Large Language Models: A Survey of Surveys

MAX HORT, Simula Research Laboratory, Norway
FERNANDO VALLECILLOS-RUIZ, Simula Research Laboratory, Norway
LEON MOONEN*, Simula Research Laboratory, Norway

Not only did the growing interest in Large Language Models (LLMs) lead to a multitude of applications, news articles, social media posts, and new products, but it also resulted in a significant increase in research publications. To gain a better understanding of this vast number of publications, surveys help provide practitioners and researchers with much-needed overviews.

However, we have reached a point with tens of thousands of LLM publications, and the number of surveys on LLM publications has grown into hundreds. Ironically, the same surveys that set out to bring order and structure now contribute to the convolution of the space. For example, someone interested in the use of LLMs for the health sector has more than 80 potential surveys to choose from. To address this challenge, we carry out a tertiary literature review to gather and analyze LLM-related surveys, reviews, and mapping studies. By doing so, we aim to help practitioners and researchers navigate the vast array of existing surveys.

In total, we found 424 LLM surveys that have been published up to September 2024 that are included in this study. We devise a taxonomy and categorise surveys according to their main focus (e.g., fine-tuning of LLMs, application for software engineering tasks). To further support the navigation of LLM surveys and keep up to date, we created a GitHub repository that extends our scope to a total of 984 publications published up to August 2025, which is available from https://github.com/dataSED-condenSE/LLM-Survey-Survey.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Computing methodologies → Natural language processing**.

Additional Key Words and Phrases: large language model, literature survey, tertiary study

1 Introduction

Summaries are useful tools for providing overviews that help facilitate the understanding of diverse topics. In research, summaries are typically conducted through *secondary studies* (e.g., survey, mapping study, literature review) [83, 126]. Going one step further, *tertiary reviews* systematically analyze and summarize secondary studies [148]. Such tertiary reviews have proven useful across various domains, such as software engineering [43, 83], medical deep learning [62], economics [46], machine learning [150], requirements engineering [12], and sentiment analysis [181].

The surging popularity of Large Language Models (LLMs) over the recent years has led to thousands of research articles and, in turn, hundreds of secondary studies. This volume of secondary studies creates a need for a structured overview, and we argue that it is time to conduct a tertiary study of the field of LLMs. By creating an overview, we support researchers and practitioners in navigating the field of surveys, finding relevant ones when learning about LLMs, and understanding which aspects have been studied when designing new surveys.

* Corresponding author.

Authors' Contact Information: Max Hort, maxh@simula.no, Simula Research Laboratory, Oslo, Norway; Fernando Vallecillos-Ruiz, fernando@simula.no, Simula Research Laboratory, Oslo, Norway; Leon Moonen, leon.moonen@computer.org, Simula Research Laboratory, Oslo, Norway.

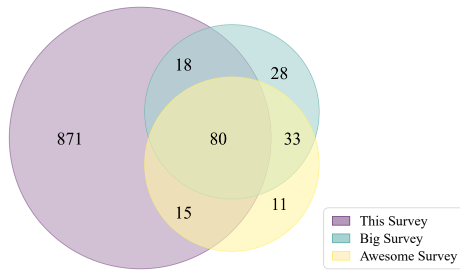

Fig. 1. Venn diagram illustrating the overlap between surveys from our systematic search and two repositories.

To the best of our knowledge, no tertiary study on LLMs has been published. However, we are aware of two GitHub repositories that provide a list of LLMs surveys. These are: ABigSurveyOfLLMs[1] and Awesome-LLM-Survey,[2] with 159 and 139 listed surveys on LLMs respectively. While valuable resources, these surveys did not follow a systematic search methodology. We extend far beyond their scope by carrying out a systematic literature search and collecting a total of 424 surveys that are included in this article, as well as an additional 560 that are included in our GitHub repository (see Section 2 for more details). The overlap of these two repositories and our collected surveys is shown in Figure 1, highlighting the additional studies collected.

Figure 2 shows the structure of our survey and how we categorize the existing surveys.

2 Survey Methodology and Search Results

2.1 Search Procedure

To search the literature for secondary studies about LLMs, we make use of the publication database dblp.[3] dblp contains publications from more than 1,800 journals and 6,000 conferences from the computer science domain, as well as non-peer-reviewed papers from arXiv. In particular, we use dblp to carry out a search of relevant publications based on filtering their titles.

To ensure that we obtain relevant search results, we define two sets of keywords. The first set contains terms related to language models, while the second contains terms related to literature collection (inspired by Kotti et al. [150]):

- LLM keywords: LLM, Language Model.
- Survey keywords: Survey, Overview, Literature, Review, Background, Research, Taxonomy, Systematic.

In addition to requiring that publication titles contain both keyword types, we treat them as inclusion criteria for our paper collection:

(1) LLMs: The paper focuses on language models.
(2) Literature overview: The paper represents a secondary study by collecting and presenting other works.

We exclude all studies that do not match these criteria, and omit studies that are not written in English. We check inclusion in two stages. First, we determine relevance of the search results based on their title. For instance, this removes literature on the study of "language modeling". Second, we read each paper with a suitable title and make a final inclusion decision based on its content.

[1] https://github.com/NiuTrans/ABigSurveyOfLLMs, last updated on 19th February 2025
[2] https://github.com/HqWu-HITCS/Awesome-LLM-Survey, last updated on 25th of May 2025
[3] https://dblp.org

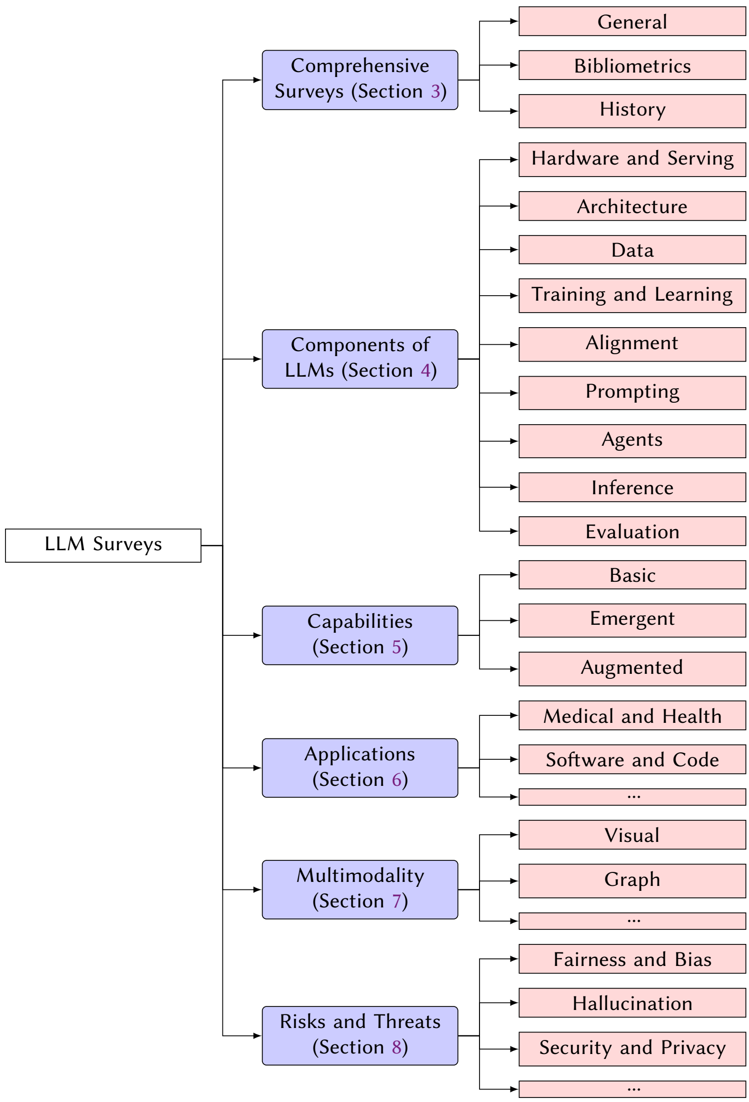

Fig. 2. Structure of this study.

2.2 Selection and Search Results

Table 1 summarizes the results of our search, which we carried out on 10th of September 2024. We start with a total of 1,173 unique publications from dblp, which fit at least one of the keyword combinations. 461 of these agree with our inclusion criteria according to their titles. After examining the 461 papers, we exclude 37 and end up with a total of 424 studies that are included in our survey and presented in the following sections. To ensure the timeliness of our work, we carried out an identical search on the 15th of August 2025, to find surveys that have been published in the last year. This resulted in an additional 560 surveys. Due to the large number of recent publications, we

Table 1. Summary of search results. The search was carried out on the 10th of September 2024 on dblp. Results for the updated search carried out on the 15th of August 2025 are shown in (blue). The additional 560 studies can be found in our GitHub repository.

	LLM Keyword		# Papers		
Survey Keyword	"LLM"	"Language Model"	Unique	Title	Content
"Survey"	63 (+181)	427 (+476)	1,173 (+1,672)	461	424 (+560)
"Overview"	6 (+12)	43 (+35)			
"Literature"	18 (+54)	65 (+91)			
"Review"	43 (+146)	231 (+291)			
"Background"	0 (+2)	8 (+9)			
"Research"	33 (+146)	201 (+241)			
"Taxonomy"	14 (+40)	48 (+51)			
"Systematic"	25 (+87)	134 (+169)			

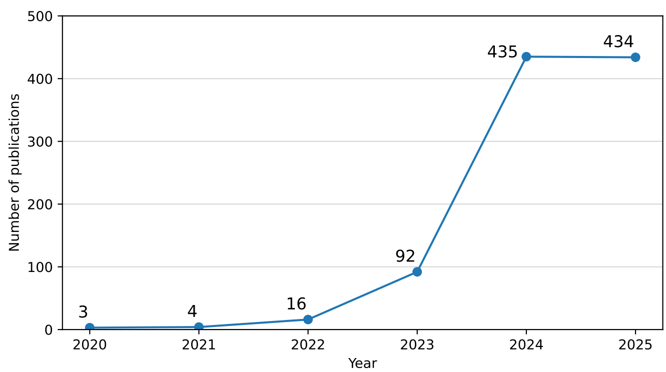

Fig. 3. Number of publications per year. The count for 2025 is based on a cut-off date of 15th of August 2025.

decided to only include them in our supplementary online repository.[4] The temporal distribution of these publications is shown in Figure 3.

3 Overviews and Comprehensive Surveys

We start the presentation of existing surveys by presenting general overviews that are helpful as an introduction to learn about LLMs. These surveys stand out by their comprehensiveness and the consideration of several aspects of LLMs. In addition to comprehensive surveys, we include surveys about bibliometrics and the history of LLMs to provide extensive background details.

Table 1 shows the included surveys for this section. For each, we list basic information (Authors, venue, year of publication), as well as the number of references they have and how often they have been cited. These can be useful indicators for their comprehensiveness and popularity. Lastly, we give a Unique Selling Point (USP), a point of focus which differentiates them from other surveys.

3.1 Comprehensive

Zhao et al. [417] created the most comprehensive survey on LLMs to date. This can not only be seen by the high number of references included (946), but also its popularity (over 5000 citations). This scale allows the survey to cover all important aspects of LLMs and is the only survey to consider aspects such as "scaling laws", which have not been covered by the other surveys.

[4] https://github.com/dataSED-condenSE/LLM-Survey-Survey

Table 2. Overview of comprehensive surveys and their unique selling point. The number of citations was collected from Google Scholar on the 21st of August 2025. The # studies shows how many publications are covered by the respective surveys.

Authors	Venue	Year	# Studies	Citations	Focus	USP
Movva et al. [222]	NAACL	2024	59	24	Bibliometrics	Industry and Academia (roles)
Fan et al. [65]	arXiv	2023	86	199	Bibliometrics	Research topics
Naveed et al. [223]	arXiv	2023	487	1496	Comprehensive	Architecture details
Raiaan et al. [254]	IEEE Access	2024	187	667	Comprehensive	Datasets per model
Zhao et al. [417]	arXiv	2023	946	5658	Comprehensive	Detailed settings
Yang et al. [368]	ACM TKDD	2024	143	1212	Comprehensive	NLP tasks
Minaee et al. [218]	arXiv	2024	243	1263	Comprehensive	Capabilities
Liu et al. [196]	arXiv	2024	175	144	Comprehensive	Training & Inference
Ling et al. [185]	arXiv	2024	297	57	Comprehensive	Specialization
Guo and Yu [93]	arXiv	2022	175	34	Comprehensive	Domain Adaptation
Wang et al. [317]	arXiv	2024	305	35	Comprehensive	Challenges and Opportunities
Miao et al. [216]	arXiv	2023	375	103	Comprehensive	Systems and Serving
Wei et al. [330]	arXiv	2023	223	74	History	Conventional models and linguistic units
Chu et al. [38]	arXiv	2024	88	93	History	Advancement of LLMs
Kumar [153]	Artif. Intell. Rev.	2024	249	138	History	Word embeddings, Deep Learning

There are several other surveys that provide a comprehensive overview of LLMs. While some of their contents naturally overlap, we outline their unique viewpoints. Raiaan et al. [254] provided an overview of the different sources for datasets (e.g., webpages, books, code). Naveed et al. [223] listed details on the architecture of LLMs. This includes information such as training objective, vocabulary size, type of attention, number of layers, attention heads, and hidden states. Minaee et al. [218] provided an overview of the capabilities of language models. Moreover, they survey the components necessary for building LLMs. Miao et al. [216] covered the serving of LLMs and optimization for faster inference time via modifying the models themselves or the hosting system.

Yang et al. [368] include the most comprehensive description of NLP tasks for LLMs. The survey by Liu et al. [196] focused on training and inference, ranging from the data processing stage to different fine-tuning paradigms and methods for speeding up the inference. Ling et al. [185] addressed the adaptation of LLMs to different domains in their survey. These techniques range from augmentation with external knowledge to fine-tuning. Similarly, Guo and Yu [93] described domain adaptation via data augmentation, model optimization (training) and model personalization.

3.2 Bibliometric

Fan et al. [65] carried out a bibliometric study covering 5752 publications from the Web of Science (WoS) Core Collection, collected from 2017 to early 2023. They investigated topics addressed by these publications and divided them into five categories: algorithm and NLP tasks, medical and engineering applications, social and humanitarian applications, critical studies, and infrastructure. Among these, “Algorithm and NLP tasks” span the majority of publications (54%), while “Infrastructure” and “Critical studies” cover less than 2% each. The countries which produced the highest number of research in this period are China and the USA. In terms of the collaboration among institutes, USA and UK have the highest centrality score.

Movva et al. [222] performed a study to reveal the influence of LLMs on AI research, and analyzed 16,979 LLM-related papers from arXiv during the period of January 2018 to September of 2023. They observed that many authors have not previously published NLP-related research, and a growing interest on the societal impact of LLMs. Similar to the findings by Fan et al. [65], US and China-based institutes contributed the highest number of publications. Overall, Movva et al. [222] observed few collaborations across countries.

3.3 History

Another three surveys outline current advances in language models while providing information on the history and early approaches [38, 153, 330]. For instance, Wei et al. [330] started their survey with an overview on conventional language models (e.g., structural and bidirectional language models), while also describing various linguistic units (i.e., characters, words, subwords, phrases, sentences). The role of word embeddings and deep learning for language models is addressed by Kumar [153]. Chu et al. [38] considered approaches ranging from 1990 (statistical language models) to 2023 (large language models).

4 Components of LLMs

This section outlines surveys that addressed the different components required for the training and use of LLMs. We structure our review around nine key components as shown in Figure 4. This taxonomy is inspired by the work of Naveed et al. [223] and Minaee et al. [218].

4.1 Hardware and Serving

LLMs are compute-intensive machine learning models and therefore require a certain degree of compute power and hardware infrastructure to be used. For instance, the running of LLMs can benefit from the use of GPUs [305] and high-performance computing [31]. The efficiency of training and applying LLMs has been improved from a diverse range of components [138, 308], such as processing units, storage systems, scheduling, and memory management [55, 61, 163, 305, 356, 386, 386, 428]. In addition, there are two dedicated surveys on improving efficiency via the key-value (KV) cache [274] (during training and inference) and compute-in-memory (i.e., reduces overhead of memory access by performing computations in memory) [334].

A frequently mentioned approach for accelerating the training process is parallelization [9, 18, 55, 61, 305]. Here, Duan et al. [61] mentioned three different types (Hybrid, Auto, Heterogeneous) and descriptions on optimizing communication. Another hardware consideration is the device on which LLMs are run. These can be edge devices [14, 250, 356] or in the cloud [9, 163, 260, 356, 386, 428].

4.2 Architecture

In this section, we present surveys that describe existing model types and information on their architectures. For instance, Gao et al. [79] listed models and provided details, such as their number of parameters and underlying base models. In addition, they evaluated 32 of them in various settings (e.g., zero-shot, few-shot, multi-modal) and presented tools that support the development with and for LLMs. Pahune and Chandrasekharan [229] showed the different available versions for each of the models and hardware details for their implementation.

Other surveys focus on specific model families. Kukreja et al. [152] considered open-source models, with particular focus on FALCON, BLOOM, and Llama2, for which data collection, architecture, and training stages are described. Kalyan [139] focused on GPT language models, in particular models ranging from GPT-3 to GPT-4, and collected their application to downstream tasks (e.g., text classification, information extraction, coding). Alipour et al. [5] focused on ChatGPT and OpenAI (e.g., the OpenAI playground). Other models were introduced as alternatives to ChatGPT. Lu et al. [201] considered different methods for LLM collaboration. For instance, LLM responses can be merged, or one can create an ensemble of multiple LLMs.

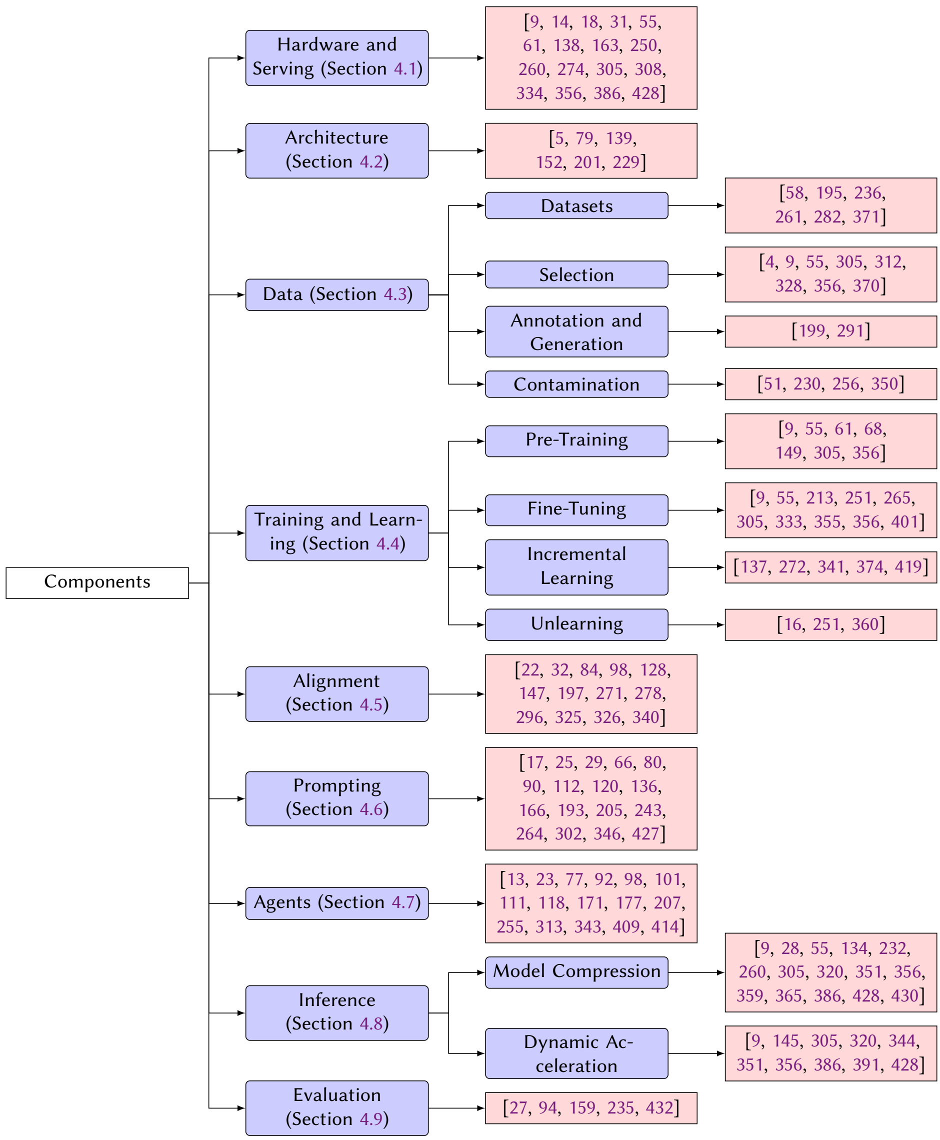

Fig. 4. Taxonomy of surveys on LLM components.

