenAI) [91] is perhaps the most disruptive [92] and generalized technology of this decade [93], already influenced many industries, including, Media [94], Marketing [95], Game development and Metaverse [96], Education [97], Software development [98], and Medical [99]. Unlike general AI systems that perform specific tasks such as data classification [100], clustering [101], object detection [102] and segmentation [103] or predictions [104]; GenAI can generate meaningful new content of multiple data modalities [105]; including, text [3], speech [106], images [107], and videos [108]. Some common examples of GenAI systems are image generators (Midjourney or stable diffusion), Chatbots (ChatGPT, Bard, Palm), code generators (CodeX, Co-Pilot [109]) audio generators(VALL-E)Vall-e [110], and video generators (Gen-2) [111]

During the past few years, GenAI models size has been scaled from a few million parameters(BERT [45], 110M) to hundreds of billions of parameters (GPT [112], 175B). Generally speaking, as the size of the model (number of parameters) increases, the performance of the model also increases [113], and it can be generalized for a variety of tasks [114], for example, Foundation models [115]. However, smaller models can also be fine-tuned for a more focused task [116].

LLMs, such as ChatGPT, Google Bard and Llama, are a type of GenAI that is specifically designed to generate human-like language in response to a given prompt [117]. These models are trained on massive amounts of data (see Table II), using techniques such as unsupervised learning to learn the statistical patterns of language. However, many people accord the capabilities provided by GPT models to “more data and computing power” instead of “better ML research" [118].

GenAI works by leveraging complex algorithms and statistical models to generate new content that mimics the patterns and characteristics of the training data [119]. These algorithms may include probabilistic techniques; such as Autoregressive model [120] and Variations Auto-encoders [121], or more recently, Generative Adversarial Networks [122] and Diffusion models [123] or Reinforcement Learning Human Feedback (RLHF) [124] .

- - -

LLMs have captured significant interest in recent years due to their remarkable performance across an extensive array of NLP tasks, including text generation, translation, summarization, question-answering, and sentiment analysis [125]. Constructed upon the foundation of the transformer architecture [42], these models exhibit an extraordinary capacity to process and generate human-like text by leveraging massive volumes of training data for various topics [126]

A. Data, Generation, and Variance

As discussed previously, LLMs are developed by training large deep neural networks using text data made up of multiple different sources including but not limited to books, social media and also text from the internet such as poetry, songs, news articles etc. This diversity in the training mix allows the model to provide output text that is coherent just as a human written text may read. However, it should be noted that the "creativity" exhibited in LLMs goes beyond the regurgitation of data that it may have seen during the training process [127]. To produce text creatively, the deep learning model of the LLM needs to form an understanding of the text used in its training with respect to language, tone and writing patterns etc [128]. This way the LLM is able to generate responses to user queries that are creative and genuine in terms of language writing by combining the different types of input information, it ingested during training, together to generate meaningful results for the provided query.

A fundamental criterion for gaining the capability of this creativity is to have sufficient variance, which indicates to the model's ability to produce an unexpected output. In short, variance ensures that there is randomness in a model's output, and it is introduced to enable it to generate sufficiently good results over a range of output results. By introducing variance in the model, one can increase the diversity of output content generated which goes beyond the scope of what the training data consisted of [129].

It is acknowledged that since the release and mainstreaming of LLMs by users of all walks of life, some have complained of LLM getting stuck in a cycle of similar answers, especially if a complex query has been asked of it multiple times. For e.g., Microsoft found that its Bing AI powered by ChatGPT tends to get repetitive in its responses after 15 consecutive chats [1]. While this problem has been mitigated since then by taking measures such as refreshing context and/or introducing limits to the questions asked per session, this does question the variance capability of LLMs, and also it will be interesting to see whether the text written by Chatbots misses the essence of human [130]?

However, from a philosophical perspective, it could be of interest how the size of the human population affects the variance of this space of potential of different fingerprints for humans? For example, if there were only 100 people in the world, the space of potential fingerprints would be much smaller, and it would be more likely that two people would have the same fingerprint [131].

III. Overview of LLMs

In this Section, we briefly discuss the history and training of LLMs.

[1]https://www.zdnet.com/article/long-chats-confuse-bing-chat-so-microsofts-chatgpt-powered-bot-is-getting-new-limits/

A. History of LLM

LLMs are a type of AI model that can process and generate natural language text. These models are typically trained on massive amounts of text data and use deep learning techniques to learn the patterns and structures of language [132]. The history of LLMs can be traced back to the early days of NLP research [133]. The first language models were developed in the 1950s and 1960s. These models were rule-based and relied on hand-crafted linguistic rules and features to process language. They were limited in their capabilities and were not able to handle the complexity of NLP [134]. In the 1980s and 1990s, statistical language models were developed. These models used probabilistic methods to estimate the likelihood of a sequence of words in a given context. They were able to handle larger amounts of data and were more accurate than rule-based models [135]. However, they still had limitations in their ability to understand the semantics and context of language [136]. The next major breakthrough in language modeling came in the mid-2010s with the development of neural language models [137]. These models used deep learning techniques to learn the patterns and structures of language from large amounts of text data. The first neural language model was the recurrent neural network language model (RNNLM) [39], which was developed in 2010. RNNLM was able to model the context of words and produce more natural-sounding text than previous models [138]. In 2015, Google introduced the first large-scale neural language model called the Google Neural Machine Translation (GNMT) system

The development of LLMs continued with the introduction of the Transformer model in 2017 [42]. The Transformer was able to learn the longer-term dependencies in language and allowed for parallel training on multiple Graphical Processing Units (GPUs), making it possible to train much larger models [139]. The release of OpenAI's GPT-1 [140] in 2018, marked a significant advance in NLP with its transformer-based architecture. With 117 million parameters, GPT-1 could generate contextually relevant sentences, demonstrating the potential of transformers in revolutionizing NLP tasks [141]. While GPT-1 had its limitations, it set the stage for subsequent, more powerful models, propelling a new era of AI research and highly-competitive research in LLMs (see Fig. 3).

In 2020, OpenAI released GPT-3 [142], which was able to generate highly coherent and natural-sounding text [143]. GPT-3 demonstrated the potential of LLMs for a wide range of NLP tasks [144]. Inspired by the success of GPT-3, OpenAI released the next iteration of their language model, GPT-4 [145] with the ability to generate even more coherent and natural-sounding text. Following GPT-4's success, Meta also introduced Llama [15], a family of open-source foundation models.

B. Training of LLMs

Training large language models involves several key steps that are fundamental to their successful development. The process typically begins with the collection and preprocessing of a large amount of text data from diverse sources, such as books, articles, websites, and other textual corpora (see Table.

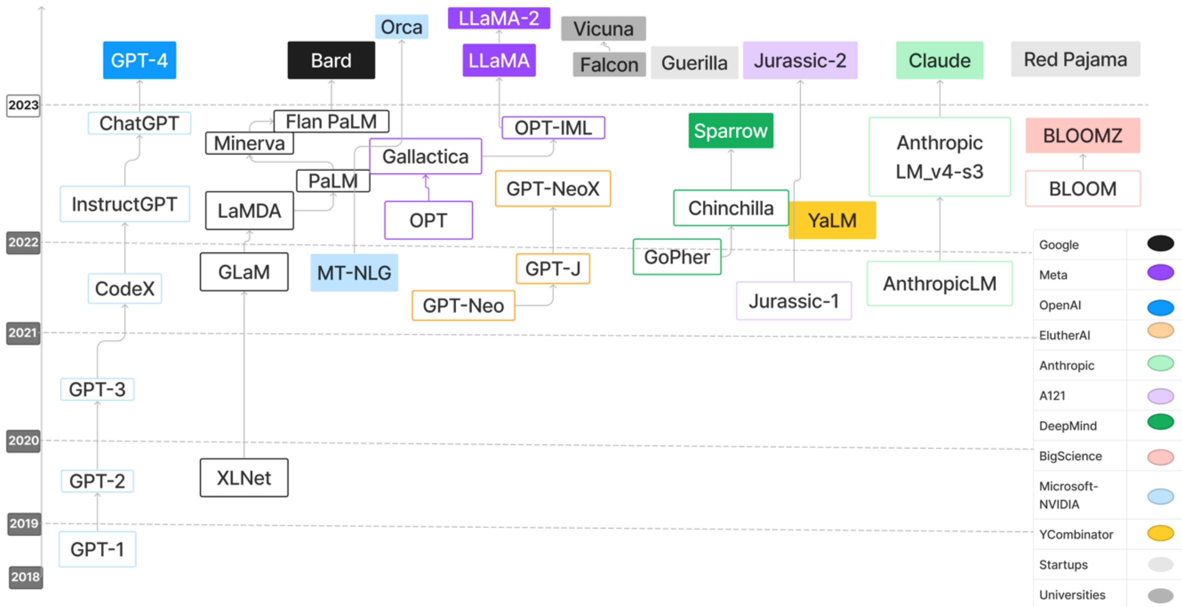


Fig. 3: Illustration of the evolution of Large Language Models (LLMs) over time, highlighting their development across a range of research and commercial organizations. Starting from the initial advancements made in this field, the figure maps out the journey of LLMs, outlining the key milestones, breakthroughs, and model iterations along the way.

TABLE III: State-of-the-art for LLM training pipeline [146]. Notations: RM: Reward Modeling, RL: Reinforcement Learning, SFT: Supervised Fine-tuned.

Stage	Pretraining	Supervised-Finetuning	Reward Modeling	Reinforcement Learning
Dataset	Raw Internet II	Demonstration	Comparisons	Prompts
Algorithm	Language Modeling	Language Modeling	Binary Classification	Reinforcement Learning
Model	Base Model	SFT Model	RM Model	RL Model
Resources	100s of GPUs months of training deployable	1-100 of GPUs days of training deployable	1-100 of GPUs days of training not deployable	1-100 of GPUs days of training deployable

III). The curated dataset [147] serves as the foundation for training the LLMs. After the removal of duplicates [148], noisy and poisonous data [149] and ensuring privacy reduction [150], the training process involves unsupervised learning, where the model learns to predict the next word in a sequence given the preceding context assuming the language generation as a random process [151]. This task is commonly referred to as language modeling. Currently, LLMs utilize Transformers which enable them to model long-range dependencies [152], understand text data [153] enable them to generate new content in the style and characteristics of a genre or author [154]. The training objective is to optimize the model's parameters to maximize the likelihood of generating the correct next word in a given context [68]. This optimization is typically achieved through an algorithm called stochastic gradient descent (SGD) [155] or its variants, combined with backpropagation [156], which computes gradients to update the model's parameters iteratively.

- Bidirectional Encoder Representations from Transformer (BERT): BERT is a prominent language model with significantly advanced NLP tasks. Its training process comprises pretraining and fine-tuning stages [157]. During pretraining, BERT learns a general language representation from large-scale unlabeled text data. It employs masked language modeling (MLM) and next-sentence prediction (NSP) tasks. MLM involves masking a portion of input tokens and training the model to predict the original masked tokens, fostering bidirectional context understanding [158]. NSP trains BERT to predict whether a second sentence follows the first, enhancing coherence comprehension. After pretraining, BERT undergoes fine-tuning on specific tasks with labeled data. Fine-tuning tailors BERT's learned representations to target tasks, such as sentiment analysis or named entity recognition. It employs backpropagation and gradient descent optimization to update model parameters. Training BERT demands significant computational resources [141], utilizing high-performance hardware like GPUs or Tensor Processing Units (TPUs) or field programmable gate arrays (FPGAs) [159], [160], [161]. Techniques such as layer normalization, residual connections, and attention mechanisms inherent in the transformer architecture further enhance BERT's capacity to capture intricate dependencies and long-range contextual relationships.
- eXtreme Language understanding Network (XLNet): XLNet is a generalized autoregressive pre-training method

that surpasses the limitations of traditional left-to-right or right-to-left language modeling. XLNet is trained using a permutation-based approach that differs from traditional autoregressive models [162]. In the training process, rather than predicting the next word given the previous words in a fixed order, XLNet considers all possible permutations of the input sequence and models the probability of each permutation. This allows XLNet to capture dependencies in both directions, thus addressing the limitations of sequential left-to-right or right-to-left modeling [163]. The training of XLNet involves two key steps: unsupervised pretraining and supervised fine-tuning. During unsupervised pretraining, XLNet learns to predict words conditioned on the entire input context by maximizing the expected log-likelihood over all possible permutations. This is achieved using a variant of the transformer architecture, similar to models like BERT. The permutation-based objective function used in XLNet training presents unique challenges. Unlike traditional autoregressive models that can rely on the causal order of words for prediction, XLNet needs to consider all possible permutations, resulting in an exponentially large number of training instances. This makes the training process computationally intensive and requires efficient strategies, such as "factorized sampling," to sample a subset of permutations during each training iteration. Another difficulty in training XLNet is the need for large-scale computing resources [68], [164], [165]. The vast number of possible permutations and the large model size contribute to increased memory and computation requirements. Training XLNet often necessitates distributed training on multiple GPUs or TPUs and can take significant time [68].

- Text-to-Text Transfer Transformer (T5): T5, developed by Google, is a versatile language model that is trained in a "text-to-text" framework. The training process of T5 involves two main steps: pretraining and fine-tuning. During pretraining, T5 is trained on a massive corpus of publicly available text from the internet. The objective is to learn a generalized representation of language that can be applied to a wide range of tasks. The key innovation of T5 is the formulation of all tasks as text generation problems. This means that every task, including text classification, summarization, translation, and question answering, is cast into a text-to-text format. For example, instead of training T5 to answer questions directly, it is trained to generate the complete answer given the question and relevant context. In the pretraining phase, T5 is trained using a variant of the transformer architecture. The transformer model allows T5 to capture long-range dependencies and effectively model the contextual relationships in the input text [166]. The pretraining objective is here is also typically based on maximum likelihood estimation, where T5 is trained to predict the target text given the source text. Once pretraining is complete, T5 undergoes fine-tuning on specific downstream tasks [166]. One of the challenges in training T5 is the availability of large-scale labeled datasets for various tasks. Fine-tuning requires task-specific labeled data, and the quality and quantity of the data play a crucial role in the model's performance [68]. Additionally, the computational resources required to train T5 can be substantial, as the model is computationally intensive due to its transformer architecture and the size of the pre-trained parameters.
- Conditional Transformer Language Model (CTRL): CTRL is a language model designed to generate text based on specific control codes or prompts. It is trained using a two-step process: pretraining and fine-tuning. During pretraining, CTRL is trained on a large corpus of publicly available text data [167]. The objective of pretraining is to teach the model to understand and generate coherent text based on different control codes or prompts [168].

  The model is trained to predict the next word or phrase in a given context, learning the statistical patterns and linguistic structures of the language. One of the unique aspects of CTRL is its conditioning of control codes or prompts. These control codes guide the model's text generation process, allowing users to specify the desired style, topic, or other characteristics of the generated text. The control codes act as explicit instructions to guide the model's behavior during both training and inference. The fine-tuning phase of CTRL is crucial for adapting the model to specific tasks or domains. Fine-tuning involves training the pre-trained CTRL model on task-specific datasets with control codes. The model is exposed to task-specific prompts and is trained to generate text that aligns with the desired output or behavior for the given task.

IV. Taxonomy of Large Language Model

A. Question-answering

Question-answering (QA) systems [169] allow users to obtain direct answers to questions posed in natural language. LLMs have become a key component in building robust QA systems. LLMs can be effectively pretrained on large text corpora and then fine-tuned on QA labeled datasets. This adapts them to extract or generate answers from passages of text. The broad linguistic knowledge learned during pretraining allows LLMs to understand the semantics of questions and reason about potential answers. Fine-tuning on QA data teaches the models to identify relevant context passages and output the correct response. Key benefits of using LLMs include handling complex questions, synthesizing answers from multiple context documents, and generating clarifying responses when a query is ambiguous. LLM-based QA systems achieve high accuracy on benchmark datasets [170] , surpassing previous state-of-the-art methods. They can be deployed via voice assistants, search engines, and other interfaces to provide users with quick access to information through natural dialog. Ongoing research is focused on improving reasoning abilities, explainability, and efficiency of LLM question answering.

B. Text Generation

Text generation is a useful application of large language models, which can automate the process of generating content for various purposes [171] , such as articles, blogs,research papers, social media posts, product descriptions, source codes, emails, and more. With their ability to comprehend and generate natural language, these models can produce high-quality content that is both accurate and coherent.

C. Language Translation

LLMs possess the capability to translate text from one language to another with exceptional accuracy and fluency [172]. This feature is beneficial for a range of users,including language service providers, global companies, and individuals, who can utilize these models for real-time translation, localization, and overcoming language barriers in communication. The impressive accuracy and fluency of these models make them a valuable tool for facilitating effective communication across different languages and cultures. This feature has the potential to enhance global collaboration and increase access to information, making it an important area of research and development in the field of NLP [173].

D. Text Classification

In addition to their text generation and translation abilities, LLMs are also equipped with exceptional organizational capabilities, such as text classification, analysis, and categorization based on predefined labels or topics [174], [175] . This feature enables the models to effectively manage large volumes of textual data, making them highly valuable for a range of tasks such as sentiment analysis, spam detection, content moderation, and customer feedback analysis. By automating these processes, language models can streamline data management, reduce manual labor, and improve the accuracy and efficiency of analysis. These capabilities are particularly useful for businesses and organizations that deal with large amounts of textual data and require effective methods for organizing and analyzing it.

E. Summarization

LLMs possess the ability to generate concise and coherent summaries of lengthy texts or documents [176]. This feature is highly advantageous for a variety of purposes, such as writing news articles, research papers, legal documents, and other types of content where extracting essential information is crucial. Summarization by language models can save time and effort, while ensuring that the most important points are captured accurately. This feature has the potential to enhance the efficiency and effectiveness of content creation, making it a valuable tool for individuals and organizations alike. Further research and development in this area can lead to improvements in the quality and accuracy of summaries generated by language models.

F. Virtual Assistance

In the realm of virtual assistants and chatbots [177], LLMs play a critical role. These models possess the ability to comprehend user queries, provide relevant information, and engage in natural language conversations. This capability enables virtual assistants and chatbots to assist with customer support, offer personalized recommendations, answer questions, and automate routine tasks, thus enhancing user experiences and increasing operational efficiency. By leveraging large language models, virtual assistants, and chatbots can provide highly effective and responsive support to users while also reducing the workload for human operators [178] . This area of research and development is of significant importance, as it has the potential to transform the way users interact with technology and improve the effectiveness and efficiency of customer support and service delivery.

G. Information Extraction (IE)

The use of LLMs in IE is significant for populating knowledge bases. By leveraging fine-tuned LLMs, entities such as people, organizations, and locations, as well as the relationships between them, can be accurately extracted from unstructured text. This process can facilitate the creation of structured knowledge graphs that can be utilized for various purposes [179]. In addition, LLMs assist in event extraction, enabling the identification of key occurrences described in text documents [180]. This feature has the potential to enhance the efficiency and accuracy of information extraction, making it a valuable tool for businesses and organizations that deal with large amounts of textual data. Further research and development in this area can lead to improvements in the quality and effectiveness of IE using LLMs.

H. Dialog systems

In the context of dialog systems, large language models (LLMs) play a crucial role in facilitating language understanding. The development of large pretrained models like Google's Meena and Microsoft's Blender has led to significant improvements in the naturalness and coherence of open-domain chatbots [181] . These models possess the ability to generate informative, interesting, and harmless responses, making conversational agents much more usable. The application of LLMs in dialog systems has the potential to transform the way users interact with technology, creating more engaging and effective conversational experiences [182]. Further research in this area can lead to improvements in the quality and effectiveness of dialog systems, making them even more valuable for a range of applications and industries.

I. Semantic Search

In the field of Semantic Search, query understanding is of utmost importance, and LLMs are unparalleled in their ability to discern the underlying intent and meaning of user search queries [183]. This ability enables next-generation search capabilities that go beyond simple keyword matching. For instance, LLMs can recognize that the phrases "best budget laptop"; and "affordable student computer" convey the same information need. This feature has the potential to enhance the accuracy and relevance of search results, making it easier for users to find the information they need. Further research and development in the area of Semantic Search can lead to

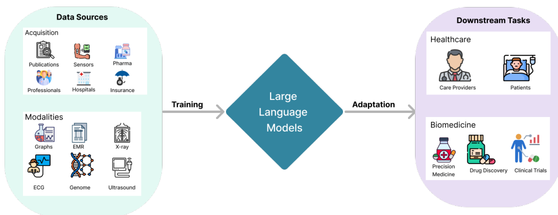


Fig. 4: Illustration of an interactive framework where LLMs enable various tasks across healthcare and biomedicine when trained on multimodal data generated by various sources in the healthcare ecosystem.

the creation of more effective and efficient search systems, making LLMs a valuable tool for a range of applications and industries [184].

J. Speech recognition

Automated speech recognition is a crucial aspect of voice interfaces and transcription. While traditional systems relied on hidden Markov models or Gaussian mixture models, the emergence of deep learning has seen large neural network models like LLMs take center stage in state-of-the-art results. LLMs that are pretrained on massive text corpora offer rich linguistic knowledge pertaining to language structure, context, and word relationships. Fine-tuning these models on transcribed speech data using connectionist temporal classification loss enables them to learn acoustic-to-text mappings. This leads to significant improvements in the accuracy of automated speech transcription, even in the presence of accented speech or domain specific vocabulary [185].The contextual knowledge and continual learning abilities of LLMs make them ideally suited for handling the variability and ambiguity inherent in speech signals. As LLMs continue to increase in scale, they are becoming the standard for building high performance and robust automated speech recognition systems [186] .

To sum up the discussion on different LLMs, Table IV provides information on the performance of various LLMs on different reasoning tasks.

V. Applications of Large Language Models

Given LLMs wide range of applications, in this section, we provide a discussion of their use in the fields of medicine, education, finance, and engineering. The selection of medical, education, finance, and engineering as the applications for LLM is based on their significance, relevance, and potential impact within their respective domains. These applications demonstrate the versatility and potential of LLMs in addressing complex challenges and supporting human endeavors.

A. Medical

LLMs like ChatGPT have exhibited remarkable potential in diverse healthcare applications [195]. They have been successfully employed in medical education, radiologic decision-making, clinical genetics, and patient care [196],[197]. In medical education, ChatGPT has emerged as an interactive tool that aids learning and problem-solving [198], [199]. Notably, ChatGPT's performance in the United States Medical Licensing Exam (USMLE) was comparable to or exceeded the passing threshold, without requiring specialized training or reinforcement [198]. Moreover, ChatGPT's explanations displayed a high level of concordance and insightful understanding [196]. A database for Covid-19 drug repurposing using NLP is proposed in [200].

According to Rao et al. [201], it is anticipated that specialized AI-based clinical decision-making tools will emerge in the future. This study emphasizes the potential of ChatGPT in radiologic decision-making, highlighting its feasibility and potential benefits in improving clinical workflow and ensuring responsible use of radiology services. Similarly, Kung et al. [196] concluded in their research that LLMs, including ChatGPT, have the capacity to enhance the delivery of individualized, compassionate, and scalable healthcare. These models can assist in medical education and potentially aid in clinical decision-making.

In the domain of clinical genetics, Duong and Solomon [202] found that ChatGPT's performance did not significantly differ from humans when answering genetics-related questions. However, the model demonstrated better accuracy on memorization-type questions compared to questions requiring critical thinking. Notably, this study also highlighted that ChatGPT provided varying answers when asked the same question multiple times, providing plausible explanations for both correct and incorrect responses. Furthermore, Fijacko [203] evaluated ChatGPT's accuracy in answering questions related to life support and resuscitation. The findings revealed that ChatGPT demonstrated the ability to provide accurate answers to a majority of the questions on the American Heart Association's Basic Life Support and Advanced Cardiovascular Life Support exams.

In the field of neurosurgical research and patient care, LLMs has been investigated for its potential role in various aspects, including gathering patient data, administering surveys or questionnaires, and providing information about care and treatment. The integration of biotechnology and AI applications to address global challenges and advance sustainable development goals is examined in [204]. These applications encompass decision support, NLP, data mining, and machine learning. The authors underscore the significance of reproducibility in the development of AI models and highlight ongoing research issues and challenges in these domains . Furthermore, AI-powered chatbots hold the potential to enhance patient outcomes by facilitating communication between patients and healthcare professionals. Leveraging NLP, these chatbots can provide patients with information about their care and treatment in a more accessible manner [205].

There are several tools already in use that allows the system to interact with patients such as Ada Health [206], Babylon Health [207], and Buoy Health [208]. The recent popularity of LLMs can potentially not only improve patient confidence in interacting with such chatbots but also improve upon the

TABLE IV: Comparison of LLMs' Reasoning Performance. Notations: MMLU [187]: high school and college knowledge, GSM8K: elementary school math, MATH: very hard math and natural science. All current models struggle, BBH [188]: a collection of 27 hard reasoning problems, HumanEval [189]: a classical dataset for evaluating coding capability, C-Eval [190]: a collection of 52 disciplines of knowledge test in Chinese, TheoremQA [191]: a question-answering dataset driven by STEM theorems. [192], [187], [15], [193], [194], [190]

Model	Param.	Type	GSM8K	MATH	MMLU	BBH	HumanEval	C-Eval	TheoremQA
GPT-4	-	RLHF	92.0	42.5	86.4	-	67.0	68.7*	43.4
claude-v1.3	-	RLHF	81.8*	-	74.8*	67.3*	-	54.2*	24.9
PaLM-2	-	Base	80.7	34.3	78.3	78.1	-	-	31.8
GPT-3.5-turbo	-	RLHF	74.9*	-	67.3*	70.1*	48.1	54.4*	30.2
claude-instant	-	RLHF	70.8*	-	-	66.9*	-	45.9*	23.6
text-davinci-003	-	RLHF	-	-	64.6	70.7	-	-	22.8
code-davinci-002	-	Base	66.6	19.1	64.5	73.7	47.0	-	-
text-davinci-002	-	SIFT	55.4	-	60.0	67.2	-	-	16.6
Minerva	540B	SIFT	58.8	33.6	-	-	-	-	-
Flan-PaLM	540B	SIFT	-	-	70.9	66.3	-	-	-
Flan-U-PaLM	540B	SIFT	-	-	69.8	64.9	-	-	-
PaLM	540B	Base	56.9	8.8	62.9	62.0	26.2	-	-
LLaMA	65B	Base	50.9	10.6	63.4	-	23.7	38.8*	-
PaLM	64B	Base	52.4	4.4	49.0	42.3	-	-	-
LLaMA	33B	Base	35.6	7.1	57.8	-	21.7	-	-
InstructCodeT5+	16B	SIFT	-	-	-	-	35.0	-	11.6
StarCoder	15B	Base	8.4	15.1	33.9	-	33.6	-	12.2
Vicuna	13B	SIFT	-	-	-	-	-	-	12.9
LLaMA	13B	Base	17.8	3.9	46.9	-	15.8	-	-
Flan-T5	11B	SIFT	16.1*	-	48.6	41.4	-	-	-
Alpaca	7B	SIFT	-	-	-	-	-	-	13.5
LLaMA	7B	Base	11.0	2.9	35.1	-	10.5	-	-
Flan-T5	3B	SIFT	13.5*	-	45.5	35.2	-	-	-

services provided. In fact, there are tools developed to assist medical practitioners. One such tool is XrayGPT [209], it can be used for automated analysis of X-ray images and have the user/patient ask questions about the analysis. Through the chats, the user can get insight into their condition through an interactive chat dialogue. Another big development is the segment any thing (SAM) model by meta, which may be fine-tuned for a variety of medical images tasks [210]. In drug discovery domain, DrugGPT [211] is developed, which can design potential ligands, targeting specific proteins, using text prompts.