4.3 Data

The characteristics of an LLM are fundamentally determined by the data used in its creation and evaluation. Consequently, surveys in this field explore the entire data lifecycle, from the composition of datasets to the evaluation of their quality.

Datasets: Liu et al. [195] presented an exhaustive overview of datasets for large language models. They considered a total of 444 datasets from five categories: pre-training, instruction fine-tuning, preference, evaluation, and NLP. Srivastava and Memon [282] presented 52 datasets for the open-domain question-answering tasks, and the study by Yang et al. [371] reviewed datasets for causal reasoning benchmarks. Röttger et al. [261] presented 102 datasets for safety evaluation.

While large datasets can be beneficial for LLM performance, one needs to be careful when obtaining data from public sources. Challenges faced when using web-mined corpora for pre-training LLMs have been reviewed by Perelkiewicz and Poswiata [236]. Among others, they presented challenges based on sensitive information, bias or the low quality of data. Du et al. [58] gathered 32 datasets (16 for pre-training and 16 for fine-tuning) while focusing on their quality and quantity.

Data selection: Datasets for training LLMs contain an enormous amount of samples with varying characteristics. While all data samples can be used for training, one can also select the ones most suitable for one's goal. For this purpose, the amount of training data can be reduced via deduplication, sampling or selection [9, 55, 305, 356]. Albalak et al. [4] surveyed data selection for LLMs. Methods are organized based on the type of data (e.g., data selection for pre-training, in-context learning). Additionally, they provided an overview of the main objectives of data selection at each stage of the training process (e.g., the main objective of data selection for fine-tuning is bias reduction and model performance). Wang et al. [312] specialized in selecting data for instruction tuning.

Wang et al. [328] considered data collection from a data management perspective. This includes concerns regarding data quality and quantity (e.g., filtering strategies) for pre-training and fine-tuning datasets. Lastly, [370] surveyed the impact of adding source code to the training data of LLMs, and found that it can improve downstream performance.

Data annotation and generation: While datasets for training LLMs are froften obtained from human-created sources, LLMs themselves can be used to augment or enhance existing data, be it by generating new data from scratch (data generation) or providing additional information to existing data (data annotation). For instance, Long et al. [199] surveyed synthetic data generation with LLMs to outline the workflow for data generation, consisting of generation, curation, and evaluation of synthetic data. Tan et al. [291] considered different facets of the data annotation process with LLMs (generation, assessment and utilization).

Data contamination: Data contamination is a problem that arises when the training data of LLMs overlap with the evaluation benchmarks. We found four surveys summarizing approaches for detection and mitigation of data contamination. Palavalli et al. [230] considered two severities of data contamination (instance level, dataset level) and examined them in two case studies (i.e., summarization, question answering). Xu et al. [350] considered the severity of data contamination (i.e., semantic, information, data, label level) and presented several tasks where contamination has been observed (e.g., code generation, sentiment analysis). Deng et al. [51] considered language model types (white-box, gray-box, black-box LLMs) when it comes to data contamination as well as several methods for detecting data contamination. Lastly, Ravaut et al. [256] organized contamination detection approaches based on open-data (dataset is known) and closed-data (dataset is not known).

4.4 Training and Learning

By leveraging large amounts of data, LLMs learn patterns that shape their performance across different stages. From initial pre-training to continuous adaptation, these stages allow them to acquire and refine their capabilities.

Pre-train: Pre-training describes the initial training stage of LLMs, in which models learn a general understanding of texts and language. Kotei and Thirunavukarasu [149] surveyed different pre-training techniques (from scratch, incessant pretraining, based on knowledge inheritance, multi-task pre-training). Afterwards, they discussed how this knowledge can be transferred to downstream tasks via fine-tuning. Fang et al. [68] reviewed metrics to consider for the training process and monitoring of the training success. While we found no other dedicated studies, several pre-training techniques have been covered by comprehensive surveys. For instance, the most frequently considered method for improving the efficiency of the pre-training process is mixed-precision training [9, 55, 61, 305, 356].

Fine-Tune: After pre-training, LLMs can be fine-tuned for specific tasks, which usually involves smaller datasets of higher quality. Weng [333] considered several fine-tuning paradigms, such as multi-task learning, knowledge distillation, transfer learning, and few-shot learning. Other surveys considered specific learning paradigms, such as federated learning [251], multi-task learning [265], or instruction-tuning [401]. A larger subset of surveys addressed the efficiency of the fine-tuning process via Parameter Efficient Fine-Tuning (PEFT) [9, 55, 305, 355, 356].

Xu et al. [355] covered the efficiency of the training of LLMs by PEFT methods. Rather than tuning the entire model (all parameters), a limited subset is fine-tuned to save time and memory. They categorized PEFT methods into 5 types: additive fine-tuning, partial fine-tuning, reparameterized fine-tuning, hybrid fine-tuning, and unified fine-tuning. In addition to the collection and description of a multitude of PEFT methods, Xu et al. carried out an empirical comparison of fine-tuning a RoBERTa model and 11 PEFT methods. Another PEFT method that received a survey of its own is LoRA (Low-Rank Adaptation) [213].

Incremental learning: To make sure that LLMs keep up with an evolving knowledge base, it is often not enough to train them once, but update them over time. Jovanovic et al. [137] considered different strategies for an incremental learning of LLMs. These include continual learning (CL), meta-learning, parameter-efficient learning, and mixture-of-experts learning.

Shi et al. [272] conducted a comprehensive survey on CL. Here, approaches are divided in two categories: vertical and horizontal continuity. Vertical continuity addresses approaches that specialize capabilities from a general set of knowledge. Horizontal continuity describes approaches that adapt capabilities across time and domains. In addition to outlining CL approaches, they included background information on CL, training objectives, as well as an overview of benchmarks.

Wu et al. [341] showed that CL can be used to update several dimensions: facts, domains, language, tasks, skills, values, preferences. Yang et al. [374] took pre-trained, fine-tuned, and vision-language models in account and CL methods are split into offline and online methods. In addition to internal methods for CL, such as the updating of parameters, Zheng et al. [419] included external approaches in their survey. External knowledge can either be incorporated by retrieving information from websites (e.g., Wikipedia), or the use of tools to allow LLMs to carry out additional tasks.

Unlearning: Learning can help LLMs attain valuable capabilities but not all the information might be useful to learn. Among others, LLMs might learn biases or access private information of individuals in the training data, which should not be replicated. Unlearning approaches are proposed to help LLMs forget about undesired information. The survey by Blanco-Justicia et al. [16] presented different types of unlearning approaches with regard to global weight modification, local weight, architecture modification, and input or output modification. They also showed datasets, models, and metrics used for evaluation. Xu [360] considered unlearning traditional ML models and LLMs, while Qu [251] surveyed unlearning approaches for federated learning.

4.5 Alignment

Via pre-training and fine-tuning, LLMs are capable of learning from data and generating sensible responses for a variety of tasks. However, such responses can be factually incorrect or harmful due to undesired biases in the training data [84, 326]. To combat this, alignment approaches are proposed not only to to align LLM responses with human values but also restrict their misuse in sensitive or potentially harmful contexts.

Wang et al. [326] focused on alignment techniques, such as reinforcement learning from human feedback. They surveyed different stages of the reinforcement learning process and included equations to explain the respective techniques. Shen et al. [271] divided alignment approaches into "outer" and "inner" approaches. Outer alignment describes the alignment of LLMs to human values. Inner alignment concerns the optimization of the objectives humans desire. Cao et al. [22] reviewed automated alignment methods, which do not require human annotations. Four method types are presented: aligning by defining constraints, imitating behavior of an aligned model, receiving feedback by existing models, and through feedback from interacting with environments.

Other than the alignment approaches, surveys considered their evaluation [197] and the collection of high-quality datasets [325]. Moreover, alignment was studied for agents [98] and the 17 United Nations' Sustainable Development Goals [340]. Other than the alignment for a general audience, the alignment of LLMs to individuals has been surveyed from the perspective of personalization [147] and preference learning [128].

In addition to aligning LLMs to human values, their adaptation of roles and pre-defined personas has been studied. Particularly, Tseng et al. [296] surveyed LLMs for role-playing and personalization. Here, LLM role-playing was performed for tasks such as software development, games, and the medical domain. The survey by Chen et al. [32] outlined how to design components for role-playing, which includes datasets and different alignment methods. Simmons and Hare [278] gave an overview of LLMs to replicate human behavior, such that public opinions can be measured.

4.6 Prompting

LLMs are able to respond to user queries, therefore, the design of such queries or prompts can be integral for achieving good results. Chen et al. [29] created a review on prompt engineering, which could be seen as a good starting point for someone trying to create prompts themselves, as they provided basics on prompting models such as GPT-4. Luo et al. [205] considered prompting from the perspective of in-context learning, which can achieve performance improvement by providing LLMs with examples. Important aspects include the number of demonstrations as well as their diversity and order [427]. The surveys by Vatsal and Dubey [302] and Sahoo et al. [264] set out to provide an overview of diverse sets of prompts used in the literature. In total, Vatsal and Dubey [302] presented 39 different prompting methods, Sahoo et al. [264] showed 29.

Focusing only on a subset of approaches, Xia et al. [346] created a survey on chain-of-thought prompts and its variations (chain-of-X). Approaches applied in literature can be divided into four groups: chain-of-intermediate, chain-of-augmentation, chain-of-feedback, chain-of-models. Other than chain-of-thought approaches, five surveys focused on Retrieval-augmented generation (RAG) prompt approaches [66, 80, 112, 120, 136]. Huang and Huang [120] created a taxonomy of RAG approaches based on four stages. First, pre-retrieval approaches deal with the data, followed by the retrieval stage which ranks the relevant pieces of information. Subsequently, a post-retrieval stage can re-rank and filter the retrieved information. In the end, the generation stage creates the prompt. Jing et al. [136] carried out a survey with a focus on the pre-retrieval stage, in particular the organization of data in vector databases to support the subsequent retrieval stages. The survey by Gao et al. [80] considered the later three stages of the RAG process (retrieval, generation and augmentation). Their survey included RAG approaches at different stages of LLMs (pre-training, fine-tuning, inference) and progression of RAG from naive to complex approaches. Moreover, they extracted useful information from the approaches, such as the data sources considered, data type and retrieval granularity (e.g., sentences). Hu and Lu [112] pointed out approaches which combined multiple RAG techniques (i.e., hybrid approaches). Additionally, surveys also cover prompting for specific domains, such as reasoning [243], software engineering tasks [17], visual-language models [90], and goal-oriented tasks [166].

While prompt engineering can lead to performance improvements, it also requires additional efforts, either from the developers to design the prompts or from additional computations carried out by the LLMs. For this purpose, Chang et al. [25] surveyed the efficiency of prompt engineering from two perspectives: prompting with efficient computation and prompting with efficient design. Lastly, Liu et al. [193] surveyed prompting frameworks, or prompt-based tools, which can simplify the prompt engineering process for users. Their overview includes and compares 28 such frameworks.

4.7 Agents

An LLM-based agent employs an LLM as the primary component of the controller module in an attempt to achieve goals autonomously [343]. Several general surveys aim to establish a common taxonomy for this field. Some researchers converge on a core set of components such as reasoning, memory, and action modules for tool use and interaction with the environment [13, 414]. Others, such as Wang et al. [313] proposed four modules: profiling, memory, planning, and action, while Xi et al. [343] conceptualized agents through a simplified brain-perception-action framework. From a different perspective, Li [171] proposed to integrate the previously separated paradigms of tool-using, planning, and feedback learning, into a task-agnostic taxonomy through three universal roles and four workflows.

Building upon these general frameworks, some researchers survey specific modules or capabilities. Huang et al. [118] surveyed and proposed a taxonomy on the planning module classifying them into five directions. Similarly, Zhang et al. [409] reviewed memory mechanisms and how they enable self-improvement and long-term interactions. The capabilities further extend when these systems encompass multiple agents. Guo et al. [92] reviewed LLM-based multi-agent systems, highlighting communication and planning strategies in different environments, while Händler [98] proposed a taxonomy for classifying these systems based on alignment, autonomy, and architectural design.

The applications of LLM-based agents are rapidly expanding through different tasks and fields. Hu et al. [111] proposed a general taxonomy for agents into six modules and further reviewed their applications in the gaming field. Cao et al. [23] explored how LLM-based agents can be used to enhance reinforcement and classified them according to their function in the framework. Gao et al. [77] focused on their application on simulation and modeling of scenarios in four domains: cyber, physical, social, and hybrid. Social simulations are also reviewed by Ma et al. [207], who focused on computational experiments to generate more realistic social behaviours. Ramos et al. [255] explored applications in chemistry, discussing automation of tasks such as literature review to controlling robotic labs. Li et al. [177] reviewed the surge of personal LLM-based agents and discussed the security threats when handling personal data. Similarly, the risks and threats introduced by the integration of agents in multiple fields are further explored by He et al. [101].

4.8 Inference

The inference process has in general been covered by surveys under the lens of efficiency. Addressed techniques can be divided into two categories: model compression and dynamic acceleration. They are summarized in Table 3.

Model compression aims for efficiency improvements by lowering the number of parameters and thereby achieving quicker responses. Park et al. [232] surveyed compression methods, which include four common methods (pruning, quantization, knowledge distillation, low-rank approximation), and they are one of two studies which considered parameter sharing. They distinguished compression methods based on their cost (i.e., low-cost and high-cost) depending on how much time and memory their use requires. For example, high-cost quantization methods might require a full retraining of an LLM while low-cost quantization can be performed without any retraining or fine-tuning. Lastly, they provided additional evidence on the usability of pruning and quantization methods by evaluating them and summarizing their cost, compression rate and performance. Chavan et al. [28] included empirical results on several quantization types and their impact on memory consumption and number of tokens that get generated per second.

Table 3. Overview of inference techniques and surveys which describe them.

Model Compression	Pruning [9, 28, 55, 134, 232, 260, 305, 320, 351, 356, 386, 428, 430]
	Quantization [9, 28, 134, 232, 260, 305, 320, 351, 356, 386, 428, 430]
	Knowledge distillation [9, 28, 55, 134, 232, 305, 320, 351, 356, 365, 386, 428, 430]
	Low rank approximation [9, 28, 232, 305, 351, 356, 386, 428, 430]
	Parameter sharing [232, 351]
Dynamic acceleration	Early exit [9, 145, 351, 386, 428]
	Input pruning, filtering and compression [9, 356, 428]
	Token parallelism [9, 145]
	Token skipping [351]
	Speculative decoding [145, 305, 344, 386, 391, 428]
	Output organization [428]
	MoE [320, 386]

Commonly, surveys considered at least three of the five compression methods, while we encountered two surveys that covered a single method. In both cases, this was Knowledge Distillation (KD) [359, 365]. Xu et al. [359] provided extensive information on the KD process to help understanding. Moreover, they described how KD can be used to distill skills (e.g., instruction following, alignment) and domain knowledge (e.g, law, science). Yang et al. [365] categorised approaches as white-box and black-box, and also presented KD approaches for multi-modal LLMs. Additionally, they included experiments with KD methods on different LLMs and compared their performance.

Dynamic acceleration refers to techniques that adapt computations based on the input the LLMs receive [9]. While it is common for surveys to not be focused on any of the dynamic acceleration techniques in detail, we found two surveys on speculative decoding, a method that predicts several future tokens alongside the current token at each decoding step [344, 391].

4.9 Evaluation

Evaluating the performance of LLMs is a critical step in assessing their quality. Chang et al. [27] investigated existing evaluation approaches based on three perspectives: what to evaluate (e.g., tasks), where to evaluate (e.g., datasets, benchmarks), how to evaluate (e.g., evaluation process). Here, tasks include NLP, social science, natural science, engineering, medical applications, robustness, and others. Benchmarks are divided into three categories: general, specific, and multi-modal. Lastly, the evaluation process can follow automatic or human evaluation criteria.

Guo et al. [94] defined three categories of LLM evaluation: knowledge and capability (e.g., reasoning), alignment evaluation (e.g., bias), and safety (e.g., robustness). They reviewed each category and provided benchmarks and evaluation methodologies. Moreover, they present LLMs specialized for various domains (e.g., biology, medicine, finance, education, computer science).

Zhuang et al. [432] structured the evaluation of LLMs based on four core competencies (i.e., knowledge, reasoning, reliability, and safety) and provided their definitions, benchmarks, and metrics. In addition to core competencies, it is important to consider sentiment, planning, and code. Beyond the evaluation of competencies, Peng et al. [235] considered the evaluation of agentic approaches (e.g., reasoning, domain knowledge).

A different focus was set by Laskar et al. [159], who put emphasis on challenges and limitations of the evaluation caused by the diversity of setups followed by existing works. In particular, they considered three types of challenges: reproducibility (e.g., missing details on data and model), reliability (e.g., prompt hacking), and robustness (e.g., lacking evaluation of generalizability). Additionally, Laskar et al. [159] gave guidelines on how these limitations can be addressed.

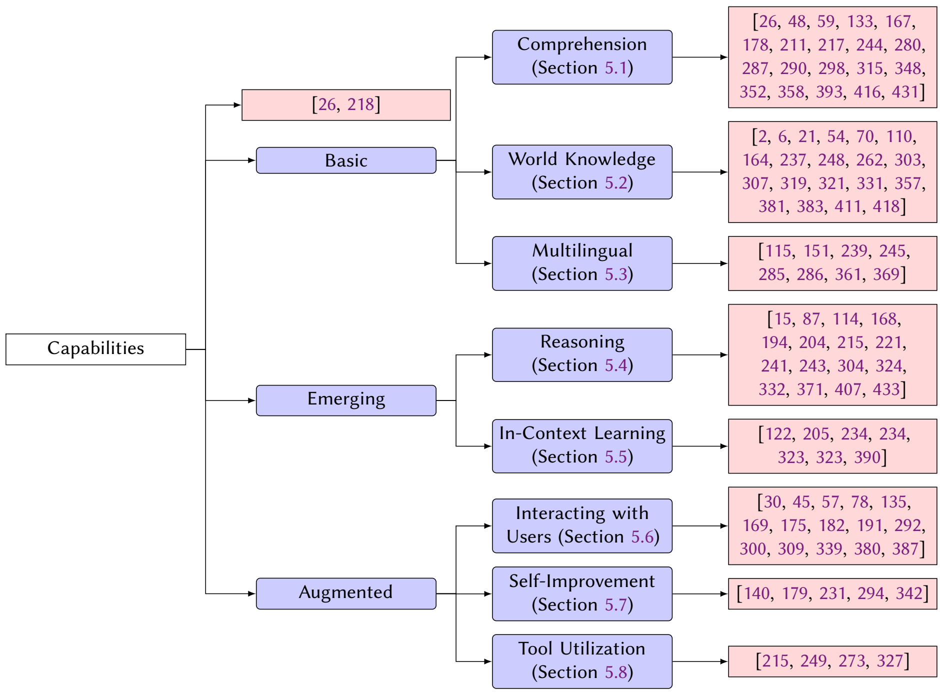

Fig. 5. Taxonomy of surveys on capabilities.

5 Capabilities

The proficiency of LLMs in diverse tasks and fields can be understood as a result of multiple abilities, or capabilities. These capabilities range from foundational language skills to more complex behaviors. While Chang and Bergen [26] surveyed capabilities of language models, this section adopts a taxonomy inspired by the work of Minaee et al. [218], as it further covers how models are integrated into larger ecosystems. This framework categorizes these abilities into three groups: Basic, Emerging, and Augmented.

Basic capabilities represent foundational skills acquired directly from their pre-training. Emerging capabilities are not specifically trained for, but appear as the scale of the model increases. Augmented capabilities regard how the model integrates with external systems, usually to overcome its own limitations and improve the outcomes.

5.1 Comprehension

The comprehension capability of LLMs refers to their ability to understand and interpret the meaning, context, and possibly the nuances of input. It is crucial for their application to various NLP tasks [26, 217, 244], such as information retrieval [416, 431], relation [358], keyphrase extraction [280], semantic parsing [211], question-answering [48], and text generation [167, 178, 393].

Several surveys focus on more specific types of LLM comprehension. For instance, the models' ability to understand the content is essential to produce accurate summaries. General text summarization surveys [133, 290] indicate that abstractive summarization requires the model to interpret and rephrase information, which allows for testing the depth of the comprehension. Xie et al. [348] further corroborated this in the domain of biomedical texts. Other authors use different tasks to test this comprehension. Umair et al. [298] surveyed this capacity with keyphrase prediction, Xu et al. [352] focused on information extraction, and Sun et al. [287] used sentiment analysis.