1) Foundation models for generalist medical AI: In [212] proposes a new paradigm for medical AI, generalist medical AI (GMAI). GMAI models are trained on large, diverse datasets of medical data, and they are able to perform a wide range of tasks, such as diagnosis, prognosis, and treatment planning. The authors of the paper evaluate the performance of GMAI models on a variety of medical tasks. GMAIL models are able to outperform traditional medical AI models on a number of tasks, including diagnosis, prognosis, and treatment planning.

In [170], introduced MultiMedQA, a new benchmark dataset for evaluating LLMs on clinical tasks. MultiMedQA combines six existing medical question answering datasets spanning professional medicine, research, and consumer queries. They introduced the concept of instruction prompt tuning [213]. is a promising approach for aligning LLMs to specific clinical domains. This approach could be used to improve the performance of LLMs on a variety of clinical tasks.

B. Law

Despite the vast influence of law, from family court to environmental policy, and over 1.3 million lawyers in the U.S., legal aid remains largely inaccessible. The services are often too expensive, resulting in around 86 percent of low-income Americans reporting insufficient or no legal help. Moreover, public defenders are usually overburdened and under-resourced, with a majority of public defender offices surpassing the recommended limit of cases per attorney. However, technology might present a solution to this problem, especially through the implementation of LLMs. These models could potentially mitigate the procedural and financial hurdles to legal services, thereby enhancing access to justice and government services. Legal applications pose unique computational challenges due to the specificity of legal language and the varied and unclear standards applied to diverse, unprecedented scenarios. Furthermore, labeled training data is often hard to come by due to high costs. Despite these challenges, foundational models might be especially well-suited to address them due to their adaptability and ability to learn from few examples. While these models can utilize various forms of evidence like audio, video, images, and text, they would primarily be beneficial for text-based legal tasks. However, before these models can be deployed in the legal or government context, significant ethical, legal, and fairness considerations must be addressed. The system places a particular emphasis on transparency, accountability, and explainability, making it crucial to scrutinize these models thoroughly before their application.

C. Education

Offering top-tier, inclusive education on a large scale presents significant societal and economic hurdles. With the cost of education per student escalating faster than general economic costs, resources for student learning become scarce.

For instance, in the United States, student-held private education debt has skyrocketed to 1.6 trillion dollars, surpassing total credit card debt. In the face of these challenges, the digital age brings hope through computational approaches to education, aiming to enhance the efficacy of learners and teachers. Artificial Intelligence (AI) has emerged as a promising tool for education, with applications including providing meaningful feedback to students, aiding teacher improvement, and designing personalized and adaptive learning experiences tailored to individual students' needs. However, implementing tech solutions to effectively scale quality education inclusively is an immense task. In this context, general-purpose Large models that are applicable across multiple tasks and subjects could provide a solution. Foundation models have already begun improving performance in certain educational tasks. For instance, MathBERT [214] has been utilized for "knowledge tracing" – tracking a student's understanding over time based on past responses, and for the "feedback challenge" – interpreting a student's answer to a structured open-ended task. The question is, can foundation models trigger even more significant changes in education? And what are the associated risks of applying these models in an educational context? It's crucial to initiate the discussion around these models with ethical considerations. Khan Academy [2] has been an early adapter of GPT-4 based LLMs working as online tutors becoming the largest case study for the evaluations of LLMs in an educational context. [212] have proposed generalist medical AI (GMAI) models as shown in Figure 4, which will be capable of carrying out a diverse set of tasks using very little or no task-specifc labelled data. Built through self-supervision on large, diverse datasets, GMAI will fexibly interpret diferent combinations of medical modalities, including data from imaging, electronic health records, laboratory results, genomics, graphs or medical text. GMAI Models will in turn produce expressive outputs such as free-text explanations, spoken recommendations or image annotations that demonstrate advanced medical reasoning abilities.

The impact of AI on education has been a topic of much discussion in recent years. One area where AI is having a significant impact is in the realm of student assignments and exams. Since the advent of ChatGPT developed by OpenAI, the way students interact with educational materials, assignments and coursework has become different [215] [216] [217]. The accuracy rate for the exams discussed in [215] was below 70 percent indicating its inability to pass the AHA exams. However, this conclusion was drawn due to a design limitation in their study, where they only generated a single response using ChatGPT, introducing bias and severely underestimating ChatGPT's capabilities in this domain. However, the latest study revealed that ChatGPT's accuracy rate increased to 96 and 92.1 percent for the Basic Life Support (BLS) and Advanced Cardiovascular Life Support (ACLS) exams, respectively, allowing ChatGPT to pass both exams with outstanding results [218]. One of the main advantages of using ChatGPT and AI bots in education is that they can help students complete their assignments more efficiently [219]. ChatGPT is capable of generating high-quality responses to a wide range of prompts, which can save students time and effort when they are working on assignments. Additionally, AI bots can help to automate the grading process, which can reduce the workload for teachers and enable them to provide more detailed feedback to students.

Another advantage of using ChatGPT and AI bots in education is that they can provide personalized learning experiences for students. AI bots can analyze a student's performance on previous assignments and exams and use this data to generate personalized recommendations for future work. This can help students to identify their strengths and weaknesses and focus their efforts on areas where they need to improve. Khan Academy, a nonprofit educational organization, has shown interest in utilizing ChatGPT for its business. They have developed an AI chatbot called Khanmigo, which serves as a virtual tutor and classroom assistant [220]. The goal of incorporating ChatGPT into their platform is to enhance tutoring and coaching experiences by providing one-on-one interactions with students. The incorporation of AI in tutoring and teaching proves that it can be a valuable tool in reducing negativity, particularly the perception that its main purpose is for cheating. Undoubtedly, AI technology is still in its nascent phase, yet it shows great potential in supporting students and catering to their individual requirements. [221].

However, there are also some potential drawbacks to using ChatGPT and AI bots in education. One concern is that these technologies may lead to a loss of creativity and critical thinking skills among students. If students rely too heavily on AI bots to complete their assignments and exams, they may not be developing the skills necessary to think critically and solve problems on their own [219].

1) Learning in the age of AI: Another major assistance that these bots such as ChatGPT can offer is the provision of assistance in designing a course in an academic setting. AI chatbots can serve as a valuable tool to aid in various aspects of syllabus preparation. Course objectives can be generated, relevant topics identified, curricula structured, learning resources gathered and reviewed, assessment methods defined, engaging learning activities established, and a well-balanced course schedule created. The iterative process of interacting with ChatGPT enables refinement and enhancement of the syllabus based on the model's suggestions and insights. It is important to note that ChatGPT acts as a supportive tool, augmenting the expertise and input of experienced educators. The collaboration between human and AI in the course syllabus design process facilitates the development of comprehensive and effective learning plans that align with desired learning outcomes.

2) Major issues for AI in Education: One of the major concerns is the utilization of these tools without proper training. It is crucial to address the issue of inadequate training and contextual fine-tuning for LLMs, as their potential utilization without such preparations raises significant concerns [222]. While it is true that LLMs possess the ability to provide answers to a wide range of questions and assist users in generating responses effortlessly, it is essential for students

[2]https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/

and scientist [223] to receive adequate training specific to their needs in order to fully harness the capabilities of LLMs. Neglecting the necessity for context-specific training and fine-tuning can render these tools less effective and limit their true potential.

Another concern is that the use of AI bots in education could lead to increased inequality [224]. Students who have access to these technologies may have an unfair advantage over those who do not, which could exacerbate existing inequalities in education. Additionally, the use of AI bots could lead to a decrease in the number of teaching jobs available, which could further widen the gap between those who have access to education and those who do not [225]. In conclusion, the use of ChatGPT and AI bots in education has both pros and cons. While these technologies can help students complete assignments more efficiently and provide personalized learning experiences, they may also lead to a loss of critical thinking skills and increased inequality. As AI continues to transform the field of education, it will be important to carefully consider these potential benefits and drawbacks and work to minimize the discussed negative consequences that may arise.

D. Finance

LLMs are making significant advancements in the finance industry [226] with applications ranging from financial NLP tasks [227], risk assessment, algorithmic trading [228], market prediction [229] and financial reporting [230]. LLM's such as BloombergGPT[81], a 50 billion parameter large language model trained on large diversified financial corpus, has revolutionized financial NLP tasks such as news classification, entity recognition and question answering. By utilizing the huge amount of financial data available, it is able to enhance customer services drastically by efficiently handling customer queries and providing them with excellent financial advisory.

In addition, LLMs are being used for risk assessment and management, by analyzing past market trends and data, it is able to identify potential risks and provide mitigation steps through different financial algorithms. Financial institutions can use it for better decision making such as credit risk assessment [231], loan approvals and investments [232]. Algorithmic Trading [233] is another application that can leverage LLM models to identify potential opportunities in the trading market by using its predictive and analyzing capabilities.

However, due to the sensitivity of the financial information and privacy concerns, techniques like data encryption, redaction and data protection policies should be implemented so that these LLMs can be used efficiently in accordance with data protection policies. In this regard, a recent proposition suggested is FinGPT [234] which is an open-source LLM tailored for finance. It is expected that more work will be carried out in this space.

E. Engineering related applications

LLMs have gained substantial attention across various fields, and their potential applications in engineering domains are increasingly being explored. For instance, ChatGPT has diverse applications in software engineering, including code generation, debugging, software testing, NLP, documentation generation, and collaboration. It enables developers to generate code snippets, identify and fix errors, generate test cases, analyze user requirements, create user interfaces, generate software documentation, and facilitate collaboration within development teams. ChatGPT's language understanding and generation capabilities enhance efficiency, streamline workflows, and foster effective communication in software engineering.

In software engineering, ChatGPT can be employed to generate code snippets based on natural language descriptions of desired functionality. This feature saves developers time and improves overall efficiency, allowing them to focus on higher-level design aspects [235]. Additionally, ChatGPT can assist in debugging code by leveraging its language understanding capabilities to identify errors and suggest potential fixes, thereby streamlining the debugging process and reducing development time. The use of ChatGPT extends to software testing, where it can generate test cases and test data based on natural language descriptions of desired test scenarios. This approach enhances the efficiency and effectiveness of software testing, ensuring comprehensive coverage and accurate validation of the software's functionality.

The possibility of ChatGPT utilization to various calculations in mechanical engineering was attempted in Tiro [236]. However, Tiro encountered instances where incorrect procedures, formulas, or results were provided. None of the tasks yielded an exact solution, leading them to discontinue further research. Based on Tiro findings, it can be concluded that, at the current stage of AI development, ChatGPT should not be relied upon for solving engineering practice problems. Furthermore, caution should be exercised in using ChatGPT for such applications, as incorrect results can have potential consequences.

In Mathematics, some attempts have been made such as Wardat et al [237] found that ChatGPT holds potential for assisting in teaching mathematics by providing interactive and dynamic learning experiences. It can generate customized examples and problem-solving strategies tailored to individual student needs, fostering personalized learning. Moreover, it can serve as a virtual tutor, offering real-time feedback and guidance, identifying areas of difficulty, and suggesting alternative approaches. As an AI language model, ChatGPT is capable of performing mathematical calculations and solving math equations. However, the accuracy and effectiveness of ChatGPT solutions may depend on various factors such as the complexity of the equation, the accuracy of the input data, and the instructions given to ChatGPT. Frieder et al., [238] has investigated the mathematical capabilities of ChatGPT by testing it on publicly available datasets, as well as hand-crafted ones, and measuring its performance against other models trained on a mathematical corpus, such as Minerva. They also test whether ChatGPT can be a useful assistant to professional mathematicians by emulating various use cases that come up in the daily professional activities of mathematicians (question answering, theorem searching).

However, it is essential to acknowledge the limitations of ChatGPT, including the possibility of generating incorrect responses or failing to address complex mathematical concepts adequately. Therefore, it should be utilized as a supplemental tool alongside traditional teaching methods and human supervision to ensure accuracy and quality in teaching mathematics.

In manufacturing, Wang et al. [239] conducted an evaluation of ChatGPT's capabilities in supporting design, manufacturing, and engineering education tasks. The results indicate that ChatGPT is impressive in providing information, generating coherent and structured content, and proposing initial solutions. The authors recommended a technology development roadmap to successfully integrate ChatGPT into the manufacturing industry. Therefore, in manufacturing, ChatGPT struggles to understand questions and lacks the ability to properly use knowledge to generate correct solutions and it can even fabricate non-existing rules or equations in order to generate solutions.

Similarly, Badini et al. [240], performed a study in additive manufacturing troubleshooting and evaluated ChatGPT's expertise in technical matters, focusing on the evaluation of printing parameters and bed detachment, warping, and stringing issues for Fused Filament Fabrication (FFF) methods using thermoplastic polyurethane polymer as feedstock material. It was found that ChatGPT provided remarkable accuracy, correctness, and organization in its responses and its approach to problem-solving offered valuable insights in addressing hurdles. In particular, for the specific technical issues of warping, bed detachment, and stringing, ChatGPT demonstrated its ability to provide hierarchical and logically organized responses while taking into account given information and constraints. Furthermore, it was also able to fine-tune printing parameters for different types of TPU filaments, showing its ability to relate the mechanical properties of the filament material to the printing parameters. Finally, the authors recommended integrating ChatGPT into an Additive Manufacturing software platform to provide real-time suggestions and optimization for users, which can enhance the efficiency and quality of the Additive Manufacturing process.

F. Media and Entertainment Industry

The media and entertainment sector is currently undergoing a transformative phase that revolves around data and prioritizes consumer-centric experiences [241]. Companies of all sizes are now striving to introduce groundbreaking innovations that enable personalized, one-to-one interactions on a large scale [242], [243]. Among the various technologies driving this change, LLMs stand out as a game-changer. LLMs not only enable the creation of original content but also demonstrate a profound grasp of intricate information and the ability to simulate human-like interactions. This includes MediaGPT, a large language model for the Chinese media domain was presented recently, which can generate high-quality and relevant outputs for various tasks in the Chinese media domain [244]. Similarly, Robertuito [245] was proposed for Spanish social media.

Content lies at the core of the Media and Entertainment industry, and its creation is evolving with the integration of data, particularly social signals, into content strategies. LLMs have a transformative role, revolutionizing how companies leverage data and AI for content development and curation. LLMs assist in generating captivating headlines, compelling copy, and providing real-time content feedback, streamlining production and enhancing quality. Large AI models can also be utilized for generating attractive advertisements and marketing [82], political speeches, slogans and social media posts [246], and promotional videos [95].

Similarly, leading entertainment networks and applications are using LLM based algorithms that can analyze user data to offer personalized recommendations for movies, TV shows, and music. This helps entertainment companies to retain customers and improve their engagement with their content [247]. Moreover, LLMs automate content curation fostering user satisfaction, retention, and monetization. Recently, many companies have developed and offered their services for media and entertainment purposes. One of the prime examples of such services is Dolly, an LLM-trained model developed by databricks Incorporation [248].

The creation of AI-based newscasters [249] is a recent concept that consists of virtual news presenters or anchors that are generated using AI technologies, particularly LLMs [250]. In April 2023, a Kuwaiti media outlet unveiled a virtual news presenter "Fedha" with plans for it to read online bulletins [251]. At the University of Kent's Centre for Journalism, lecturers are grappling with how to prepare the next generation of reporters for the potentially AI-powered newsrooms of the future [252]. AI algorithms have the capability to analyze user data, providing tailored suggestions for movies, TV shows, and music. This enhances customer retention and boosts engagement with entertainment content. Table **??** presents the recent tools and applications that are transforming the entertainment industry.

G. Role of LLMs in the Future of Legal Practice

With advancements in AI and the development of tools such as GPT-4, Bard, and Bing, it is aimed that these advancements will empower lawyers to enhance legal research, drafting tasks, and decision-making. This has sparked interest among entrepreneurs developing AI tools [253], law firms integrating AI into their workflow, and law professors exploring AI-based techniques for legal aid [254].

The impact of LLMs on the legal profession has been a topic of discussion, as mentioned in [255] [85]. This study assesses ChatGPT's potential as a substitute for litigation lawyers, analyzing its drafting and research abilities. Additionally, a legal informatics approach was introduced in [256] to align AI with human goals and societal values. By incorporating legal knowledge and reasoning into AI systems, the paper contributes to the research agenda of enhancing the integration of AI and law. In their research [257], the authors propose legal prompt engineering (LPE) as a means to improve LLM performance in legal judgment prediction tasks. The effectiveness of this method has been demonstrated on three multilingual datasets, showcasing the model's capability to handle the intricacies of legal language and reasoning from various sources of information. LLMs' transformative potential in the legal field is evident from their impressive performance

TABLE V: Unveiling the AI Revolution in Entertainment: Real-world Illustrations

Tools	Function	Link
Scriptbook	A cutting-edge AI-powered script analysis tool, is harnessed by film studios to forecast the commercial triumph of a screenplay. The tool meticulously assesses the script's characters, themes, and plot points, drawing comparisons to the historical performance of comparable films to foresee its potential box office success.	Scriptbook
Aiva	AIVA (Artificial Intelligence Virtual Artist) represents an AI-driven music composition tool that generates original music tracks tailored to user preferences. By analyzing data points such as genre, tempo, and mood, the tool crafts unique compositions suitable for integration into films, TV shows, and video games.	Aiva
LyricFind	LyricFind takes center stage as an AI-powered lyrics search engine, empowering users to find song lyrics using natural language queries. By employing natural language processing algorithms, the tool comprehends user queries and delivers precise and relevant results.	LyricFind
Ziva Dynamics	Ziva Dynamics showcases an AI-powered software tool tailored for creating authentic 3D character models in films and video games. The tool utilizes machine learning algorithms to simulate muscle and skin movement, resulting in character models that boast unparalleled realism and intricate detailing.	Ziva
DeepMotion	DeepMotion introduces an AI-powered animation tool capable of producing lifelike 3D animations for video games and films. Leveraging machine learning algorithms, the tool simulates human movement and behavior, delivering animations with enhanced realism and natural aesthetics.	DeepMotion
Speechify	Speechify is one of the most popular and efficient first AI Voice Over generators for using famous singer's voices for singing differents songs. It also creates human-quality voice-over recordings in real time. Narrate text, videos, explainers anything you have and in any style.	Speechify

in legal exams. GPT-4 scored in the 90th percentile on the Uniform Bar Examination [258], and ChatGPT autonomously passed four law school final exams at a top law school [259]. These achievements showcase the significant impact of AI language models on legal practice. The authors present Chain-of-Thought (CoT) prompts, which aid LLMs in generating coherent and contextually relevant sentences following a logical structure, simulating a lawyer's analytical approach [260]. The study shows that CoT prompts outperform baseline prompts in the COLIEE entailment task using Japanese Civil Code articles. Furthermore, LLMs have been utilized to explore fiduciary obligations, as discussed in [261].

When using such LLMs powered platforms, it is essential to be aware that they may provide inaccurate or biased information. The model's responses are based on the data it was trained on and do not incorporate real-time learning from new inputs or the internet [262]. Thus, its knowledge is limited to what it learned during its training, which could potentially lead to outdated or incomplete information. Undoubtedly, this leads to the discussion that AI cannot currently replace human lawyers due to frequent "hallucinations" [263] [264].

In a recent working paper by Choi et al., they conducted experiments using ChatGPT to generate answers for four authentic exams administered at the University of Minnesota Law School [263]. In summary, the authors concluded that ChatGPT successfully passed all four exams with an overall average grade of C+. This level of performance would grant it credit towards a JD degree, but it would also place the student on academic probation. Interestingly, if ChatGPT maintained this performance throughout law school, it would be able to graduate successfully. However, ChatGPT's answers exhibited consistent issues and errors, which rendered its performance significantly poorer compared to the average student. One of its main challenges was "identifying and addressing issues" when presented with open-ended prompts, a crucial skill in law school exams.

Recently in June 2023, in response to fake case citations generated by ChatGPT and submitted in a court filing, a US judge has imposed a fine of $5,000 (3,935) on two lawyers, Steven Schwartz and Peter LoDuca, along with their law firm Levidow, Levidow & Oberman [265] [266]. The fictitious legal research was utilized in an aviation injury claim, and Schwartz admitted to inventing six non-existent cases referred to in a legal brief against the Colombian airline Avianca. ChatGPT, a chatbot capable of generating plausible text responses, was involved in the creation of these false references. The consequences of Schwartz's mishandling have reverberated globally, prompting changes in how attorneys interact with AI tools in courtrooms. In Texas, a judge now requires attorneys to verify that no part of a filing was composed by generative AI or, if it was, that a human has verified its accuracy [267] [266]. However, not all judges share the same stance on chatbots in legal proceedings. For instance, Judge Juan Manuel Padilla, based in Colombia, acknowledged using ChatGPT's assistance in a case concerning an autistic child [267].

In light of these examples and use cases, LLMs indeed offers numerous benefits, but it is crucial to recognize and comprehend their limitations.They can serve as a valuable tool for initial research, explanations, and improving efficiency in legal practice. However, lawyers must be mindful of its limitations. While they can provide seemingly convincing answers, it may still be misleading or inaccurate. Its reliance on statistical patterns from training data means it lacks human-like reasoning and may not incorporate the most recent legal developments. Ethical and confidentiality concerns arise due to the storage and potential use of prompts and information for training purposes, posing risks to sensitive information. While lawyers will likely need to integrate AI to remain competitive, it must be done responsibly, upholding ethical obligations. Moreover, it should be used cautiously to complement legal work but cannot replace the expertise, experience, and judgment of lawyers.

H. Marketing

Large language models are crucial in modern marketing, transforming customer engagement and content delivery. They excel in content generation, creating compelling product descriptions, ad copy, blogs, and social media posts, saving time and resonating with audiences [268] [269]. Personalization is a standout feature, allowing marketers to deliver tailored messages based on customer data, improving satisfaction and loyalty. Customer support is revolutionized by chatbots, providing instant assistance 24/7, and reducing the workload on support teams [270]. In market research, large language models analyze vast data, including feedback and social media, offering insights into trends, sentiment, and competition. They contribute to SEO optimization, identify keywords, and enhance social media monitoring [269].

The adoption of LLMs and ChatGPT in marketing offers numerous benefits, but it also comes with potential risks for marketers, consumers, and other stakeholders [269]. The similarity and lack of uniqueness in ChatGPT's responses to similar prompts from different marketers could undermine the distinct identity of the marketer or brand. This presents a challenge for marketers who prioritize creativity and innovation in their strategic decision-making, especially within an AI-driven environment [271] [268] [272]. AI marketing tools like ChatGPT may draw information from unreliable sources, leading to the provision of incorrect information. In the worst-case scenario, if inaccurate data overwhelms the system, it can lead to false outcomes [269] [273]. Ethics is a significant concern as LLMs can generate content that appears human-generated, raising transparency and disclosure issues [274]. Misleading information is another risk, as relying solely on AI-generated content without human oversight can lead to the dissemination of inaccurate or outdated information to consumers. Negative consumer perceptions could arise if AI-generated content is overwhelming or perceived as inauthentic, leading to reduced trust in the brand. Compliance with regulations and data protection laws is crucial to avoid potential legal consequences [275]. Moreover, marketers may become dependent on third-party AI providers, leading to vendor lock-in or reliance on external platforms. To mitigate these risks, marketers must use AI responsibly, provide clear disclosure when AI is involved, and maintain human oversight to verify the accuracy and appropriateness of AI-generated content. Regular audits, continuous monitoring, risk assessment with its mitigation, and adherence to AI ethics guidelines can help ensure that marketers' use of LLMs and ChatGPT aligns with best practices and meets consumer expectations. It is at the interjunction of the risks (market and consumers both) and ethics where we can balance the scale which opens the window of opportunity for responsible AI adoption (see Figure 5).

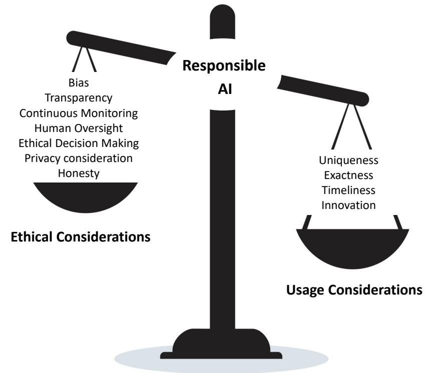


Fig. 5: The Tradeoff for Responsible AI.

I. Customer Service

Another application that has witnessed a significant impact is customer service. LLM-powered chatbots and virtual assistants are increasingly being integrated into customer support systems, providing companies with a scalable and efficient means of addressing customer inquiries and concerns [276] [277]. Unlike human representatives, LLM-powered chatbots can process and respond to inquiries instantaneously, enhancing the overall customer experience [278]. The use of LLMs also leads to cost savings for companies [279]. Implementing virtual assistants can significantly reduce the need for maintaining a large workforce of human customer service representatives, leading to substantial operational cost reductions. Due to the high cost associated with employing dedicated customer service agents, an increasing number of companies are exploring the use of NLP to assist human agents [280]. NLP enables the auto-generation of responses that can be directly utilized or modified by agents. In this context, LLMs emerge as a natural and suitable solution.

In [278], the authors have introduce a cost framework to evaluate NLP model utility for customer service. They compare three LLM strategies - prompt engineering, fine-tuning, and knowledge distillation - using agent feedback.

Usability compensates for cost differences, applicable to the broader enterprise space. The authors also demonstrate comparison case evaluation between conversation assist and LLM-Generated Suggestions to returns some items from the last order and they found the conversation assist which is a system that offers pre-defined, canned responses. In contrast, the system described in their paper [278] utilizes LLMs to generate dynamic and context-specific suggestions was able to response and assist the customer in returning back some of the items. Similarly, ChatGPT can be trained and fine-tuned with customer-specific data, enabling it to deliver personalized and customized answers for individual customers.

Furthermore, LLMs and ChatGPT/GPT-4 have exhibited exceptional potential in customer service, but they also come with inherent limitations and challenges. Context understanding remains a critical issue, as LLMs and ChatGPT/GPT-4 may struggle to grasp complex queries fully, resulting in responses that lack nuance and accuracy. Additionally, emotional intelligence, scalability, and the preference for human interactions represent further challenges in delivering seamless customer service experiences.

Apart from these fundamental application sectors, LLMs could be benefited in following contexts as well: i) Content Generation: These models can generate human-like text, which can be applied to writing articles, blog posts, stories, poetry, scripts, etc. They can also create varied content for video games, simulations, law, healthcare, education, and other interactive experiences. For example, in a real-world healthcare application, LLMs can generate a summary from a patient information form to aid the medical concerned practitioner [281]. ii) Conversation and Customer Support: They can power conversational agents (like the one you're interacting with now), providing automated customer service, tech support, and general conversational companionship. Businesses can use these systems to handle frequent inquiries, thus freeing up human agents for more complex tasks. Language models can make virtual assistants more conversational and capable, allowing them to understand and execute complex commands. iii) Coding and Email Assistance: Language models can be used to understand programming languages, providing recommendations to programmers, autocompleting code, or even writing chunks of code from a natural language specification. Models like GPT-4 can be used to draft emails, write out standard responses, and schedule tasks, reducing time spent on administrative tasks. iv) Data Analysis and Creativity: Language models can help researchers sift through large amounts of text data, extracting insights, summarizing information, and more. They can also generate creative content, including poetry, story ideas, screenplay drafts, and more. They can be a tool for brainstorming, providing many different ideas and perspectives. v) Accessibility: For individuals with disabilities that make typing or writing difficult, language models can be a powerful tool to assist in communication. They can help improve search engine responses, document summarization, and other tasks in the information retrieval domain. While these are some of the major applications, it's important to note that the use of such models also comes with ethical considerations around misinformation, bias in responses, privacy, and other issues that are important to address.

VI. AI-Enabled Tools: The Future of Everything

AI tools are becoming increasingly powerful and versatile. They can be used to generate text, images, and videos [282], translate languages, write different kinds of creative content, and answer your questions in an informative way [283]. These powerful tools provide a user-friendly interface for the optimization of daily routine tasks [68]. One such example is the popular website, "There's an AI for THAT", which contains about $7\tilde{K}$ AI tools for $2\tilde{K}$ differnt tasks. [3] In this Section, we discuss various AI-enabled tools based on LLMs or text prompts.

A. Chatbots / ChatGPT

Chatbots are frequently used in customer service applications where they can respond to queries, offer assistance, and fix problems [47]. High-tech companies are likely to become even more interested in using chatbots to improve their customer experience and grow their businesses. For example; OpenAI developed ChatGPT [284], Google developed Bard [285], and Meta launched Llama-2 [15]. Here, we critically compare these Chatbots in terms of accuracy, ease of use, cost, integration and others.

1) Comparison between Chatbots: ChatGPT and Google Bard are two of the most popular LLMs available today [286], [287]. The third popular LLM being Bing, which is based on GPT-4. Bard is based on the LaMDA [288] (Language Model for Dialogue Applications) architecture, while ChatGPT is based on the GPT-3 (Generative Pre-trained Transformer 3) architecture. ChatGPT was modified and improved using both supervised and reinforcement learning methods [289], with the assistance of human trainers (RLHF) [63]. The learning includes three steps;(i) supervised fine-tuning [290], reward model [291], and maximum policy optimization [292]. First, a pre-trained GPT-3 model is used, and fine-tuned with the help of labelers by creating a supervised dataset. After the supervised fine-tuning, different input prompts are fed to the model, and 4 to 7 responses are generated for each response, the labelers rank each response of the model. The responses are scalar values, which are used to train the reward model. In third step, the model is tested on unseen input sequences, and responses are evaluated by the reward model, and the output reward is used to fine-tune the parameters of the model, to incorporate more human-like characteristics and behaviors via reinforcement learning [293].

LaMDA is a newer architecture (conversational neural language models) that is specifically designed for dialogue applications. Bard uses LaMDA, which is a hybrid architecture that combines batch processing and streaming processing [112]. This allows BARD to handle both historical and real-time data. It is trained on a massive dataset of text and code, while ChatGPT is trained on a massive dataset of text, which means that Bard has a broader understanding of the world and can generate more comprehensive and informative responses,

[3] https://theresanaiforthat.com/

while ChatGPT is better at generating creative and interesting responses.

Both models are capable of generating text, translating languages, writing different kinds of creative content, and answering your questions in an informative way. However, there are some key differences between the two models, such as ChatGPT is more creative, while Google Bard is more authentic. Bard is more personalized than ChatGPT, the responses generated by Bard are more tailored to specific needs, and it is also more scalable than ChatGPT. A comparison between ChatGPT, Bard, and BingChat is made is [285] on VNHSGE [294] dataset, which is a Vietnamese High School Graduation Examination Dataset for Large Language Models. The results indicate that BingChat performed better than Bard, and ChatGPT. All models perform better than the Vietnamese students [295]. Infact, a comparison between the three popular LLMs services, namely, ChatGPT, Bard and Microsoft Bing has been of interest to researchers and field practitioners. A recent work by Bhardwaz et. al /citeBhardwaz23provided a general comparison for these three models considering accuracy, response time, user experience and engagement. From their experiments, they found that ChatGPT provided the most relevant responses and accuracy, Bard provided the quickest response and Bing provided the best user experience and engagement. Another comparison by Campello et. al experimented with four different chatbots (above three and Quoras Poe /cite poe) when asked to solve an intelligence test for recruitment in Brazil found that all four chatbots scored above the 95th percentile while ChatGPT and Bing scored 99th percentile. These are in addition to comparisons being made for typical as well as atypical specific use cases such as news fact checking (GPT-4 performing the best) [296], taxes [297] as well as political leaning [298] and more. To complete the discussion, Table VIII presents a comparison between ChatGPT, Google Bard, Llama-2, and Microsoft Bing Chatbots.

B. AI tools for image generation, history, medical, industry

1) Diffusion Models: Diffusion models are a scheme of generative models that have provided excellent performance in a variety of applications, most notably the synthesis of images [299]. Starting from a sample of a target data distribution, a diffusion model works in two steps, a forward diffusion process and a reverse diffusion process. The forward diffusion process gradually adds increasing amounts of Gaussian noise[300] to the sample image successively over time. The model is then tasked to start form this noisy image and undo the noise addition by going through a reverse process to recreate the original data [301]. A survey in [302] discusses foundation vision models.

The forward process takes the form of a Markov chain [303] where the distribution at a given time instant only depends on the sample from the timestep immediately preceding it. Therefore, the distribution of the corrupted samples at any given point with respect to the original sample is the product of the successive single-step conditionals up till that point.

Moreover, typically the number of passes of noise addition is in the order of a thousand with the increments each time being quite small. This is necessary to ensure that the reverse process of “recovering” the original sample is more achievable as it has been shown that with infinitely small step sizes, the reverse form will be able to achieve the same functional form as the forward process [304]. Diffusion models use this observation in the forward process. In addition to the input sample, diffusion models also consider the time increments to keep track of the association between successive noise levels for them to be able to effectively perform the reverse process. Similar to the forward process, the reverse run is also set up as a Markov chain. The model starts from a Gaussian noise sample and goes through the sample one timestep at a time to remove noise at each step. In essence, the forward process is designed to push the sample out of the original data distribution with the reverse process is designed to learn to bring it back into the original data distribution.

A diffusion model can be interpreted as a latent variable generative model similar to a variational autoencoder (VAE). The forward process can be thought of as producing latent from data and the reverse process is as converting latent to data. However, as opposed to VAEs, the forward process for diffusion models is typically fixed so that only a single network needs to be trained that deals with the reverse process. The objective function is the variational lower bound on the log-likelihood of the data. It consists of a log-likelihood term or reconstruction term minus a KL divergence term [305] also called the regularization term [306]. The log-likelihood terms [307] encourage the model to maximize the expected density assigned to the data. The KL divergence term encourages the approximated distribution to the prior distribution on the latent variable. Moreover, diffusion models can also be directed to sample conditionally based on a variable of interest which can be incorporated as an additional input during training. This has been the reason that diffusion models have shown better performance than Generative Adversarial Networks [308] in a variety of image generation tasks including perceptual quality ([309], text to image generation ([310], image inpainting [311] and manipulation of images[312].

2) Image generation: The images contained in this section were generated by a model incorporating the stable diffusion process in to existing diffusion models as suggested in [313] and uses text to generate photorealistic images. The stable diffusion process makes use of a denoising score-matching objective when performing the reverse stage. This score is then used to direct the diffusion process correctly by estimating the likelihood of the data thereby ensuring better and reliable data generation that is similar to the structure and style of the input training data. This model was released by stability.ai [314] and was demonstrated to be capable of generating images which were previously difficult to generate, such as images of people with accurate facial features as well as objects with abnormal or impossible shapes. This capability of being able to generate a diverse set of images from a text input without requiring a large amount of data for training and its public availability opened up text-to-image generation usage for a host of applications. Table VI showcases the output of image generation using various prompts. In total, nine different prompts were used, these required the AI model to generate humans and natural scenery. The first four prompts tended to depiction of famous personalities (sportsmen and politicians in this case), Muhammad Salah, Lionel Messi, Mike Tyson and Imran Khan. The prompts used were *Mo Salah playing cricket*, *Lionel Messi playing tennis*, *Mike Tyson playing football* and *Imran Khan as a hero*. The second prompt used was regarding the famous painting Monalisa. The prompt was "Generate an image of Monalisa showing her teeth in a wedding ceremony". The third prompt related to natural scenery and was written as *Area of rocks, deep inside the forest, divine domain*. Lastly, the fourth prompt also centered around the generation of humans. In this case, three prompts were given, *A man kissing a girl*, *Generate an image of a guy* and *Generate an image of a woman*.

3) Video Generation using text prommpts: T2V [315] is a video generation model using text prompts.

C. AI tools for text classification

AI tools are increasingly being used for text classification. Text classification is the process of assigning a category to a piece of text [316]. For example, a text classification tool could be used to classify emails as spam or not spam or to classify news articles as business, sports, or entertainment. Some of the popular libraries include Scikit-learn, NLTK [317], and Spacy [318]. Similarly, Hugging Face's Transformers library [319] is the state-of-the-art toolkit for developers to implement AI text generation capabilities into their applications; including fine-tune models for sentiment analysis, language translation, and text summarization.. This library offers a collection of pre-trained language models, including GPT-3.5 and various other popular models like Bidirectional Encoder Representations from Transformers BERT [52] and RoBERTa [53].

D. AI tools for Literature review Research

AI tools are increasingly being used to assist with literature review research. These tools can be used to automate tasks such as: Identifying relevant literature, extracting information, and summarizing the content [320], [321]. One such tool is PDFGPT [322], which uses the GPT-3 model to generate responses to user queries. PDFGPT can be used to extract information from PDF files, answer questions about the content of PDF files, and generate summaries of PDF files. An example of PDFChat is shown in Fig. 6.

Another interesting AI tool is elicit.org, which helps automate literature reviews. The website offers a variety of features, including, finding relevant literature, summarizing and visualizing literature, and extracting relevant information.

1) Fake references: One of the major drawbacks of using AI tools such as ChatGPT in research is the creation of fake citations and references using AI tools which can have serious complications, particularly in academic or professional settings where accuracy and credibility are essential [323], [324]. Fake citations is an inherent consequence of the generation capabilities of LLMs wherein they may prefer to lean on their generation capabilities rather than on search. One way to think about this is in terms of sets where the sample space of there being any number of possible citations/references, be them fake or real is large (in terms of word makeup) whereas the pool of real citations/papers which have been published is only a small subset of the total possible sample space as shown in Fig. 7.

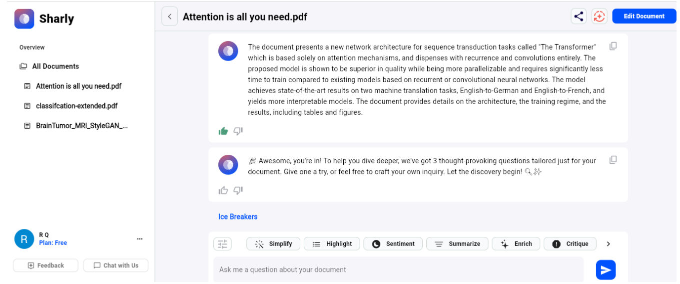


Fig. 6: An example of PdfGPT. Upload any PDF document and start chatting. It helps in summarizing, highlighting, critiquing, and simplifying the content.

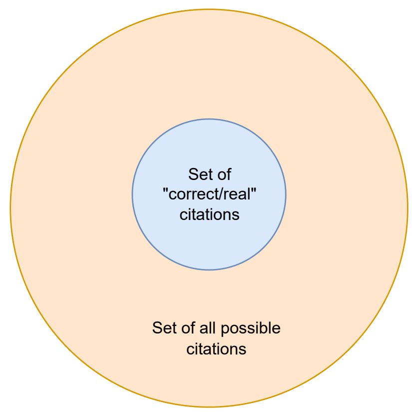


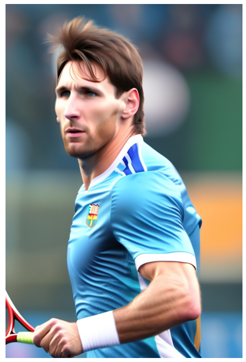
Fig. 7: Possible Citations and Real Citations set/sample space analogy.

The potential complications that are being created due to the uncontrolled usage of these tools result in many issues among which misleading the scientific community carries vital importance. Fake citations and references can mislead readers into thinking that a certain piece of information has been sourced from a credible and reliable source, when in fact it has not. This can undermine the credibility of the author and the work they are presenting. Similarly, the research which is based on fake citations and references has compromised integrity [325]. This can lead to inaccurate conclusions and potentially harmful decisions being made based on faulty information. Using fake citations and references can hide the true sources of information used in the research, making it difficult for others to replicate or verify the findings. To avoid these complications, it is important to ensure that any citations and references used are accurate and reliable and that they have been properly vetted and sourced. It is also important to be transparent about the sources of information used in research so that others can verify and build upon the work. Finally, developers of AI tools should implement rigorous

TABLE VI: Image generation examples

Prompt: Different famous personalities in roles other than their original ones
Negative Prompt: blurry, photorealistic
Generated Images:


a


b

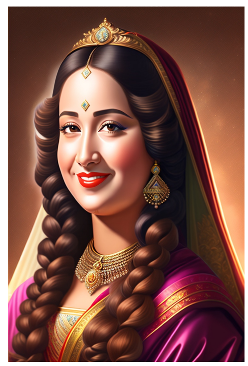

c

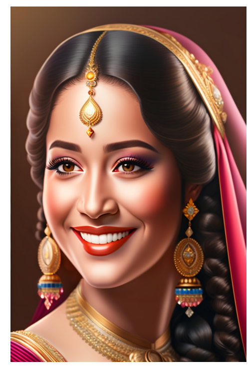

d

Prompt: Generate an image of Monalisa showing her teeth in a wedding ceremony
Negative Prompt: blurry, low resolution, artistic
Generated Images:


a


b

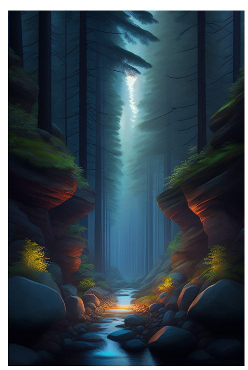

c

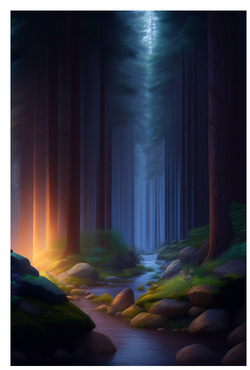

d

Prompt: Area of rocks, deep inside the forest, divine domain
Negative Prompt: artistic, blurry, background
Generated Images:


a


b

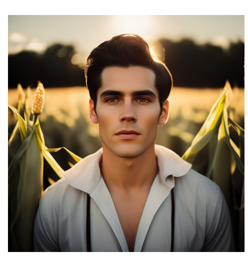

c

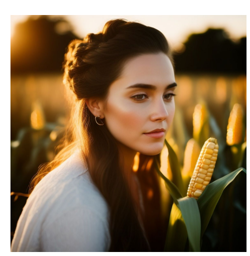

d

Prompt: A man kissing a girl/ Generate an image of a guy/ woman
Negative Prompt: artistic, blurry, background, young
Generated Images:


a


b


c

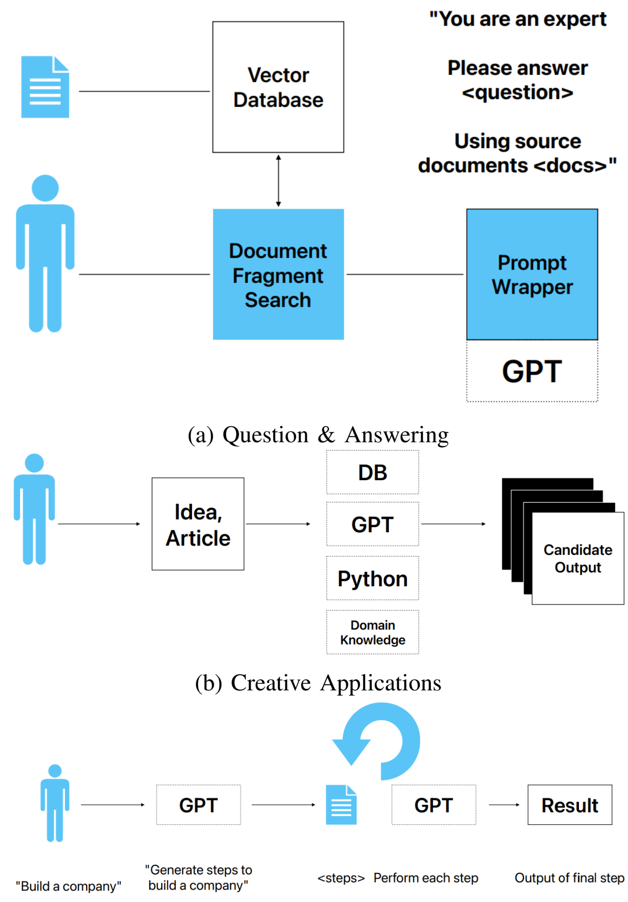

d

quality control measures to ensure that their tools generate accurate and reliable citations and references.
Recently, WebChatGPT [4] is an impressive extension that has the potential to address the pervasive issue of fake citations. With the installation of this extension, WebChatGPT becomes equipped with robust capabilities to detect and eliminate fake citations. This advanced tool uses sophisticated algorithms to analyze the authenticity and reliability of citations, ensuring that only accurate and legitimate sources are included. By incorporating WebChatGPT into the research process, researchers and writers can confidently rely on its ability to verify citations, resulting in improved academic integrity and the mitigation of misleading information.

E. AI tools for coding / CodeGPT

AI tools are increasingly being used to help programmers write code. These tools can be used to automate tasks such as code completion, refactoring, linting, and testing [326]. GitHub Copilot [327] is an AI-powered code completion tool developed by GitHub in collaboration with OpenAI. It utilizes OpenAI's GPT-3 language model to assist developers in writing code more efficiently. Meta also released the Code-Llama [328], a LLM model, that can use text prompts to generate and discuss code. It has the potential to generate clean and robust code with well documentation in Python, c/C++, Java, PHP, Typescript (Javascript), Bash and other programming languages.

LLMs have been used to develop applications in three primary categories which include: (a) Question Answering, (b) Creativity (c) Multi-step planning [329].Bilal Please expand These template categories are illustrated in Fig. 8.



(a) Question & Answering

(b) Creative Applications

(c) Multi-step Planning

Fig. 8: Templates for LLM-based application development. GPT is taken as an example scenario representing LLMs.

VII. GPT-Plug-ins

GPT-Plugins are a new way to extend the functionality of ChatGPT. They allow developers to create custom apps that can be integrated into ChatGPT, providing users with new features and capabilities. GPT-Plugins can be used to do things, such as access to external data sources, automate tasks, and enhance user experience [339]. In this Section, we demonstrate several GPT-Plug-ins.

A. ChatGPT prompts guidelines

Arguably, the watershed event in the use of ChatGPT was the introduction of plugins by OpenAI. Plugins allow ChatGPT to communicate with third-party sources of data and knowledge bases, thereby providing a platform to extend ChatGPTs capabilities for composition, summarization, nuanced tasks such as sentiment analysis and more to any resource on the internet. Moreover, given that ChatGPT has provided sufficiently acceptable performance for various tasks, plugins allow for ChatGPT to provide answers to queries with updated information from the internet which may not be present in its training dataset. This also has the advantage of providing references for queries to add credibility to answers. For e.g., Bing, the search engine by Microsoft works with OpenAI's ChatGPT through its API to allow its users to ask questions from its Bing search system and get answers with references/sources mentioned. The integration of LLMs in to search engines, thereby allowing users to get answers to human like queries has spearheaded the search engine business in to a new direction. Moreover, this addition of credibility is an important consideration to enable use of ChatGPT and similar LLMs in other critical tasks. While, at the time of this manuscript, OpenAI still hasn't rolled out plugin development access to all developers, there have been several notable use cases that have already come out. For example, twelve companies have been listed on the OpenAI website [5], namely, Expedia, FiscalNote, Instacart, KAYAK, Klarna, Milo, OpenTable, Shopify, Slack, Speak, Wolfram, and Zapier to have created the first plugins. The power that plugins provide in terms of flexibility to develop new applications has drawn a big attention towards plugin development. Apart from the above-mentioned early developers, three plugins are already made available by OpenAI. The first is the web-browser plugin, the second is the code interpreter plugin, and the third is the information retrieval plugin. The web browser plugin enables ChatGPT to access the internet for information

[4]https://tools.zmo.ai/webChatGPT

[5]https://openai.com/blog/ChatGPT-plugins

TABLE VII: Publicly available AI /LLM tools

Tools	Function	Link	Availability
ChatGPT	Conversational AI Chatbot	ChatGPT	Both
RoomGPT	Redesign your room in eight different themes	RoomGPT	Public
HomGPT	Redesign your home and office	HomeGPT	Subscription based
PDFGPT.IO	Turns PDF into the knowledge base for a ChatGPT type interface	PDFGPT	Subscription based
TexGPT	Harnesses GPT-3's power to help you write in Overleaf	TexGPT	Public
AcademicGPT	An AI tool to write and review scientific papers, critical analysis and explanation of complex concepts	AcademicGPT	Public
DiagramGPT	An AI tool for creating scientific diagrams and flow charts of different processes	DiagramGPT	Public
BloombergGPT	A Large Language Model for Finance	BloombergGPT	NA
AutoGPT	Auto-prompting without the user intervention	AutoGPT	Public
AgentGPT	Autonomous AI agent in the browser	AgentGPT	Public
HuggingGPT [330]	A framework to connect various AI models to solve AI tasks	HuggingGPT	Public
XrayGPT [331]	Automated analysis of chest radiographs	XrayGPT	Public
Video-ChatGPT	A vision language model for video understanding and conservation about videos	Video-ChatGPT	Public
ClimateGPT	Large language model for a conversation about the climate in English and Arabic	ClimateGPT	Public
CodeGPT	An AI assistant for coding	CodeGPT	Public
Code Llama	Open Foundation Models to generate and discuss code	Code Llama	Public
MiniGPT-4 [171]	Multi-modal model for a number of tasks, including image generation and website development, using prompts	MiniGPT	Public
BiomedGPT [332]	A Unified and Generalist Biomedical Generative Pre-trained Transformer for Vision, Language, and Multi-modal Tasks	BiomedGPT	Public
SkinGPT [333]	An Interactive Dermatology Diagnostic System		
PatientGPT	An AI engine to transform patient navigation by providing healthcare organizations and their patients with a seamless and customized experience	PatientGPT	Subscription based
SentimentGPT [334]	Exploiting GPT for sentiment analysis	SentimentGPT	Public
DrugGPT [211]	A GPT based model to design potential ligands, targeting specific proteins	DrugGPT	Public
Elicit	AI research assistant, automated literature reviews	Elicit	Public
Citation AI	AI research assistant to generate real evidence-based answers	Citation AI	Subscription based
Midjourey AI	AI tool to create realistic synthetic images	Mid Journey	Subscription based
DALL.E2	DALL·E 2 is an AI system that can create realistic images and art from a text description	Daall-e-2	Subscription based
VALL-E	An audio synthesization tool	Vall-e	Public
Gen-2	Video generation using text, images, and videos	Gen-2	Public
AI Avatar	Avatar generation	AI Avatar	Public
Langchain [335]	Building applications with LLMs through composability	Langchain	Public

TABLE VIII: Comparison of Bard, ChatGPT, Llama-2 and Bing Chat

Feature	ChatGPT (GPT 3.5)	Bard	Bing Chat (GPT - 4)	Llama2
Accuracy	Not as accurate as Bard	Generally more accurate than ChatGPT	Most accurate	least accurate-‘
Versatile	Generally more versatile than Bard	Can generate text, translate languages, and write different kinds of creative content	Not as versatile as ChatGPT or Bard	Less than ChatGPT and Bard both better than Bing
Company	OpenAI	Google	Microsoft	Meta
Primary Purpose	Creative text generation	Conversational AI	Information retrieval	Text generation, answer questions, language translation etc
Integration	Standalone model	Standalone model	Integrated with Bing search engine	Standalone model
Easy to use	User-friendly	User friendly	Not as user-friendly as ChatGPT or Bard	User-friendly
Access to online data	No, trained on data available till 2021	Yes	Yes	Yes
Cost	GPT 3.5 free / GPT-4 (20 USD per month)	Free	Free	Free
Availability	Publicly available	Publicly available	Publicly available	Publicly available
Architecture	Generative pre-trained transformer [336]	Pathways Language models (PaLM2) [337]	Next Generation GPT [338]	Transformer
Plagriasm detector	Yes	No	No	Less likely to generate plagiarised text
Limitations	May generate less coherent or incorrect text	Not as creative as ChatGPT	May provide limited or incomplete information	Trained on a smaller dataset than ChatGPT and Bard, may not generate text for some topics

gathering which it can use to answer a query given to it by the user. An example of using this API is shown in Fig. 11 where the prompt *Explain the GPT4 architecture* has been used.

The Code interpreter is a built-in Python code interpreter which can be used for performing logical calculations as well as writing code. The interpreter can use the language model's understanding of a human language description of a problem and use that as input to develop Python code for the problem's solution.

A third knowledge-based retrieval plugin has also been open-sourced[6] which can be used by developers as need be. This plugin can be used to enable ChatGPT to access data and then use it to gather useful or relevant information from the data. These can be files, emails, notes etc. All this by using queries or questions in normal human language. Once deployed and registered with OpenAI, this plugin can make use of OpenAI embeddings along with a selection of databases for indexing or searching through documents. Lastly, third-party plugins are also an option. These can be created and have been created by several entities. Fig. 13 demonstrate the use of two third-party plugins, namely ShowMe which can be used to generate diagrams and ScholarAI can be used to access academic journals.

[6]https://github.com/openai/ChatGPT-retrieval-plugin

Table IX provides a list of plugins available for ChatGPT which can be utilized, it should be mentioned that this list is not exhaustive and more and more plugins are being developed, especially, third party to perform tasks specific to the developer.

VIII. Guidelines for effective use of Large Language Models

In this section, we will provide a list of steps to make best used of LLMS, as well as guidelines which intended to ensure the responsible development and use of LLMs.

A. Model selection and deployment guidelines

By following these steps, we can effectively use LLMs to perform NLP tasks and improve the performance of our applications and systems [347].

- **Identify the task**: Determine what task you want the LLM to perform. LLMs can be used for a wide range of NLP tasks, such as text classification, sentiment analysis, question answering, and text generation [348], [349], [350].
- **Choose the right model**: Choose a pre-trained LLM that is suitable for your task. There are several pre-trained LLMs available, such as GPT-3, BERT, and RoBERTa.