The underlying mechanisms enabling LLMs to comprehend texts based on previous knowledge are also an ongoing research field. Wang et al. [315] discussed how true comprehension allows for high-level reasoning and problem-solving skills to develop. They further pointed out the fragility of comprehension in LLMs, agreeing with the main challenge presented by Du et al. [59].

However, the depth of this comprehension remains ongoing research. Du et al. [59] indicated that LLMs can perform "shortcut learning", where they overfit to dataset biases rather than achieve proper understanding of the text, leading to poor performance in out-of-distribution scenarios.

5.2 World Knowledge

We refer to world knowledge capability in LLMs as their ability to store and accurately recall factual information acquired during their pre-training phase. This capability is fundamental to the use of LLMs, but also involves challenges regarding its representation, robustness, and maintenance.

The current understanding indicates that models learn universal language representations that capture knowledge about the world during their pre-training phase and are stored implicitly in the model's parameters [248, 381, 411] and can be used as knowledge bases [6]. Cao et al. [21] conceptualized the knowledge in an LLM through a life cycle consisting of five phases: acquisition, representation, probing, editing, and application. On the other hand, Safavi and Koutra [262] conceptualized this knowledge through different levels of Knowledge Base (KB) supervision.

The accuracy of the knowledge assimilated by the LLM is still an ongoing concern. Adilazuarda et al. [2] noted that knowledge assimilated by LLMs often exhibits biases. Wang et al. [307] focused on a different issue, factuality, i.e., generating content contradicting facts. Zhang et al. [411] argued that since facts learned during pre-training can become outdated, updating this knowledge also poses a problem. Similarly, the meaning of words can change over time, as shown in the survey by Periti and Montanelli [237]. Updating the models with contextual information is a possible solution; however, Xu et al. [357] pointed out that models also struggle with discrepancies between learned facts and contextual information.

Researchers are exploring different techniques to address these limitations by integrating the models with desired knowledge [70, 110], which can come from different sources and formats [418]. Wang et al. [319] reviewed knowledge editing techniques, aiming to update knowledge without complete retraining. On the other hand, Yin et al. [381] and Wei et al. [331] advocated for augmenting LLMs with external sources of knowledge. Wang et al. [321] further explored this synergy with a survey of LLMs enhanced with knowledge representation learning.

In addition to teaching facts to LLMs, probing can be used to extract facts [383], and LLMs have been studied for fact-checking [54, 303]. To support the ability of fact-checking, the generated texts can be attributed with respective sources [164].

5.3 Multilingual

The multilingual capabilities of LLMs enable them to understand and generate text in multiple languages. This ability facilitates knowledge transfer from high-resource to low-resource languages [115, 151, 361]. Generally, multilingualism in a model is achieved through multilingual corpora during pre-training [239, 361]. Xu et al. [361] surveyed multilingual LLMs and gathered valuable resources for corpora (training and downstream tasks) as well as LLMs. For the datasets, they showed the languages in the corpora, the sizes, and sources. For LLMs, they extracted information such as the base model, number of parameters and pre-training information (dataset and loss function). Moreover, they carried out a performance comparison of several models over three tasks (bilingual lexicon induction, cross-lingual classification, machine translation) and ten languages.

Qin et al. [245] focused on data resources and alignment techniques needed to achieve this capability, distinguishing between parameter-tuning and parameter-frozen approaches. The survey by Huang et al. [115] extended beyond the contents of Qin et al. [245] and Xu et al. [361]. While Qin et al. [245] provided a comprehensive taxonomy for alignment techniques, Huang et al. [115] devised a taxonomy spanning training, inference, security, domains (medical, legal), and datasets. Similar to Xu et al. [361], Huang et al. [115] provided comprehensive overviews for datasets (training and benchmarking) and available models. In contrast to Xu et al. [361], only models with 7B and more parameters have been listed. Lastly, Huang et al. [115] provided insights on security via attack (e.g, jailbreaking) and defense methods.

In contrast to the broad scope of multiple languages, other surveys have focused on a single language. For high-resource languages such as Chinese, Sun et al. [286] provided an extensive review of LLMs in different downstream applications such as sentiment analysis. Similarly, Subies et al. [285] focused on surveying clinical LLMs for Spanish. On the other hand, surveys on lower resource languages tend to focus on establishing a benchmark for foundational capabilities. Yang [369] surveyed and benchmarked the performance of available LLMs in Korean. Addressing the challenges of languages with even fewer resources, Kryeziu and Shehu [151] reviewed techniques to build LLMs for low-resource languages, using Albanian as a key example.

While the previous surveys have focused on how this capability is acquired, Philippy et al. [239] underscored the lack of explanation for this capability and performed a review on possible factors. In total, they considered five factors: linguistic similarity, lexical overlap, model architecture, pre-training settings, and pre-training data.

5.4 Reasoning

We refer to reasoning as the capability to perform deductions, make decisions, and conduct multi-step problem solving. This capability is crucial for LLMs to move from simple pattern matching and produce more refined inferences able to tackle complex tasks.

This reasoning has significantly increased through prompting techniques such as Chain-of-Thought (CoT) [215, 241, 243]. Li et al. [168] discussed general fundamental capabilities of LLMs, further decomposing the reasoning step. Huang and Chang [114] surveyed a general method to elicit and evaluate reasoning abilities, while Qiao et al. [243] and Plaat et al. [241] proposed taxonomies for prompt-based reasoning.

Similarly to previously discussed capabilities, evaluating reasoning still poses a challenge [221]. Some surveys have researched specific branches of this capability; Luo et al. [204] focused on logical reasoning, Wang et al. [324] researched multimodal reasoning, and Zhang et al. [407] studied strategic reasoning. Some authors have evaluated reasoning through tasks: Giadikiaroglou et al. [87] explored evaluations through puzzles, Zong and Lin [433] used categorical syllogisms, Wen et al. [332] studied answer refusals, and Bhargava and Ng [15] involved commonsense knowledge. However, Yang et al. [371] argued that these benchmarks may be solved through knowledge retrieval. Lastly, Liu et al. [194] surveyed the applicability of LLMs and their reasoning capability to aid with causal inference, or causal discovery [304].

5.5 In-context Learning

In-Context Learning (ICL) refers to the ability to adapt to new tasks by processing examples in the LLM input. This allows for quick, few-shot adaptation to new tasks without modifying the model's parameters. Luo et al. [205] provided a general survey focusing on retrieval-based ICL.

The limited context window of LLMs poses a challenge for ICL, by limiting the number of examples that can be used in the prompt [234, 323]. Huang et al. [122] surveyed progress in LLMs aimed at increasing their context. Similarly, Pawar et al. [234] and Wang et al. [323] surveyed techniques for extending context length. However, Zeng et al. [390] underscored that context length is only one of the three conflicting goals, the other two being accuracy and performance.

5.6 Interacting with Users

LLMs are increasingly deployed in applications that require direct user interaction. This interaction requires abilities to engage in dynamic conversation, understand the intentions from the user, and generate helpful responses [309, 380]. Zaib et al. [387] performed a general survey on dialogue systems with LLMs, exploring how they can be leveraged for conversational agents. Similarly, Dam et al. [45] analyzed LLM-based chatbots and their impact on diverse fields. Gao et al. [78] devised four stages for the interaction between humans and LLMs. They consist of: planning, facilitating, iterating, and testing. Focusing on the progression of these systems, Wang et al. [309] provided a deeper assessment on the evolution and trends of LLM-based dialogue systems. One key aspect in this evolution, multi-turn dialogues, was surveyed by Yi et al. [380]. However, Dong et al. [57] stated that as these models become more integrated into user-facing applications, ensuring their safety and robustness becomes essential.

Beyond the focus on dialogue, LLMs are used to understand user needs and characteristics. This area of research is known as user modeling. Tan and Jiang [292] described how LLMs are used to model and understand user-generated content. Jin et al. [135] surveyed how LLMs can infer the user background based on cues in the prompts and tailor their responses accordingly. Another avenue of interaction between LLMs and users is recommendation systems. In this context, the survey by Li et al. [175] provided background details on ML and DL-based recommendation, comparing them to LLM-based approaches. Lin et al. [182] and Vats et al. [300] presented how and what parts of a recommendation system can be supported by LLMs. Wu et al. [339] categorized LLM approaches for recommendation in two paradigms: discriminative (generating embeddings for users and items) and generative. Other than computing scores for items, generative recommendation directly generates recommendations. This can be achieved by representing user and item IDs via tokens. Generative recommendation was examined in more detail by two more surveys [169, 175].

Other surveys have focused on practical aspects of building these LLM-based recommender systems. For instance, Liu et al. [191] examined the training strategies for LLMs in recommendation tasks, describing learning objectives, data types, and datasets used in each publication. Lastly, Chen [30] surveyed how to generate explanations for recommendations and accompanying challenges.

5.7 Self-Improvement

Self-Improvement encompasses the capability of LLMs to learn from feedback to autonomously enhance their results. Two surveys offer a broad view: Pan et al. [231] classified self-correction strategies according to when the correction occurs (training, generation, post-hoc). Tao et al. [294] used the concept of "self-evolution" and broke it down into a four-phase iterative cycle: experience acquisition, experience refinement, updating, and evaluation.

Other surveys go into a deeper analysis of the self-improvement process. Kamoi et al. [140] claimed self-correction results are being overstated due to unfair evaluation. They concluded that reliable feedback is often the bottleneck, indicating that self-correction without external tools generally fails except for suitable tasks. The unreliability of self-feedback is further discussed by Liang et al. [179], who connected the success of self-improvement to internal consistency. They concluded that since LLMs are trained on mostly correct data, improving the consistency of their outputs tends to increase the probability of a correct output more than an incorrect one. Lastly, Wu et al. [342] considered how evolutionary algorithms can be used to enhance LLMs by supporting them with search capabilities, while LLMs can be used to enhance evolutionary algorithms by guiding the search process with domain knowledge.

5.8 Tool Utilization

LLMs are capable of using tools with the goal of interacting and leveraging external programs, such as external software and APIs, to overcome limitations and perform additional functionalities. This capability allows the LLMs to solve more complex problems and interact with the environment.

Wang et al. [327] performed a general survey and proposed a taxonomy for tools based on their functionality. Qu et al. [249] surveyed tool utilization and proposed a four-stage workflow for tool learning. Other authors survey specific applications in this field. For example, Shi et al. [273] examined the use of tools after the content is generated by focusing on Text-to-SQL tasks, and Mialon et al. [215] studied the use of other models, search engines, and the web as tools.

6 Applications

LLMs have shown promise in various applications and industries [299], ranging from critical fields (e.g., finance, health, law) [35], to niche topics such as fitness or climate modeling [142]. This section outlines the main application domains in which LLMs have been used, and their respective surveys.

6.1 Medical and Health

Medical and health applications are the most popular domain for LLM surveys we encountered, with a total of 32 surveys carried out up to September'24.

Xiao et al. [347] and Zhou et al. [422] created surveys containing information about the training, data and applications for LLMs in the medical domain, as well as challenges and areas for future research. Both surveys provided helpful overviews of datasets and models, with information such as the base model and data source, where Xiao et al. [347] also took multimodal LLMs into account. In total, Xiao et al. [347] considered six applications: medical diagnosis, clinical report generation, medical education, mental health services, medical language translation, and surgical assistance. The set of applications studied by Zhou et al. [422] shows some overlap; however, the fields of medical robotics, clinical coding, medical inquiry, and response are novel.

Similarly, Wang et al. [306] considered vision and standard LLMs for pre-trained models and fine-tuning for downstream tasks. Luo et al. [206] focused their survey on pre-trained LLMs for NLP tasks. Their overview included English and Chinese LLMs used for various tasks, such as question-answering, machine translation, sentiment analysis, and named entity recognition. For each task, they provided details on datasets and metrics used.

He et al. [102] transitioned from PLMs to LLMs. This included details on training and datasets. Similarly, Wang et al. [311] covered the data acquisition process and different training paradigms to adapt general LLMs for the medical domain. Their survey also included concerns about fairness, accountability, transparency and ethics. Park et al. [233] considered ethical implications in their review, as well as legal and socioeconomic concerns. In addition to a comprehensive overview, Liu et al. [190] put emphasis on trustworthiness and safety of LLMs, which includes a discussion of their fairness, accountability, privacy, and robustness. Several other surveys considered privacy and ethical concerns in the medical domain [96, 224, 252, 420].

Huang et al. [121] focused on the evaluation of medical LLMs. This included evaluation approaches and metrics for different applications: departments and specific diseases, medical research, medical education and public awareness, and medical text processing. LLMs in the medical domain have been evaluated by three different evaluators: human experts, automated metrics, and AI-driven assessments. Automated metrics can be categorized in four groups: correctness, completeness, usability, and consistency. AI-driven assessments are in the minority. Chen et al.[33] also considered the evaluation of LLMs for medical tasks such as image processing and information extraction.

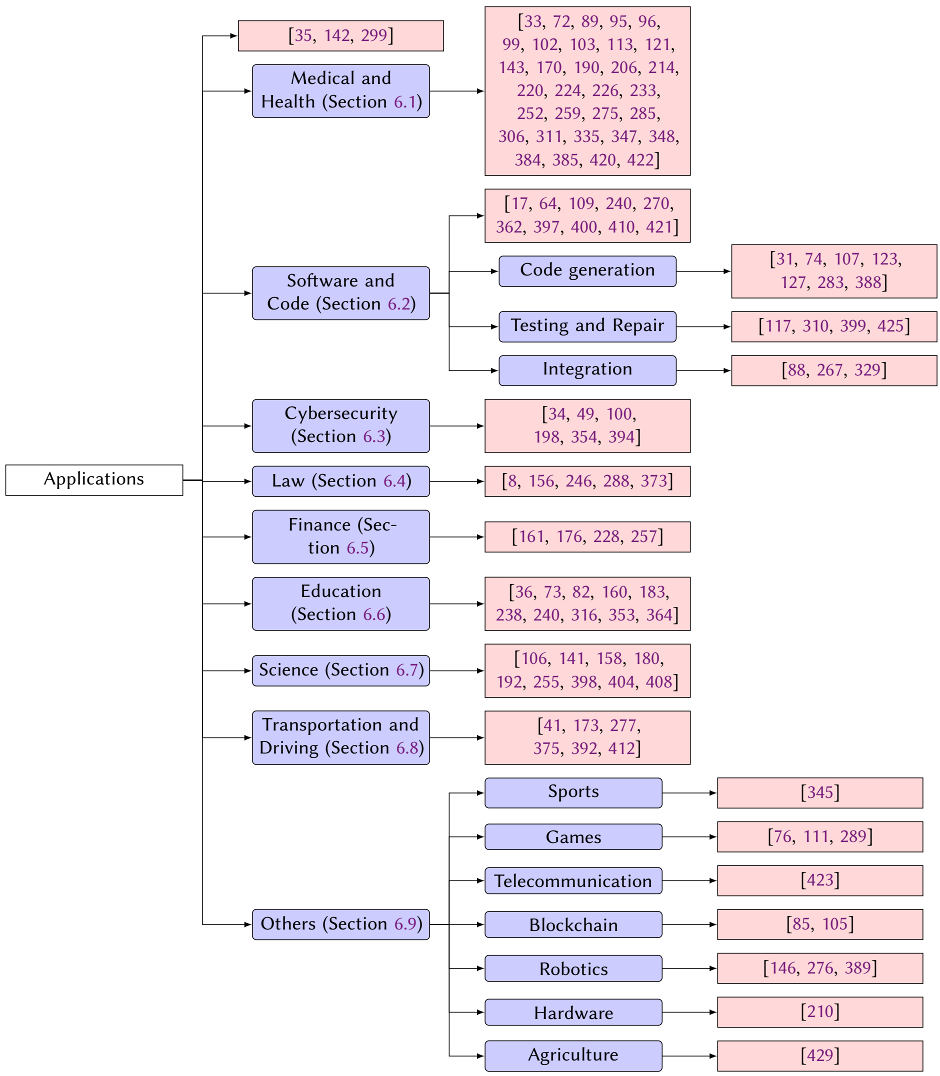

Fig. 6. Taxonomy of surveys on applications.

While a lot of the surveys described tasks based on texts, the field of medicine is multimodal and several types of data have been surveyed [99, 306, 347, 385]. Ferrara [72] studied data collected by wearable sensors and the survey by Nerella et al. [226] covered data types such as NLP, medical imaging, structured Electronic Health Records (EHR), social media, biophysiological signals, and biomolecular sequences. Particularly, electronic health records have been of interest for surveys [170, 335, 348]. Li et al. [170] surveyed LLMs working with Electronic Health Records, in particular with regards to seven tasks: named entity recognition, information extraction, text summarization, text similarity, text classification, dialogue system, diagnosis, and prediction. Xie et al. [348] only considered the task of text summarization, which has been applied for EHR and biomedical literature, medical conversation, and questions.

Another set of surveys included bibliometric analysis. For instance, Restrepo et al. [259] analyzed metadata such as author affiliations, countries, and funding source to assess diversity. Yu et al. [384]

considered information such as collaboration networks. The remaining surveys covered areas ranging from only considering Spanish language models [285] to LLMs in medical examinations[220], psychology [103, 143], mental health [89, 95, 113, 214], and critical care medicine [275].

6.2 Software and Code

In the software engineering domain, we found several surveys which provided comprehensive overviews. The earliest survey is by Xu and Zhu [362], from 2022. They surveyed datasets, tasks, and architectures for pre-trained LLMs as well as their training procedures.

Subsequent surveys increased in comprehensiveness, with the survey by Ziyin Zhang et al. [410] covering more than 900 works. They created both a taxonomy for code LLMs as well as a taxonomy for more than 40 tasks according to the software development stages. The survey by Quanjun Zhang et al. [400], which also entails more than 900 references, provided another comprehensive overview. Interesting aspects they considered included an overview of pre-training tasks as well as the integration of LLMs for SE activities (e.g., their security or size).

Zheng et al. [421] gave information about organizations which developed the LLMs (e.g., Company-led, University-led, Research teams & Open-source community-led). Also, their survey put emphasis on the performance of LLMs. One research question was aimed at finding whether code LLMs perform better than general LLMs for SE tasks. Moreover, they presented the performance reported in collected works for several tasks, to find which LLM is most suitable. Hou et al. [109] provided valuable insights on the datasets used for SE tasks, including data collection, selection, and processing steps. She et al. [270] surveyed pitfalls which could hinder the performance of LLMs in practice. These are divided into four categories: data collection and labeling, system design and learning, performance evaluation, deployment, and maintenance. For each of these pitfalls, implications and solutions are outlined. Similarly, Fan et al. [64] listed open problems for each stage of the software development lifecycle.

Other surveys investigated how LLMs have been prompted for various SE tasks [17], how LLMs can be used in an educational setting to help with code related tasks (e.g., explaining error messages) [240], or support failure management for Artificial Intelligence for IT Operations [397].

Code generation: Jiang et al. [127] created a comprehensive survey on the generation of code from natural language descriptions. Collected works are structured given a taxonomy in: data curation, recent advances (e.g., training and prompting), evaluation, and application (e.g., GitHub Copilot). They also provided an overview of existing LLMs and a performance comparison of several LLMs on two popular benchmarking datasets: HumanEval and MBPP. Zan et al. [388] also provided a comparison of LLMs on the HumanEval benchmark, where they included a larger quantity of small LLMs (smaller than 1 billion parameters). Additionally, they presented 17 benchmarks with statistics, such as the number of tests available. In contrast, Hong et al. [107] surveyed approaches for generating SQL queries from natural language.

Husein et al. [123] surveyed the completion of code rather than generating code from natural language descriptions. They considered different granularities (token, line, API calls, Block level) and performance metrics for evaluation. Other than generating code itself, LLMs have been used to generate programming exercises [74], infrastructure configurations [283], and support HPC [31].

Testing and Repair: The survey by Wang et al. [310] discussed the field of software testing and different associated tasks. The most commonly addressed tasks include program repair as well as the generation of tests (e.g., unit tests, system tests). For these, Wang et al. extracted the most common prompts (e.g., zero-shot) and the LLMs used for these tasks.

The survey by Zhang et al. [399] focused on APR and found 127 APR papers covering 18 bug types that used LLMs. Zhou et al. [425] considered both vulnerability detection and repair. They

investigated how LLMs have been adapted to these tasks and found that the majority of approaches perform fine-tuning. Huang et al. [117] covered the use of LLMs for fuzzing as a testing activity.
Integration: While previously outlined surveys covered the use of LLMs for software engineering activities, they can also be treated as components of software itself. In this regard, Weber [329] created a taxonomy for LLM-integrated software systems, and Sergeyuk et al. [267] studied the use of LLMs in Integrated Development Environments (IDEs). Gorissen et al. [88] considered the use of LLMs in Low-Code Development Platforms.