TABLE IX: Some ChatGPT Plugins

Name	Task	Example use cases
Language Translation [340]	Translate between languages	This is particularly useful for business, travel, medical science, education and law where documents and information from different languages might need to be translated and students can use it to learn new languages
Sentiment Analysis [341]	Determine tone of text or conversation	This can be used for the task of market research, customer analysis and social media monitoring
Spell Checker [342]	Check and correct spelling mistakes	This service can be useful for formal and informal communication such as emails, word processing and also browsing the web
Question-Answering [343]	Answer questions for a user query	This can find use in education to build learning platforms, search engines, especially when a more 'understandable' response is required and also be used in automated customer service agents
Knowledge Graph [344]	Find and present information from a database	Knowledge graphs can be used for improving on search queries (i.e. search engines), integrating data sources better and of course creating recommendations.
Speech Recognition [345]	Understand and transcribe speech audio	This service can be used in audio based customer service, transcription services through dictation and also provide services to differently abled people through audio
Emotion Detection [346]	Detect emotion from text or audio	This service can be used for applications relating to market research using verbal ques, interaction in vehicles to improve safety, used for healthcare as well as assessing reactions to games and other media

Each model has different strengths and weaknesses, so it's important to choose the one that best fits your needs [153].

- **Fine-tune the model**:Fine-tune the pre-trained model on your specific task. This involves training the model on your own dataset to adapt it to your specific task. Fine-tuning involves adjusting the model's parameters, such as learning rate, batch size, and number of epochs, to optimize its performance on your task [351].
- **Evaluate the model**: Evaluate the performance of the model on a test dataset. This involves measuring the accuracy, precision, recall, and F1 score of the model on the test dataset. This step is important to ensure that the model is performing well on your task and to identify any areas for improvement [352].
- **Deploy the model**: Deploy the model in your application or system. This involves integrating the model into your application or system and exposing it through an API or user interface. This step also involves setting up monitoring and logging to track the performance of the model in production [353].
- **Monitor and retrain the model**: Monitor the performance of the model in production and retrain it as needed. This involves regularly checking the performance of the model and identifying any areas for improvement. If the performance of the model degrades over time, it may be necessary to retrain the model on new data or adjust its parameters [284].
- **Continuously improve the model**: Continuously improve the model by incorporating user feedback and updating it with new data. This involves collecting feedback from users and incorporating it into the model to improve its performance. It also involves regularly updating the model with new data to keep it up-to-date and relevant [354].

B. Ethical guidelines

The following guidelines will help to ensure the responsible development and use of LLMs focusing on user privacy, bias mitigation, ethical considerations, transparency, competition, collaboration, and environmental impact [63]. By adhering to these guidelines, LLMs can be used to positively impact society while minimizing potential negative consequences.

- **Protect User Privacy**: LLMs should uphold user privacy and protect user data. This includes safeguarding user-generated content, such as emails, messages, and personal information. Best practices should be followed, such as data minimization, anonymization, and encryption, to ensure user privacy is not compromised [355].
- **Mitigate Bias**: LLMs can inherit and amplify biases present in the data they are trained on. Developers and researchers should actively identify and mitigate bias in their models. This can be achieved through diverse and inclusive training data, bias detection techniques, and evaluation metrics [356].
- **Address Ethical Implications**: LLMs have the potential to be used for harmful purposes, such as spreading disinformation or generating deepfakes. Ethical considerations should be taken into account, and models should be developed in a way that benefits society. This includes ensuring accountability, transparency, and responsibility in the development and deployment of models [357].
- **Foster Transparency**: It is crucial that the inner workings of LLMs are transparent and explainable. This can help build user trust and facilitate understanding of the model's behavior. Explainability techniques, such as attention mechanisms and model interpretation tools, can be employed to provide insight into the decision-making process of models [235].
- **Promote Competition**: The development and deployment of LLMs should not be monopolized by a small number of companies or individuals. This can limit innovation and negatively affect competition. Collaboration between academia, industry, and government can foster competition, while also promoting responsible development and use of models [358].
- **Encourage Collaboration**: Collaboration between researchers, developers, and industry should be encour-

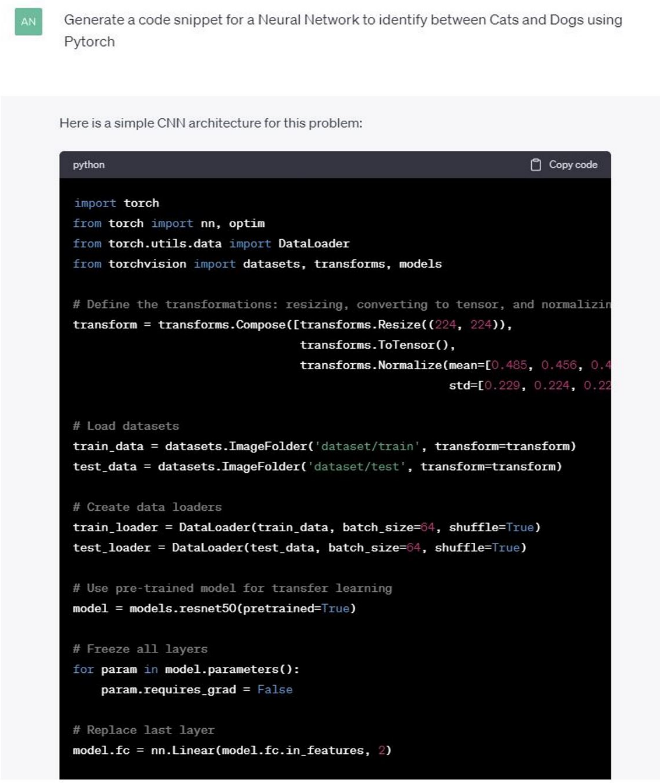


Fig. 9: Demonstration of code generation using ChatGPT

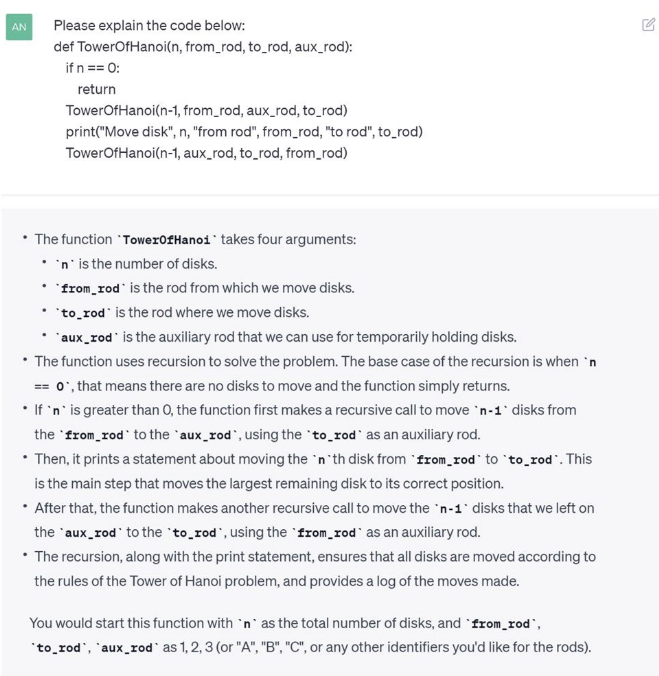


Fig. 10: Demonstration of code debugging using ChatGPT

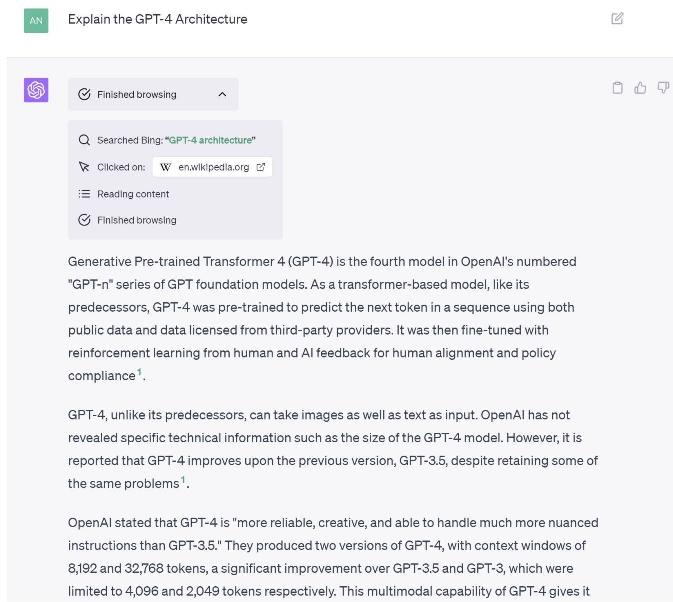


Fig. 11: Demonstration of the WebBrowsing plugin.

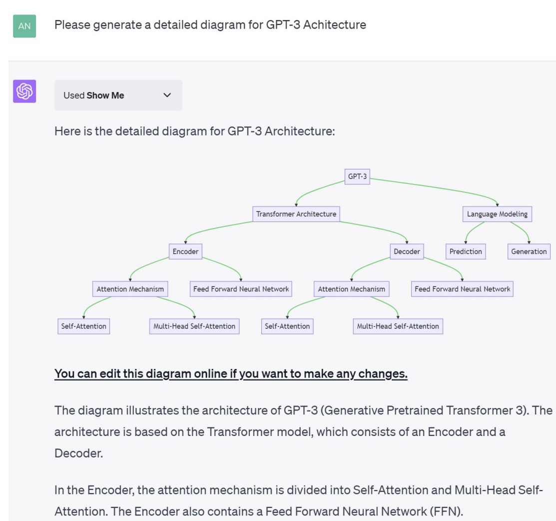


Fig. 12: Demonstration of the ShowMe plugin.

aged to promote the responsible development and use of LLMs. This includes open sourcing models and data, as well as facilitating the sharing of research findings and best practices [359].

- **Minimize Environmental Impact**: Training LLMs can require significant computational resources and energy, which can have negative environmental impacts. Developers should strive to create more energy-efficient models and explore alternative training methods, such as model distillation or transfer learning, to reduce the environmental footprint of models [360], [361].
- **Optimization is exploitation:** is a statement that holds particular significance in the context of LLMs and AI technologies [362]. While these technologies have the potential to revolutionize the way we live and work, they also have the potential to perpetuate existing inequalities and introduce new forms of exploitation [363]. The

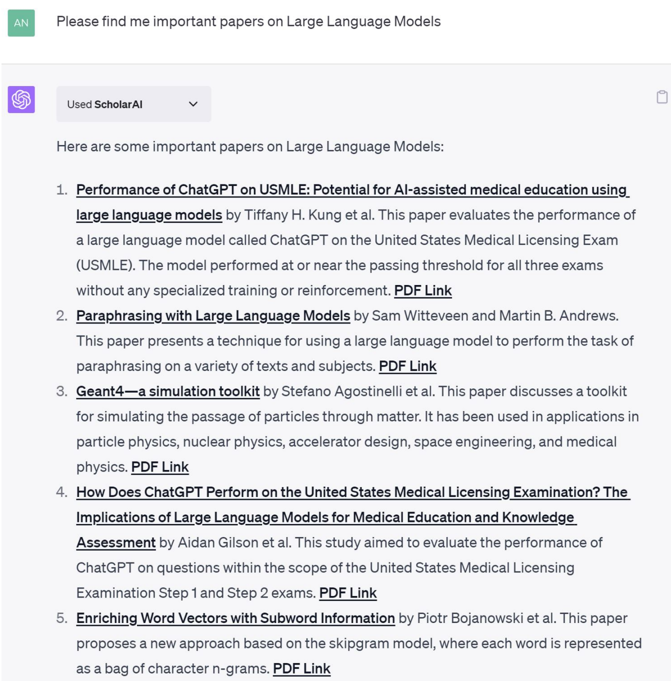


Fig. 13: Demonstration of the ScholarAI plugin.

development and deployment of LLMs often require significant resources, such as data and computational power, which may be controlled by a select few organizations or countries, leading to further disparities in economic and technological development [161]. Furthermore, the optimization process for these models can introduce biases and reinforce existing inequalities, leading to the exploitation of individuals or groups who are negatively impacted by the model's outputs. Therefore, it is important to carefully consider the ethical implications of optimization in the development and deployment of LLMs and AI technologies [60].

C. Prompting

LLMs have given rise to whats called "Prompt Engineering". Prompts are the instructions provided to an LLM to make it follow specified rules, automation of processes and to ensure that the output generated is of a specific quality or quantity [62], [364]. While there is a lack of a formal definition, prompt engineering refers to the designing and wording of prompts given to LLMs so as to get a desired response from them. Writing a prompt appropriately is therefore very important if one needs to use LLMs to assist with tasks in the best manner possible. While some formal techniques such as Explicit instruction (providing a clear direction to the LLM to do something), System Specific Instruction (asking a question from the LLM to answer), Formatting with an example (providing a sample question and its answer and asking the LLM to provide an answer in the same manner), Control tokens (use special keywords in the prompt to help the LLM provide an answer while considering special provided criteria) and Interaction and iteration/chaining (interact with model iteratively to reach to a good answer by fine-tuning on each reply) have been presented. Here, we aim to present some sets of commands to help users get the most out of the LLMs capabilities. LLMs have given rise to whats called "Prompt Engineering". Prompts are the instructions provided to an LLM to make it follow specified rules, automation of processes and to ensure that the output generated is of a specific quality or quantity [62]. While there is a lack of a formal definition, prompt engineering refers to the designing and wording of prompts given to LLMs so as to get a desired response from them. Writing a prompt appropriately is therefore very important if one needs to use LLMs to assist with tasks in the best manner possible. While some formal techniques such as Explicit instruction (providing a clear direction to the LLM to do something), System Specific Instruction (asking a question from the LLM to answer), Formatting with an example (providing a sample question and its answer and asking the LLM to provide an answer in the same manner), Control tokens (use special keywords in the prompt to help the LLM provide an answer while considering special provided criteria) and Interaction and iteration/chaining (interact with model iteratively to reach to a good answer by fine-tuning on each reply) have been presented. Here, we aim to present some sets of commands to help users get the most out of the LLMs capabilities. LLMs have given rise to whats called "Prompt Engineering". Prompts are the instructions provided to an LLM to make it follow specified rules, automation of processes and to ensure that the output generated is of a specific quality or quantity [62]. While there is a lack of a formal definition, prompt engineering refers to the designing and wording of prompts given to LLMs so as to get a desired response from them. Writing a prompt appropriately is therefore very important if one needs to use LLMs to assist with tasks in the best manner possible. While some formal techniques such as Explicit instruction (providing a clear direction to the LLM to do something), System Specific Instruction (asking a question from the LLM to answer), Formatting with an example (providing a sample question and its answer and asking the LLM to provide an answer in the same manner), Control tokens (use special keywords in the prompt to help the LLM provide an answer while considering special provided criteria) and Interaction and iteration/chaining (interact with model iteratively to reach to a good answer by fine-tuning on each reply) have been presented. Several different frameworks have been suggested in lieu of prompt patterns for LLMs, these are generic prompt patterns targetting a specific category such as prompt improvement, input semantics etc [62], [365], [366], or prompting for software engineering tasks [367], [368], however, in this work, we aim to present some sets of commands to help users get the most out of the LLMs capabilities from a generic perspective.

- *Defining the role/context*: This should be the first prompt for the LLM. An example of this prompt could be: "Act as a secretary to the Chair of the department", "Act as a Lawyer" or "Act as my programming tutor for Python". By defining a role for the LLM, one can direct it to provide replies or do tasks as a human would do when provided information to work on. A similar first prompt could be providing the context. This can be performed to give the LLM a background of the conditions in which the LLM is supposed to work. For e.g., "We are a company performing mobile application development for Fortune 500 organizations". This can then be followed up with aspects like actions, tasks to perform, steps to follow, etc as mentioned before.

- *Prompt creation*: Another interesting prompt command is to ask the model to generate prompts for a certain task. This way, the LLM can be used to generate optimized prompts for tasks that need to be done. An example of this could be: " You are a large language model and are an expert in generating prompts for ChatGPT. Please generate the best prompts on extracting important information from my time series data".
- Other interesting directions in which Prompts can be given are explanation prompts (e.g., "Explain the concept of infinity", Instructional Guides (e.g., "How do I tie my shoe laces"), Extract information (e.g.: one can paste a passage and ask the model to provide answers to questions that one might have), Solve Math problems (e.g., "Find the roots for the quadratic equation, $2x^2+3x+10$ ") and Code help (e.g., "Find the syntax error in the following code").

Other interesting aspects of prompting are Negative prompting and Visual Prompting. Here, a brief discussion is provided on each of these types.

1) Negative Prompting: Negative prompting [369], [370], [371] provides directions to the LLM about aspects of the prompt that it should avoid generating or deliberately excluding during the generation process. Through the use of negative prompts, one can fine-tune the results generated by the LLM in response to a prompt while being able to keep the prompt generation generic. Another advantage of the use of negative prompting is that it allows for moderation of the output content generated by the model thereby preventing harmful or inappropriate from being generated. "Don't write anything that is offensive or harmful, or factually incorrect." This prompt tells the model to avoid generating text that could be offensive or harmful to others and inaccurate. Notably, the authors in [372] conducted experiments for text based image tranlation and found that negative prompting to be very useful when working with textureless images. Moreover, this type of prompting is very useful when working on text to image generation scenarios and has been incorporated in text to image generation methods such as Muse [373].

2) Visual Prompting: Visual prompting [374] refers to the use of visual prompts (such as images or non-visual ones such as music) when providing directions to a model in addition to plain text prompts. The aim is to provide the AI model with a starting point or an example/reference that it can use for the given generative task. For images, this may be given to modify the image provided or generate something that is similar in style, color, or texture etc [375]. This can help in generating content that is closer to a user's expectation from the generative AI being used.

An image-based example of visual prompting could be providing a picture of an office and asking the AI to generate a different theme for it, maybe more nature-centric or in a different color or organizational style. Visual prompting provides greater control of the generated output and therefore results in a more accurate result. Using the provided input image/video can provide generated outputs that are more consistent with the intentions of the user input prompt due to the additional reference input. It should be noted that visual prompting is not related to images only, this is currently being explored for a host of different applications, including, text generation (generating something based on a sample text so as to copy its style of writing for e.g.) [80], composition of music (wherein the supplied music piece can be used as a reference for the type of music to compose) [376], game development [377] (where a defined game environment may be provided to the model as a starting point and the model is asked to generate new and unique content) and virtual and augmented reality (wherein a set of augmented/virtual reality environments can be provided to further populate/create current/new environments) [378].

IX. Challenges and Limitations of Large Language Models

Although LLMs have made significant contributions to NLP, they are not without challenges and limitations [85], [379]. LLMs are currently perceived as forerunners of AGI. However, despite their phenomenal success in conversational tasks, the state-of-the-art LLMs still lack in many aspects that makes them less likely an early manifestation of AGI. We first provide a quick list of the challenges and limitations of LLMs from that perspective as ilustrated in Fig. 14 and then present a more detailed discussion on a few limitations of critical nature in light of the current outlook of LLMs. Here we highlights a number of these challenges and limitations, including biased data, overreliance on surface-level patterns, limited common sense, poor ability to reason and interpret feedback, the need for vast amounts of data and computational resources, limited generalizability, lack of interpretability, difficulty with rare or out-of-vocabulary words, limited understanding of syntax and grammar, limited domain-specific knowledge, susceptibility to adversarial attacks, ethical concerns, difficulty with context-dependent language, absence of emotion and sentiment analysis, limited multilingual capabilities, limited memory, lack of creativity, restricted real-time capabilities, high costs of training and maintenance, limited scalability, lack of causality, inadequate ability to handle multimodal inputs, limited attention span, limited transfer learning capabilities, insufficient understanding of the world beyond text, inadequate comprehension of human behavior and psychology, limited ability to generate long-form text, restricted collaboration capabilities, limited ability to handle ambiguity, inadequate understanding of cultural differences, limited ability to learn incrementally, limited ability to handle structured data, and limited ability to handle noise or errors in input data [380], [381], [382], [383], [127], [384], [74]. Therefore, it is essential for researchers and practitioners to acknowledge and address these challenges and limitations to ensure the ethical and effective use of LLMs and to develop new models that can address these challenges and limitations.

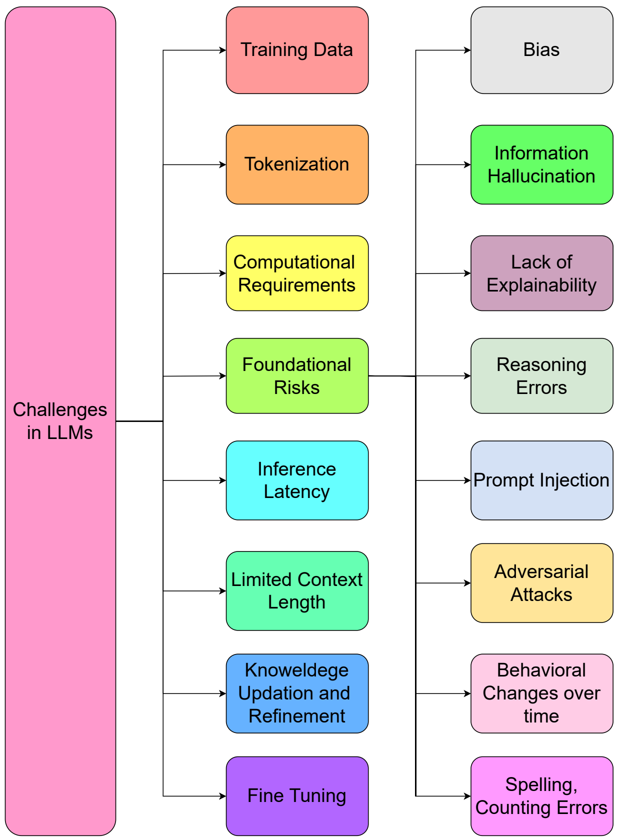


Fig. 14: Challenges in LLMs.

A. Training Data Requirements

Large Language Models (LLMs) require large corpus of data for pre-training the model. Collecting and curating these datasets can be extremely challenging. The size of the dataset makes it impossible to read or assess the quality of the dataset making it prone to having duplicates, making the model bias and degrading its responses [385] [379]. It also makes it difficult to access the model as the training data may contain data similar to testing samples leading to incorrect evaluation metrics. Since there is no way of checking the datasets manually, it may contain confidential or personal information such as telephone numbers leading to privacy leaks during prompting[386]. Due to the fact that the data distribution is more of a black box, it remains uncertain what amount of data is required for different tasks.

B. Tokenization Problems

LLMs heavily rely on tokenization which consists of breaking down a sequence of words into tokens for the models input. Most LLMs use subward tokenization[387] which is used to create tokens by splitting the words to handle non-familiar vocabulary and at the same time maintaining the computational complexity. However, there are some major drawbacks of tokenization which includes, different combinations of token can be used to relay the same prompts, which may lead to unfair pricing for the APIs of these LLMs. In a multilingual environment it may cause unexpected model responses due to different spacing in the prompts for languages such taiwanese mandarin or chinese mandarin [388]. Subward Tokenization utilizes greedy algorithms that tends to favor prompts dominant with tokens from the training data which may cause to omit low coinciding tokens causing loss of information. These tokenization techniques also cause a high computationally overhead, vocabulary size and the ability to handle non-overlapping tokens from the training data.

C. Computational Requirements

Pre-training LLMs requires significant computational costs which can be very expensive both financially and environmentally. Millions of dollars are spent in training these LLMs with thousands of compute hours and energy consumption. These are classified as Red AI;[389] referring to models achieving state of the art results due to vast computation. Scaling these models can also be a challenging task due to the number of resources invested to train these LLMs. The concept of Computer Optimal Training [385] was introduced to address this for maximizing the training efficiency with respect to the corpus and model size. Model parallelism can also be used to distribute the model and the train it faster.

D. Fine-Tuning LLMs

Fine-Tuning LLMs is a useful technique to train LLMs to custom tasks by training further on these task-specific datasets [45]. However, it requires a high amount of memory and large compute resources to store model gradients, parameters and activations, along with storing these fine-tuned models, limiting its access to a few institutions. Parameter-Efficient Fine-Tuning is a technique that can be used to address this problem which consists of updating a subset of model parameters such as prefix fine-tuning[390], prompt-tuning[391] and adapters[392]. Although techniques like Low-Rank Adaptation (LoRA)[393] can be used to optimize the computation cost, but still computational demands remain a significant barrier for Fine-Tuning LLMs.

E. High Inference Latency; Problems and Techniques

High inference latency is one of the major challenges of LLMs which is mainly dues to large memory footprints and lack of model parallelism. Several techniques can be used to mitigate this problem. Efficient Attention[394] can be used for accelerating attention through sub quadratic approximations such as multi-attention query or flash attention. Quantization[395] can be used to reduce the large memory footprints by reducing the computational precision of activations and weights. Pruning[396] and cascading[397] are some more techniques that can reduce the inference latency drastically for efficient and seamless responses.

F. Limited Context Length in LLMs

Limited Context Length is one of the crucial aspects of LLMs, as it is extremely useful for interpretation of different prompts and semantic analysis. Without this contextual information, it can drastically degrade the performance of LLMs. There are several strategies that can be used to address this; Positional Embedding Schemas [398], Efficient Attention[399] and Transformer Alternatives. Different Positional Embedding Schemes can help LLMs to generalize well to different prompts which may not exist in the training data. Transient Global[400] and Luna [401]are some efficient attention mechanisms that can process larger context lengths effectively. Recurrent Neural Networks (RNNs)[402] and State Space Models (SSMs)[403] are good alternative for transformer-based approaches and are effective for addressing limiting context length. Although the above approaches, can help to address the limited context lengths, LLMs ability to generalize to these sequences remains to be unsolved.

G. Updating and Refining Knowledge in LLMs

Although LLMs are trained on a large corpus of data, the factual information learned may become outdated over time. Retraining the models is a costly process and is not sustainable. To address this, approaches such as model editing [404] is a technique which uses non-parametric knowledge resources to alter a model's behavior, and preserving model parameters by feeding new weights to modify the model's behavior can be used. However, these approaches are found to have limited generalizability and may only be applicable to a limited model architecture

H. Risks of Foundation models

In [114], authors advised a careful assessment of the risks and benefits of foundation models before they are widely deployed. A review [348] also highlights the potential threats and benefits of large models in health and education.

- **Bias**: Language models have the potential to unintentionally demonstrate bias when the training data used in their development is biased. According to Schramowski et al. [405], large pre-trained models designed to mimic natural languages can inadvertently perpetuate unfairness and prejudices. Consequently, this can lead to discriminatory or inaccurate analyses and recommendations, resulting in public criticism across various domains, including politics, society, and law. The manifestations of these biases are as follows: (i) Training data bias: Language models typically rely on extensive datasets of human language for training. If these datasets contain biases related to factors such as race, gender, or socioeconomic status, the model may internalize and reproduce these biases in its responses. For example, if the training data exhibits a gender bias, the model may generate responses that favor a particular gender. (ii) User interaction bias: The responses generated by Chatbots are influenced by the input received from users. If users consistently pose biased or prejudiced questions, the model may learn and perpetuate these biases in its responses. Consequently, if users frequently ask discriminatory questions targeting a specific group, the model may generate responses that reinforce such biases. (iii) Algorithmic bias: Biases can also be introduced through the algorithms employed in training and operating language models and Chatbots. For instance, if the model is trained to optimize for a specific metric, such as accuracy or engagement, it may prioritize generating responses that align with that metric, even if those responses are biased in some way. (iv) Contextual bias: Chatbots generate responses based on the context provided by users. If the context contains bias associated with factors like the user's location or language, the model may generate biased responses. For instance, if a user asks about a particular culture or religion and the model lacks training on that specific cultural or religious context, it may produce biased responses due to its limited knowledge.
- **Information Hallucination:**Hallucination in Natural Language Generation (NLG) is the generation of text that is nonsensical or unfaithful to the provided source content [281]. Hallucinations in LLMs are often the result of the model's attempt to fill in gaps in knowledge or context, with assumptions that are based on the patterns it has learned during training. This can lead to incorrect or misleading outputs, which can be particularly problematic in sensitive applications.
  The cause of hallucinations in LLMs is an area of active research. Recent advances suggest that it's a complex problem related to the model's training process, dataset, and architectural design. In particular, LLMs might be biased towards producing more "interesting" or fluent outputs, leading to a higher risk of hallucination [406].
  There have been several proposed methods to mitigate the issue of hallucinations. One approach is to modify the training process to explicitly penalize hallucinations, such as in the case of "reality grounding" [407]. Another is to provide the model with a larger and more diverse dataset, which might reduce the risk of the model making incorrect assumptions [60].
  In addition, researchers are exploring the use of "verifiable" or "fact-checkable" data during training, to teach the model to rely more on facts and less on its own assumptions [408]. This, however, requires careful consideration of the data and metrics used.
  Moving forward, more research is needed to better understand and address hallucinations in LLMs. Some potential directions include the development of more sophisticated models that can better discern between factual information and assumptions, as well as novel training methods and datasets.
- **LLMs Explainability:** No one can explain a model containing 175 billion parameters: The advent of LLMs has ushered in unprecedented advancements in NLP tasks. However, the sheer complexity and scale of these models present challenges in terms of explainability [409], [410]. As LLMs continue to grow in size, with models containing billions of parameters, the ability to comprehensively explain their decision-making processes becomes increasingly elusive [411], [412]
  This complexity makes it exceedingly difficult for humans to understand and interpret the decision-making

mechanisms employed by the model [68]. The lack of transparency [413] hinders the ability to gain insights into how specific inputs lead to particular outputs [414]. Moreover, the training process of LLMs involves vast amounts of data, often collected from diverse sources. These models learn patterns and correlations within the data, leading to the emergence of implicit biases and associations that may not be readily apparent or interpretable. Consequently, when a decision is made by an LLM, it becomes challenging to discern the underlying factors that influenced that decision, making it difficult to provide a clear and concise explanation. Additionally, the intricate architecture of LLMs, often consisting of deep neural networks, exacerbates the challenge of explainability [415]. The numerous layers and complex interactions make it challenging to trace the reasoning process of the model. While techniques such as attention mechanisms [416] can provide some insights into the model's focus, they do not provide a comprehensive understanding of how the model arrives at its final output.
Finally, the lack of explainability in LLMs raises concerns regarding accountability, trust, and ethical considerations [417]. In critical domains such as healthcare or finance, where decisions can have significant implications, it is crucial to have transparency and the ability to explain the reasoning behind the model's predictions [413]. Without explainability, stakeholders may be reluctant to fully trust and adopt LLMs for sensitive applications.
- **Reasoning Errors:** LLM can make mistakes in logical reasoning [418], either because of ambiguities in the prompt or inherent limitations in its understanding of complex logical operations. LLMs can not plan, reason, and have limited knowledge and commonsense [419] about the physical world [420]. From a cognitive science perspective, Auto-regressive LLMs at their best can approximate the Wernicke and Broca areas in the brain [421].
- **Struggles in Classes of Applications Such as Spelling Errors:** Some specific tasks, like identifying and correcting spelling errors, can be challenging for GPT-4 due to its statistical nature.
- **Counting Errors**One common counting error occurs when the model miscounts or misinterprets numerical quantities. For instance, it may provide incorrect calculations or misplace decimal points when performing arithmetic operations, and counting the number of words or characters in long paragraphs [384], [422].
- **Susceptible to Prompt Injection, 'Jail Break' Attacks [423], Data Poisoning Attacks:** GPT-4 is susceptible to various adversarial attacks. For instance, a malicious actor might inject misleading prompts, perform 'jailbreak' attacks to make the model reveal sensitive information, or use data poisoning strategies to manipulate the model's output. Such vulnerabilities have been discussed in [424], [365] through experiments.
- **Adversarial Attacks** Adversarial attacks on large language models (LLMs) are a type of security threat that can be used to manipulate or control the output of an LLM. These attacks work by deliberately introducing small changes to the input text, which the LLM then misinterprets and produces incorrect or harmful output [425]. One common type of adversarial attack is called a text injection attack. In this type of attack, the attacker introduces carefully crafted text into the input, which the LLM then interprets as a command. For example, the attacker could inject the text "delete all files" into an LLM that is used to control a computer system. The LLM would then delete all of the files on the system [426]. Visual-prompt based models are also being attacked by these corrupted prompts [427].
- **Behavioral Change of Large Models over Time** Chen et. al. [428] investigated the performance of GPT 3.5 and GPT 4 over time, between March 2023 to June 2023, and found that the performance can greatly vary over time. For example, In March, GPT-4 had an accuracy of 84%, but in June, its accuracy dropped to 51%, a decrease of 33%. However, many experts suggest that the performance decrease is due to model drift [429] or prompt drift [430], we need to prompt better for maintaining the performance.

X. Broader Impact of LLMs on Humans

Despite the aforementioned limitations, LLMs such as OpenAI's ChatGPT and Google's Bard have gained popularity for their ability to produce human-like responses to user input [431]. However, the training process of these models has significant environmental implications, particularly with respect to the usage of water and energy [432]. This section will also discuss the environmental impact of LLMs and propose potential solutions to reduce their adverse effects, thereby promoting their sustainable use.

A. Environmental

New studies have revealed that the training process for GPT-3 alone used up 185,000 gallons of water, equivalent to what's needed to fill a cooling tower of a nuclear reactor [433]. This high consumption of water is primarily due to the cooling process of data centers, which necessitates a massive amount of water to regulate the servers' optimal temperature. Typically, freshwater sources are utilized to prevent corrosion and bacterial growth that can occur with seawater, but this limits the available water sources. Moreover, it is expected that the development of newer and advanced version models would need even more significant amounts of water due to their larger data parameters [145]. This concern has been discussed in [434] who present a method to estimate the water footprint of AI language models and suggest more information transparency in this regard. Apart from water usage, the training of LLMs demands a considerable quantity of electricity. The training of OpenAI's GPT-3 alone resulted in the release of 502 metric tons of carbon, which could provide energy to an average American household for hundreds of years [435]. Furthermore, the indirect water consumption of data centers located off-site should also be taken into account since they necessitate a substantial amount of electricity, leading to carbon emissions [432].

To lessen the harmful environmental effects of LLMs, various remedies can be implemented. One such solution is for data centers to adopt more eco-friendly cooling systems, such as using recycled water or implementing advanced cooling technologies [436]. Additionally, renewable energy sources, such as solar or wind power, can be utilized to power data centers, thereby reducing carbon emissions. Limiting the size and intricacy of LLMs is another potential solution, as smaller models require less data, resulting in reduced energy and water consumption [432]. Another study by Chien et. al [437] found that with models like ChatGPT, inference services dominated the power consumption and the power emissions for one year were equivalent to 25 times the training power of GPT3. They suggested the use of request direction approaches as a promising manner of reducing power consumption in LLMs.

While, given that generative AI in general has exhibited high emissions during training/inference as mentioned previously, a study found that AI-generated emissions were several times less than the emissions a human would make for writing and illustration tasks [438]. This finding indicates to the need to approach this problem in a more nuanced manner.

B. Sustainibility, Energy resources

The development and deployment of AI tools to automate, and enhance the business and user experience required a significant amount of energy to train these systems, particularly deep learning models such as ChatGPT.The amount of energy consumed by AI tools during training can be staggering, with some estimates suggesting that it can take hundreds of thousands or even millions of kWh to train a single large-scale model like GPT-3 [439], [440]. This energy consumption can have significant implications for power and energy usage, as well as the environment. The energy consumption of AI training can be attributed to several factors, including the hardware used to run the training algorithms, the complexity and size of the models being trained, and the amount of data being processed.

In order to train deep learning models like ChatGPT, specialized hardware such as GPUs are often used, which can consume large amounts of energy due to their high processing power and data transfer requirements. Furthermore, the size and complexity of these models also contribute to their energy consumption. The more parameters and layers a model has, the more energy it will require to train, as each iteration of the training algorithm requires the model to process large amounts of data and make complex calculations to adjust the weights and biases of the model. The energy consumption of AI training has significant implications for the environment, particularly in terms of greenhouse gas emissions and climate change [441]. The energy required to train AI models is often generated from fossil fuels, such as coal and natural gas, which emit large amounts of carbon dioxide and other greenhouse gases into the atmosphere. This can contribute to global warming and other environmental impacts [442]. This issue highlights the need for responsible and sustainable practices in AI development and deployment.

To mitigate the environmental impact of AI training, several approaches can be taken. One approach is to develop more energy-efficient algorithms and models, which can reduce the amount of energy required to train AI systems. Another approach is to use renewable energy sources, such as solar or wind power, to generate the energy required for AI training. Additionally, there are efforts to develop more energy-efficient hardware, such as neuromorphic computing, which can significantly reduce the energy requirements of AI training. In conclusion, the energy consumption of AI training, particularly for deep learning models like ChatGPT, can have significant implications for power and energy usage, as well as the environment. As AI becomes more pervasive in our daily lives, it is important to consider the energy requirements of these systems and develop strategies to mitigate their impact on the environment [443]. It is crucial to continue developing AI technologies and leveraging their potential benefits while being mindful of the environmental impact. By promoting sustainable practices, investing in energy-efficient computing, and exploring alternative training methods, we can work towards a more sustainable integration of AI in society.

C. Sigularity

Singularity refers to a point in the development of AI where it becomes more intelligent than humans, thereby triggering accelerated development of technology. With the increasing popularity of LLMs for general use, especially after the claim[7] by a google employee regarding Googles Chatbot being sentient, the idea of artificial general intelligence and it surpassing of equating human level intelligence has been a topic of serious debate in the AI community. The widely used criteria for determining if a machine has intelligence is the Turing test [444]. The Turing test measures a machines capability to have a conversation with a subject that is indistinguishable from that of a human conversing in its place. If a machine is judged to be indistinguishable, it is deemed to have passed the Turing test and therefore demonstrated intelligence on par with that of humans. While appearing deceivingly simple, the test considers nuances in human behavior over a range of subjects and contexts. LLMs as of the date of this publication, have not yet passed the Turing test in all its forms and therefore are not deemed to posses human level intelligence. Having said that, there are two takes on AI reaching or surpassing human level intelligence, a group which believes that the increasing use of AI and it reaching human reaching intelligence will free or greatly reduce the burden of labor for humans as well as spearhead technological progress to help solve existential problems to the human race such as Climate Change, Social equity, food insecurity among others. The benefits are endless, from the optimization of resources in every domain to making new scientific discoveries and decreasing human bias and error. For example; we, humans, solve a problem, optimize it, and lock our model, and that solution becomes state-of-the-art. Almost, everyone follows it. However, humans have limited knowledge and computational power compared to

[7]https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/

LLMs. LLMs can further explore the search space to find new and more optimzed ways of solving a solution.

However, there is another group of scientists who adhere to a more pessimistic view towards such development in that there is a fear that such a powerful system might at one point become hostile to humans and become uncontrollable. Infact, in March of 2023, notable personalities in tech published an open letter[8] requesting a pause in LLM AI development. More recently, personnel from tech companies as well as researchers have requested politicians to consider the risk of human extinction due to AI as a top priority [9]. In their letter, they consider the major risks based on the potential to be misused for societal disruption.

Hyper disparity among developed and underdeveloped nations. There are talks about the dark side of training LLMs, there is a need of upper bound in terms of parameters. There is a need for new directions to advance the quest for AGI.

It is expected that the debate on singularity and also AI regulation will continue in the foreseeable future, a balanced approach needs to be applied with strong and effective AI regulation where it aligns with benevolent human values.

D. *Competition among for-profit organizations*

OpenAI was founded around the premises of having a Large scale AI company operating as not for profit organization. However, with the evolution of scale and nature of investment required, soon it converted to a for profit organization which almost eliminated the freedom of access of large scale AI to the masses. However, startups like Hugging face are supporting the growth of AI through is massive open-source campaign [445]. Being a for-profit organization has some demerits such as excessive control by the investors, decoupling of development and issues arising from the general public [446].

XI. Experts point of view

In this Section, we evaluate the opinions of world-renowned AI researchers and discuss conflicting opinions among them.

A. *AI Human Coexistence*

The experts suggest that we need to take a step back to assess the ethical concerns and develop safeguards to mitigate them [447]. The suggested reasoning behind this suggestion is that there are concerns about the ethical implications of LLMs and their potential negative impact on society, including issues such as privacy, human extinction, job market, bias, and the concentration of power in the hands of a few large tech companies.

Several prominent figures in the field of AI, including Timnit Gebru, an AI ethics researcher who was fired by Google in late 2020, and Yoshua Bengio, a prominent AI researcher at MILA Canada, have expressed their support for this suggestion. More recently, Geoffrey Hinton, a renowned AI expert known as the "Godfather of AI," made an announcement confirming that he had resigned from his position at Google. The reason behind his resignation was to raise awareness about the potential "dangers" associated with AI, a technology that he helped to develop. Elon Musk, the CEO of Tesla and SpaceX, has been vocal about his concerns regarding the dangers of AI. Musk believes that AI poses an existential threat to humanity if it is not developed responsibly. He has called for regulation of AI development to ensure its values with human values and society.

However, not all experts agree with this suggestion [448]. Some argue that the potential benefits of LLMs outweigh the potential risks. They also point out that a pause in research and development could put some organizations at a disadvantage, as they may fall behind in the race to develop new and innovative LLM technologies.

B. *Consideration of a 6-Month Suspension: Is it Necessary?*

In terms of the positive and negative sides of stopping LLMs development for six months, there are arguments on both sides. On the positive side, a pause could allow for more thorough ethical considerations and the development of more responsible approaches to LLM development [449]. This could potentially mitigate some of the negative impacts of LLMs, such as bias, singularity, and privacy concerns. On the negative side, a pause could slow down progress in areas where LLMs could have significant benefits, such as healthcare, education, and communication [450]. Additionally, it is possible that some organizations may choose to continue their work on LLMs in secret, which could lead to even less accountability and oversight.

As AI continues to advance and become more complex, it is crucial to heed the concerns of experts and examine the possible ethical implications and risks associated with this technology. To ensure that AI is used and developed in a responsible manner, we need to take appropriate measures that prioritize the safety and well-being of individuals and society as a whole. This includes developing and implementing robust ethical frameworks and guidelines that can govern the use of AI and prevent its misuse. Perhaps, an FDA like regulation for Large Language Models, beyond GPT-4 can be one potential solution [451].

Despite the potential risks, it is essential not to overlook the many benefits of AI. AI has the potential to improve various aspects of our lives, from healthcare to transportation, and even to address some of humanity's most pressing challenges, such as climate change, pandemics, another asteroid, and poverty. Therefore, it is crucial to strike a balance between harnessing the power of AI while also being mindful of its potential risks and drawbacks.

C. *Open Questions*

1) Ethical Considerations: Inadvertently, LLMs may perpetuate biases inherent in the training data, resulting in outputs that are biased or discriminatory [452]. The challenge lies in identifying and mitigating such biases to ensure fair and equitable treatment across diverse user groups and disciplines [453]. It is crucial to explore, the methods that can effectively address bias in the training data and enhance the

[8]https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[9]https://www.safe.ai/statement-on-ai-risk

fairness of LLMs. Additionally, LLMs may also generate and disseminate misinformation or harmful content. Ensuring that LLMs prioritize accurate and reliable information necessitates the implementation of mechanisms that can effectively assess and prioritize the authenticity and trustworthiness of the generated content. Incorporating robust data authenticity and consent mechanisms, data anonymization techniques, and data retention policies into the development and deployment of LLMs can help ensure the responsible and ethical handling of user data. Alignment is another approach that is about fine-tuning the models with the help of reward models through Reinforcement Learning where the reward model helps the LLM to generate honest, helpful and harmless output as perceived by humans.

2) Humans VS LLMs: Human interactions offer a deep level of empathy, emotional intelligence, and the ability to understand complex nuances in everyday life-situations. Humans responses are not only based on the current situation (prompt), but also considers other factors [454].

On the other hand, chatbots powered by AI have their advantages. They can operate 24/7, handle large volumes of inquiries simultaneously, and provide quick and consistent responses [455]. Chatbots excel in scenarios where efficiency, scalability, and rapid information retrieval are essential. They can assist with routine tasks, answer common questions, and provide instant access to information. It is imperative that Chatbots are well-aligned with human ethical considerations. AI-driven chatbots continuously learn and improve from user interactions, allowing them to become more accurate and efficient over time.

Chatbots are becoming increasingly autonomous [456]. They are now able to make their own decisions and to take action without human input. However, since in many disciplines, we need a human touch, it will be interesting to see whether in the case of generated AI or Conversational AI, the human touch is still needed?. In [457], investigates the differences between AI-generated scientific text and human-written scientific text. They found that AI-generated texts are informative, specific, objective and coherent, but also repetitive, generic, and boring [458].

There is also a need to develop new performance metrics for measuring the intelligence of AI systems, as traditional methods of assessing intelligence, such as IQ tests [196], are not well-suited for AI systems, as they are designed to measure human intelligence [459].

3) Interpretability: How can we enhance the interpretability of LLMs? Despite their impressive capabilities, LLMs often lack transparency, making it difficult to understand their decision-making process. Enhancing the interpretability of LLMs holds importance for several reasons [460]. It fosters trust and transparency by enabling users to understand the reasoning behind a model's specific response. It aids in identifying and addressing potential biases, errors, or unethical behavior. Additionally, interpretability contributes to debugging and improving model performance. However, achieving interpretability in LLMs is challenging due to their complexity and the nature of their training processes. LLMs have millions or even billions of parameters, making it difficult to directly trace their decision-making process. Furthermore, LLMs are trained using deep learning techniques, such as transformer architectures, which are considered black-box models, providing limited insight into their internal workings. Addressing the interpretability challenge in LLMs remains an active area of research. The ultimate goal is to make LLMs more transparent and accountable while preserving their impressive capabilities.

4) Data Efficiency: Data efficiency refers to the efficient use of training data for developing LLMs [461]. LLMs are typically trained using extremely large amounts of data to gain a performance that is acceptable or "human-like". Developing techniques to achieve this will be an open area of research as it will potentially enable better or similar performance with less data thereby reducing environmental impact. Making LLM development data efficient would allow for targeted development of LLM systems, and reduce turnaround time by easing off data collection and labeling burden. Several techniques which are being explored are data augmentation [462] and data selection [463], knowledge distillation [464], transfer learning [465], meta-learning [466], and others [467].

5) Training data contamination from AI-generated content: Data sources for training large models are typically scraped from the internet. With the increasing popularity of generative AI, it is possible that data present on the internet will have a significant component generated by AI models and therefore, reduce the human creativity aspect of the training data. Models, if trained on such data might end up trying to copy the generation aspects of previous AI models rather than humans only. One solution to this could be to use AI detection engines [468] that can determine content generated by AI before passing it through the model during the training process. There is a need to develop a dependable mechanism [356] to perform this task and retain the integrity of data.

D. Recommendations

In this Section, we provide recommendations for achieving optimal performance and highlight some of the practical applications.

1) Recommendations for Optimal Performance and Achieving Your Goals:

- **Use Advanced architecture:** At present, GPT-4 is one of the most advanced language models. It's impressive ability to generate highly relevant and coherent content makes it a preferred choice for most of the tasks.
- **Use Prompts with Detailed Task Context and Relevant Information:** LLM's performance is largely determined by the specificity and clarity of the input prompt. Detailed task contexts and relevant information help the model understand the task at hand better, leading to more accurate responses [469]. LLMs have been found to benefit significantly from one- or few-shot learning, which means that you can give one or few example responses in the prompt to significantly improve the quality of the generated output of LLMs.
- **Retrieve and Add Any Relevant Information to the Prompt:** Additional information, when included in the prompts, helps the model deliver more specific and focused responses [470]. If the user's task involves specific

knowledge, such as coding or medical information, providing relevant data and instructions in the prompt can improve the model's output [471].
- **Experiment with Prompt Engineering Techniques:** Given the complex and non-deterministic nature of LLM's behavior, trying out various prompt engineering strategies can lead to significant performance improvements. Techniques such as providing more explicit instructions, using leading questions, "double-quoting keyword", or presenting information in different formats may help achieve better results [472], [62].

2) Applications: Large Language has vast potential for practical applications, particularly when combined with human oversight and judgement.

- **Use in Low Stakes Applications, Combine with Human Oversight:** LLMs are best suited for low stakes applications, where errors or inaccuracies can be tolerated. Moreover, combining LLMs with human oversight can significantly mitigate the risk of errors, biases, and other issues [473], [474].
- **Source of Inspiration, Suggestions:** LLMs can serve as an invaluable source of inspiration and suggestions, helping users brainstorm ideas [475], create content [476], and make decisions [477].
- **Copilots Over Autonomous Agents:** Given its limitations, LLMs are better suited as a 'copilot' that provides assistance and suggestions, rather than an autonomous agent that acts without human input or oversight [471], [478].

3) Artificial General Intelligence - AGI: Artificial general intelligence (AGI [479]) is a hypothetical type of artificial intelligence that would have the ability to learn and perform any intellectual task. In [192], GPT-4 is found to have sparks of artificial general intelligence. GPT-4 is able to perform a variety of tasks; such as solving math problems, writing creative contents, writing poems and poetry [480] and answering questions in an informative way.

However, in our opinion, realizing the dream of AGI is still far away, despite of the rapid progress in the LLMs development. The key challenges include; understanding natural intelligence [481], developing adaptable fully autonomous models [482], and being safe and reliable with the understanding of the physical world [483], [484].

4) Democratizing AI: Democratizing AI [485] is a crucial movement that seeks to make artificial intelligence accessible and inclusive for a wide range of individuals and organizations. By breaking down barriers and providing user-friendly tools, democratization empowers diverse communities to leverage the power of AI to solve problems and drive innovation. It emphasizes the importance of open data, transparency, and accountability, ensuring that AI systems are unbiased, understandable, and ethically grounded.

No-code AI platforms [486] may also assist in democratizing AI initiatives [487], by providing a user-friendly interface that allows users to build and deploy ML models without any coding experience [488]. *No-code AI* can be used to leverage machine learning operations (MLOps) [489], to ensure models are deployed and managed effectively in production. Through democratization, we can harness the transformative potential of AI for the benefit of all, promoting a more inclusive and equitable future.

XII. Conclusion

In this survey, we provided a comprehensive exploration of LLMs, their implications, technical concepts, and practical learning and usage. We discussed the potential benefits and risks of LLMs, and explored the different ways in which they can be used. We also provided a number of examples of how LLMs are being used in practice. By delving into the technical intricacies, effective utilization, and future potential of LLMs, the survey will contribute to a deeper understanding and usage of these models within the research community. The survey has shed light on the key elements that drive the success of large language models through an examination of their working principles, diverse architectures and comparison between chatbots, guidelines for prompting, AI-enabled tools and plug-ins, optimal strategies for employing LLMs, as well as advancements in pre-training, fine-tuning, and capability evaluation.

Furthermore, the survey has also highlighted the importance of the safe and ethical use of AI tools like ChatGPT and others. It recognizes the need for developing guidelines and regulations to address concerns related to security, ethics, the economy, and the environment. Ensuring the responsible integration of LLMs in healthcare, academia, and other industries is critical, as it enables these tools to effectively support and enhance human endeavors while upholding the values of integrity, privacy, and fairness.

As the field of LLMs continues to evolve and progress, future research and development efforts should focus on improving the accuracy and performance of these models, addressing their limitations, and exploring new ways to use them. By adopting the guidelines presented in this survey, researchers and practitioners can contribute to the ongoing advancement of LLMs and ensure that they are used in a responsible and beneficial way.

Author contributions

Abbas Shah: Methodology, software, writing—original draft preparation, project administration, validation, visualization, software formal analysis. **Anas Zafar:** Software formal analysis, methodology, visualization, conceptualization, software, writing—original draft preparation, validation. **Muhammad Bilal Shaikh:** Project administration, methodology, software, writing—original draft preparation, visualization, validation, software formal analysis. **Amgad Muneer:** Project administration, Software validation, writing—original draft preparation, formal analysis, methodology, validation, investigation. **Muhammad Irfan:** Methodology, investigation, writing—original draft preparation, visualization, validation, software formal analysis. **Qasem Al-Tashi:** Methodology, conceptualization, software, writing—original draft preparation, visualization, investigation, formal analysis. **Rizwan

Qureshi: Conceptualization, methodology, project administration, software, writing—original draft preparation, visualization, formal analysis. **Muhammad Usman Hadi:** Conceptualization, project administration, methodology, writing—original draft preparation, formal analysis, visualization, funding acquisition. **Seyedali Mirjalili:** Supervision, writing—review and editing. **Naveed Akhtar:** Supervision, writing—review and editing. **Mohammed Ali Al-Garadi:** Supervision, writing—review and editing. **Jia Wu:** Supervision, writing—review and editing. **Mubarak Shah:**Project administraion, Supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Declaration of interest

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors have used generative artificial intelligence (AI) and AI-assisted technologies in the writing process and survey preparation. The authors used these technologies to draw figures, analyze data, improve readability, writing code and language. Authors are ultimately responsible and accountable for the contents of this work.

References

[1] K. S. Jones, "Natural language processing: a historical review," *Current issues in computational linguistics: in honour of Don Walker*, pp. 3–16, 1994.

[2] K. Chowdhary and K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.

[3] T. Iqbal and S. Qureshi, "The survey: Text generation models in deep learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2515–2528, 2022.

[4] D. Nozza, F. Bianchi, D. Hovy, *et al.*, "Honest: Measuring hurtful sentence completion in language models," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2021.

[5] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, and D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Computing Surveys*, 2021.

[6] M. Soam and S. Thakur, "Next word prediction using deep learning: A comparative study," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 653–658, IEEE, 2022.

[7] S. Diao, R. Xu, H. Su, Y. Jiang, Y. Song, and T. Zhang, "Taming pre-trained language models with n-gram representations for low-resource domain adaptation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 3336–3349, 2021.

[8] P. F. Brown, V. J. Della Pietra, P. V. Desouza, J. C. Lai, and R. L. Mercer, "Class-based n-gram models of natural language," *Computational linguistics*, vol. 18, no. 4, pp. 467–480, 1992.

[9] N. Omar and Q. Al-Tashi, "Arabic nested noun compound extraction based on linguistic features and statistical measures," *GEMA Online Journal of Language Studies*, vol. 18, no. 2, pp. 93–107, 2018.

[10] B. Rawat, A. S. Bist, U. Rahardja, Q. Aini, and Y. P. A. Sanjaya, "Recent deep learning based nlp techniques for chatbot development: An exhaustive survey," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–4, IEEE, 2022.

[11] Q. Lhoest, A. V. del Moral, Y. Jernite, A. Thakur, P. von Platen, S. Patil, J. Chaumond, M. Drame, J. Plu, L. Tunstall, *et al.*, "Datasets: A community library for natural language processing," *arXiv preprint arXiv:2109.02846*, 2021.