6.3 Cybersecurity

Four studies created comprehensive overviews of the field of cybersecurity [34, 100, 354, 394]. Hassanin and Moustafa [100] covered diverse cyber defense strategies such as vulnerability assessment, intrusion detection, or anonymization, while others put emphasis on vulnerability assessments [49] and threat detection [34]. Zhang et al. [394] not only outlined defense activities but also indicated ways to use LLMs for attacks. Xu et al. [354] provided insights on how to construct LLMs for the security domain via means of fine-tuning, prompting, or augmentation with external tools. Lastly, Liu [198] gave an overview of available pre-trained models for cybersecurity, and Xu et al. [354] addressed the data collection process and available datasets.

6.4 Law

LLMs have been used to automate various legal tasks, but their adoption also raised challenges [288]. Anh et al. [8] researched the impact of LLMs on NLP, focusing on legal text processing. They explained how NLP addresses different challenges in the field, such as ambiguity and sentence complexity. They performed an empirical analysis that suggests that encoder-decoder models outperform encoder-only architectures, advocating for their use in legal NLP tasks.

Lai et al. [156] provided a general survey on the applications of LLMs within the judicial systems. They included the impact on common users as well as experts (e.g., judges and lawyers). The authors indicated limitations and issues of LLMs that can affect judicial practices. They gave practical recommendations for improving the use of LLMs in the legal system and highlighted the importance of understanding the societal impacts of these technologies. Similarly, Qin and Sun [246] covered the practical application of LLMs in the legal system, such as case retrieval and legal analysis. They indicated potential challenges such as biases, interpretability issues, and data privacy concerns. This study emphasized the need for fine-tuned models and presented an overview of datasets for their training in different languages. Lastly, Yang et al. [373] presented a systematic review of legal LLMs focusing on fine-tuning for question-answering tasks. They provided a practical view focusing on the implementation of these systems and the techniques that they could use (e.g., Low-Rank Adaptation). They used a bottom-up approach to examine how existing models can be adapted to the legal domain.

6.5 Finance

Nie et al. [228] provided a comprehensive survey on LLMs for finance. They first categorized existing works according to application areas in the financial domain, including, among others, time series forecasting, reasoning, and sentiment analysis. Further information on datasets, benchmarks, and challenges is presented. In addition to providing an overview of finance applications, Li et al. [176] developed a decision framework to help practitioners select an LLM based on their task. For this, they also provided a comparison with estimated costs of different LLM options (e.g., zero-shot, fine-tuning, training from scratch). Lee et al. [161] put emphasis on presenting benchmark tasks and datasets. Moreover, they showed a timeline of LLMs and financial LLMs. Ren et al. [257] addressed the use of LLMs in an e-commerce setting. In this context, LLMs have been used for tasks such as product recommendations, question answering and analysis of customer feedback.

6.6 Education

Wang et al. [316] created a comprehensive survey on how LLMs can assist teachers, students, and different tools that are available. Additionally, they provided an overview of datasets and benchmarks, as well as discussed risks and challenges of LLMs in education. Pester et al. [238] addressed the use of LLMs for immersive learning activities. The survey by Xu et al. [353] provided more background information on education, as well as how to integrate LLMs in the process, while García-Méndez et al. [82] considered LLMs used for different education activities. This focus on integration also extends to specific disciplines, with dedicated surveys exploring the use the integration of LLMs in subjects such as computer science [240] or engineering [73].

Yan et al. [364] covered a total of 53 educational tasks from nine categories (e.g., grading, content generation) and put emphasis on practical and ethical challenges. In a similar fashion, Chhina et al. [36] looked at both the challenges and benefits of LLMs in education. Lee et al. [160] focused their survey on different types of biases when using LLMs in an educational setting. Biases were investigated at different stages of the LLM lifecycle (e.g., data collection, training, and deployment). Lin et al. [183] listed available open-source LLMs for use in education activities.

6.7 Science

Ho et al. [106] provided an overview of scientific LLMs applied to text, and presented different tasks, datasets, and existing models. In addition to scientific LLMs for text, Zhang et al. [404] surveyed more than 260 LLMs, not only taking different scientific fields but also different modalities into account. Complementing this broad overviews, other surveys focus on LLM applications in specific fields such as chemistry [180, 255, 398], biology [398], and mathematics [192], as well as for specialized sub-domains like single-cell biology [158] and computational neuroscience [141].

A trait of scientific texts is the presence or use of citations, to give credit to relevant sources. Here, Zhang et al. [408] created a survey to show the relation between LLMs and citations. Their survey provided an overview of four different citation tasks LLMs can be applied to: citation classification, citation-based summarization, citation sentence generation, and citation recommendation. Additionally, they discussed how citations can be incorporated in the training of LLMs.

6.8 Transportation and Driving

In the realm of Intelligent Transportation Systems (ITS), LLMs have been used to advance transportation intelligence and traffic management. The surveys by Shoaib et al. [277] covered tasks such as traffic prediction and transportation management, while Zhang et al. [392] considered traffic management, transportation safety, and autonomous driving. Moreover, they provided a list of datasets for the ITS domain. Zhang et al. [412] focused on travel behavior prediction as a time series forecasting problem and provided an overview of LLM-based approaches.

Autonomous driving was covered by three dedicated surveys [41, 173, 375]. Cui et al. [41] addressed the use of LLMs for autonomous driving from a multimodal perspective (vision and language). They provided a holistic overview, considering the use of multimodal LLMs for autonomous driving, transportation, and maps. Furthermore, they presented datasets for autonomous driving and traffic scene understanding, and extracted information from existing approaches, such as the LLMs used. In contrast, Yang et al. [375] provided a more fine-grained view on tasks and metrics used for evaluation. They distinguish four categories, based on the respective tasks: planning, perception, question answering, and generation. Li et al. [173] covered the use of LLMs in autonomous driving either as part of the pipeline, to support existing systems, or as end-to-end systems.

6.9 Others

Sports: Xia et al. [345] investigated datasets and applications for LLMs in a sports setting. Here, LLMs can be applied to different input types (text, video, audio) and have addressed a diverse range of tasks (e.g., hate speech detection, fan engagement, game summarization).

Games: Sweetser [289] carried out a scoping review on 76 papers on LLMs for video games, to provide an overview and support future research. In the gameplaying context, LLMs have been used as parts of the game (e.g., agents, dialogue generation) or part of the development and analysis process (e.g., content generation, analysis of reviews).

Gallotta et al. [76] addressed the different roles of LLMs in games in their survey. In total, they identified nine roles that LLMs have taken: Player, NPC (non-player character), player assistant, commentator, analyst, game master, game mechanic, automated designer, and design assistant. Additionally, they presented a roadmap for future applications of LLMs for games, as well as limitations and ethical implications of their use.

Hu et al. [111] focused their survey on LLM-based game agents, for which they found 62 approaches. These were categorized based on game type (text, video) and genre (e.g., adventure, cooperation, simulation). Moreover, agents were discussed from six perspectives: perception, memory, thinking, role-playing, action, and learning.

Telecommunication: Zhou et al. [423] presented a comprehensive overview of LLMs in the field of telecommunications. In particular, LLM activities (generation, classification, optimization, prediction) were mapped to telecommunication applications. Such applications include network issue troubleshooting, network defect detection, and traffic load level prediction.

Blockchain: Geren et al. [85] surveyed how blockchain techniques can support the security and safety of LLMs, for example by verifying training data authenticity and privacy preservation. On the other hand, He et al. [105] surveyed LLMs for supporting blockchain security. They showed that LLMs can support the blockchain community by detecting vulnerabilities in the source code of smart contracts, detecting irregular transaction patterns or the generating of smart contracts.

Robotics: Kim et al. [146] explored the use of LLMs in robotics. Their focus is on recent LLMs (after GPT-3.5) and text-based LLMs, while still allowing the inclusion of relevant multimodal approaches. They distinguish four main categories of LLM use: communication, perception, planning, and control. Additionally, they provided guidelines for prompting LLMs for four robotic tasks: interactive grounding, scene-graph generation, few-shot planning, reward function generation.

Similarly, the survey by Zeng et al. [389] presented LLM applications in robotics with regards to control, perception, decision-making and path planning. Different from Kim et al. [146], they put more emphasis on LLMs and transformer architectures, as well as challenges. Lastly, Shi et al. [276] addressed the use of LLMs in socially assistive robots (SARs) with a short survey. Herein, challenges and opportunities of using LLMs in SAR were discussed.

Hardware: Similar to their use in detecting software vulnerabilities (Section 6.2), LLMs can support the security of hardware components. Makhzan and Kamali [210] compared 10 such studies.

Agriculture: Zhu et al. [429] reviewed how LLMs and vision models can be applied in agriculture.

7 Multimodality

Generally, LLMs are applied to textual data and excel at language-based tasks. However, their application has been extended beyond texts to further modalities, for which we discuss relevant surveys. Wu et al. [337] outlined the history of multimodal approaches, from single modality to recent large-scale multimodal systems. Yin et al. [382] presented information on the architecture of Multimodal Large Language Models (MLLMs) as well as their training and evaluation. Song et

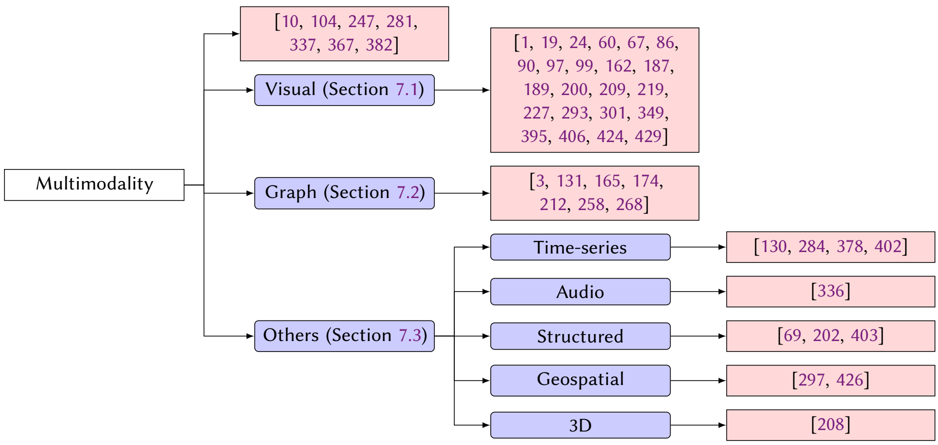

Fig. 7. Taxonomy of surveys on multimodality.

al. [281] described how different modalities can be aligned. Other surveys studied the generation and editing across modalities [104], training data [10, 247], or analysis of sentiments [367].

7.1 Visual

The most frequent application of multimodal language models we found is for visual tasks, with multiple surveys presenting comprehensive overviews [19, 395]. For example, Zhang et al. [395] gave background information on the visual paradigm as well as a summary of characteristics such as downstream tasks of Vision Language Models (VLMs) and their architecture. Among others, they outlined datasets and pre-training methods. There are several other comprehensive surveys, which put different foci, such as datasets [97], models [86], or details on regular LLMs [24].

The surveys by Du et al. [60] and Long et al. [200] focused on pre-trained vision-language models. First, data is transformed into desired representations. Afterwards, an architecture is designed to model the interaction between text and image. Further surveys took prompting [90], fine-tuning [349], and the detection of out-of-distribution samples and anomalies [219] into account. These advancements enabled the application of VLMs in diverse domains with surveys describing their applications in agriculture [429], medicine [99], autonomous navigation [209, 406], document understanding [1], and video analysis [227, 293, 424]

While VLMs offer advantages in several tasks, they can be vulnerable to attacks, which affects their usability in real-world applications [67]. Here, Liu et al. [187] surveyed four types of attack methods (adversarial attacks, jailbreak, prompt injection, and data poisoning) as well as potential defense methods. Fan et al. [67] considered different attack scenarios based on the type of model access (i.e., white-box, gray-box, black-box). Ethical AI has been further taken into account by Vatsa et al. [301] who surveyed bias, robustness, and interpretability of VLMs. Lee et al.[162] solely focused on biases and their mitigation. Another shortcoming of VLMs are hallucinations, which was surveyed by Liu et al. [189]. They collected methods and benchmarks for evaluating hallucinations and mitigate them. In total, there are five areas that have been addressed for mitigation: data, vision encoder, connection module, LLM, post-processing.

7.2 Graph

Jin et al.[131] created a comprehensive survey on different ways LLMs can interact with the structured information provided in graphs. Hereby, there are three types of graphs to consider: pure graphs, text-attributed graphs (e.g., nodes have texts), text-paired graphs (a complete graph is paired with text). Additionally, LLMs can be used in three different manners for graph tasks: as predictors, encoders (e.g., encoding node texts as vectors), or for aligning text encoding with Graph Neural Networks (GNNs). Their taxonomy considered the intersection of two dimensions: the application scenario (graph type) and the LLM technique. Moreover, they created an overview of datasets from different domains (e.g., academia, e-commerce, books, Wikipedia) and graph problems studied (e.g., shortest path, neighbor detection).

Other taxonomies are included in the works by Ren et al. [258] and Li et al. [174]. The taxonomy by Ren et al. [258] considered four aspects: GNNs as Prefix, LLMs as Prefix, LLMs-Graphs Integration, LLMs-Only. For GNN as prefix, data is first processed by GNNs and then fed into LLMs. LLMs as Prefix does the opposite, processing data with LLMs to improve GNNs. LLMs-Graphs Integration entails methods that improve the ability of LLMs to handle graph data, while LLMs-only describes work that applies LLMs to graph tasks via prompting. Li et al. [174] devised a taxonomy with three categories: enhancer (enhancing quality of node embeddings), predictor (using LLMs for prediction in graph-tasks), and alignment (aligning embedding spaces of LLMs and Graph Neural Networks).

Mao et al. [212] studied the integration of LLMs for Graph Representation Learning (GRL). They outlined existing approaches for using LLMs to improve GRL tasks. Approaches are investigated with regards to four components: knowledge extraction, knowledge organization, integration strategies, and training strategies. Shang and Huang [268] surveyed the use of LLMs for graph analytics tasks. Their survey considered three aspects, the processing of graph queries with LLMs, inference and learning over graphs, and applications. Ample visual examples are provided for the tasks (graph understanding, graph learning, graph-formed reasoning) and prompts.

While the outlined surveys addressed graphs in general, we found two surveys focused on **knowledge graphs**. Such graphs are used to model and structure knowledge bases. On one hand, Agrawal et al. [3] surveyed how knowledge graphs have been used to combat hallucinations in LLMs. For this, they defined three groups: inference (e.g., RAG), training (e.g., pre-training, fine-tuning), and validation (e.g., fact-checking LLMs). On the other hand, Li and Xu [165] addressed both, how LLMs can enhance knowledge graphs and how knowledge graphs can enhance LLMs.

7.3 Others

Beyond the extensively researched domains of vision and graphs, MLLMs are expanding to a broader range of formats. The following surveys cover these emerging modalities, each presenting unique challenges and opportunities when integrated with LLMs.

Time-series: Jiang et al. [130] created a survey on time-series analysis with LLMs. LLMs can model time-series via querying, tokenization, prompting, fine-tuning, or the integration of LLM output in existing models. Overall, this survey includes 21 studies, over various applications (e.g., CV, mobility, healthcare, finance), for which the modeling approaches, tasks, and underlying LLMs are extracted. The survey by Ye et al. [378] contains time-series studies for similar application domains, however, their analysis focused on three dimensions: effectiveness, efficiency, and explainability. Su et al. [284] included a discussion of anomaly detection for time series.

Beyond this scope, Zhang et al. [402] considered visual representations of time series as well as LLM-based tools to support the processing of time-series, for example by creating code.

Audio: By converting audio into discrete codes, they can be processed by language models. Wu et al. [336] provided an overview of six neural models and 11 language models for processing audio. For each language model, they presented the addressed tasks as well as input and output format.

Structured: Tables represent data in a structured, two-dimensional manner and can be processed with LLMs. Fang et al. [69] reviewed techniques, metrics, datasets, and models for four techniques for applying LLMs to tables: serialization, table manipulation, prompt engineering, and end-to-end systems. Emphasis is also put on the use of LLMs to generate tabular data. In addition to discussing training approaches for LLMs and Visual language models, Lu et al. [202] described prompting techniques and the use of agents.

Zhang et al. [403] focused their survey on techniques to improve the performance of LLMs for different table processing tasks (QA, fact verification, table to text, text to SQL). For five popular improvement techniques, they showed a performance comparison over four datasets.

Geospatial: Zhou et al. [426] surveyed LLMs with geo-perceptive capabilities to handle multiple modalities of geospatial data. They focused on a specific family of language models, Vision-language geo-foundation models (VLGFM). These VLGFM incorporate diverse data modalities to (satellite images, geo-tagged text, remote sensing images) to address a wide range of geospatial tasks (e.g, image captioning, visual grounding) The survey includes an overview of tasks, datasets and metrics for evaluation as well as a description of model architectures. Tucker [297] reviewed LLMs for Geospatial Location Embeddings (GLE) to represent and express space.

3D: LLMs have seen use in spatial tasks, which require the consideration of three dimensions. In particular, Ma et al. [208] investigated how LLMs can understand and interact with 3D data. Their survey provided information on different 3D data representations (e.g., point cloud, grid, mesh), tasks (captioning, grounding, conversation (question answering), agent, generation), and datasets. Additionally, the LLMs and 3D components for 37 publications are extracted and described.

8 Risks and Mitigation

While prior sections outlined the benefits in various domains, LLMs can be susceptible to bias and safety issues or share private information [42, 197]. These concerns propagate to various fields [47], such as healthcare [96] or education [364], and are major challenges that need to be overcome to achieve trust [71, 119, 184, 197, 301] and transparency (e.g., by explaining responses) [20, 203, 253, 413]. Researchers showed interest in the different types of risks and their mitigation [266].

In the following, we discuss the main concerns pointed out and covered by existing surveys: fairness, hallucinations, security, and privacy [40, 56, 81, 119, 144, 154].

8.1 Fairness and Bias

LLMs can propagate social biases from the training data, causing fairness and bias issues, which has been covered by several studies [39, 75, 172]. The survey by Gallegos et al. [75] is the most comprehensive with three taxonomies, one for metrics, datasets, and bias mitigation methods each. Chu et al. [39] presented toolkits in addition to datasets, while Li et al. [172] took model size into account. They distinguished fairness studies based on LLM size, as smaller models allow for fine-tuning, while large models are prompted instead.

Surveys have also focused on a specific aspect, such as metrics [50] or the debiasing of LLMs [184]. Another set of works surveyed specific fields for biases, such as education [160], e-commerce [257], information retrieval [44], vision-language models [162], or recommender systems [263].

Lastly, Wang et al. [314] collected human perspectives on LLM bias from several studies and summarized their perspectives. Among other things, people perceived bias more when they failed to receive desired responses.

8.2 Hallucination

At times, the outputs generated by LLMs are inconsistent with the actual answer or the user input itself, which is called “hallucination”. There are three comprehensive surveys addressing this issue [116, 377, 405]. They contain details on causes, benchmarks, and mitigation approaches. We have also found two surveys addressing hallucinations for vision-language models [11, 189].

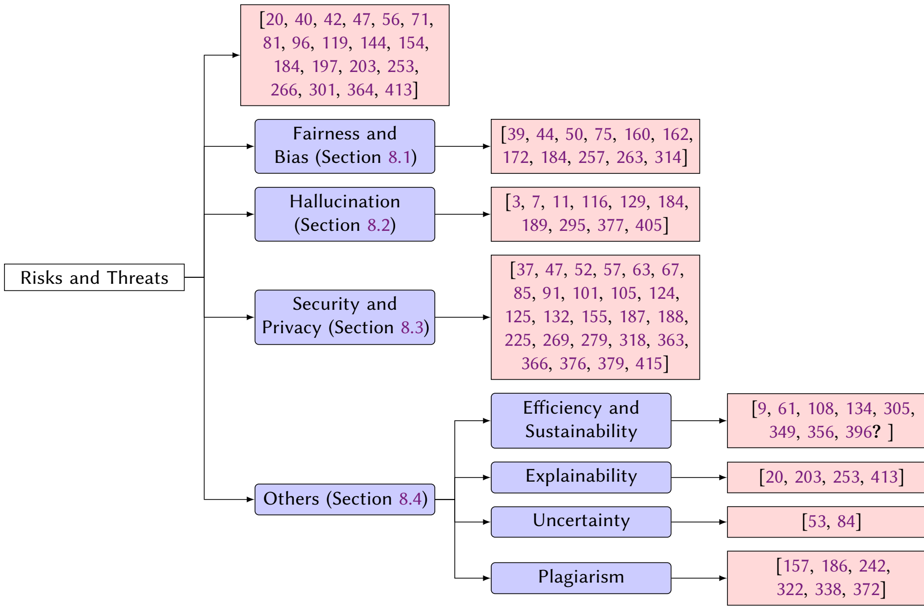

Fig. 8. Taxonomy of surveys on risks and threats.