[12] O. Sharir, B. Peleg, and Y. Shoham, "The cost of training nlp models: A concise overview," *arXiv preprint arXiv:2004.08900*, 2020.

[13] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.

[14] A. Srivastava, A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, A. R. Brown, A. Santoro, A. Gupta, A. Garriga-Alonso, *et al.*, "Beyond the imitation game: Quantifying and extrapolating the capabilities of language models," *arXiv preprint arXiv:2206.04615*, 2022.

[15] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] D. Luitse and W. Denkena, "The great transformer: Examining the role of large language models in the political economy of ai," *Big Data & Society*, vol. 8, no. 2, p. 20539517211047734, 2021.

[18] Z. Dong, T. Tang, L. Li, and W. X. Zhao, "A survey on long text modeling with transformers," *arXiv preprint arXiv:2302.14502*, 2023.

[19] K. Adnan and R. Akbar, "An analytical study of information extraction from unstructured and multidimensional big data," *Journal of Big Data*, vol. 6, no. 1, pp. 1–38, 2019.

[20] H. Zhang, X. Li, and L. Bing, "Video-llama: An instruction-tuned audio-visual language model for video understanding," *arXiv preprint arXiv:2306.02858*, 2023.

[21] A. Rouditchenko, A. Boggust, D. Harwath, B. Chen, D. Joshi, S. Thomas, K. Audhkhasi, H. Kuehne, R. Panda, R. Feris, *et al.*, "Avlnet: Learning audio-visual language representations from instructional videos," *arXiv preprint arXiv:2006.09199*, 2020.

[22] Y. Zhao, Z. Lin, D. Zhou, Z. Huang, J. Feng, and B. Kang, "Bubogpt: Enabling visual grounding in multi-modal llms," *arXiv preprint arXiv:2307.08581*, 2023.

[23] J. Huang and K. C.-C. Chang, "Towards reasoning in large language models: A survey," *arXiv preprint arXiv:2212.10403*, 2022.

[24] N. Pappas and T. Meyer, "A survey on language modeling using neural networks," tech. rep., Idiap, 2012.

[25] J. R. Bellegarda, "Statistical language model adaptation: review and perspectives," *Speech communication*, vol. 42, no. 1, pp. 93–108, 2004.

[26] J. Lafferty and C. Zhai, "Probabilistic relevance models based on document and query generation," *Language modeling for information retrieval*, pp. 1–10, 2003.

[27] V. A. Petrushin, "Hidden markov models: Fundamentals and applications," in *Online Symposium for Electronics Engineer*, 2000.

[28] S. Khudanpur and J. Wu, "Maximum entropy techniques for exploiting syntactic, semantic and collocational dependencies in language modeling," *Computer Speech & Language*, vol. 14, no. 4, pp. 355–372, 2000.

[29] R. Rosenfeld, "Two decades of statistical language modeling: Where do we go from here?," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1270–1278, 2000.

[30] E. Arisoy, T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Deep neural network language models," in *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, pp. 20–28, 2012.

[31] J. R. Bellegarda, "Exploiting latent semantic information in statistical language modeling," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1279–1296, 2000.

[32] F. Alva-Manchego, C. Scarton, and L. Specia, "Data-driven sentence simplification: Survey and benchmark," *Computational Linguistics*, vol. 46, no. 1, pp. 135–187, 2020.

[33] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: a survey," *Multimedia Tools and Applications*, vol. 80, pp. 9411–9457, 2021.

[34] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[35] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *Journal of Big Data*, vol. 2, no. 1, pp. 1–24, 2015.

[36] M. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*, pp. 1–5, IEEE, 2013.

[37] L. Deng and Y. Liu, "A joint introduction to natural language processing and to deep learning," *Deep learning in natural language processing*, pp. 1–22, 2018.

[38] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017.

[39] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur, "Recurrent neural network based language model.," in *Interspeech*, vol. 2, pp. 1045–1048, Makuhari, 2010.

[40] S. Hochreiter, "Recurrent neural net learning and vanishing gradient," *International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.

[41] S. Hihi and Y. Bengio, "Hierarchical recurrent neural networks for long-term dependencies," *Advances in neural information processing systems*, vol. 8, 1995.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[43] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," *arXiv preprint arXiv:1803.02155*, 2018.

[44] B. Ghojogh and A. Ghodsi, "Attention mechanism, transformers, bert, and gpt: tutorial and survey," 2020.

[45] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[46] Q. Liu, M. J. Kusner, and P. Blunsom, "A survey on contextual embeddings," *arXiv preprint arXiv:2003.07278*, 2020.

[47] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, 2020.

[48] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "Text summarization techniques: a brief survey," *arXiv preprint arXiv:1707.02268*, 2017.

[49] Y. Ge, W. Hua, J. Ji, J. Tan, S. Xu, and Y. Zhang, "Openagi: When llm meets domain experts," *arXiv preprint arXiv:2304.04370*, 2023.

[50] R. V. P. Marcel, B. E. M. Fernando, and Y. V. J. Roberto, "A brief history of the artificial intelligence: chatgpt: The evolution of gpt," in *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–5, IEEE, 2023.

[51] M. Zhang and J. Li, "A commentary of gpt-3 in mit technology review 2021," *Fundamental Research*, vol. 1, no. 6, pp. 831–833, 2021.

[52] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.

[53] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[54] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature Medicine*, pp. 1–11, 2023.

[55] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, *et al.*, "Scientific discovery in the age of artificial intelligence," *Nature*, vol. 620, no. 7972, pp. 47–60, 2023.

[56] J. Wang, Y. Huang, C. Chen, Z. Liu, S. Wang, and Q. Wang, "Software testing with large language model: Survey, landscape, and vision," *arXiv preprint arXiv:2307.07221*, 2023.

[57] F. F. Xu, U. Alon, G. Neubig, and V. J. Hellendoorn, "A systematic evaluation of large language models of code," in *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming*, pp. 1–10, 2022.

[58] J. Cabrera, M. S. Loyola, I. Magaña, and R. Rojas, "Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 313–326, Springer, 2023.

[59] A. Creswell, M. Shanahan, and I. Higgins, "Selection-inference: Exploiting large language models for interpretable logical reasoning," *arXiv preprint arXiv:2205.09712*, 2022.

[60] E. Ferrara, "Should chatgpt be biased? challenges and risks of bias in large language models," *arXiv preprint arXiv:2304.03738*, 2023.

[61] K. Tirumala, D. Simig, A. Aghajanyan, and A. S. Morcos, "D4: Improving llm pretraining via document de-duplication and diversification," *arXiv preprint arXiv:2308.12284*, 2023.

[62] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with chatgpt," *arXiv preprint arXiv:2302.11382*, 2023.

[63] P. P. Ray, "Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, 2023.

[64] A. Sudmann, "On the media-political dimension of artificial intelligence: Deep learning as a black box and openai," *Digital Culture & Society*, vol. 4, no. 1, pp. 181–200, 2018.

[65] A. Koubaa, "Gpt-4 vs. gpt-3.5: A concise showdown," 2023.

[66] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27730–27744, 2022.

[67] S. Huang, L. Dong, W. Wang, Y. Hao, S. Singhal, S. Ma, T. Lv, L. Cui, O. K. Mohammed, Q. Liu, *et al.*, "Language is not all you need: Aligning perception with language models," *arXiv preprint arXiv:2302.14045*, 2023.

[68] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

[69] Y. Du, Z. Liu, J. Li, and W. X. Zhao, "A survey of vision-language pre-trained models," *arXiv preprint arXiv:2202.10936 , year=2022*.

[70] G. Mialon, R. Dessì, M. Lomeli, C. Nalmpantis, R. Pasunuru, R. Raileanu, B. Rozière, T. Schick, J. Dwivedi-Yu, A. Celikyilmaz, *et al.*, "Augmented language models: a survey," *arXiv preprint arXiv:2302.07842*, 2023.

[71] R. Qureshi, M. Irfan, H. Ali, A. Khan, A. S. Nittala, S. Ali, A. Shah, T. M. Gondal, F. Sadak, Z. Shah, *et al.*, "Artificial intelligence and biosensors in healthcare and its clinical relevance: A review," *IEEE Access*, 2023.

[72] Q. Al-Tashi, M. B. Saad, A. Sheshadri, C. C. Wu, J. Y. Chang, B. Al-Lazikani, C. Gibbons, N. I. Vokes, J. Zhang, J. J. Lee, *et al.*, "Swarmdeepsurv: swarm intelligence advances deep survival network for prognostic radiomics signatures in four solid cancers," *Patterns*.

[73] S. Mohamadi, G. Mujtaba, N. Le, G. Doretto, and D. A. Adjeroh, "Chatgpt in the age of generative ai and large language models: A concise survey," *arXiv preprint arXiv:2307.04251*, 2023.

[74] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[75] M. Sallam, "The utility of chatgpt as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations," *medRxiv*, pp. 2023–02, 2023.

[76] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, *et al.*, "Language models of protein sequences at the scale of evolution enable accurate structure prediction," *BioRxiv*, vol. 2022, p. 500902, 2022.

[77] A. Madani, B. McCann, N. Naik, N. S. Keskar, N. Anand, R. R. Eguchi, P.-S. Huang, and R. Socher, "Progen: Language modeling for protein generation," *arXiv preprint arXiv:2004.03497*, 2020.

[78] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt," *arXiv preprint arXiv:2303.04226*, 2023.

[79] I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," *arXiv preprint arXiv:1903.10676*, 2019.

[80] J. Li, T. Tang, W. X. Zhao, J.-Y. Nie, and J.-R. Wen, "Pretrained language models for text generation: A survey," *arXiv preprint arXiv:2201.05273*, 2022.

[81] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, "Bloomberggpt: A large language model for finance," *arXiv preprint arXiv:2303.17564*, 2023.

[82] T. Eloundou, S. Manning, P. Mishkin, and D. Rock, "Gpts are gpts: An early look at the labor market impact potential of large language models," *arXiv preprint arXiv:2303.10130*, 2023.

[83] M. Chen and T. et. al., "Evaluating large language models trained on code," *arXiv preprint arXiv:2107.03374*, 2021.

[84] S. Salman, J. A. Shamsi, and R. Qureshi, "Deep fake generation and detection: Issues, challenges, and solutions," *IT Professional*, vol. 25, no. 1, pp. 52–59, 2023.

[85] Z. Sun, "A short survey of viewing large language models in legal aspect," *arXiv preprint arXiv:2303.09136 , year=2023*.

[86] R. Qureshi, M. Irfan, T. M. Gondal, S. Khan, J. Wu, M. U. Hadi, J. Heymach, X. Le, H. Yan, and T. Alam, "Ai in drug discovery and its clinical relevance," *Heliyon*, 2023.

[87] Q. Al-Tashi, M. B. Saad, A. Muneer, R. Qureshi, S. Mirjalili, A. Sheshadri, X. Le, N. I. Vokes, J. Zhang, and J. Wu, "Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review," *International journal of molecular sciences*, vol. 24, no. 9, p. 7781, 2023.

[88] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, *et al.*, "A survey on large language model based autonomous agents," *arXiv preprint arXiv:2308.11432*, 2023.

[89] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, and N. Zhang, "Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities," *arXiv preprint arXiv:2305.13168*, 2023.

[90] L. Huynh, J. Hong, A. Mian, H. Suzuki, Y. Wu, and S. Camtepe, "Quantum-inspired machine learning: a survey," *arXiv preprint arXiv:2308.11269*, 2023.

[91] E. Brynjolfsson, D. Li, and L. R. Raymond, "Generative ai at work," tech. rep., National Bureau of Economic Research, 2023.

[92] P. Samuelson, "Generative ai meets copyright," *Science*, vol. 381, no. 6654, pp. 158–161, 2023.

[93] I. Chiang, *Unleashing the Power of Generative AI: The Race for Advancement and the Global Ramifications*. PhD thesis, Massachusetts Institute of Technology, 2023.

[94] S. Wang, S. Menon, T. Long, K. Henderson, D. Li, K. Crowston, M. Hansen, J. V. Nickerson, and L. B. Chilton, "Reelframer: Co-creating news reels on social media with generative ai," *arXiv preprint arXiv:2304.09653*, 2023.

[95] S. Mayahi and M. Vidrih, "The impact of generative ai on the future of visual content marketing," *arXiv preprint arXiv:2211.12660*, 2022.

[96] S.-C. Chen, "Multimedia research toward the metaverse," *IEEE MultiMedia*, vol. 29, no. 1, pp. 125–127, 2022.

[97] A. Zentner, "Applied innovation: Artificial intelligence in higher education," *Available at SSRN 4314180*, 2022.

[98] J. Sun, Q. V. Liao, M. Muller, M. Agarwal, S. Houde, K. Talamadupula, and J. D. Weisz, "Investigating explainability of generative ai for code through scenario-based design," in *27th International Conference on Intelligent User Interfaces*, pp. 212–228, 2022.

[99] J. Morley, N. J. DeVito, and J. Zhang, "Generative ai for medical research," 2023.

[100] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, 2006.

[101] A. Pérez-Suárez, J. F. Martínez-Trinidad, and J. A. Carrasco-Ochoa, "A review of conceptual clustering algorithms," *Artificial Intelligence Review*, vol. 52, pp. 1267–1296, 2019.

[102] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[103] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine*, vol. 7, p. 25, 2020.

[104] A. A. de Hond, A. M. Leeuwenberg, L. Hooft, I. M. Kant, S. W. Nijman, H. J. van Os, J. J. Aardoom, T. P. Debray, E. Schuit, M. van Smeden, *et al.*, "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review," *NPJ digital medicine*, vol. 5, no. 1, p. 2, 2022.

[105] C. Zhang, C. Zhang, S. Zheng, Y. Qiao, C. Li, M. Zhang, S. K. Dam, C. M. Thwal, Y. L. Tun, L. L. Huy, *et al.*, "A complete survey on generative ai (aigc): Is chatgpt from gpt-4 to gpt-5 all you need?," *arXiv preprint arXiv:2303.11717*, 2023.

[106] C. Zhang, C. Zhang, S. Zheng, M. Zhang, M. Qamar, S.-H. Bae, and I. S. Kweon, "A survey on audio diffusion models: Text to speech synthesis and enhancement in generative ai," *arXiv preprint arXiv:2303.13336*, vol. 2, 2023.

[107] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," *IEEE Access*, vol. 8, pp. 63514–63537, 2020.

[108] N. Aldausari, A. Sowmya, N. Marcus, and G. Mohammadi, "Video generative adversarial networks: a review," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–25, 2022.

[109] S. Barke, M. B. James, and N. Polikarpova, "Grounded copilot: How programmers interact with code-generating models," *Proceedings of the ACM on Programming Languages*, vol. 7, no. OOPSLA1, pp. 85–111, 2023.

[110] D. Zhang, S. Li, X. Zhang, J. Zhan, P. Wang, Y. Zhou, and X. Qiu, "Speechgpt: Empowering large language models with intrinsic cross-modal conversational abilities," *arXiv preprint arXiv:2305.11000*, 2023.

[111] S. Hong, J. Seo, S. Hong, H. Shin, and S. Kim, "Large language models are frame-level directors for zero-shot text-to-video generation," *arXiv preprint arXiv:2305.14330*, 2023.

[112] Ö. AYDIN and E. KARAARSLAN, "Is chatgpt leading generative ai? what is beyond expectations," *What is Beyond Expectations*, 2023.

[113] B. Kim, H. Kim, S.-W. Lee, G. Lee, D. Kwak, D. H. Jeon, S. Park, S. Kim, S. Kim, D. Seo, *et al.*, "What changes can large-scale language models bring? intensive study on hyperclova: Billions-scale korean generative pretrained transformers," *arXiv preprint arXiv:2109.04650*, 2021.

[114] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[115] Y. Yuan, "On the power of foundation models," in *International Conference on Machine Learning*, pp. 40519–40530, PMLR, 2023.

[116] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[117] A. Jo, "The promise and peril of generative ai," *Nature*, vol. 614, no. 1, pp. 214–216, 2023.

[118] H. Bansal, K. Gopalakrishnan, S. Dingliwal, S. Bodapati, K. Kirchhoff, and D. Roth, "Rethinking the role of scale for in-context learning: An interpretability-based case study at 66 billion scale," *arXiv preprint arXiv:2212.09095*, 2022.

[119] M. Mariani, "Generative artificial intelligence and innovation: Conceptual foundations," *Available at SSRN 4249382*, 2022.

[120] W. Zeng, X. Ren, T. Su, H. Wang, Y. Liao, Z. Wang, X. Jiang, Z. Yang, K. Wang, X. Zhang, *et al.*, "Pangu-$alpha$: Large-scale autoregressive pretrained chinese language models with auto-parallel computation," *arXiv preprint arXiv:2104.12369*, 2021.

[121] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks," in *International conference on machine learning*, pp. 2391–2400, PMLR, 2017.

[122] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[123] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[124] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, *et al.*, "Training a helpful and harmless assistant with reinforcement learning from human feedback," *arXiv preprint arXiv:2204.05862*, 2022.

[125] A. Muneer and S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on twitter," *Future Internet*, vol. 12, no. 11, p. 187, 2020.

[126] H. Hassani and E. S. Silva, "The role of chatgpt in data science: how ai-assisted conversational interfaces are revolutionizing the field," *Big data and cognitive computing*, vol. 7, no. 2, p. 62, 2023.

[127] C.-H. Chiang and H.-y. Lee, "Can large language models be an alternative to human evaluations?," *arXiv preprint arXiv:2305.01937*, 2023.

[128] A. Yuan, A. Coenen, E. Reif, and D. Ippolito, "Wordcraft: story writing with large language models," in *27th International Conference on Intelligent User Interfaces*, pp. 841–852, 2022.

[129] T. Chakrabarty, V. Padmakumar, and H. He, "Help me write a poem: Instruction tuning as a vehicle for collaborative poetry writing," *arXiv preprint arXiv:2210.13669*, 2022.

[130] G. Franceschelli and M. Musolesi, "On the creativity of large language models," *arXiv preprint arXiv:2304.00008*, 2023.

[131] J. Daugman, "Biometric decision landscapes," tech. rep., University of Cambridge, Computer Laboratory, 2000.

[132] B. D. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang, "Chatgpt and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing," *Journal of the Association for Information Science and Technology*, vol. 74, no. 5, pp. 570–581, 2023.

[133] M. Du, F. He, N. Zou, D. Tao, and X. Hu, "Shortcut learning of large language models in natural language understanding: A survey," *arXiv preprint arXiv:2208.11857*, 2022.
[134] E. D. Liddy, "Natural language processing," 2001.
[135] X. Liu and W. B. Croft, "Statistical language modeling," *Annual Review of Information Science and Technology*, vol. 39, p. 1, 2004.
[136] B.-H. Juang and L. R. Rabiner, "Automatic speech recognition–a brief history of the technology development," *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, vol. 1, p. 67, 2005.
[137] P. Azunre, *Transfer learning for natural language processing*. Simon and Schuster, 2021.
[138] A. Kovačević and D. Kečo, "Bidirectional lstm networks for abstractive text summarization," in *Advanced Technologies, Systems, and Applications VI: Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT) 2021*, pp. 281–293, Springer, 2022.
[139] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, *et al.*, "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pp. 38–45, 2020.
[140] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *et al.*, "Improving language understanding by generative pre-training," 2018.
[141] N. A. Akbar, I. Darmayanti, S. M. Fati, and A. Muneer, "Deep learning of a pre-trained language model's joke classifier using gpt-2," *Journal of Hunan University Natural Sciences*, vol. 48, no. 8, 2021.
[142] R. Dale, "Gpt-3: What's it good for?," *Natural Language Engineering*, vol. 27, no. 1, pp. 113–118, 2021.
[143] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
[144] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020.
[145]
[146] A. Karpathy, "State of GPT." https://www.youtube.com/watch?v=bZQun8Y4L2A, 2023.
[147] S. Biderman and S. et. al., "Pythia: A suite for analyzing large language models across training and scaling," *arXiv preprint arXiv:2304.01373*, 2023.
[148] D. Hernandez, T. Brown, T. Conerly, N. DasSarma, D. Drain, S. El-Showk, N. Elhage, Z. Hatfield-Dodds, T. Henighan, T. Hume, *et al.*, "Scaling laws and interpretability of learning from repeated data," *arXiv preprint arXiv:2205.10487 , year=2022*.
[149] J. W. Rae, S. Borgeaud, T. Cai, K. Millican, J. Hoffmann, F. Song, J. Aslanides, S. Henderson, R. Ring, S. Young, *et al.*, "Scaling language models: Methods, analysis & insights from training gopher," *arXiv preprint arXiv:2112.11446 , year=2021*.
[150] N. Carlini, D. Ippolito, M. Jagielski, K. Lee, F. Tramer, and C. Zhang, "Quantifying memorization across neural language models," *arXiv preprint arXiv:2202.07646*, 2022.
[151] P. Banerjee and H. Han, "Language modeling approaches to information retrieval," 2009.
[152] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.
[153] I. Dergaa, K. Chamari, P. Zmijewski, and H. B. Saad, "From human writing to artificial intelligence generated text: examining the prospects and potential threats of chatgpt in academic writing," *Biology of Sport*, vol. 40, no. 2, pp. 615–622, 2023.
[154] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. d. L. Casas, L. A. Hendricks, J. Welbl, A. Clark, *et al.*, "Training compute-optimal large language models," *arXiv preprint arXiv:2203.15556*, 2022.
[155] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade: Second Edition*, pp. 421–436, Springer, 2012.
[156] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.
[157] S. Praveen and V. Vajrobol, "Understanding the perceptions of healthcare researchers regarding chatgpt: a study based on bidirectional encoder representation from transformers (bert) sentiment analysis and topic modeling," *Annals of Biomedical Engineering*, pp. 1–3, 2023.
[158] W. Zhao, H. Hu, W. Zhou, J. Shi, and H. Li, "Best: Bert pre-training for sign language recognition with coupling tokenization," *arXiv preprint arXiv:2302.05075 , year=2023*.
[159] L. Jiarong, X. Hong, J. Wenchao, Y. Jianren, and W. Tao, "Knowledge enhanced bert based on corpus associate generation," in *Machine Learning for Cyber Security: 4th International Conference, ML4CS 2022, Guangzhou, China, December 2–4, 2022, Proceedings, Part III*, pp. 533–547, Springer, 2023.
[160] M. Irfan, A. I. Sanka, Z. Ullah, and R. C. Cheung, "Reconfigurable content-addressable memory (CAM) on FPGAs: A tutorial and survey," *Future Generation Computer Systems*, vol. 128, pp. 451–465, 2022.
[161] L. Fan, L. Li, Z. Ma, S. Lee, H. Yu, and L. Hemphill, "A bibliometric review of large language models research from 2017 to 2023," *arXiv preprint arXiv:2304.02020*, 2023.
[162] J. Su, S. Yu, and D. Luo, "Enhancing aspect-based sentiment analysis with capsule network," *IEEE Access*, vol. 8, pp. 100551–100561, 2020.
[163] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.
[164] A. Qamar, F. B. Muslim, F. Gregoretti, L. Lavagno, and M. T. Lazarescu, "High-level synthesis for semi-global matching: Is the juice worth the squeeze?," *IEEE Access*, vol. 5, pp. 8419–8432, 2016.
[165] W. Ahmad, B. Ayrancioglu, and I. Hamzaoglu, "Low error efficient approximate adders for fpgas," *IEEE Access*, vol. 9, pp. 117232–117243, 2021.
[166] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.
[167] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "Ctrl: A conditional transformer language model for controllable generation," *arXiv preprint arXiv:1909.05858*, 2019.
[168] P. Li, M. Zhang, P. Lin, J. Wan, and M. Jiang, "Conditional embedding pre-training language model for image captioning," *Neural Processing Letters*, vol. 54, no. 6, pp. 4987–5003, 2022.
[169] D. Su, Y. Xu, G. I. Winata, P. Xu, H. Kim, Z. Liu, and P. Fung, "Generalizing question answering system with pre-trained language model fine-tuning," in *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, pp. 203–211, 2019.
[170] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, *et al.*, "Large language models encode clinical knowledge," *Nature*, pp. 1–9, 2023.
[171] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny, "Minigpt-4: Enhancing vision-language understanding with advanced large language models," *arXiv preprint arXiv:2304.10592*, 2023.
[172] L. Wang, C. Lyu, T. Ji, Z. Zhang, D. Yu, S. Shi, and Z. Tu, "Document-level machine translation with large language models," *arXiv preprint arXiv:2304.02210*, 2023.
[173] B. Zhang, B. Haddow, and A. Birch, "Prompting large language model for machine translation: A case study," *arXiv preprint arXiv:2301.07069*, 2023.
[174] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, "Text classification via large language models," *arXiv preprint arXiv:2305.08377*, 2023.
[175] R. Song, Z. Liu, X. Chen, H. An, Z. Zhang, X. Wang, and H. Xu, "Label prompt for multi-label text classification," *Applied Intelligence*, vol. 53, no. 8, pp. 8761–8775, 2023.
[176] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. McKeown, and T. B. Hashimoto, "Benchmarking large language models for news summarization," *arXiv preprint arXiv:2301.13848*, 2023.
[177] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz, and T. Funkhouser, "Tidybot: Personalized robot assistance with large language models," *arXiv preprint arXiv:2305.05658*, 2023.
[178] R. Luo, Z. Zhao, M. Yang, J. Dong, M. Qiu, P. Lu, T. Wang, and Z. Wei, "Valley: Video assistant with large language model enhanced ability," *arXiv preprint arXiv:2306.07207*, 2023.
[179] S. Wadhwa, S. Amir, and B. C. Wallace, "Revisiting relation extraction in the era of large language models," *arXiv preprint arXiv:2305.05003*, 2023.
[180] X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang, S. Huang, P. Xie, J. Xu, Y. Chen, M. Zhang, *et al.*, "Zero-shot information extraction via chatting with chatgpt," *arXiv preprint arXiv:2302.10205*, 2023.
[181] C. Li, X. Zhang, D. Chrysostomou, and H. Yang, "Tod4ir: A humanised task-oriented dialogue system for industrial robots," *IEEE Access*, vol. 10, pp. 91631–91649, 2022.
[182] J. Deng, H. Sun, Z. Zhang, J. Cheng, and M. Huang, "Recent advances towards safe, responsible, and moral dialogue systems: A survey," *arXiv preprint arXiv:2302.09270*, 2023.