As hallucinations can be detrimental to LLM usage, their mitigation ("dehallucination") is studied. We encountered surveys that either gave an overview of mitigation techniques [184, 295], or elaborated on a single approach in detail, such as knowledge graphs [3] or knowledge augmentation [7].

A different point of view is provided by the survey of Jiang et al. [129], who considered hallucinations as a chance to foster creativity in LLMs.

8.3 Security and Privacy

Yao et al. [376] conducted a holistic survey on the security and privacy of LLMs, including the use of LLMs to support privacy by, for example, generating secure code or testing applications. Security has further been considered for multimodal LLMs [67, 132, 187, 269], agents [101, 318], and blockchain [85, 105]. In addition, Derner et al. [52] considered the use of LLMs for malicious intents, such as writing spam messages or sharing fake news.

Remaining surveys consider the security and safety risks of LLMs themselves. Particularly, security and privacy issues can be exploited passively by leaking personally identifiable information via regular interactions with a model, or dedicated attacks [47, 279, 363]. In terms of attacks, existing surveys have summarized them into the categories listed in Table 4.

The majority of these surveys not only describe the attacks, but also provide details on their mitigation. While attacks are usually automatic, some surveys list methods with human interference [37, 57, 269]. Attacks can be carried out on the provided service, by assessing or changing the prompt or responses, or slowing the model down with intensive prompts [52, 318]. Likewise, attacks can be performed by the service provider, for instance to recompute user queries [91].

8.4 Others

In addition to core risks detailed previously, a holistic assessment requires considering broader challenges related to their development and use in society. The following surveys address several of these considerations.

Table 4. Overview of attack techniques and surveys which describe them.

Security Attacks	Backdoor/data poisoning attacks [37, 47, 52, 57, 63, 124, 187, 188, 318, 363, 366, 415]
	Jailbreak [37, 47, 132, 187, 188, 269, 318, 379]
	Adversarial [37, 57, 91, 187, 269, 318]
	Prompt injection [37, 47, 52, 57, 63, 155, 187, 188, 269, 318]
Privacy Attacks - Data	Membership inference [47, 52, 63, 125, 225, 279, 318, 363]
	Attribute inference [91, 225, 279, 318, 363]
	Corpus inference [63, 91, 318, 363]
	Data extraction attack [47, 225, 279, 318]
	Knowledge stealing attacks (With RAG)[318]
Privacy Attacks - Model	Model extraction/stealing [52, 63, 318, 363]
	Accuracy extraction [91]
	Fidelity extraction [91]

Efficiency and Sustainability: The training and fine-tuning are compute-intensive activities that require time, compute resources, and energy, which introduces a critical challenge.

Efficiency concerns have been addressed for both text-based language models and multimodal language models [134, 349, 356]. For instance, Xu et al.'s [356] focus was on resource-efficiency, which entails the improvement of the efficiency of LLMs but also making better use of the available resources (e.g., data). Thereby, some of the methods would increase performance and not necessarily efficiency (e.g., better performance via data augmentation). Bai et al. [9] categorized resource efficiency methods based on the resources they optimized: computational, memory, energy, financial, and network. Wan et al. [305] listed frameworks which provide efficiency methods. Duan et al. [61] did not only consider efficiency but also fault tolerance. This is important to ensure that the computations are reliable and do not get lost. Further efficiency considerations for the different components of LLMs can be found in Section 4.

In terms of sustainability, Hort et al. [108] considered the sustainability of training and fine-tuning LLMs, as well as sharing them. By making trained LLMs publicly available, users do not need to train them themselves and save energy. Zhang and Chen [396] presented how LLMs can support in improving the energy efficiency of buildings and aid in decarbonization activities.

Explainability: In the context of LLMs, explainability refers to methods used to understand the reasoning behind their generation. These models are often categorized as "black-box" systems since their inner reasoning is challenging to interpret. Explainability allows to understand the model's behavior, which increases user trust, allows debugging, and helps guide their development.

Some authors provide a more general overview of the field. Zhao et al. [413] proposed a taxonomy classifying these methods based on two training paradigms: traditional fine-tuning-based and prompting-based. For each paradigm, they further discussed the goals to achieve local and global explanations. Luo and Specia [203] classified methods into local analysis and global analysis. Local analysis includes feature attribution and transformer block analysis, while global analysis focuses on mechanistic interpretability and probing-based methods. Meanwhile, Cambria et al. [20] explored the balance between interpretability and functional advancements. They labeled each work based on one or more primary objectives: comparison of models, explanation, improvement, interpretability, and reasoning. However, a common thread in these overviews is the challenges of evaluating explainability [203, 413]. In this landscape, Rai et al. [253] provided insights and challenges in the subfield of mechanistic interpretability which focuses on reverse engineering operations to understand LLMs. This approach analyzes features and circuits of a model to provide insights and enable practical applications in fields such as knowledge editing and AI safety.

Uncertainty: To determine the confidence of LLMs with their generated responses, their uncertainty can be determined [53, 84]. This metric can indicate the likelihood that a response is correct, which is critical for establishing model reliability.
Plagiarism: The content generated by LLMs can also be seen as a potential threat, as one could misuse them for plagiarism [242]. Therefore, other studies set out to detect LLM-generated content [338, 372]. Another approach to identify as well as protect the ownership of LLM generated texts is watermarking [157, 186, 322].

9 Conclusion

A lot of secondary reviews have been written for LLMs. We have found 984 secondary studies with our systematic search procedure, 424 of which are included in this article (all studies can be found in our GitHub repository). We hope that this provides researchers with a useful resource for navigating the growing field of surveys, to which we are adding another one.

Acknowledgments

This work is supported by the Research Council of Norway through the secureIT project (#288787), and by the European Union through Horizon Europe Marie Skłodowska-Curie Actions (#101151798).

References

[1] Abdelrahman Abdallah, Daniel Eberharter, Zoe Pfister, and Adam Jatowt. 2024. Transformers and Language Models in Form Understanding: A Comprehensive Review of Scanned Document Analysis. https://doi.org/10.48550/ARXIV.2403.04080 arXiv:2403.04080
[2] Muhammad Farid Adilazuarda, Sagnik Mukherjee, Pradhyumna Lavania, Siddhant Singh, Ashutosh Dwivedi, Alham Fikri Aji, Jacki O'Neill, Ashutosh Modi, and Monojit Choudhury. 2024. Towards Measuring and Modeling "Culture" in Llms: A Survey. https://doi.org/10.48550/ARXIV.2403.15412 arXiv:2403.15412
[3] Garima Agrawal, Tharindu Kumarage, Zeyad Alghamdi, and Huan Liu. 2024. Can Knowledge Graphs Reduce Hallucinations in Llms? : A Survey. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 3947–3960. https://doi.org/10.18653/V1/2024.NAACL-LONG.219
[4] Alon Albalak, Yanai Elazar, Sang Michael Xie, Shayne Longpre, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, Colin Raffel, Shiyu Chang, Tatsunori Hashimoto, and William Yang Wang. 2024. A Survey on Data Selection for Language Models. https://doi.org/10.48550/ARXIV.2402.16827 arXiv:2402.16827
[5] Hanieh Alipour, Nick Pendar, and Kohinoor Roy. 2024. ChatGPT Alternative Solutions: Large Language Models Survey. https://doi.org/10.48550/ARXIV.2403.14469 arXiv:2403.14469
[6] Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona T. Diab, and Marjan Ghazvininejad. 2022. A Review on Language Models as Knowledge Bases. https://doi.org/10.48550/ARXIV.2204.06031 arXiv:2204.06031
[7] Konstantinos Andriopoulos and Johan Pouwelse. 2023. Augmenting LLMs with Knowledge: A Survey on Hallucination Prevention. https://doi.org/10.48550/ARXIV.2309.16459 arXiv:2309.16459
[8] Dang Hoang Anh, Dinh-Truong Do, Vu Tran, and Nguyen Le Minh. 2023. The Impact of Large Language Modeling on Natural Language Processing in Legal Texts: A Comprehensive Survey. In *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, Hanoi, Vietnam, 1–7. https://doi.org/10.1109/KSE59128.2023.10299488
[9] Guangji Bai, Zheng Chai, Chen Ling, Shiyu Wang, Jiaying Lu, Nan Zhang, Tingwei Shi, Ziyang Yu, Mengdan Zhu, Yifei Zhang, Carl Yang, Yue Cheng, and Liang Zhao. 2024. Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models. https://doi.org/10.48550/ARXIV.2401.00625 arXiv:2401.00625
[10] Tianyi Bai, Hao Liang, Binwang Wan, Ling Yang, Bozhou Li, Yifan Wang, Bin Cui, Conghui He, Binhang Yuan, and Wentao Zhang. 2024. A Survey of Multimodal Large Language Model from A Data-Centric Perspective. https://doi.org/10.48550/ARXIV.2405.16640 arXiv:2405.16640
[11] Zechen Bai, Pichao Wang, Tianjun Xiao, Tong He, Zongbo Han, Zheng Zhang, and Mike Zheng Shou. 2024. Hallucination of Multimodal Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2404.18930 arXiv:2404.18930
[12] Muneera Bano, Didar Zowghi, and Naveed Ikram. 2014. Systematic Reviews in Requirements Engineering: A Tertiary Study. In *2014 IEEE 4th International Workshop on Empirical Requirements Engineering (EmpiRE)*. IEEE, Karlskrona,

Sweden, 9–16. https://doi.org/10.1109/EmpiRE.2014.6890110
[13] Saikat Barua. 2024. Exploring Autonomous Agents through the Lens of Large Language Models: A Review. https://doi.org/10.48550/ARXIV.2404.04442 arXiv:2404.04442
[14] Sarthak Bhardwaj, Pardeep Singh, and Mohammad Khalid Pandit. 2024. A Survey on the Integration and Optimization of Large Language Models in Edge Computing Environments. In *16th International Conference on Computer and Automation Engineering, ICCAE 2024, Melbourne, Australia, March 14-16, 2024*. IEEE, Melbourne, Australia, 168–172. https://doi.org/10.1109/ICCAE59995.2024.10569285
[15] Prajjwal Bhargava and Vincent Ng. 2022. Commonsense Knowledge Reasoning and Generation with Pre-Trained Language Models: A Survey. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, the Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, Vol. 36. AAAI Press, Virtual Event, 12317–12325. https://doi.org/10.1609/AAAI.V36I11.21496
[16] Alberto Blanco-Justicia, Najeeb Jebreel, Benet Manzanares-Salor, David Sánchez, Josep Domingo-Ferrer, Guillem Collell, and Kuan Eeik Tan. 2024. Digital Forgetting in Large Language Models: A Survey of Unlearning Methods. https://doi.org/10.48550/ARXIV.2404.02062 arXiv:2404.02062
[17] Víctor A. Braberman, Flavia Bonomo-Braberman, Yiannis Charalambous, Juan Gabriel Colonna, Lucas C. Cordeiro, and Rosiane de Freitas. 2024. Tasks People Prompt: A Taxonomy of LLM Downstream Tasks in Software Verification and Falsification Approaches. https://doi.org/10.48550/ARXIV.2404.09384 arXiv:2404.09384
[18] Felix Brakel, Uraz Odyurt, and Ana Lucia Varbanescu. 2024. Model Parallelism on Distributed Infrastructure: A Literature Review from Theory to LLM Case-Studies. https://doi.org/10.48550/ARXIV.2403.03699 arXiv:2403.03699
[19] Davide Caffagni, Federico Cocchi, Luca Barsellotti, Nicholas Moratelli, Sara Sarto, Lorenzo Baraldi, Marcella Cornia, and Rita Cucchiara. 2024. The Revolution of Multimodal Large Language Models: A Survey. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and Virtual Meeting, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 13590–13618. https://doi.org/10.18653/v1/2024.findings-acl.807
[20] Erik Cambria, Lorenzo Malandri, Fabio Mercorio, Navid Nobani, and Andrea Seveso. 2024. XAI Meets Llms: A Survey of the Relation between Explainable AI and Large Language Models. https://doi.org/10.48550/ARXIV.2407.15248 arXiv:2407.15248
[21] Boxi Cao, Hongyu Lin, Xianpei Han, and Le Sun. 2024. The Life Cycle of Knowledge in Big Language Models: A Survey. *Machine Intelligence Research* 21, 2 (2024), 217–238. https://doi.org/10.1007/S11633-023-1416-X
[22] Boxi Cao, Keming Lu, Xinyu Lu, Jiawei Chen, Mengjie Ren, Hao Xiang, Peilin Liu, Yaojie Lu, Ben He, Xianpei Han, Le Sun, Hongyu Lin, and Bowen Yu. 2024. Towards Scalable Automated Alignment of Llms: A Survey. https://doi.org/10.48550/ARXIV.2406.01252 arXiv:2406.01252
[23] Yuji Cao, Huan Zhao, Yuheng Cheng, Ting Shu, Guolong Liu, Gaoqi Liang, Junhua Zhao, and Yun Li. 2024. Survey on Large Language Model-Enhanced Reinforcement Learning: Concept, Taxonomy, and Methods. https://doi.org/10.48550/ARXIV.2404.00282 arXiv:2404.00282
[24] Kilian Carolan, Laura Fennelly, and Alan F. Smeaton. 2024. A Review of Multi-Modal Large Language and Vision Models. https://doi.org/10.48550/ARXIV.2404.01322 arXiv:2404.01322
[25] Kaiyan Chang, Songcheng Xu, Chenglong Wang, Yingfeng Luo, Tong Xiao, and Jingbo Zhu. 2024. Efficient Prompting Methods for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2404.01077 arXiv:2404.01077
[26] Tyler A. Chang and Benjamin K. Bergen. 2024. Language Model Behavior: A Comprehensive Survey. *Comput. Linguistics* 50, 1 (2024), 293–350. https://doi.org/10.1162/COLI_A_00492
[27] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 39:1–39:45. https://doi.org/10.1145/3641289
[28] Arnav Chavan, Raghav Magazine, Shubham Kushwaha, Mérouane Debbah, and Deepak K. Gupta. 2024. Faster and Lighter Llms: A Survey on Current Challenges and Way Forward. https://doi.org/10.48550/ARXIV.2402.01799 arXiv:2402.01799
[29] Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. 2023. Unleashing the Potential of Prompt Engineering in Large Language Models: A Comprehensive Review. https://doi.org/10.48550/ARXIV.2310.14735 arXiv:2310.14735
[30] Junyi Chen. 2023. A Survey on Large Language Models for Personalized and Explainable Recommendations. https://doi.org/10.48550/ARXIV.2311.12338 arXiv:2311.12338
[31] Le Chen, Nesreen K. Ahmed, Akash Dutta, Arijit Bhattacharjee, Sixing Yu, Quazi Ishtiaque Mahmud, Waqwoya Abebe, Hung Phan, Aishwarya Sarkar, Branden Butler, Niranjan Hasabnis, Gal Oren, Vy A. Vo, Juan Pablo Muñoz, Theodore L. Willke, Tim Mattson, and Ali Jannesari. 2024. The Landscape and Challenges of HPC Research and Llms.

https://doi.org/10.48550/ARXIV.2402.02018 arXiv:2402.02018

[32] Nuo Chen, Yang Deng, and Jia Li. 2024. The Oscars of AI Theater: A Survey on Role-Playing with Language Models. https://doi.org/10.48550/ARXIV.2407.11484 arXiv:2407.11484

[33] Xiaolan Chen, Jiayang Xiang, Shanfu Lu, Yexin Liu, Mingguang He, and Danli Shi. 2024. Evaluating Large Language Models in Medical Applications: A Survey. https://doi.org/10.48550/ARXIV.2405.07468 arXiv:2405.07468

[34] Yiren Chen, Mengjiao Cui, Ding Wang, Yiyang Cao, Peian Yang, Bo Jiang, Zhigang Lu, and Baoxu Liu. 2024. A Survey of Large Language Models for Cyber Threat Detection. *Computers & Security* 145 (2024), 104016. https://doi.org/10.1016/J.COSE.2024.104016

[35] Zhiyu Zoey Chen, Jing Ma, Xinlu Zhang, Nan Hao, An Yan, Armineh Nourbakhsh, Xianjun Yang, Julian J. McAuley, Linda R. Petzold, and William Yang Wang. 2024. A Survey on Large Language Models for Critical Societal Domains: Finance, Healthcare, and Law. https://doi.org/10.48550/ARXIV.2405.01769 arXiv:2405.01769

[36] Shipra Chhina, Bhavna Josephine Antony, and Selena Firmin. 2023. Navigating the Terrain of Large Language Models in Higher Education- A Systematic Literature Review. In *Australasian Conference on Information Systems, ACIS 2023, Wellington, New Zealand, December 5-8, 2023*. Association for Information Systems (AIS), Wellington, New Zealand.

[37] Arijit Ghosh Chowdhury, Md Mofijul Islam, Vaibhav Kumar, Faysal Hossain Shezan, Vaibhav Kumar, Vinija Jain, and Aman Chadha. 2024. Breaking down the Defenses: A Comparative Survey of Attacks on Large Language Models. https://doi.org/10.48550/ARXIV.2403.04786 arXiv:2403.04786

[38] Zhibo Chu, Shiwen Ni, Zichong Wang, Xi Feng, Chengming Li, Xiping Hu, Ruifeng Xu, Min Yang, and Wenbin Zhang. 2024. History, Development, and Principles of Large Language Models-an Introductory Survey. https://doi.org/10.48550/ARXIV.2402.06853 arXiv:2402.06853

[39] Zhibo Chu, Zichong Wang, and Wenbin Zhang. 2024. Fairness in Large Language Models: A Taxonomic Survey. *SIGKDD Explor.* 26, 1 (2024), 34–48. https://doi.org/10.1145/3682112.3682117

[40] Jaymari Chua, Yun Li, Shiyi Yang, Chen Wang, and Lina Yao. 2024. AI Safety in Generative AI Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2407.18369 arXiv:2407.18369

[41] Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, Tianren Gao, Erlong Li, Kun Tang, Zhipeng Cao, Tong Zhou, Ao Liu, Xinrui Yan, Shuqi Mei, Jianguo Cao, Ziran Wang, and Chao Zheng. 2024. A Survey on Multimodal Large Language Models for Autonomous Driving. In *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, WACVW 2024 - Workshops, Waikoloa, HI, USA, January 1-6, 2024*. IEEE, Waikoloa, HI, USA, 958–979. https://doi.org/10.1109/WACVW60836.2024.00106

[42] Tianyu Cui, Yanling Wang, Chuanpu Fu, Yong Xiao, Sijia Li, Xinhao Deng, Yunpeng Liu, Qinglin Zhang, Ziyi Qiu, Peiyang Li, Zhixing Tan, Junwu Xiong, Xinyu Kong, Zujie Wen, Ke Xu, and Qi Li. 2024. Risk Taxonomy, Mitigation, and Assessment Benchmarks of Large Language Model Systems. https://doi.org/10.48550/ARXIV.2401.05778 arXiv:2401.05778

[43] Fabio Q.B. Da Silva, André L.M. Santos, Sérgio Soares, A. César C. França, Cleviton V.F. Monteiro, and Felipe Farias Maciel. 2011. Six Years of Systematic Literature Reviews in Software Engineering: An Updated Tertiary Study. *Information and Software Technology* 53, 9 (Sept. 2011), 899–913. https://doi.org/10.1016/j.infsof.2011.04.004

[44] Sunhao Dai, Chen Xu, Shicheng Xu, Liang Pang, Zhenhua Dong, and Jun Xu. 2024. Unifying Bias and Unfairness in Information Retrieval: A Survey of Challenges and Opportunities with Large Language Models. https://doi.org/10.48550/ARXIV.2404.11457 arXiv:2404.11457

[45] Sumit Kumar Dam, Choong Seon Hong, Yu Qiao, and Chaoning Zhang. 2024. A Complete Survey on LLM-based AI Chatbots. https://doi.org/10.48550/ARXIV.2406.16937 arXiv:2406.16937

[46] Cinzia Daraio, Kristiaan Kerstens, Thyago Nepomuceno, and Robin C. Sickles. 2020. Empirical Surveys of Frontier Applications: A Meta-review. *International Transactions in Operational Research* 27, 2 (March 2020), 709–738. https://doi.org/10.1111/itor.12649

[47] Badhan Chandra Das, M. Hadi Amini, and Yanzhao Wu. 2024. Security and Privacy Challenges of Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.00888 arXiv:2402.00888

[48] Xavier Daull, Patrice Bellot, Emmanuel Bruno, Vincent Martin, and Elisabeth Murisasco. 2023. Complex QA and Language Models Hybrid Architectures, Survey. https://doi.org/10.48550/ARXIV.2302.09051 arXiv:2302.09051

[49] Gabriel de Jesus Coelho da Silva and Carlos Becker Westphall. 2024. A Survey of Large Language Models in Cybersecurity. https://doi.org/10.48550/ARXIV.2402.16968 arXiv:2402.16968

[50] Pieter Delobelle, Ewoenam Kwaku Tokpo, Toon Calders, and Bettina Berendt. 2021. Measuring Fairness with Biased Rulers: A Survey on Quantifying Biases in Pretrained Language Models. https://doi.org/10.48550/arXiv.2112.07447 arXiv:2112.07447

[51] Chunyuan Deng, Yilun Zhao, Yuzhao Heng, Yitong Li, Jiannan Cao, Xiangru Tang, and Arman Cohan. 2024. Unveiling the Spectrum of Data Contamination in Language Model: A Survey from Detection to Remediation. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and Virtual Meeting, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok,

Thailand and virtual meeting, 16078–16092. https://doi.org/10.18653/v1/2024.findings-acl.951
[52] Erik Derner, Kristina Batistic, Jan Zahálka, and Robert Babuska. 2023. A Security Risk Taxonomy for Large Language Models. https://doi.org/10.48550/ARXIV.2311.11415 arXiv:2311.11415
[53] Dhruv Dhamani and Mary Lou Maher. 2023. The Tyranny of Possibilities in the Design of Task-Oriented LLM Systems: A Scoping Survey. https://doi.org/10.48550/ARXIV.2312.17601 arXiv:2312.17601
[54] Laurence Dierickx, Arjen van Dalen, Andreas L. Opdahl, and Carl-Gustav Linden. 2024. Striking the Balance in Using Llms for Fact-Checking: A Narrative Literature Review. In *Disinformation in Open Online Media - 6th Multidisciplinary International Symposium, MISDOOM 2024, Münster, Germany, September 2-4, 2024, Proceedings (Lecture Notes in Computer Science, Vol. 15175)*, Mike Preuss, Agata Leszkiewicz, Jean-Christopher Boucher, Ofer Fridman, and Lucas Stampe (Eds.). Springer, Cham, 1–15. https://doi.org/10.1007/978-3-031-71210-4_1
[55] Tianyu Ding, Tianyi Chen, Haidong Zhu, Jiachen Jiang, Yiqi Zhong, Jinxin Zhou, Guangzhi Wang, Zhihui Zhu, Ilya Zharkov, and Luming Liang. 2023. The Efficiency Spectrum of Large Language Models: An Algorithmic Survey. https://doi.org/10.48550/ARXIV.2312.00678 arXiv:2312.00678
[56] Yi Dong, Ronghui Mu, Yanghao Zhang, Siqi Sun, Tianle Zhang, Changshun Wu, Gaojie Jin, Yi Qi, Jinwei Hu, Jie Meng, Saddek Bensalem, and Xiaowei Huang. 2024. Safeguarding Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2406.02622 arXiv:2406.02622
[57] Zhichen Dong, Zhanhui Zhou, Chao Yang, Jing Shao, and Yu Qiao. 2024. Attacks, Defenses and Evaluations for LLM Conversation Safety: A Survey. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 6734–6747. https://doi.org/10.18653/V1/2024.NAACL-LONG.375
[58] Fei Du, Xin-Jian Ma, Jing-Ru Yang, Yi Liu, Chao-Ran Luo, Xue-Bin Wang, Hai-Ou Jiang, and Xiang Jing. 2024. A Survey of LLM Datasets: From Autoregressive Model to AI Chatbot. *Journal of Computer Science and Technology* 39, 3 (2024), 542–566. https://doi.org/10.1007/S11390-024-3767-3
[59] Mengnan Du, Fengxiang He, Na Zou, Dacheng Tao, and Xia Hu. 2022. Shortcut Learning of Large Language Models in Natural Language Understanding: A Survey. https://doi.org/10.48550/ARXIV.2208.11857 arXiv:2208.11857
[60] Yifan Du, Zikang Liu, Junyi Li, and Wayne Xin Zhao. 2022. A Survey of Vision-Language Pre-Trained Models. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, Vienna, Austria, 5436–5443. https://doi.org/10.24963/IJCAI.2022/762
[61] Jiangfei Duan, Shuo Zhang, Zerui Wang, Lijuan Jiang, Wenwen Qu, Qinghao Hu, Guoteng Wang, Qizhen Weng, Hang Yan, Xingcheng Zhang, Xipeng Qiu, Dahua Lin, Yonggang Wen, Xin Jin, Tianwei Zhang, and Peng Sun. 2024. Efficient Training of Large Language Models on Distributed Infrastructures: A Survey. https://doi.org/10.48550/ARXIV.2407.20018 arXiv:2407.20018
[62] Jan Egger, Christina Gsaxner, Antonio Pepe, Kelsey L. Pomykala, Frederic Jonske, Manuel Kurz, Jianning Li, and Jens Kleesiek. 2022. Medical Deep Learning—A Systematic Meta-Review. *Computer Methods and Programs in Biomedicine* 221 (June 2022), 106874. https://doi.org/10.1016/j.cmpb.2022.106874
[63] Aysan Esmradi, Daniel Wankit Yip, and Chun-Fai Chan. 2023. A Comprehensive Survey of Attack Techniques, Implementation, and Mitigation Strategies in Large Language Models. In *Ubiquitous Security - Third International Conference, UbiSec 2023, Exeter, UK, November 1-3, 2023, Revised Selected Papers (Communications in Computer and Information Science, Vol. 2034)*, Guojun Wang, Haozhe Wang, Geyong Min, Nektarios Georgalas, and Weizhi Meng (Eds.). Springer, Singapore, 76–95. https://doi.org/10.1007/978-981-97-1274-8_6
[64] Angela Fan, Beliz Gokkaya, Mark Harman, Mitya Lyubarskiy, Shubho Sengupta, Shin Yoo, and Jie M. Zhang. 2023. Large Language Models for Software Engineering: Survey and Open Problems. In *IEEE/ACM International Conference on Software Engineering: Future of Software Engineering, ICSE-FoSE 2023, Melbourne, Australia, May 14-20, 2023*. IEEE, Melbourne, Australia, 31–53. https://doi.org/10.1109/ICSE-FOSE59343.2023.00008
[65] Lizhou Fan, Lingyao Li, Zihui Ma, Sanggyu Lee, Huizi Yu, and Libby Hemphill. 2023. A Bibliometric Review of Large Language Models Research from 2017 to 2023. https://doi.org/10.48550/ARXIV.2304.02020 arXiv:2304.02020
[66] Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. A Survey on RAG Meeting Llms: Towards Retrieval-Augmented Large Language Models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, Ricardo Baeza-Yates and Francesco Bonchi (Eds.). ACM, 6491–6501. https://doi.org/10.1145/3637528.3671470
[67] Yihe Fan, Yuxin Cao, Ziyu Zhao, Ziyao Liu, and Shaofeng Li. 2024. Unbridled Icarus: A Survey of the Potential Perils of Image Inputs in Multimodal Large Language Model Security. https://doi.org/10.48550/ARXIV.2404.05264 arXiv:2404.05264
[68] Wenyi Fang, Hao Zhang, Ziyu Gong, Longbin Zeng, Xuhui Lu, Biao Liu, Xiaoyu Wu, Yang Zheng, Zheng Hu, and Xun Zhang. 2023. A Survey of Metrics to Enhance Training Dependability in Large Language Models. In *34th IEEE International Symposium on Software Reliability Engineering, ISSRE 2023 - Workshops, Florence, Italy, October 9-12,*

2023. IEEE, Florence, Italy, 180–185. https://doi.org/10.1109/ISSREW60843.2023.00071

[69] Xi Fang, Weijie Xu, Fiona Anting Tan, Jiani Zhang, Ziqing Hu, Yanjun Qi, Scott Nickleach, Diego Socolinsky, Srinivasan H. Sengamedu, and Christos Faloutsos. 2024. Large Language Models(Llms) on Tabular Data: Prediction, Generation, and Understanding - A Survey. https://doi.org/10.48550/ARXIV.2402.17944 arXiv:2402.17944

[70] Zhangyin Feng, Weitao Ma, Weijiang Yu, Lei Huang, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. Trends in Integration of Knowledge and Large Language Models: A Survey and Taxonomy of Methods, Benchmarks, and Applications. https://doi.org/10.48550/ARXIV.2311.05876 arXiv:2311.05876

[71] Md Meftahul Ferdaus, Mahdi Abdelguerfi, Elias Ioup, Kendall N. Niles, Ken Pathak, and Steven Sloan. 2024. Towards Trustworthy AI: A Review of Ethical and Robust Large Language Models. https://doi.org/10.48550/ARXIV.2407.13934 arXiv:2407.13934

[72] Emilio Ferrara. 2024. Large Language Models for Wearable Sensor-Based Human Activity Recognition, Health Monitoring, and Behavioral Modeling: A Survey of Early Trends, Datasets, and Challenges. *Sensors* 24, 15 (2024), 5045. https://doi.org/10.3390/S24155045

[73] Stefano Filippi and Barbara Motyl. 2024. Large Language Models (Llms) in Engineering Education: A Systematic Review and Suggestions for Practical Adoption. *Inf.* 15, 6 (2024), 345. https://doi.org/10.3390/INFO15060345

[74] Eduard Frankford, Ingo Höhn, Clemens Sauerwein, and Ruth Breu. 2024. A Survey Study on the State of the Art of Programming Exercise Generation Using Large Language Models. https://doi.org/10.48550/ARXIV.2405.20183 arXiv:2405.20183

[75] Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md. Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2023. Bias and Fairness in Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2309.00770 arXiv:2309.00770

[76] Roberto Gallotta, Graham Todd, Marvin Zammit, Sam Earle, Antonios Liapis, Julian Togelius, and Georgios N. Yannakakis. 2024. Large Language Models and Games: A Survey and Roadmap. https://doi.org/10.48550/ARXIV.2402.18659 arXiv:2402.18659

[77] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2023. Large Language Models Empowered Agent-Based Modeling and Simulation: A Survey and Perspectives. https://doi.org/10.48550/ARXIV.2312.11970 arXiv:2312.11970

[78] Jie Gao, Simret Araya Gebreegziabher, Kenny Tsu Wei Choo, Toby Jia-Jun Li, Simon Tangi Perrault, and Thomas W. Malone. 2024. A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, and Corina Sas (Eds.). ACM, Honolulu HI USA, 24:1–24:11. https://doi.org/10.1145/3613905.3650786

[79] Kaiyuan Gao, Sunan He, Zhenyu He, Jiacheng Lin, Qizhi Pei, Jie Shao, and Wei Zhang. 2023. Examining User-Friendly and Open-Sourced Large GPT Models: A Survey on Language, Multimodal, and Scientific GPT Models. https://doi.org/10.48550/ARXIV.2308.14149 arXiv:2308.14149

[80] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. 2023. Retrieval-Augmented Generation for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2312.10997 arXiv:2312.10997

[81] Zhengjie Gao, Xuanzi Liu, Yuanshuai Lan, and Zheng Yang. 2024. A Brief Survey on Safety of Large Language Models. *Journal of Computing and Information Technology* 32, 1 (2024), 47–64. https://doi.org/10.20532/cit.2024.1005778

[82] Silvia García-Méndez, Francisco de Arriba Pérez, and Maria del Carmen Lopez-Perez. 2024. A Review on the Use of Large Language Models as Virtual Tutors. https://doi.org/10.48550/ARXIV.2405.11983 arXiv:2405.11983

[83] Vahid Garousi and Mika V. Mäntylä. 2016. A Systematic Literature Review of Literature Reviews in Software Testing. *Information and Software Technology* 80 (Dec. 2016), 195–216. https://doi.org/10.1016/j.infsof.2016.09.002

[84] Jiahui Geng, Fengyu Cai, Yuxia Wang, Heinz Koeppl, Preslav Nakov, and Iryna Gurevych. 2024. A Survey of Confidence Estimation and Calibration in Large Language Models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 6577–6595. https://doi.org/10.18653/V1/2024.NAACL-LONG.366

[85] Caleb Geren, Amanda Board, Gaby G. Dagher, Tim Andersen, and Jun Zhuang. 2024. Blockchain for Large Language Model Security and Safety: A Holistic Survey. https://doi.org/10.48550/ARXIV.2407.20181 arXiv:2407.20181

[86] Akash Ghosh, Arkadeep Acharya, Sriparna Saha, Vinija Jain, and Aman Chadha. 2024. Exploring the Frontier of Vision-Language Models: A Survey of Current Methodologies and Future Directions. https://doi.org/10.48550/ARXIV.2404.07214 arXiv:2404.07214

[87] Panagiotis Giadikiaroglou, Maria Lymperaiou, Giorgos Filandrianos, and Giorgos Stamou. 2024. Puzzle Solving Using Reasoning of Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.11291 arXiv:2402.11291

[88] Simon Cornelius Gorissen, Stefan Sauer, and Wolf G. Beckmann. 2024. A Survey of Natural Language-Based Editing of Low-Code Applications Using Large Language Models. In *Human-Centered Software Engineering - 10th IFIP WG 13.2 International Working Conference, HCSE 2024, Reykjavik, Iceland, July 8-10, 2024, Proceedings (Lecture Notes in Computer Science, Vol. 14793)*, Marta Kristín Lárusdóttir, Bilal Naqvi, Regina Bernhaupt, Carmelo Ardito, and Stefan Sauer (Eds.). Springer, Cham, 243–254. https://doi.org/10.1007/978-3-031-64576-1_15

[89] Candida Maria Greco, Andrea Simeri, Andrea Tagarelli, and Ester Zumpano. 2023. Transformer-Based Language Models for Mental Health Issues: A Survey. *Pattern Recognition Letters* 167 (2023), 204–211. https://doi.org/10.1016/J.PATREC.2023.02.016

[90] Jindong Gu, Zhen Han, Shuo Chen, Ahmad Beirami, Bailan He, Gengyuan Zhang, Ruotong Liao, Yao Qin, Volker Tresp, and Philip H. S. Torr. 2023. A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models. https://doi.org/10.48550/ARXIV.2307.12980 arXiv:2307.12980

[91] Shangwei Guo, Chunlong Xie, Jiwei Li, Lingjuan Lyu, and Tianwei Zhang. 2022. Threats to Pre-Trained Language Models: Survey and Taxonomy. arXiv:2202.06862

[92] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large Language Model Based Multi-Agents: A Survey of Progress and Challenges. https://doi.org/10.48550/ARXIV.2402.01680 arXiv:2402.01680

[93] Xu Guo and Han Yu. 2022. On the Domain Adaptation and Generalization of Pretrained Language Models: A Survey. https://doi.org/10.48550/ARXIV.2211.03154 arXiv:2211.03154

[94] Zishan Guo, Renren Jin, Chuang Liu, Yufei Huang, Dan Shi, Supryadi, Linhao Yu, Yan Liu, Jiaxuan Li, Bojian Xiong, and Deyi Xiong. 2023. Evaluating Large Language Models: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2310.19736 arXiv:2310.19736

[95] Zhijun Guo, Alvina Lai, Johan Hilge Thygesen, Joseph Farrington, Thomas Keen, and Kezhi Li. 2024. Large Language Model for Mental Health: A Systematic Review. https://doi.org/10.48550/ARXIV.2403.15401 arXiv:2403.15401

[96] Joschka Haltaufderheide and Robert Ranisch. 2024. The Ethics of ChatGPT in Medicine and Healthcare: A Systematic Review on Large Language Models (LLMs). *npj Digit. Medicine* 7, 1 (2024), 183. https://doi.org/10.1038/S41746-024-01157-X

[97] Raby Hamadi. 2023. Large Language Models Meet Computer Vision: A Brief Survey. https://doi.org/10.48550/ARXIV.2311.16673 arXiv:2311.16673

[98] Thorsten Händler. 2023. Balancing Autonomy and Alignment: A Multi-Dimensional Taxonomy for Autonomous LLM-powered Multi-Agent Architectures. https://doi.org/10.48550/ARXIV.2310.03659 arXiv:2310.03659

[99] Iryna Hartsock and Ghulam Rasool. 2024. Vision-Language Models for Medical Report Generation and Visual Question Answering: A Review. https://doi.org/10.48550/ARXIV.2403.02469 arXiv:2403.02469

[100] Mohammed Hassanin and Nour Moustafa. 2024. A Comprehensive Overview of Large Language Models (Llms) for Cyber Defences: Opportunities and Directions. https://doi.org/10.48550/ARXIV.2405.14487 arXiv:2405.14487

[101] Feng He, Tianqing Zhu, Dayong Ye, Bo Liu, Wanlei Zhou, and Philip S. Yu. 2024. The Emerged Security and Privacy of LLM Agent: A Survey with Case Studies. https://doi.org/10.48550/ARXIV.2407.19354 arXiv:2407.19354

[102] Kai He, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. 2023. A Survey of Large Language Models for Healthcare: From Data, Technology, and Applications to Accountability and Ethics. https://doi.org/10.48550/ARXIV.2310.05694 arXiv:2310.05694

[103] Tianyu He, Guanghui Fu, Yijing Yu, Fan Wang, Jianqiang Li, Qing Zhao, Changwei Song, Hongzhi Qi, Dan Luo, Huijing Zou, and Bing Xiang Yang. 2023. Towards a Psychological Generalist AI: A Survey of Current Applications of Large Language Models and Future Prospects. https://doi.org/10.48550/ARXIV.2312.04578 arXiv:2312.04578

[104] Yingqing He, Zhaoyang Liu, Jingye Chen, Zeyue Tian, Hongyu Liu, Xiaowei Chi, Runtao Liu, Ruibin Yuan, Yazhou Xing, Wenhai Wang, Jifeng Dai, Yong Zhang, Wei Xue, Qifeng Liu, Yike Guo, and Qifeng Chen. 2024. LLMs Meet Multimodal Generation and Editing: A Survey. https://doi.org/10.48550/ARXIV.2405.19334 arXiv:2405.19334

[105] Zheyuan He, Zihao Li, and Sen Yang. 2024. Large Language Models for Blockchain Security: A Systematic Literature Review. https://doi.org/10.48550/ARXIV.2403.14280 arXiv:2403.14280

[106] Xanh Ho, Anh-Khoa Duong Nguyen, Tuan-An Dao, Junfeng Jiang, Yuki Chida, Kaito Sugimoto, Huy Quoc To, Florian Boudin, and Akiko Aizawa. 2024. A Survey of Pre-Trained Language Models for Processing Scientific Text. https://doi.org/10.48550/ARXIV.2401.17824 arXiv:2401.17824

[107] Zijin Hong, Zheng Yuan, Qinggang Zhang, Hao Chen, Junnan Dong, Feiran Huang, and Xiao Huang. 2024. Next-Generation Database Interfaces: A Survey of LLM-based Text-to-SQL. https://doi.org/10.48550/ARXIV.2406.08426 arXiv:2406.08426

[108] Max Hort, Anastasiia Grishina, and Leon Moonen. 2023. An Exploratory Literature Study on Sharing and Energy Use of Language Models for Source Code. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM 2023, New Orleans, LA, USA, October 26-27, 2023*. IEEE, New Orleans, LA, USA, 1–12. https://doi.org/10.1109/ESEM56168.2023.10304803

[109] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John C. Grundy, and Haoyu Wang. 2023. Large Language Models for Software Engineering: A Systematic Literature Review. https://doi.org/10.48550/ARXIV.2308.10620 arXiv:2308.10620

[110] Linmei Hu, Zeyi Liu, Ziwang Zhao, Lei Hou, Liqiang Nie, and Juanzi Li. 2024. A Survey of Knowledge Enhanced Pre-Trained Language Models. *IEEE Transactions on Knowledge and Data Engineering* 36, 4 (2024), 1413–1430. https://doi.org/10.1109/TKDE.2023.3310002

[111] Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim F. Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. 2024. A Survey on Large Language Model-Based Game Agents. https://doi.org/10.48550/ARXIV.2404.02039 arXiv:2404.02039

[112] Yucheng Hu and Yuxing Lu. 2024. RAG and RAU: A Survey on Retrieval-Augmented Language Model in Natural Language Processing. https://doi.org/10.48550/ARXIV.2404.19543 arXiv:2404.19543

[113] Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li, Yi-han Sheu, Peilin Zhou, Lauren V. Moran, Sophia Ananiadou, and Andrew Beam. 2024. Large Language Models in Mental Health Care: A Scoping Review. https://doi.org/10.48550/ARXIV.2401.02984 arXiv:2401.02984

[114] Jie Huang and Kevin Chen-Chuan Chang. 2023. Towards Reasoning in Large Language Models: A Survey. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 1049–1065. https://doi.org/10.18653/V1/2023.FINDINGS-ACL.67