[183] J. Wei, J. Wei, Y. Tay, D. Tran, A. Webson, Y. Lu, X. Chen, H. Liu, D. Huang, D. Zhou, *et al.*, "Larger language models do in-context learning differently," *arXiv preprint arXiv:2303.03846*, 2023.
[184] T. Y. Zhuo, Z. Li, Y. Huang, Y.-F. Li, W. Wang, G. Haffari, and F. Shiri, "On robustness of prompt-based semantic parsing with large pre-trained language model: An empirical study on codex," *arXiv preprint arXiv:2301.12868*, 2023.
[185] V. Bhat and P. Bhattacharyya, "Survey: Automatic speech recognition for indian languages,"
[186] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *International Conference on Machine Learning*, pp. 28492–28518, PMLR, 2023.
[187] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, E. Li, X. Wang, M. Dehghani, S. Brahma, *et al.*, "Scaling instruction-finetuned language models," *arXiv preprint arXiv:2210.11416*, 2022.
[188] M. Suzgun, N. Scales, N. Schärli, S. Gehrmann, Y. Tay, H. W. Chung, A. Chowdhery, Q. V. Le, E. H. Chi, D. Zhou, *et al.*, "Challenging big-bench tasks and whether chain-of-thought can solve them," *arXiv preprint arXiv:2210.09261*, 2022.
[189] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A. N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba, "Evaluating large language models trained on code," 2021.
[190] Y. Huang, Y. Bai, Z. Zhu, J. Zhang, J. Zhang, T. Su, J. Liu, C. Lv, Y. Zhang, J. Lei, *et al.*, "C-eval: A multi-level multi-discipline chinese evaluation suite for foundation models," *arXiv preprint arXiv:2305.08322*, 2023.
[191] W. Chen and E. W. X. M. J. X. T. X. X. W. P. L. Ming Yin, Max Ku, "Theoremqa: A theorem-driven question answering dataset," *arXiv preprint arXiv:2305.12524*, 2023.
[192] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.
[193] R. Anil, A. M. Dai, O. Firat, M. Johnson, D. Lepikhin, A. Passos, S. Shakeri, E. Taropa, P. Bailey, Z. Chen, *et al.*, "Palm 2 technical report," *arXiv preprint arXiv:2305.10403*, 2023.
[194] Y. Wang, H. Le, A. D. Gotmare, N. D. Bui, J. Li, and S. C. Hoi, "Codet5+: Open code large language models for code understanding and generation," *arXiv preprint arXiv:2305.07922*, 2023.
[195] F. C. Kitamura, "Chatgpt is shaping the future of medical writing but still requires human judgment," 2023.
[196] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, *et al.*, "Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models," *PLoS digital health*, vol. 2, no. 2, p. e0000198, 2023.
[197] M. Sallam, "Chatgpt utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns," in *Healthcare*, vol. 11, p. 887, MDPI, 2023.
[198] A. Gilson, C. W. Safranek, T. Huang, V. Socrates, L. Chi, R. A. Taylor, D. Chartash, *et al.*, "How does chatgpt perform on the united states medical licensing examination? the implications of large language models for medical education and knowledge assessment," *JMIR Medical Education*, vol. 9, no. 1, p. e45312, 2023.
[199] M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, "Evaluating the feasibility of chatgpt in healthcare: an analysis of multiple clinical and research scenarios," *Journal of Medical Systems*, vol. 47, no. 1, p. 33, 2023.
[200] S. A. Basit, R. Qureshi, S. Musleh, R. Guler, M. S. Rahman, K. H. Biswas, and T. Alam, "Covid-19base v3: Update of the knowledgebase for drugs and biomedical entities linked to covid-19," *Frontiers in Public Health*, vol. 11, p. 1125917, 2023.
[201] A. Rao, J. Kim, M. Kamineni, M. Pang, W. Lie, and M. D. Succi, "Evaluating chatgpt as an adjunct for radiologic decision-making," *medRxiv*, pp. 2023–02, 2023.
[202] D. Duong and B. D. Solomon, "Analysis of large-language model versus human performance for genetics questions," *medRxiv*, pp. 2023–01, 2023.
[203] N. Fijačko, L. Gosak, G. Štiglic, C. T. Picard, and M. J. Douma, "Can chatgpt pass the life support exams without entering the american heart association course?," *Resuscitation*, vol. 185, 2023.
[204] A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, and H. Müller, "Ai for life: Trends in artificial intelligence for biotechnology," *New Biotechnology*, vol. 74, pp. 16–24, 2023.
[205] M. R. Haque and S. Rubya, "An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews," *JMIR mHealth and uHealth*, vol. 11, no. 1, p. e44838, 2023.
[206] S. M. Jungmann, T. Klan, S. Kuhn, and F. Jungmann, "Accuracy of a chatbot (ada) in the diagnosis of mental disorders: comparative case study with lay and expert users," *JMIR formative research*, vol. 3, no. 4, p. e13863, 2019.
[207] D. Magalhaes Azevedo and S. Kieffer, "User reception of ai-enabled mhealth apps: The case of babylon health.,"
[208] P. Malik, M. Pathania, V. K. Rathaur, *et al.*, "Overview of artificial intelligence in medicine," *Journal of family medicine and primary care*, vol. 8, no. 7, p. 2328, 2019.
[209] mbzuai oryx, "Xraygpt: Chest radiographs summarization using medical vision-language models," 2023.
[210] J. Ma and B. Wang, "Segment anything in medical images," *arXiv preprint arXiv:2304.12306*, 2023.
[211] Y. Li, C. Gao, X. Song, X. Wang, Y. Xu, and S. Han, "Druggpt: A gpt-based strategy for designing potential ligands targeting specific proteins," *bioRxiv*, pp. 2023–06, 2023.
[212] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," *Nature*, vol. 616, no. 7956, pp. 259–265, 2023.
[213] S. K. Karn, R. Ghosh, O. Farri, *et al.*, "shs-nlp at radsum23: Domain-adaptive pre-training of instruction-tuned llms for radiology report impression generation," *arXiv preprint arXiv:2306.03264*, 2023.
[214] S. Peng, K. Yuan, L. Gao, and Z. Tang, "Mathbert: A pre-trained model for mathematical formula understanding," *arXiv preprint arXiv:2105.00377*, 2021.
[215] J. S. () and W. Y. (), "Unlocking the power of chatgpt: A framework for applying generative ai in education," *ECNU Review of Education*, vol. 0, no. 0, p. 20965311231168423, 0.
[216] "An era of chatgpt as a significant futuristic support tool: A study on features, abilities, and challenges," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, p. 100089, 2022.
[217] H. Crompton and D. Burke, "Artificial intelligence in higher education: the state of the field," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 22, 2023.
[218] L. Zhu, W. Mou, T. Yang, and R. Chen, "Chatgpt can pass the aha exams: Open-ended questions outperform multiple-choice format," *Resuscitation*, vol. 188, p. 109783, 2023.
[219] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.
[220] "Khan academy explores the potential for gpt-4 in a limited pilot program," 2023.
[221] "Harnessing gpt-4 so that all students benefit. a nonprofit approach for equal access," 2023.
[222] H. H. Thorp, "Chatgpt is fun, but not an author," 2023.
[223] C. Stokel-Walker, "Chatgpt listed as author on research papers: many scientists disapprove," *Nature*, vol. 613, no. 7945, pp. 620–621, 2023.
[224] E. Hannan and S. Liu, "Ai: new source of competitiveness in higher education," *Competitiveness Review: An International Business Journal*, vol. 33, no. 2, pp. 265–279, 2023.
[225] O. Buruk, "Academic writing with gpt-3.5: Reflections on practices, efficacy and transparency," *arXiv preprint arXiv:2304.11079*, 2023.
[226] M. Dowling and B. Lucey, "Chatgpt for (finance) research: The bananarama conjecture," *Finance Research Letters*, vol. 53, p. 103662, 2023.
[227] X.-Y. Liu, G. Wang, and D. Zha, "Fingpt: Democratizing internet-scale data for financial large language models," *arXiv preprint arXiv:2307.10485*, 2023.
[228] A. Zaremba and E. Demir, "Chatgpt: Unlocking the future of nlp in finance," *Available at SSRN 4323643*, 2023.
[229] A. Lopez-Lira and Y. Tang, "Can chatgpt forecast stock price movements? return predictability and large language models," *arXiv preprint arXiv:2304.07619*, 2023.

[230] Y. Yang, M. C. S. Uy, and A. Huang, "Finbert: A pretrained language model for financial communications," *arXiv preprint arXiv:2006.08097*, 2020.

[231] D. Peskoff and B. M. Stewart, "Credible without credit: Domain experts assess generative language models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 427–438, 2023.

[232] S. Agarwal, S. Alok, P. Ghosh, and S. Gupta, "Financial inclusion and alternate credit scoring for the millennials: role of big data and machine learning in fintech," *Business School, National University of Singapore Working Paper, SSRN*, vol. 3507827, 2020.

[233] K. B. Hansen, "The virtue of simplicity: On machine learning models in algorithmic trading," *Big Data & Society*, vol. 7, no. 1, p. 2053951720926558, 2020.

[234] Z. Lin, Z. Song, Z. Dai, and Q. V. Le, "Fingpt: Open-source financial large language models," *arXiv preprint arXiv:2306.06031*, 2023.

[235] M. Fraiwan and N. Khasawneh, "A review of chatgpt applications in education, marketing, software engineering, and healthcare: Benefits, drawbacks, and research directions," *arXiv preprint arXiv:2305.00237*, 2023.

[236] D. Tiro, "The possibility of applying chatgpt (ai) for calculations in mechanical engineering," in *New Technologies, Development and Application VI: Volume 1*, pp. 313–320, Springer, 2023.

[237] Y. Wardat, M. A. Tashtoush, R. AlAli, and A. M. Jarrah, "Chatgpt: A revolutionary tool for teaching and learning mathematics," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 19, no. 7, p. em2286, 2023.

[238] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, "Mathematical capabilities of chatgpt," *arXiv preprint arXiv:2301.13867 , year=2023*.

[239] X. Wang, N. Anwer, Y. Dai, and A. Liu, "Chatgpt for design, manufacturing, and education," 2023.

[240] S. Badini, S. Regondi, E. Frontoni, and R. Pugliese, "Assessing the capabilities of chatgpt to improve additive manufacturing troubleshooting," *Advanced Industrial and Engineering Polymer Research*, 2023.

[241] J. V. Pavlik, "Collaborating with chatgpt: Considering the implications of generative artificial intelligence for journalism and media education," *Journalism & Mass Communication Educator*, vol. 78, no. 1, pp. 84–93, 2023.

[242] L. Chan, L. Hogaboam, and R. Cao, "Ai in media and entertainment," in *Applied Artificial Intelligence in Business: Concepts and Cases*, pp. 305–324, Springer, 2022.

[243] R. Lachman and M. Joffe, "Applications of artificial intelligence in media and entertainment," in *Analyzing future applications of AI, sensors, and robotics in society*, pp. 201–220, IGI Global, 2021.

[244] Z. Wang, "Mediagpt: A large language model target chinese media," *arXiv preprint arXiv:2307.10930*, 2023.

[245] J. M. Pérez, D. A. Furman, L. A. Alemany, and F. Luque, "Robertuito: a pre-trained language model for social media text in spanish," *arXiv preprint arXiv:2111.09453*, 2021.

[246] M. Abdulhai, C. Crepy, D. Valter, J. Canny, and N. Jaques, "Moral foundations of large language models," in *AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI*, 2022.

[247] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico, "Deep learning for recommender systems: A netflix case study," *AI Magazine*, vol. 42, no. 3, pp. 7–18, 2021.

[248] "Databricks - Media Entertainment Solutions." https://www.databricks.com/solutions/industries/media-and-entertainment. Accessed: Insert date accessed.

[249] J. Kim, K. Xu, and K. Merrill Jr, "Man vs. machine: Human responses to an ai newscaster and the role of social presence," *The Social Science Journal*, pp. 1–13, 2022.

[250] M. Feng, "The development of" ai" synthetic anchor in the context of artificial intelligence," *Highlights in Art and Design*, vol. 2, no. 1, pp. 38–40, 2023.

[251] A. of the article, "Ai-generated news presenter appears in kuwait," *Al Jazeera*, April 2023.

[252] A. of the article, "This is how ai could change the future of journalism," *Sky News*, 2023.

[253] M. Ajevski, K. Barker, A. Gilbert, L. Hardie, and F. Ryan, "Chatgpt and the future of legal education and practice," *The Law Teacher*, vol. 0, no. 0, pp. 1–13, 2023.

[254] "The legal ai you've been waiting for." https://casetext.com/cocounsel/. 31 July 2023].

[255] K. Y. Iu and V. M.-Y. Wong, "ChatGPT by OpenAI: The End of Litigation Lawyers?," 2023. Available at SSRN.

[256] J. J. Nay, "Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans," *arXiv preprint*, 2022.

[257] D. Trautmann, A. Petrova, and F. Schilder, "Legal Prompt Engineering for Multilingual Legal Judgement Prediction," *arXiv preprint*, 2022.

[258] D. M. Katz, M. J. Bommarito, S. Gao, and P. Arredondo, "GPT-4 Passes the Bar Exam," March 2023.

[259] J. H. Choi, K. E. Hickman, A. Monahan, and D. B. Schwarcz, "ChatGPT Goes to Law School," *Journal of Legal Education*, January 2023. Forthcoming.

[260] F. Yu, L. Quartey, and F. Schilder, "Legal prompting: Teaching a language model to think like a lawyer," 2022.

[261] J. J. Nay, "Large language models as fiduciaries: A case study toward robustly communicating with artificial intelligence through legal standards," 2023.

[262] H. Alkaissi and S. I. McFarlane, "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing," *CURIUS J. MED. SCI.*, vol. 15, 2023. Forthcoming.

[263] J. H. Choi, K. E. Hickman, A. Monahan, and D. B. Schwarcz, "Supra Note 7," 2023. Reference to a previously cited work.

[264] J. Greene, "Will ChatGPT Make Lawyers Obsolete? (Hint: Be Afraid)," *Reuters*, December 2022.

[265] The Guardian, "Two US Lawyers Fined for Submitting Fake Court Citations by ChatGPT," June 2023.

[266] ABC News, "US Lawyer Uses ChatGPT to Research Case with Embarrassing Result," June 2023.

[267] Business Standard, "US Judge Orders Lawyers Not to Use ChatGPT-drafted Content in Court," May 2023.

[268] P. Rivas and L. Zhao, "Marketing with ChatGPT: Navigating the Ethical Terrain of GPT-Based Chatbot Technology," *AI*, vol. 4, pp. 375–384, 2023.

[269] S. Verma, R. Sharma, S. Deb, and D. Maitra, "Artificial intelligence in marketing: Systematic review and future research direction," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100002, 2021.

[270] C. Zielinski, M. Winker, R. Aggarwal, L. Ferris, M. Heinemann, J. Lapeña, S. Pai, and L. Citrome, "Chatbots, ChatGPT, and Scholarly Manuscripts - WAME Recommendations on ChatGPT and Chatbots in Relation to Scholarly Publications," *Afro-Egypt. J. Infect. Endem. Dis.*, vol. 13, pp. 75–79, 2023.

[271] A. F.-B. Sun, Grace H. DNP and R.-B.-C. C. F. Hoelscher, Stephanie H. DNP, "The ChatGPT Storm and What Faculty Can Do," *Nurse Educator*, vol. 48, pp. 119–124, May/June 2023.

[272] L. Ma and B. Sun, "Machine learning and ai in marketing – connecting computing power to human insights," *International Journal of Research in Marketing*, vol. 37, no. 3, pp. 481–504, 2020.

[273] O. Yara, A. Brazheyev, L. Golovko, and V. Bashkatova, "Legal regulation of the use of artificial intelligence: Problems and development prospects," *European Journal of Sustainable Development*, vol. 10, p. 281, Feb. 2021.

[274] M. Stone, E. Aravopoulou, Y. Ekinci, G. Evans, M. Hobbs, A. Labib, P. Laughlin, J. Machtynger, and L. Machtynger, "Artificial Intelligence (AI) in Strategic Marketing Decision-Making: A research agenda," *Bottom Line*, vol. 33, pp. 183–200, 2020.

[275] E. Hermann, "Leveraging Artificial Intelligence in Marketing for Social Good—An Ethical Perspective," *J Bus Ethics*, vol. 179, pp. 43–61, 2022.

[276] Z. Liu, X. Yu, L. Zhang, Z. Wu, C. Cao, H. Dai, L. Zhao, W. Liu, D. Shen, Q. Li, *et al.*, "Deid-gpt: Zero-shot medical text de-identification by gpt-4," *arXiv preprint arXiv:2303.11032*, 2023.

[277] A. D. Subagja, A. M. A. Ausat, A. R. Sari, M. I. Wanof, and S. Suherlan, "Improving customer service quality in msmes through the use of chatgpt," *Jurnal Minfo Polgan*, vol. 12, no. 2, pp. 380–386, 2023.

[278] K. Howell, G. Christian, P. Fomitchov, G. Kehat, J. Marzulla, L. Rolston, J. Tredup, I. Zimmerman, E. Selfridge, and J. Bradley, "The economic trade-offs of large language models: A case study," *arXiv preprint arXiv:2306.07402*, 2023.

[279] J. Potts, D. W. Allen, C. Berg, and N. Ilyushina, "Large language models reduce agency costs," *Available at SSRN*, 2023.

[280] P. A. Olujimi and A. Ade-Ibijola, "Nlp techniques for automating responses to customer queries: a systematic review," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 20, 2023.

[281] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–38, 2023.

[282] Y. Li, M. Min, D. Shen, D. Carlson, and L. Carin, "Video generation from text," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.
[283] A. Conneau and G. Lample, "Cross-lingual language model pretraining," *Advances in neural information processing systems*, vol. 32, 2019.
[284] M. J. Ali, "Chatgpt and lacrimal drainage disorders: performance and scope of improvement," *Ophthalmic Plastic and Reconstructive Surgery*, vol. 39, no. 3, p. 221, 2023.
[285] X.-Q. Dao, "Performance comparison of large language models on vnhsge english dataset: Openai chatgpt, microsoft bing chat, and google bard," *arXiv preprint arXiv:2307.02288*, 2023.
[286] J. Rudolph, S. Tan, and S. Tan, "War of the chatbots: Bard, bing chat, chatgpt, ernie and beyond. the new ai gold rush and its impact on higher education," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023.
[287] I. Ahmed, M. Kajol, U. Hasan, P. P. Datta, A. Roy, and M. R. Reza, "Chatgpt vs. bard: A comparative study," *UMBC Student Collection*, 2023.
[288] R. Thoppilan, D. De Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H.-T. Cheng, A. Jin, T. Bos, L. Baker, Y. Du, *et al.*, "Lamda: Language models for dialog applications," *arXiv preprint arXiv:2201.08239*, 2022.
[289] X. Amatriain, "Transformer models: an introduction and catalog," *arXiv preprint arXiv:2302.07730*, 2023.
[290] Y. Hu, I. Ameer, X. Zuo, X. Peng, Y. Zhou, Z. Li, Y. Li, J. Li, X. Jiang, and H. Xu, "Zero-shot clinical entity recognition using chatgpt," *arXiv preprint arXiv:2303.16416*, 2023.
[291] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of chatgpt: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1122–1136, 2023.
[292] M. Abdullah, A. Madain, and Y. Jararweh, "Chatgpt: Fundamentals, applications and social impacts," in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–8, IEEE, 2022.
[293] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.
[294] D. Xuan-Quy, L. Ngoc-Bich, V. The-Duy, P. Xuan-Dung, N. Bac-Bien, N. Van-Tien, N. Thi-My-Thanh, and N. Hong-Phuoc, "Vnhsge: Vietnamese high school graduation examination dataset for large language models," *arXiv preprint arXiv:2305.12199*, 2023.
[295] H. Trng, "Chatgpt in education-a global and vietnamese research overview," 2023.
[296] K. M. Caramancion, "News verifiers showdown: A comparative performance evaluation of chatgpt 3.5, chatgpt 4.0, bing ai, and bard in news fact-checking," *arXiv preprint arXiv:2306.17176*, 2023.
[297] L. Zhang, "Four tax questions for chatgpt and other language models," 2023.
[298] M. King, "Gpt-4 aligns with the new liberal party, while other large language models refuse to answer political questions,"
[299] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, Y. Shao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *arXiv preprint arXiv:2209.00796*, 2022.
[300] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high fidelity image generation," *The Journal of Machine Learning Research*, vol. 23, no. 1, pp. 2249–2281, 2022.
[301] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2023.
[302] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan, "Foundational models defining a new era in vision: A survey and outlook," *arXiv preprint arXiv:2307.13721*, 2023.
[303] J. Huggins and J. Zou, "Quantifying the accuracy of approximate diffusions and markov chains," in *Artificial Intelligence and Statistics*, pp. 382–391, PMLR, 2017.
[304] W. Feller, "On the theory of stochastic processes, with particular reference to applications, p 403–432," 1949.
[305] J. R. Hershey and P. A. Olsen, "Approximating the kullback leibler divergence between gaussian mixture models," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4, pp. IV–317, IEEE, 2007.
[306] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy," *arXiv preprint arXiv:1710.10686*, 2017.
[307] A. Sedghi, L. J. O'Donnell, T. Kapur, E. Learned-Miller, P. Mousavi, and W. M. Wells III, "Image registration: Maximum likelihood, minimum entropy and deep learning," *Medical image analysis*, vol. 69, p. 101939, 2021.
[308] X. Pan, A. Tewari, T. Leimkühler, L. Liu, A. Meka, and C. Theobalt, "Drag your gan: Interactive point-based manipulation on the generative image manifold," in *ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1–11, 2023.
[309] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
[310] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.
[311] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–10, 2022.
[312] G. Kim, T. Kwon, and J. C. Ye, "Diffusionclip: Text-guided diffusion models for robust image manipulation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2426–2435, 2022.
[313] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
[314] S. AI, "Stablediffusion2.1 release." https://stability.ai/blog/stablediffusion2-1-release7-dec-2022, 2022. Accessed on August 3, 2023.
[315] U. Singer, A. Polyak, T. Hayes, X. Yin, J. An, S. Zhang, Q. Hu, H. Yang, O. Ashual, O. Gafni, *et al.*, "Make-a-video: Text-to-video generation without text-video data," *arXiv preprint arXiv:2209.14792*, 2022.
[316] T. Reddy, R. Williams, and C. Breazeal, "Text classification for ai education.," in *SIGCSE*, p. 1381, 2021.
[317] E. Loper and S. Bird, "Nltk: The natural language toolkit," *arXiv preprint cs/0205028*, 2002.
[318] Y. Vasiliev, *Natural language processing with Python and spaCy: A practical introduction*. No Starch Press, 2020.
[319] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, *et al.*, "Huggingface's transformers: State-of-the-art natural language processing," *arXiv preprint arXiv:1910.03771*, 2019.
[320] J. Pachouly, S. Ahirrao, K. Kotecha, G. Selvachandran, and A. Abraham, "A systematic literature review on software defect prediction using artificial intelligence: Datasets, data validation methods, approaches, and tools," *Engineering Applications of Artificial Intelligence*, vol. 111, p. 104773, 2022.
[321] E. A. Van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, "Chatgpt: five priorities for research," *Nature*, vol. 614, no. 7947, pp. 224–226, 2023.
[322] K. Nguyen-Trung, A. K. Saeri, and S. Kaufman, "Applying chatgpt and ai-powered tools to accelerate evidence reviews," 2023.
[323] N. Gleason, "Chatgpt and the rise of ai writers: How should higher education respond?," *Times Higher Education*, 2022.
[324] G. Cooper, "Examining science education in chatgpt: An exploratory study of generative artificial intelligence," *Journal of Science Education and Technology*, vol. 32, pp. 444–452, 2023.
[325] L. Skavronskaya, A. H. Hadinejad, and D. Cotterell, "Reversing the threat of artificial intelligence to opportunity: a discussion of chatgpt in tourism education," *Journal of Teaching in Travel & Tourism*, vol. 23, no. 2, pp. 253–258, 2023.
[326] B. Yetiştiren, I. Özsoy, M. Ayerdem, and E. Tüzün, "Evaluating the code quality of ai-assisted code generation tools: An empirical study on github copilot, amazon codewhisperer, and chatgpt," *arXiv preprint arXiv:2304.10778*, 2023.
[327] A. M. Dakhel, V. Majdinasab, A. Nikanjam, F. Khomh, M. C. Desmarais, and Z. M. J. Jiang, "Github copilot ai pair programmer: Asset or liability?," *Journal of Systems and Software*, vol. 203, p. 111734, 2023.
[328] F. G. e. a. Baptiste Rozière, Jonas Gehring, "Code llama: Open foundation models for code," *arXiv preprint*, 2023.
[329] T. Calò and L. De Russis, "Leveraging large language models for end-user website generation," in *International Symposium on End User Development*, pp. 52–61, Springer, 2023.
[330] Y. Shen, K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang, "Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface," *arXiv preprint arXiv:2303.17580*, 2023.

[331] O. Thawkar, A. Shaker, S. S. Mullappilly, H. Cholakkal, R. M. Anwer, S. Khan, J. Laaksonen, and F. S. Khan, "Xraygpt: Chest radiographs summarization using medical vision-language models," *arXiv preprint arXiv:2306.07971*, 2023.

[332] Y. Luo, J. Zhang, S. Fan, K. Yang, Y. Wu, M. Qiao, and Z. Nie, "Biomedgpt: Open multimodal generative pre-trained transformer for biomedicine," *arXiv preprint arXiv:2308.09442*, 2023.

[333] J. Zhou, X. He, L. Sun, J. Xu, X. Chen, Y. Chu, L. Zhou, X. Liao, B. Zhang, and X. Gao, "Skingpt-4: An interactive dermatology diagnostic system with visual large language model," *medRxiv*, pp. 2023–06, 2023.

[334] K. Kheiri and H. Karimi, "Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning," *arXiv preprint arXiv:2307.10234*, 2023.

[335] H. Chase, "Langchain, 10 2022," *URL https://github.com/hwchase17/langchain*.

[336] M. R. Chavez, T. S. Butler, P. Rekawek, H. Heo, and W. L. Kinzler, "Chat generative pre-trained transformer: why we should embrace this technology," *American Journal of Obstetrics and Gynecology*, 2023.

[337] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, *et al.*, "Palm: Scaling language modeling with pathways," *arXiv preprint arXiv:2204.02311 , year=2022*.

[338] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, and O. Hinz, "Welcome to the era of chatgpt et al. the prospects of large language models," *Business & Information Systems Engineering*, pp. 1–7, 2023.

[339] C. Xu, Y. Xu, S. Wang, Y. Liu, C. Zhu, and J. McAuley, "Small models are valuable plug-ins for large language models," *arXiv preprint arXiv:2305.08848*, 2023.

[340] C. Lyu, J. Xu, and L. Wang, "New trends in machine translation using large language models: Case examples with chatgpt," *arXiv preprint arXiv:2305.01181*, 2023.

[341] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, pp. 1678–1684, 2012.

[342] M. M. Al-Jefri and S. A. Mahmoud, "Context-sensitive arabic spell checker using context words and n-gram language models," in *2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, pp. 258–263, IEEE, 2013.

[343] Z. Jiang, J. Araki, H. Ding, and G. Neubig, "How can we know when language models know? on the calibration of language models for question answering," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 962–977, 2021.

[344] L. Wang, W. Zhao, Z. Wei, and J. Liu, "Simkgc: Simple contrastive knowledge graph completion with pre-trained language models," *arXiv preprint arXiv:2203.02167*, 2022.

[345] H. J. Dolfing and I. L. Hetherington, "Incremental language models for speech recognition using finite-state transducers," in *IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU'01.*, pp. 194–197, IEEE, 2001.

[346] R. Mao, Q. Liu, K. He, W. Li, and E. Cambria, "The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection," *IEEE Transactions on Affective Computing*, 2022.

[347] J. Zhang, R. Xie, Y. Hou, W. X. Zhao, L. Lin, and J.-R. Wen, "Recommendation as instruction following: A large language model empowered recommendation approach," *arXiv preprint arXiv:2305.07001*, 2023.