[115] Kaiyu Huang, Fengran Mo, Hongliang Li, You Li, Yuanchi Zhang, Weijian Yi, Yulong Mao, Jinchen Liu, Yuzhuang Xu, Jinan Xu, Jian-Yun Nie, and Yang Liu. 2024. A Survey on Large Language Models with Multilingualism: Recent Advances and New Frontiers. https://doi.org/10.48550/ARXIV.2405.10936 arXiv:2405.10936

[116] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. https://doi.org/10.48550/ARXIV.2311.05232 arXiv:2311.05232

[117] Linghan Huang, Peizhou Zhao, Huaming Chen, and Lei Ma. 2024. Large Language Models Based Fuzzing Techniques: A Survey. https://doi.org/10.48550/ARXIV.2402.00350 arXiv:2402.00350

[118] Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024. Understanding the Planning of LLM Agents: A Survey. https://doi.org/10.48550/ARXIV.2402.02716 arXiv:2402.02716

[119] Xiaowei Huang, Wenjie Ruan, Wei Huang, Gaojie Jin, Yi Dong, Changshun Wu, Saddek Bensalem, Ronghui Mu, Yi Qi, Xingyu Zhao, Kaiwen Cai, Yanghao Zhang, Sihao Wu, Peipei Xu, Dengyu Wu, André Freitas, and Mustafa A. Mustafa. 2024. A Survey of Safety and Trustworthiness of Large Language Models through the Lens of Verification and Validation. *Artificial Intelligence Review* 57, 7 (2024), 175. https://doi.org/10.1007/S10462-024-10824-0

[120] Yizheng Huang and Jimmy Huang. 2024. A Survey on Retrieval-Augmented Text Generation for Large Language Models. https://doi.org/10.48550/ARXIV.2404.10981 arXiv:2404.10981

[121] Yining Huang, Keke Tang, and Meilian Chen. 2024. A Comprehensive Survey on Evaluating Large Language Model Applications in the Medical Industry. https://doi.org/10.48550/ARXIV.2404.15777 arXiv:2404.15777

[122] Yunpeng Huang, Jingwei Xu, Zixu Jiang, Junyu Lai, Zenan Li, Yuan Yao, Taolue Chen, Lijuan Yang, Zhou Xin, and Xiaoxing Ma. 2023. Advancing Transformer Architecture in Long-Context Large Language Models: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2311.12351 arXiv:2311.12351

[123] Rasha Ahmad Husein, Hala Aburajouh, and Cagatay Catal. 2025. Large Language Models for Code Completion: A Systematic Literature Review. *Computer Standards & Interfaces* 92 (2025), 103917. https://doi.org/10.1016/J.CSI.2024.103917

[124] Aftab Hussain, Md. Rafiqul Islam Rabin, Toufique Ahmed, Bowen Xu, Premkumar T. Devanbu, and Mohammad Amin Alipour. 2024. Trojans in Large Language Models of Code: A Critical Review through a Trigger-Based Taxonomy. https://doi.org/10.48550/ARXIV.2405.02828 arXiv:2405.02828

[125] Shotaro Ishihara. 2023. Training Data Extraction from Pre-Trained Language Models: A Survey. https://doi.org/10.48550/ARXIV.2305.16157 arXiv:2305.16157

[126] Jill Jesson, Fiona M Lacey, and Lydia Matheson. 2011. Doing your literature review: Traditional and systematic techniques. (2011).

[127] Juyong Jiang, Fan Wang, Jiasi Shen, Sungju Kim, and Sunghun Kim. 2024. A Survey on Large Language Models for Code Generation. https://doi.org/10.48550/ARXIV.2406.00515 arXiv:2406.00515

[128] Ruili Jiang, Kehai Chen, Xuefeng Bai, Zhixuan He, Juntao Li, Muyun Yang, Tiejun Zhao, Liqiang Nie, and Min Zhang. 2024. A Survey on Human Preference Learning for Large Language Models. https://doi.org/10.48550/ARXIV.2406.11191 arXiv:2406.11191

[129] Xuhui Jiang, Yuxing Tian, Fengrui Hua, Chengjin Xu, Yuanzhuo Wang, and Jian Guo. 2024. A Survey on Large Language Model Hallucination via a Creativity Perspective. https://doi.org/10.48550/ARXIV.2402.06647 arXiv:2402.06647

[130] Yushan Jiang, Zijie Pan, Xikun Zhang, Sahil Garg, Anderson Schneider, Yuriy Nevmyvaka, and Dongjin Song. 2024. Empowering Time Series Analysis with Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.03182 arXiv:2402.03182

[131] Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, and Jiawei Han. 2023. Large Language Models on Graphs: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2312.02783 arXiv:2312.02783

[132] Haibo Jin, Leyang Hu, Xinuo Li, Peiyan Zhang, Chonghan Chen, Jun Zhuang, and Haohan Wang. 2024. JailbreakZoo: Survey, Landscapes, and Horizons in Jailbreaking Large Language and Vision-Language Models. https://doi.org/10.48550/ARXIV.2407.01599 arXiv:2407.01599

[133] Hanlei Jin, Yang Zhang, Dan Meng, Jun Wang, and Jinghua Tan. 2024. A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-based Methods. https://doi.org/10.48550/ARXIV.2403.02901 arXiv:2403.02901

[134] Yizhang Jin, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He, Bo Zhao, Xin Tan, Zhenye Gan, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. 2024. Efficient Multimodal Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2405.10739 arXiv:2405.10739

[135] Zhijing Jin, Nils Heil, Jiarui Liu, Shehzaad Dhuliawala, Yahang Qi, Bernhard Schölkopf, Rada Mihalcea, and Mrinmaya Sachan. 2024. Implicit Personalization in Language Models: A Systematic Study. https://doi.org/10.48550/ARXIV.2405.14808 arXiv:2405.14808

[136] Zhi Jing, Yongye Su, Yikun Han, Bo Yuan, Haiyun Xu, Chunjiang Liu, Kehai Chen, and Min Zhang. 2024. When Large Language Models Meet Vector Databases: A Survey. https://doi.org/10.48550/ARXIV.2402.01763 arXiv:2402.01763

[137] Mladjan Jovanovic and Peter Voss. 2024. Towards Incremental Learning in Large Language Models: A Critical Review. https://doi.org/10.48550/ARXIV.2404.18311 arXiv:2404.18311

[138] Christoforos Kachris. 2024. A Survey on Hardware Accelerators for Large Language Models. https://doi.org/10.48550/ARXIV.2401.09890 arXiv:2401.09890

[139] Katikapalli Subramanyam Kalyan. 2024. A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4. *Natural Language Processing Journal* 6 (2024), 100048. https://doi.org/10.1016/J.NLP.2023.100048

[140] Ryo Kamoi, Yusen Zhang, Nan Zhang, Jiawei Han, and Rui Zhang. 2024. When Can Llms Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of Llms. https://doi.org/10.48550/ARXIV.2406.01297 arXiv:2406.01297

[141] Antonia Karamolegkou, Mostafa Abdou, and Anders Søgaard. 2023. Mapping Brains with Language Models: A Survey. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 9748–9762. https://doi.org/10.18653/V1/2023.FINDINGS-ACL.618

[142] Pravneet Kaur, Gautam Siddharth Kashyap, Ankit Kumar, Md. Tabrez Nafis, Sandeep Kumar, and Vikrant Shokeen. 2024. From Text to Transformation: A Comprehensive Review of Large Language Models' Versatility. https://doi.org/10.48550/ARXIV.2402.16142 arXiv:2402.16142

[143] Luoma Ke, Song Tong, Peng Cheng, and Kaiping Peng. 2024. Exploring the Frontiers of Llms in Psychological Applications: A Comprehensive Review. https://doi.org/10.48550/ARXIV.2401.01519 arXiv:2401.01519

[144] Krishnaram Kenthapadi, Mehrnoosh Sameki, and Ankur Taly. 2024. Grounding and Evaluation for Large Language Models: Practical Challenges and Lessons Learned (Survey). In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, Ricardo Baeza-Yates and Francesco Bonchi (Eds.). ACM, Barcelona Spain, 6523–6533. https://doi.org/10.1145/3637528.3671467

[145] Mahsa Khoshnoodi, Vinija Jain, Mingye Gao, Malavika Srikanth, and Aman Chadha. 2024. A Comprehensive Survey of Accelerated Generation Techniques in Large Language Models. https://doi.org/10.48550/ARXIV.2405.13019 arXiv:2405.13019

[146] Yeseung Kim, Dohyun Kim, Jieun Choi, Jisang Park, Nayoung Oh, and Daehyung Park. 2024. A Survey on Integration of Large Language Models with Intelligent Robots. https://doi.org/10.48550/ARXIV.2404.09228 arXiv:2404.09228

[147] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A. Hale. 2023. Personalisation within Bounds: A Risk Taxonomy and Policy Framework for the Alignment of Large Language Models with Personalised Feedback. https://doi.org/10.48550/ARXIV.2303.05453 arXiv:2303.05453

[148] Barbara Kitchenham, Stuart Charters, et al. 2007. Guidelines for performing systematic literature reviews in software engineering. (2007).

[149] Evans Kotei and Ramkumar Thirunavukarasu. 2023. A Systematic Review of Transformer-Based Pre-Trained Language Models through Self-Supervised Learning. *Inf.* 14, 3 (2023), 187. https://doi.org/10.3390/INFO14030187

[150] Zoe Kotti, Rafaila Galanopoulou, and Diomidis Spinellis. 2023. Machine Learning for Software Engineering: A Tertiary Study. *Comput. Surveys* 55, 12 (Dec. 2023), 1–39. https://doi.org/10.1145/3572905 arXiv:2211.09425 [cs]

[151] Labehat Kryeziu and Visar Shehu. 2022. A Survey of Using Unsupervised Learning Techniques in Building Masked Language Models for Low Resource Languages. In *11th Mediterranean Conference on Embedded Computing, MECO 2022, Budva, Montenegro, June 7-10, 2022.* IEEE, Budva, Montenegro, 1–6. https://doi.org/10.1109/MECO55406.2022.9797081

[152] Sanjay Kukreja, Tarun Kumar, Amit Purohit, Abhijit Dasgupta, and Debashis Guha. 2024. A Literature Survey on Open Source Large Language Models. In *Proceedings of the 7th International Conference on Computers in Management and Business, ICCMB 2024, Singapore, January 12-14, 2024*. ACM, Singapore Singapore, 133–143. https://doi.org/10.1145/3647782.3647803

[153] Pranjal Kumar. 2024. Large Language Models (LLMs): Survey, Technical Frameworks, and Future Challenges. *Artificial Intelligence Review* 57, 9 (2024), 260. https://doi.org/10.1007/S10462-024-10888-Y

[154] Sachin Kumar, Vidhisha Balachandran, Lucille Njoo, Antonios Anastasopoulos, and Yulia Tsvetkov. 2023. Language Generation Models Can Cause Harm: So What Can We Do about It? An Actionable Survey. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 3291–3313. https://doi.org/10.18653/V1/2023.EACL-MAIN.241

[155] Surender Suresh Kumar, Mary L. Cummings, and Alexander J. Stimpson. 2024. Strengthening LLM Trust Boundaries: A Survey of Prompt Injection Attacks Surender Suresh Kumar Dr. M.L. Cummings Dr. Alexander Stimpson. In *4th IEEE International Conference on Human-Machine Systems, ICHMS 2024, Toronto, on, Canada, May 15-17, 2024*. IEEE, Toronto, ON, Canada, 1–6. https://doi.org/10.1109/ICHMS59971.2024.10555871

[156] Jinqi Lai, Wensheng Gan, Jiayang Wu, Zhenlian Qi, and Philip S. Yu. 2023. Large Language Models in Law: A Survey. https://doi.org/10.48550/ARXIV.2312.03718 arXiv:2312.03718

[157] Harsh Nishant Lalai, Aashish Anantha Ramakrishnan, Raj Sanjay Shah, and Dongwon Lee. 2024. From Intentions to Techniques: A Comprehensive Taxonomy and Challenges in Text Watermarking for Large Language Models. https://doi.org/10.48550/ARXIV.2406.11106 arXiv:2406.11106

[158] Wei Lan, Guohang He, Mingyang Liu, Qingfeng Chen, Junyue Cao, and Wei Peng. 2024. Transformer-Based Single-Cell Language Model: A Survey. https://doi.org/10.48550/ARXIV.2407.13205 arXiv:2407.13205

[159] Md. Tahmid Rahman Laskar, Sawsan Alqahtani, M. Saiful Bari, Mizanur Rahman, Mohammad Abdullah Matin Khan, Haidar Khan, Israt Jahan, Amran Bhuiyan, Chee Wei Tan, Md. Rizwan Parvez, Enamul Hoque, Shafiq Joty, and Jimmy Xiangji Huang. 2024. A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations. https://doi.org/10.48550/ARXIV.2407.04069 arXiv:2407.04069

[160] Jinsook Lee, Yann Hicke, Renzhe Yu, Christopher Brooks, and René F. Kizilcec. 2024. The Life Cycle of Large Language Models: A Review of Biases in Education. https://doi.org/10.48550/ARXIV.2407.11203 arXiv:2407.11203

[161] Jean Lee, Nicholas Stevens, Soyeon Caren Han, and Minseok Song. 2024. A Survey of Large Language Models in Finance (FinLLMs). https://doi.org/10.48550/ARXIV.2402.02315 arXiv:2402.02315

[162] Nayeon Lee, Yejin Bang, Holy Lovenia, Samuel Cahyawijaya, Wenliang Dai, and Pascale Fung. 2023. Survey of Social Bias in Vision-Language Models. https://doi.org/10.48550/ARXIV.2309.14381 arXiv:2309.14381

[163] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. 2024. LLM Inference Serving: Survey of Recent Advances and Opportunities. https://doi.org/10.48550/ARXIV.2407.12391 arXiv:2407.12391

[164] Dongfang Li, Zetian Sun, Xinshuo Hu, Zhenyu Liu, Ziyang Chen, Baotian Hu, Aiguo Wu, and Min Zhang. 2023. A Survey of Large Language Models Attribution. https://doi.org/10.48550/ARXIV.2311.03731 arXiv:2311.03731

[165] DaiFeng Li and Fan Xu. 2024. Synergizing Knowledge Graphs with Large Language Models: A Comprehensive Review and Future Prospects. https://doi.org/10.48550/ARXIV.2407.18470 arXiv:2407.18470

[166] Haochen Li, Jonathan Leung, and Zhiqi Shen. 2024. Towards Goal-Oriented Large Language Model Prompting: A Survey. https://doi.org/10.48550/ARXIV.2401.14043 arXiv:2401.14043

[167] Junyi Li, Tianyi Tang, Wayne Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. 2024. Pre-Trained Language Models for Text Generation: A Survey. *Acm Computing Surveys* 56, 9 (2024), 230:1–230:39. https://doi.org/10.1145/3649449

[168] Jiawei Li, Yizhe Yang, Yu Bai, Xiaofeng Zhou, Yinghao Li, Huashan Sun, Yuhang Liu, Xingpeng Si, Yuhao Ye, Yixiao Wu, Yiguan Lin, Bin Xu, Ren Bowen, Chong Feng, Yang Gao, and Heyan Huang. 2024. Fundamental Capabilities of Large Language Models and Their Applications in Domain Scenarios: A Survey. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 11116–11141. https://doi.org/10.18653/v1/2024.acl-long.599

[169] Lei Li, Yongfeng Zhang, Dugang Liu, and Li Chen. 2024. Large Language Models for Generative Recommendation: A Survey and Visionary Discussions. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Nicoletta Calzolari, Min-Yen Kan, Véronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (Eds.). ELRA and ICCL, Torino, Italia, 10146–10159.

[170] Lingyao Li, Jiayan Zhou, Zhenxiang Gao, Wenyue Hua, Lizhou Fan, Huizi Yu, Loni Hagen, Yongfeng Zhang, Themistocles L. Assimes, Libby Hemphill, and Siyuan Ma. 2024. A Scoping Review of Using Large Language Models (Llms) to Investigate Electronic Health Records (Ehrs). https://doi.org/10.48550/ARXIV.2405.03066 arXiv:2405.03066

[171] Xinzhe Li. 2024. A Survey on LLM-based Agents: Common Workflows and Reusable LLM-profiled Components. https://doi.org/10.48550/ARXIV.2406.05804 arXiv:2406.05804

[172] Yingji Li, Mengnan Du, Rui Song, Xin Wang, and Ying Wang. 2023. A Survey on Fairness in Large Language Models. https://doi.org/10.48550/ARXIV.2308.10149 arXiv:2308.10149

[173] Yun Li, Kai Katsumata, Ehsan Javanmardi, and Manabu Tsukada. 2024. Large Language Models for Human-like Autonomous Driving: A Survey. https://doi.org/10.48550/ARXIV.2407.19280 arXiv:2407.19280

[174] Yuhan Li, Zhixun Li, Peisong Wang, Jia Li, Xiangguo Sun, Hong Cheng, and Jeffrey Xu Yu. 2023. A Survey of Graph Meets Large Language Model: Progress and Future Directions. https://doi.org/10.48550/ARXIV.2311.12399 arXiv:2311.12399

[175] Yongqi Li, Xinyu Lin, Wenjie Wang, Fuli Feng, Liang Pang, Wenjie Li, Liqiang Nie, Xiangnan He, and Tat-Seng Chua. 2024. A Survey of Generative Search and Recommendation in the Era of Large Language Models. https://doi.org/10.48550/ARXIV.2404.16924 arXiv:2404.16924

[176] Yinheng Li, Shaofei Wang, Han Ding, and Hang Chen. 2023. Large Language Models in Finance: A Survey. In *4th ACM International Conference on AI in Finance, ICAIF 2023, Brooklyn, NY, USA, November 27-29, 2023*. ACM, Brooklyn NY USA, 374–382. https://doi.org/10.1145/3604237.3626869

[177] Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, Rui Kong, Yile Wang, Hanfei Geng, Jian Luan, Xuefeng Jin, Zilong Ye, Guanjing Xiong, Fan Zhang, Xiang Li, Mengwei Xu, Zhijun Li, Peng Li, Yang Liu, Ya-Qin Zhang, and Yunxin Liu. 2024. Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security. https://doi.org/10.48550/ARXIV.2401.05459 arXiv:2401.05459

[178] Zhen Li, Xiaohan Xu, Tao Shen, Can Xu, Jia-Chen Gu, and Chongyang Tao. 2024. Leveraging Large Language Models for NLG Evaluation: A Survey. https://doi.org/10.48550/ARXIV.2401.07103 arXiv:2401.07103

[179] Xun Liang, Shichao Song, Zifan Zheng, Hanyu Wang, Qingchen Yu, Xunkai Li, Rong-Hua Li, Feiyu Xiong, and Zhiyu Li. 2024. Internal Consistency and Self-Feedback in Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2407.14507 arXiv:2407.14507

[180] Chang Liao, Yemin Yu, Yu Mei, and Ying Wei. 2024. From Words to Molecules: A Survey of Large Language Models in Chemistry. https://doi.org/10.48550/ARXIV.2402.01439 arXiv:2402.01439

[181] Alexander Ligthart, Cagatay Catal, and Bedir Tekinerdogan. 2021. Systematic Reviews in Sentiment Analysis: A Tertiary Study. *Artificial Intelligence Review* 54, 7 (Oct. 2021), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

[182] Jianghao Lin, Xinyi Dai, Yunjia Xi, Weiwen Liu, Bo Chen, Xiangyang Li, Chenxu Zhu, Huifeng Guo, Yong Yu, Ruiming Tang, and Weinan Zhang. 2023. How Can Recommender Systems Benefit from Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2306.05817 arXiv:2306.05817

[183] Michael Pin-Chuan Lin, Daniel Chang, Sarah Hall, and Gaganpreet Jhajj. 2024. Preliminary Systematic Review of Open-Source Large Language Models in Education. In *Generative Intelligence and Intelligent Tutoring Systems - 20th International Conference, ITS 2024, Thessaloniki, Greece, June 10-13, 2024, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 14798)*, Angelo Sifaleras and Fuhua Lin (Eds.). Springer, Cham, 68–77. https://doi.org/10.1007/978-3-031-63028-6_6

[184] Zichao Lin, Shuyan Guan, Wending Zhang, Huiyan Zhang, Yugang Li, and Huaping Zhang. 2024. Towards Trustworthy LLMs: A Review on Debiasing and Dehallucinating in Large Language Models. *Artificial Intelligence Review* 57, 9 (2024), 243. https://doi.org/10.1007/S10462-024-10896-Y

[185] Chen Ling, Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, Yun Li, Hejie Cui, Xuchao Zhang, Tianjiao Zhao, Amit Panalkar, Wei Cheng, Haoyu Wang, Yanchi Liu, Zhengzhang Chen, Haifeng Chen, Chris White, Quanquan Gu, Carl Yang, and Liang Zhao. 2023. Beyond One-Model-Fits-All: A Survey of Domain Specialization for Large Language Models. https://doi.org/10.48550/ARXIV.2305.18703 arXiv:2305.18703