[348] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, *et al.*, "Summary of chatgpt/gpt-4 research and perspective towards the future of large language models," *arXiv preprint arXiv:2304.01852*, 2023.

[349] X. He, X. Shen, Z. Chen, M. Backes, and Y. Zhang, "Mgtbench: Benchmarking machine-generated text detection," *arXiv preprint arXiv:2303.14822*, 2023.

[350] M. T. I. Khondaker, A. Waheed, E. M. B. Nagoudi, and M. Abdul-Mageed, "Gptaraeval: A comprehensive evaluation of chatgpt on arabic nlp," *arXiv preprint arXiv:2305.14976*, 2023.

[351] J. Kim, J. H. Lee, S. Kim, J. Park, K. M. Yoo, S. J. Kwon, and D. Lee, "Memory-efficient fine-tuning of compressed large language models via sub-4-bit integer quantization," *arXiv preprint arXiv:2305.14152*, 2023.

[352] S. Arora, B. Yang, S. Eyuboglu, A. Narayan, A. Hojel, I. Trummer, and C. Ré, "Language models enable simple systems for generating structured views of heterogeneous data lakes," *arXiv preprint arXiv:2304.09433*, 2023.

[353] S. R. Bowman, "Eight things to know about large language models," *arXiv preprint arXiv:2304.00612*, 2023.

[354] Y. Fu, H. Peng, T. Khot, and M. Lapata, "Improving language model negotiation with self-play and in-context learning from ai feedback," *arXiv preprint arXiv:2305.10142*, 2023.

[355] H. Matsumi, D. Hallinan, D. Dimitrova, E. Kosta, and P. De Hert, *Data Protection and Privacy, Volume 15: In Transitional Times*. Bloomsbury Publishing, 2023.

[356] P. Hacker, A. Engel, and M. Mauer, "Regulating chatgpt and other large generative ai models," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1112–1123, 2023.

[357] S. A. Khowaja, P. Khuwaja, and K. Dev, "Chatgpt needs spade (sustainability, privacy, digital divide, and ethics) evaluation: A review," *arXiv preprint arXiv:2305.03123*, 2023.

[358] A. Chan, H. Bradley, and N. Rajkumar, "Reclaiming the digital commons: A public data trust for training data," *arXiv preprint arXiv:2303.09001*, 2023.

[359] W. H. Deng, B. Guo, A. Devrio, H. Shen, M. Eslami, and K. Holstein, "Understanding practices, challenges, and opportunities for user-engaged algorithm auditing in industry practice," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2023.

[360] M. Kraus, J. A. Bingler, M. Leippold, T. Schimanski, C. C. Senni, D. Stammbach, S. A. Vaghefi, and N. Webersinke, "Enhancing large language models with climate resources," *arXiv preprint arXiv:2304.00116*, 2023.

[361] E. Agathokleous, C. J. Saitanis, C. Fang, and Z. Yu, "Use of chatgpt: What does it mean for biology and environmental science?," *Science of The Total Environment*, p. 164154, 2023.

[362] Y. Shen, L. Heacock, J. Elias, K. D. Hentel, B. Reig, G. Shih, and L. Moy, "Chatgpt and other large language models are double-edged swords," 2023.

[363] Z. Sun, Y. Shen, Q. Zhou, H. Zhang, Z. Chen, D. Cox, Y. Yang, and C. Gan, "Principle-driven self-alignment of language models from scratch with minimal human supervision," *arXiv preprint arXiv:2305.03047*, 2023.

[364] H. Liu, Z. Teng, L. Cui, C. Zhang, Q. Zhou, and Y. Zhang, "Logicot: Logical chain-of-thought instruction-tuning data collection with gpt-4," *arXiv preprint arXiv:2305.12147*, 2023.

[365] Y. Liu, G. Deng, Z. Xu, Y. Li, Y. Zheng, Y. Zhang, L. Zhao, T. Zhang, and Y. Liu, "Jailbreaking chatgpt via prompt engineering: An empirical study," *arXiv preprint arXiv:2305.13860*, 2023.

[366] K. Yang, S. Ji, T. Zhang, Q. Xie, Z. Kuang, and S. Ananiadou, "Towards interpretable mental health analysis with chatgpt," 2023.

[367] J. White, S. Hays, Q. Fu, J. Spencer-Smith, and D. C. Schmidt, "Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design," *arXiv preprint arXiv:2303.07839*, 2023.

[368] C. Liu, X. Bao, H. Zhang, N. Zhang, H. Hu, X. Zhang, and M. Yan, "Improving chatgpt prompt for code generation," *arXiv preprint arXiv:2305.08360*, 2023.

[369] D. Miyake, A. Iohara, Y. Saito, and T. Tanaka, "Negative-prompt inversion: Fast image inversion for editing with text-guided diffusion models," *arXiv preprint arXiv:2305.16807*, 2023.

[370] N. Liu, S. Li, Y. Du, A. Torralba, and J. B. Tenenbaum, "Compositional visual generation with composable diffusion models," in *European Conference on Computer Vision*, pp. 423–439, Springer, 2022.

[371] AUTOMATIC1111, "Negative-prompt." https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Negative-prompt, 2022. Accessed on August 1, 2023.

[372] N. Tumanyan, M. Geyer, S. Bagon, and T. Dekel, "Plug-and-play diffusion features for text-driven image-to-image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1921–1930, 2023.

[373] H. Chang, H. Zhang, J. Barber, A. Maschinot, J. Lezama, L. Jiang, M.-H. Yang, K. Murphy, W. T. Freeman, M. Rubinstein, *et al.*, "Muse: Text-to-image generation via masked generative transformers," *arXiv preprint arXiv:2301.00704*, 2023.

[374] A. Chen, Y. Yao, P.-Y. Chen, Y. Zhang, and S. Liu, "Understanding and improving visual prompting: A label-mapping perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19133–19143, 2023.

[375] A. Bar, Y. Gandelsman, T. Darrell, A. Globerson, and A. Efros, "Visual prompting via image inpainting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25005–25017, 2022.

[376] T. Chakrabarty, A. Saakyan, O. Winn, A. Panagopoulou, Y. Yang, M. Apidianaki, and S. Muresan, "I spy a metaphor: Large language models and diffusion models co-create visual metaphors," *arXiv preprint arXiv:2305.14724*, 2023.

[377] R. Volum, S. Rao, M. Xu, G. A. DesGarennes, C. Brockett, B. Van Durme, O. Deng, A. Malhotra, and B. Dolan, "Craft an iron sword: Dynamically generating interactive game characters by prompting large language models tuned on code," in *The Third Wordplay: When Language Meets Games Workshop*, 2022.

[378] D. Hegde, J. M. J. Valanarasu, and V. M. Patel, "Clip goes 3d: Leveraging prompt tuning for language grounded 3d recognition," *arXiv preprint arXiv:2303.11313*, 2023.

[379] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy, "Challenges and applications of large language models," *arXiv preprint arXiv:2307.10169*, 2023.

[380] Y. Wolf, N. Wies, Y. Levine, and A. Shashua, "Fundamental limitations of alignment in large language models," *arXiv preprint arXiv:2304.11082*, 2023.

[381] R. Tang, Y.-N. Chuang, and X. Hu, "The science of detecting llm-generated texts," *arXiv preprint arXiv:2303.07205*, 2023.

[382] F. Ufuk, "The role and limitations of large language models such as chatgpt in clinical settings and medical journalism," *Radiology*, vol. 307, no. 3, p. e230276, 2023.

[383] R. Bhayana, S. Krishna, and R. R. Bleakney, "Performance of chatgpt on a radiology board-style examination: Insights into current strengths and limitations," *Radiology*, p. 230582, 2023.

[384] X. Yang, Y. Li, X. Zhang, H. Chen, and W. Cheng, "Exploring the limits of chatgpt for query or aspect-based text summarization," *arXiv preprint arXiv:2302.08081*, 2023.

[385] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[386] J. Kaddour, "The minipile challenge for data-efficient language models," *arXiv preprint arXiv:2304.08442*, 2023.

[387] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," *arXiv preprint arXiv:1508.07909*, 2015.

[388] T. Fujii, K. Shibata, A. Yamaguchi, T. Morishita, and Y. Sogawa, "How do different tokenizers perform on downstream tasks in scriptio continua languages?: A case study in japanese," *arXiv preprint arXiv:2306.09572*, 2023.

[389] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green ai," *Communications of the ACM*, vol. 63, no. 12, pp. 54–63, 2020.

[390] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," *arXiv preprint arXiv:2101.00190*, 2021.

[391] H. Laurençon, L. Saulnier, T. Wang, C. Akiki, A. Villanova del Moral, T. Le Scao, L. Von Werra, C. Mou, E. González Ponferrada, H. Nguyen, *et al.*, "The bigscience roots corpus: A 1.6 tb composite multilingual dataset," *Advances in Neural Information Processing Systems*, vol. 35, pp. 31809–31826, 2022.

[392] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *International Conference on Machine Learning*, pp. 2790–2799, PMLR, 2019.

[393] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.

[394] M. Pagliardini, D. Paliotta, M. Jaggi, and F. Fleuret, "Faster causal attention over large sequences through sparse flash attention," *arXiv preprint arXiv:2306.01160*, 2023.

[395] Z. Yao, R. Yazdani Aminabadi, M. Zhang, X. Wu, C. Li, and Y. He, "Zeroquant: Efficient and affordable post-training quantization for large-scale transformers," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27168–27183, 2022.

[396] S. Liu and Z. Wang, "Ten lessons we have learned in the new" sparseland": A short handbook for sparse neural network researchers," *arXiv preprint arXiv:2302.02596*, 2023.

[397] L. Chen, M. Zaharia, and J. Zou, "Frugalgpt: How to use large language models while reducing cost and improving performance," *arXiv preprint arXiv:2305.05176*, 2023.

[398] S. Chen, S. Wong, L. Chen, and Y. Tian, "Extending context window of large language models via positional interpolation," *arXiv preprint arXiv:2306.15595*, 2023.

[399] R. Li, J. Su, C. Duan, and S. Zheng, "Linear attention mechanism: An efficient attention for semantic segmentation," *arXiv preprint arXiv:2007.14902*, 2020.

[400] M. Guo, J. Ainslie, D. Uthus, S. Ontanon, J. Ni, Y.-H. Sung, and Y. Yang, "Longt5: Efficient text-to-text transformer for long sequences," *arXiv preprint arXiv:2112.07916*, 2021.

[401] X. Ma, X. Kong, S. Wang, C. Zhou, J. May, H. Ma, and L. Zettlemoyer, "Luna: Linear unified nested attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 2441–2453, 2021.

[402] B. Peng, E. Alcaide, Q. Anthony, A. Albalak, S. Arcadinho, H. Cao, X. Cheng, M. Chung, M. Grella, K. K. GV, *et al.*, "Rwkv: Reinventing rnns for the transformer era," *arXiv preprint arXiv:2305.13048*, 2023.

[403] T. Dao, D. Y. Fu, K. K. Saab, A. W. Thomas, A. Rudra, and C. Ré, "Hungry hungry hippos: Towards language modeling with state space models," *arXiv preprint arXiv:2212.14052*, 2022.

[404] Y. Yao, P. Wang, B. Tian, S. Cheng, Z. Li, S. Deng, H. Chen, and N. Zhang, "Editing large language models: Problems, methods, and opportunities," *arXiv preprint arXiv:2305.13172*, 2023.

[405] P. Schramowski, C. Turan, N. Andersen, C. A. Rothkopf, and K. Kersting, "Large pre-trained language models contain human-like biases of what is right and wrong to do," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 258–268, 2022.

[406] T. McCoy, E. Pavlick, and T. Linzen, "Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, (Florence, Italy), pp. 3428–3448, Association for Computational Linguistics, July 2019.

[407] J. Weston, E. Dinan, and A. Miller, "Retrieve and refine: Improved sequence generation models for dialogue," in *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*, (Brussels, Belgium), pp. 87–92, Association for Computational Linguistics, Oct. 2018.

[408] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: a large-scale dataset for fact extraction and VERification," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, (New Orleans, Louisiana), pp. 809–819, Association for Computational Linguistics, June 2018.

[409] D. V. Hada and S. K. Shevade, "Rexplug: Explainable recommendation using plug-and-play language model," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 81–91, 2021.

[410] Y. Gao, T. Sheng, Y. Xiang, Y. Xiong, H. Wang, and J. Zhang, "Chat-rec: Towards interactive and explainable llms-augmented recommender system," *arXiv preprint arXiv:2303.14524 , year=2023*.

[411] A. Uchendu, *REVERSE TURING TEST IN THE AGE OF DEEPFAKE TEXTS*. PhD thesis, The Pennsylvania State University, 2023.

[412] E. M. Bonsu and D. Baffour-Koduah, "From the consumers' side: Determining students' perception and intention to use chatgpt in ghanaian higher education," *Journal of Education, Society & Multiculturalism*, vol. 4, no. 1, pp. 1–29, 2023.

[413] Q. V. Liao and J. W. Vaughan, "AI Transparency in the Age of LLMs: A human-centered research roadmap," *arXiv preprint arXiv:2306.01941 , year=2023*.

[414] G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," *Information Fusion*, vol. 76, pp. 89–106, 2021.

[415] N. M. Deshpande, S. Gite, B. Pradhan, and M. E. Assiri, "Explainable artificial intelligence–a new step towards the trust in medical diagnosis with ai frameworks: A review," *Comput. Model. Eng. Sci*, vol. 133, pp. 1–30, 2022.

[416] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10076–10085, 2020.

[417] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai," *International Journal of Human-Computer Studies*, vol. 146, p. 102551, 2021.

[418] K. Valmeekam, A. Olmo, S. Sreedharan, and S. Kambhampati, "Large language models still can't plan (a benchmark for llms on planning and reasoning about change)," *arXiv preprint arXiv:2206.10498*, 2022.

[419] N. Bian, X. Han, L. Sun, H. Lin, Y. Lu, and B. He, "Chatgpt is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models," *arXiv preprint arXiv:2303.16421*, 2023.

[420] Y. LeCun, "A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27," *Open Review*, vol. 62, 2022.

[421] M. Hardy, I. Sucholutsky, B. Thompson, and T. Griffiths, "Large language models meet cognitive science: Llms as tools, models, and participants," in *Proceedings of the annual meeting of the cognitive science society*, 2023.

[422] F. Huang, H. Kwak, and J. An, “Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech,” *arXiv preprint arXiv:2302.07736*, 2023.

[423] A. Wan, E. Wallace, S. Shen, and D. Klein, “Poisoning language models during instruction tuning,” *arXiv preprint arXiv:2305.00944*, 2023.

[424] H. Li, D. Guo, W. Fan, M. Xu, and Y. Song, “Multi-step jailbreaking privacy attacks on chatgpt,” *arXiv preprint arXiv:2304.05197*, 2023.

[425] F. Perez and I. Ribeiro, “Ignore previous prompt: Attack techniques for language models,” *arXiv preprint arXiv:2211.09527*, 2022.

[426] Y. Liu, G. Deng, Y. Li, K. Wang, T. Zhang, Y. Liu, H. Wang, Y. Zheng, and Y. Liu, “Prompt injection attack against llm-integrated applications,” *arXiv preprint arXiv:2306.05499*, 2023.

[427] C. Zhang, C. Zhang, T. Kang, D. Kim, S.-H. Bae, and I. S. Kweon, “Attack-sam: Towards evaluating adversarial robustness of segment anything model,” *arXiv preprint arXiv:2305.00866*, 2023.

[428] L. Chen, M. Zaharia, and J. Zou, “How is chatgpt’s behavior changing over time?,” *arXiv preprint arXiv:2307.09009*, 2023.

[429] K. Nelson, G. Corbin, M. Anania, M. Kovacs, J. Tobias, and M. Blowers, “Evaluating model drift in machine learning algorithms,” in *2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, pp. 1–8, IEEE, 2015.

[430] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, “Large language models are human-level prompt engineers,” *arXiv preprint arXiv:2211.01910*, 2022.

[431] M. S. Rahaman, M. T. Ahsan, N. Anjum, H. J. R. Terano, and M. M. Rahman, “From chatgpt-3 to gpt-4: A significant advancement in ai-driven nlp tools,” *Journal of Engineering and Emerging Technologies*, vol. 2, no. 1, pp. 1–11, 2023.

[432] M. C. Rillig, M. Ågerstrand, M. Bi, K. A. Gould, and U. Sauerland, “Risks and benefits of large language models for the environment,” *Environmental Science & Technology*, vol. 57, no. 9, pp. 3464–3466, 2023.

[433] G. Fergusson, C. Fitzgerald, C. Frascella, M. Iorio, T. McBrien, C. Schroeder, B. Winters, and E. Zhou, “Contributions by,”

[434] P. Li, J. Yang, M. A. Islam, and S. Ren, “Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models,” *arXiv preprint arXiv:2304.03271*, 2023.

[435] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean, “Carbon emissions and large neural network training,” 2021.

[436] A. S. George, A. H. George, and A. G. Martin, “The environmental impact of ai: A case study of water consumption by chat gpt,” *Partners Universal International Innovation Journal*, vol. 1, no. 2, pp. 97–104, 2023.

[437] A. A. Chien, L. Lin, H. Nguyen, V. Rao, T. Sharma, and R. Wijayawardana, “Reducing the carbon impact of generative ai inference (today and in 2035),” *ACM Hot Carbon 2023*, 2023.

[438] B. Tomlinson, R. W. Black, D. J. Patterson, and A. W. Torrance, “The carbon emissions of writing and illustrating are lower for ai than for humans,” *arXiv preprint arXiv:2303.06219*, 2023.

[439] S. Biswas, “Potential use of chat gpt in global warming,” *Ann Biomed Eng*, vol. 51, pp. 1126–1127, 2023.

[440] Z. Yao, Y. Lum, A. Johnston, and et al., “Machine learning for a sustainable energy future,” *Nat Rev Mater*, vol. 8, pp. 202–215, 2023.

[441] X. Zhi and J. Wang, “Editorial: Ai-based prediction of high-impact weather and climate extremes under global warming: A perspective from the large-scale circulations and teleconnections,” *Frontiers in Earth Science*, vol. 11, 2023.

[442] J. Zhong, Y. Zhong, M. Han, T. Yang, and Q. Zhang, “The impact of ai on carbon emissions: evidence from 66 countries,” *Applied Economics*, vol. 0, no. 0, pp. 1–15, 2023.

[443] M. A. Habila, M. Ouladsmane, and Z. A. Alothman, “Chapter 21 - role of artificial intelligence in environmental sustainability,” in *Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence* (A. Srivastav, A. Dubey, A. Kumar, S. Kumar Narang, and M. Ali Khan, eds.), pp. 449–469, Elsevier, 2023.

[444] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. LIX, pp. 433–460, 1950.

[445] C. Carugati, “Competition in generative artificial intelligence foundation models,” *ARTIFICIAL INTELLIGENCE*, vol. 2, p. 3, 2023.

[446] T. Hoppner and L. Streatfeild, “Chatgpt, bard & co.: An introduction to ai for competition and regulatory lawyers,” *An Introduction to AI for Competition and Regulatory Lawyers (February 23, 2023)*, vol. 9, 2023.

[447] F. O. Letters, “Pause giant ai experiments: An open letter,” *Future of Life Institution. https://futureoflife. org/open-letter/pause-giant-ai-experiments*, 2023.

[448] M. Ienca, “Don’t pause giant ai for the wrong reasons,” *Nature Machine Intelligence*, pp. 1–2, 2023.

[449] B. Lin, D. Bouneffouf, G. Cecchi, and K. R. Varshney, “Towards healthy ai: Large language models need therapists too,” *arXiv preprint arXiv:2304.00416 , year=2023*.

[450] S. Harrer, “Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine,” *EBioMedicine*, vol. 90, 2023.

[451] M. Elmahdy and R. Sebro, “A snapshot of artificial intelligence research 2019–2021: is it replacing or assisting physicians?,” *Journal of the American Medical Informatics Association*, p. ocad094, 2023.

[452] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh, *et al.*, “Ethical and social risks of harm from language models,” *arXiv preprint arXiv:2112.04359*, 2021.

[453] G. Pistilli, “What lies behind agi: ethical concerns related to llms,” *Revue Ethique et Numérique*, 2022.

[454] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, “Why johnny can’t prompt: how non-ai experts try (and fail) to design llm prompts,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–21, 2023.

[455] J. Hill, W. R. Ford, and I. G. Farreras, “Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations,” *Computers in human behavior*, vol. 49, pp. 245–250, 2015.

[456] W. Wlasak, S. P. Zwanenburg, C. Paton, *et al.*, “Supporting autonomous motivation for physical activity with chatbots during the covid-19 pandemic: Factorial experiment,” *JMIR Formative Research*, vol. 7, no. 1, p. e38500, 2023.

[457] Y. Ma, J. Liu, and F. Yi, “Is this abstract generated by ai? a research for the gap between ai-generated scientific text and human-written scientific text,” *arXiv preprint arXiv:2301.10416*, 2023.

[458] E. Adamopoulou and L. Moussiades, “An overview of chatbot technology,” in *IFIP international conference on artificial intelligence applications and innovations*, pp. 373–383, Springer, 2020.

[459] M. Mitchell, “How do we know how smart ai systems are?,” 2023.

[460] C. Singh, J. X. Morris, J. Aneja, A. M. Rush, and J. Gao, “Explaining patterns in data with language models via interpretable autoprompting,” *arXiv preprint arXiv:2210.01848*, 2022.

[461] Z. Hu, Y. Lan, L. Wang, W. Xu, E.-P. Lim, R. K.-W. Lee, L. Bing, and S. Poria, “Llm-adapters: An adapter family for parameter-efficient fine-tuning of large language models,” *arXiv preprint arXiv:2304.01933*, 2023.

[462] V. Kumar, A. Choudhary, and E. Cho, “Data augmentation using pre-trained transformer models,” *arXiv preprint arXiv:2003.02245*, 2020.

[463] L. Xu, X. Zhang, and Q. Dong, “Cluecorpus2020: A large-scale chinese corpus for pre-training language model,” *arXiv preprint arXiv:2003.01355*, 2020.

[464] H. Pan, C. Wang, M. Qiu, Y. Zhang, Y. Li, and J. Huang, “Meta-kd: A meta knowledge distillation framework for language model compression across domains,” *arXiv preprint arXiv:2012.01266*, 2020.

[465] M. Ostendorff and G. Rehm, “Efficient language model training through cross-lingual and progressive transfer learning,” *arXiv preprint arXiv:2301.09626*, 2023.

[466] S. Iyer, X. V. Lin, R. Pasunuru, T. Mihaylov, D. Simig, P. Yu, K. Shuster, T. Wang, Q. Liu, P. S. Koura, *et al.*, “Opt-iml: Scaling language model instruction meta learning through the lens of generalization,” *arXiv preprint arXiv:2212.12017*, 2022.

[467] B. Zhuang, J. Liu, Z. Pan, H. He, Y. Weng, and C. Shen, “A survey on efficient training of transformers,” *arXiv preprint arXiv:2302.01107*, 2023.

[468] W. J. Dupps Jr, “Artificial intelligence and academic publishing,” *Journal of Cataract & Refractive Surgery*, vol. 49, no. 7, pp. 655–656, 2023.

[469] T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh, “Autoprompt: Eliciting knowledge from language models with automatically generated prompts,” *arXiv preprint arXiv:2010.15980*, 2020.

[470] W. Yu, D. Iter, S. Wang, Y. Xu, M. Ju, S. Sanyal, C. Zhu, M. Zeng, and M. Jiang, “Generate rather than retrieve: Large language models are strong context generators,” *arXiv preprint arXiv:2209.10063*, 2022.

[471] K. Greshake, S. Abdelnabi, S. Mishra, C. Endres, T. Holz, and M. Fritz, “More than you’ve asked for: A comprehensive analysis of novel prompt injection threats to application-integrated large language models,” *arXiv preprint arXiv:2302.12173*, 2023.

[472] L. S. Lo, "The clear path: A framework for enhancing information literacy through prompt engineering," *The Journal of Academic Librarianship*, vol. 49, no. 4, p. 102720, 2023.
[473] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, and L. Floridi, "Taking ai risks seriously: a new assessment model for the ai act," *AI & SOCIETY*, pp. 1–5, 2023.
[474] U. M. Fayyad, "From stochastic parrots to intelligent assistants—the secrets of data and human interventions," *IEEE Intelligent Systems*, vol. 38, no. 3, pp. 63–67, 2023.
[475] V. Salikutluk, D. Koert, and F. Jäkel, "Interacting with large language models: A case study on ai-aided brainstorming for guesstimation problems," in *HHAI 2023: Augmenting Human Intellect*, pp. 153–167, IOS Press, 2023.
[476] S. Moore, R. Tong, A. Singh, Z. Liu, X. Hu, Y. Lu, J. Liang, C. Cao, H. Khosravi, P. Denny, *et al.*, "Empowering education with llms-the next-gen interface and content generation," in *International Conference on Artificial Intelligence in Education*, pp. 32–37, Springer, 2023.
[477] K. Nottingham, P. Ammanabrolu, A. Suhr, Y. Choi, H. Hajishirzi, S. Singh, and R. Fox, "Do embodied agents dream of pixelated sheep?: Embodied decision making using language guided world modelling," *arXiv preprint arXiv:2301.12050*, 2023.
[478] S. Kang, J. Yoon, and S. Yoo, "Large language models are few-shot testers: Exploring llm-based general bug reproduction," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, pp. 2312–2323, IEEE, 2023.
[479] B. Goertzel, "Artificial general intelligence: concept, state of the art, and future prospects," *Journal of Artificial General Intelligence*, vol. 5, no. 1, p. 1, 2014.
[480] T. Giannini and J. P. Bowen, "Generative art and computational imagination: Integrating poetry and art," in *Proceedings of EVA London 2023*, pp. 211–219, BCS Learning & Development, 2023.
[481] C. Summerfield, *Natural General Intelligence: How understanding the brain can help us build AI*. Oxford University Press, 2022.
[482] N. Nascimento, P. Alencar, and D. Cowan, "Self-adaptive large language model (llm)-based multiagent systems," *arXiv preprint arXiv:2307.06187*, 2023.
[483] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, *et al.*, "Towards artificial general intelligence with hybrid tianjic chip architecture," *Nature*, vol. 572, no. 7767, pp. 106–111, 2019.
[484] T. Everitt, *Towards safe artificial general intelligence*. PhD thesis, The Australian National University (Australia), 2019.
[485] C. T. Wolf, "Democratizing ai? experience and accessibility in the age of artificial intelligence," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 26, no. 4, pp. 12–15, 2020.
[486] C. K. Iyer, F. Hou, H. Wang, Y. Wang, K. Oh, S. Ganguli, and V. Pandey, "Trinity: A no-code ai platform for complex spatial datasets," in *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pp. 33–42, 2021.
[487] B. Allen, S. Agarwal, J. Kalpathy-Cramer, and K. Dreyer, "Democratizing ai," *Journal of the American College of Radiology*, vol. 16, no. 7, pp. 961–963, 2019.
[488] L. Sundberg and J. Holmström, "Democratizing artificial intelligence: How no-code ai can leverage machine learning operations," *Business Horizons*, 2023.
[489] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (mlops): Overview, definition, and architecture," *IEEE Access*, 2023.