[186] Aiwei Liu, Leyi Pan, Yijian Lu, Jingjing Li, Xuming Hu, Lijie Wen, Irwin King, and Philip S. Yu. 2023. A Survey of Text Watermarking in the Era of Large Language Models. https://doi.org/10.48550/ARXIV.2312.07913 arXiv:2312.07913

[187] Daizong Liu, Mingyu Yang, Xiaoye Qu, Pan Zhou, Yu Cheng, and Wei Hu. 2024. A Survey of Attacks on Large Vision-Language Models: Resources, Advances, and Future Trends. https://doi.org/10.48550/ARXIV.2407.07403 arXiv:2407.07403

[188] Frank Weizhen Liu and Chenhui Hu. 2024. Exploring Vulnerabilities and Protections in Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2406.00240 arXiv:2406.00240

[189] Hanchao Liu, Wenyuan Xue, Yifei Chen, Dapeng Chen, Xiutian Zhao, Ke Wang, Liping Hou, Rongjun Li, and Wei Peng. 2024. A Survey on Hallucination in Large Vision-Language Models. https://doi.org/10.48550/ARXIV.2402.00253 arXiv:2402.00253

[190] Lei Liu, Xiaoyan Yang, Junchi Lei, Xiaoyang Liu, Yue Shen, Zhiqiang Zhang, Peng Wei, Jinjie Gu, Zhixuan Chu, Zhan Qin, and Kui Ren. 2024. A Survey on Medical Large Language Models: Technology, Application, Trustworthiness, and Future Directions. https://doi.org/10.48550/ARXIV.2406.03712 arXiv:2406.03712

[191] Peng Liu, Lemei Zhang, and Jon Atle Gulla. 2023. Pre-Train, Prompt, and Recommendation: A Comprehensive Survey of Language Modeling Paradigm Adaptations in Recommender Systems. *Transactions of the Association for*

Computational Linguistics 11 (2023), 1553–1571. https://doi.org/10.1162/TACL_A_00619

[192] Wentao Liu, Hanglei Hu, Jie Zhou, Yuyang Ding, Junsong Li, Jiayi Zeng, Mengliang He, Qin Chen, Bo Jiang, Aimin Zhou, and Liang He. 2023. Mathematical Language Models: A Survey. https://doi.org/10.48550/ARXIV.2312.07622 arXiv:2312.07622

[193] Xiaoxia Liu, Jingyi Wang, Jun Sun, Xiaohan Yuan, Guoliang Dong, Peng Di, Wenhai Wang, and Dongxia Wang. 2023. Prompting Frameworks for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2311.12785 arXiv:2311.12785

[194] Xiaoyu Liu, Paiheng Xu, Junda Wu, Jiaxin Yuan, Yifan Yang, Yuhang Zhou, Fuxiao Liu, Tianrui Guan, Haoliang Wang, Tong Yu, Julian J. McAuley, Wei Ai, and Furong Huang. 2024. Large Language Models and Causal Inference in Collaboration: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2403.09606 arXiv:2403.09606

[195] Yang Liu, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin. 2024. Datasets for Large Language Models: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2402.18041 arXiv:2402.18041

[196] Yiheng Liu, Hao He, Tianle Han, Xu Zhang, Mengyuan Liu, Jiaming Tian, Yutong Zhang, Jiaqi Wang, Xiaohui Gao, Tianyang Zhong, Yi Pan, Shaochen Xu, Zihao Wu, Zhengliang Liu, Xin Zhang, Shu Zhang, Xintao Hu, Tuo Zhang, Ning Qiang, Tianming Liu, and Bao Ge. 2024. Understanding Llms: A Comprehensive Overview from Training to Inference. https://doi.org/10.48550/ARXIV.2401.02038 arXiv:2401.02038

[197] Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. 2023. Trustworthy Llms: A Survey and Guideline for Evaluating Large Language Models' Alignment. https://doi.org/10.48550/ARXIV.2308.05374 arXiv:2308.05374

[198] Zefang Liu. 2024. A Review of Advancements and Applications of Pre-Trained Language Models in Cybersecurity. In *12th International Symposium on Digital Forensics and Security, ISDFS 2024, San Antonio, TX, USA, April 29-30, 2024*. IEEE, San Antonio, TX, USA, 1–10. https://doi.org/10.1109/ISDFS60797.2024.10527236

[199] Lin Long, Rui Wang, Ruixuan Xiao, Junbo Zhao, Xiao Ding, Gang Chen, and Haobo Wang. 2024. On Llms-Driven Synthetic Data Generation, Curation, and Evaluation: A Survey. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and Virtual Meeting, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 11065–11082. https://doi.org/10.18653/v1/2024.findings-acl.658

[200] Siqu Long, Feiqi Cao, Soyeon Caren Han, and Haiqin Yang. 2022. Vision-and-Language Pretrained Models: A Survey. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, Vienna, Austria, 5530–5537. https://doi.org/10.24963/IJCAI.2022/773

[201] Jinliang Lu, Ziliang Pang, Min Xiao, Yaochen Zhu, Rui Xia, and Jiajun Zhang. 2024. Merge, Ensemble, and Cooperate! A Survey on Collaborative Strategies in the Era of Large Language Models. https://doi.org/10.48550/ARXIV.2407.06089 arXiv:2407.06089

[202] Weizheng Lu, Jiaming Zhang, Jing Zhang, and Yueguo Chen. 2024. Large Language Model for Table Processing: A Survey. https://doi.org/10.48550/ARXIV.2402.05121 arXiv:2402.05121

[203] Haoyan Luo and Lucia Specia. 2024. From Understanding to Utilization: A Survey on Explainability for Large Language Models. https://doi.org/10.48550/ARXIV.2401.12874 arXiv:2401.12874

[204] Man Luo, Shrinidhi Kumbhar, Ming Shen, Mihir Parmar, Neeraj Varshney, Pratyay Banerjee, Somak Aditya, and Chitta Baral. 2023. Towards LogiGLUE: A Brief Survey and A Benchmark for Analyzing Logical Reasoning Capabilities of Language Models. https://doi.org/10.48550/ARXIV.2310.00836 arXiv:2310.00836

[205] Man Luo, Xin Xu, Yue Liu, Panupong Pasupat, and Mehran Kazemi. 2024. In-Context Learning with Retrieved Demonstrations for Language Models: A Survey. https://doi.org/10.48550/ARXIV.2401.11624 arXiv:2401.11624

[206] Xudong Luo, Zhiqi Deng, Binxia Yang, and Michael Y. Luo. 2024. Pre-Trained Language Models in Medicine: A Survey. *Artificial Intelligence in Medicine* 154 (2024), 102904. https://doi.org/10.1016/J.ARTMED.2024.102904

[207] Qun Ma, Xiao Xue, Deyu Zhou, Xiangning Yu, Donghua Liu, Xuwen Zhang, Zihan Zhao, Yifan Shen, Peilin Ji, Juanjuan Li, Gang Wang, and Wanpeng Ma. 2024. Computational Experiments Meet Large Language Model Based Agents: A Survey and Perspective. https://doi.org/10.48550/ARXIV.2402.00262 arXiv:2402.00262

[208] Xianzheng Ma, Yash Bhalgat, Brandon Smart, Shuai Chen, Xinghui Li, Jian Ding, Jindong Gu, Dave Zhenyu Chen, Songyou Peng, Jia-Wang Bian, Philip H. S. Torr, Marc Pollefeys, Matthias Nießner, Ian D. Reid, Angel X. Chang, Iro Laina, and Victor Adrian Prisacariu. 2024. When Llms Step into the 3D World: A Survey and Meta-Analysis of 3D Tasks via Multi-Modal Large Language Models. https://doi.org/10.48550/ARXIV.2405.10255 arXiv:2405.10255

[209] Yueen Ma, Zixing Song, Yuzheng Zhuang, Jianye Hao, and Irwin King. 2024. A Survey on Vision-Language-Action Models for Embodied AI. https://doi.org/10.48550/ARXIV.2405.14093 arXiv:2405.14093

[210] Mohammad A. Makhzan and Hadi Mardani Kamali. 2024. Evolutionary Large Language Models for Hardware Security: A Comparative Survey. https://doi.org/10.48550/ARXIV.2404.16651 arXiv:2404.16651

[211] Amogh Mannekote. 2024. Towards Compositionally Generalizable Semantic Parsing in Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2404.13074 arXiv:2404.13074

[212] Qiheng Mao, Zemin Liu, Chenghao Liu, Zhuo Li, and Jianling Sun. 2024. Advancing Graph Representation Learning with Large Language Models: A Comprehensive Survey of Techniques. https://doi.org/10.48550/ARXIV.2402.05952 arXiv:2402.05952

[213] Yuren Mao, Yuhang Ge, Yijiang Fan, Wenyi Xu, Yu Mi, Zhonghao Hu, and Yunjun Gao. 2024. A Survey on LoRA of Large Language Models. https://doi.org/10.48550/ARXIV.2407.11046 arXiv:2407.11046

[214] Johana Cabrera Medina and Rodrigo Rojas Andrade. 2024. Advancements in Artificial Intelligence for Health: A Rapid Review of AI-based Mental Health Technologies Used in the Age of Large Language Models. In *Bioinformatics and Biomedical Engineering - 11th International Conference, IWBBIO 2024, Meloneras, Gran Canaria, Spain, July 15-17, 2024, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 14848)*, Ignacio Rojas, Francisco Ortuño, Fernando Rojas, Luis Javier Herrera, and Olga Valenzuela (Eds.). Springer, Cham, 318–343. https://doi.org/10.1007/978-3-031-64629-4_26

[215] Grégoire Mialon, Roberto Dessì, Maria Lomeli, Christoforos Nalmpantis, Ramakanth Pasunuru, Roberta Raileanu, Baptiste Rozière, Timo Schick, Jane Dwivedi-Yu, Asli Celikyilmaz, Edouard Grave, Yann LeCun, and Thomas Scialom. 2023. Augmented Language Models: A Survey. *Transactions on Machine Learning Research* 2023 (2023).

[216] Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Hongyi Jin, Tianqi Chen, and Zhihao Jia. 2023. Towards Efficient Generative Large Language Model Serving: A Survey from Algorithms to Systems. https://doi.org/10.48550/ARXIV.2312.15234 arXiv:2312.15234

[217] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2024. Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *Acm Computing Surveys* 56, 2 (2024), 30:1–30:40. https://doi.org/10.1145/3605943

[218] Shervin Minaee, Tomás Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. 2024. Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.06196 arXiv:2402.06196

[219] Atsuyuki Miyai, Jingkang Yang, Jingyang Zhang, Yifei Ming, Yueqian Lin, Qing Yu, Go Irie, Shafiq Joty, Yixuan Li, Hai Li, Ziwei Liu, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2024. Generalized Out-of-Distribution Detection and beyond in Vision Language Model Era: A Survey. https://doi.org/10.48550/ARXIV.2407.21794 arXiv:2407.21794

[220] Andrea Moglia, Konstantinos Georgiou, Pietro Cerveri, Luca T. Mainardi, Richard M. Satava, and Alfred Cuschieri. 2024. Large Language Models in Healthcare: From a Systematic Review on Medical Examinations to a Comparative Analysis on Fundamentals of Robotic Surgery Online Test. *Artificial Intelligence Review* 57, 9 (2024), 231. https://doi.org/10.1007/S10462-024-10849-5

[221] Philipp Mondorf and Barbara Plank. 2024. Beyond Accuracy: Evaluating the Reasoning Behavior of Large Language Models - A Survey. https://doi.org/10.48550/ARXIV.2404.01869 arXiv:2404.01869

[222] Rajiv Movva, Sidhika Balachandar, Kenny Peng, Gabriel Agostini, Nikhil Garg, and Emma Pierson. 2024. Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K arXiv Papers. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 1223–1243. https://doi.org/10.18653/V1/2024.NAACL-LONG.67

[223] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Nick Barnes, and Ajmal Mian. 2023. A Comprehensive Overview of Large Language Models. https://doi.org/10.48550/ARXIV.2307.06435 arXiv:2307.06435

[224] Zabir Al Nazi and Wei Peng. 2024. Large Language Models in Healthcare and Medical Domain: A Review. https://doi.org/10.48550/ARXIV.2401.06775 arXiv:2401.06775

[225] Seth Neel and Peter W. Chang. 2023. Privacy Issues in Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2312.06717 arXiv:2312.06717

[226] Subhash Nerella, Sabyasachi Bandyopadhyay, Jiaqing Zhang, Miguel Contreras, Scott Siegel, Aysegul Bumin, Brandon Silva, Jessica Sena, Benjamin Shickel, Azra Bihorac, Kia Khezeli, and Parisa Rashidi. 2024. Transformers and Large Language Models in Healthcare: A Review. *Artificial Intelligence in Medicine* 154 (2024), 102900. https://doi.org/10.1016/J.ARTMED.2024.102900

[227] Thong Nguyen, Yi Bin, Junbin Xiao, Leigang Qu, Yicong Li, Jay Zhangjie Wu, Cong-Duy Nguyen, See-Kiong Ng, and Anh Tuan Luu. 2024. Video-Language Understanding: A Survey from Model Architecture, Model Training, and Data Perspectives. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and Virtual Meeting, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 3636–3657. https://doi.org/10.18653/v1/2024.findings-acl.217

[228] Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren. 2024. A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges. https://doi.org/10.48550/ARXIV.2406.11903 arXiv:2406.11903

[229] Saurabh Pahune and Manoj Chandrasekharan. 2023. Several Categories of Large Language Models (Llms): A Short Survey. https://doi.org/10.48550/ARXIV.2307.10188 arXiv:2307.10188
[230] Medha Palavalli, Amanda Bertsch, and Matthew R. Gormley. 2024. A Taxonomy for Data Contamination in Large Language Models. https://doi.org/10.48550/ARXIV.2407.08716 arXiv:2407.08716
[231] Liangming Pan, Michael Saxon, Wenda Xu, Deepak Nathani, Xinyi Wang, and William Yang Wang. 2024. Automatically Correcting Large Language Models: Surveying the Landscape of Diverse Automated Correction Strategies. *Transactions of the Association for Computational Linguistics* 12 (2024), 484–506. https://doi.org/10.1162/tacl_a_00660
[232] Seungcheol Park, Jaehyeon Choi, Sojin Lee, and U Kang. 2024. A Comprehensive Survey of Compression Algorithms for Language Models. https://doi.org/10.48550/ARXIV.2401.15347 arXiv:2401.15347
[233] Ye-Jean Park, Abhinav Pillai, Jiawen Deng, Eddie Guo, Mehul Gupta, Mike Paget, and Christopher Naugler. 2024. Assessing the Research Landscape and Clinical Utility of Large Language Models: A Scoping Review. *BMC Medical Informatics Decis. Mak.* 24, 1 (2024), 72. https://doi.org/10.1186/S12911-024-02459-6
[234] Saurav Pawar, S. M. Towhidul Islam Tonmoy, S. M. Mehedi Zaman, Vinija Jain, Aman Chadha, and Amitava Das. 2024. The What, Why, and How of Context Length Extension Techniques in Large Language Models - A Detailed Survey. https://doi.org/10.48550/ARXIV.2401.07872 arXiv:2401.07872
[235] Ji-Lun Peng, Sijia Cheng, Egil Diau, Yung-Yu Shih, Po-Heng Chen, Yen-Ting Lin, and Yun-Nung Chen. 2024. A Survey of Useful LLM Evaluation. https://doi.org/10.48550/ARXIV.2406.00936 arXiv:2406.00936
[236] Michal Perelkiewicz and Rafal Poswiata. 2024. A Review of the Challenges with Massive Web-Mined Corpora Used in Large Language Models Pre-Training. https://doi.org/10.48550/ARXIV.2407.07630 arXiv:2407.07630
[237] Francesco Periti and Stefano Montanelli. 2024. Lexical Semantic Change through Large Language Models: A Survey. *Acm Computing Surveys* 56, 11 (2024), 282:1–282:38. https://doi.org/10.1145/3672393
[238] Andreas Pester, Ahmed Tammaa, Christian Gütl, Alexander Steinmaurer, and Samir Abou El-Seoud. 2024. Conversational Agents, Virtual Worlds, and beyond: A Review of Large Language Models Enabling Immersive Learning. In *IEEE Global Engineering Education Conference, EDUCON 2024, Kos Island, Greece, May 8-11, 2024*. IEEE, Kos Island, Greece, 1–6. https://doi.org/10.1109/EDUCON60312.2024.10578895
[239] Fred Philippy, Siwen Guo, and Shohreh Haddadan. 2023. Towards a Common Understanding of Contributing Factors for Cross-Lingual Transfer in Multilingual Language Models: A Review. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 5877–5891. https://doi.org/10.18653/V1/2023.ACL-LONG.323
[240] Farman Ali Pirzado, Awais Ahmed, Román A. Mendoza-Urdiales, and Hugo Terashima-Marín. 2024. Navigating the Pitfalls: Analyzing the Behavior of Llms as a Coding Assistant for Computer Science Students - A Systematic Review of the Literature. *IEEE access : practical innovations, open solutions* 12 (2024), 112605–112625. https://doi.org/10.1109/ACCESS.2024.3443621
[241] Aske Plaat, Annie Wong, Suzan Verberne, Joost Broekens, Niki van Stein, and Thomas Bäck. 2024. Reasoning with Large Language Models, a Survey. https://doi.org/10.48550/ARXIV.2407.11511 arXiv:2407.11511
[242] Shushanta Pudasaini, Luis Miralles-Pechuán, David Lillis, and Marisa Llorens-Salvador. 2024. Survey on Plagiarism Detection in Large Language Models: The Impact of ChatGPT and Gemini on Academic Integrity. https://doi.org/10.48550/ARXIV.2407.13105 arXiv:2407.13105
[243] Shuofei Qiao, Yixin Ou, Ningyu Zhang, Xiang Chen, Yunzhi Yao, Shumin Deng, Chuanqi Tan, Fei Huang, and Huajun Chen. 2023. Reasoning with Language Model Prompting: A Survey. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 5368–5393. https://doi.org/10.18653/V1/2023.ACL-LONG.294
[244] Libo Qin, Qiguang Chen, Xiachong Feng, Yang Wu, Yongheng Zhang, Yinghui Li, Min Li, Wanxiang Che, and Philip S. Yu. 2024. Large Language Models Meet NLP: A Survey. https://doi.org/10.48550/ARXIV.2405.12819 arXiv:2405.12819
[245] Libo Qin, Qiguang Chen, Yuhang Zhou, Zhi Chen, Yinghui Li, Lizi Liao, Min Li, Wanxiang Che, and Philip S. Yu. 2024. Multilingual Large Language Model: A Survey of Resources, Taxonomy and Frontiers. https://doi.org/10.48550/ARXIV.2404.04925 arXiv:2404.04925
[246] Weicong Qin and Zhongxiang Sun. 2024. Exploring the Nexus of Large Language Models and Legal Systems: A Short Survey. https://doi.org/10.48550/ARXIV.2404.00990 arXiv:2404.00990
[247] Zhen Qin, Daoyuan Chen, Wenhao Zhang, Liuyi Yao, Yilun Huang, Bolin Ding, Yaliang Li, and Shuiguang Deng. 2024. The Synergy between Data and Multi-Modal Large Language Models: A Survey from Co-Development Perspective. https://doi.org/10.48550/ARXIV.2407.08583 arXiv:2407.08583
[248] XiPeng Qiu, TianXiang Sun, YiGe Xu, YunFan Shao, Ning Dai, and XuanJing Huang. 2020. Pre-Trained Models for Natural Language Processing: A Survey. *Science China Technological Sciences* 63, 10 (Oct. 2020), 1872–1897. https://doi.org/10.1007/s11431-020-1647-3

[249] Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-Rong Wen. 2024. Tool Learning with Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2405.17935 arXiv:2405.17935
[250] Guanqiao Qu, Qiyuan Chen, Wei Wei, Zheng Lin, Xianhao Chen, and Kaibin Huang. 2024. Mobile Edge Intelligence for Large Language Models: A Contemporary Survey. https://doi.org/10.48550/ARXIV.2407.18921 arXiv:2407.18921
[251] Youyang Qu. 2024. Federated Learning Driven Large Language Models for Swarm Intelligence: A Survey. https://doi.org/10.48550/ARXIV.2406.09831 arXiv:2406.09831
[252] Md. Abdur Rahman. 2023. A Survey on Security and Privacy of Multimodal Llms - Connected Healthcare Perspective. In *IEEE Globecom Workshops 2023, Kuala Lumpur, Malaysia, December 4-8, 2023*. IEEE, Kuala Lumpur, Malaysia, 1807–1812. https://doi.org/10.1109/GCWKSHPS58843.2023.10465035
[253] Daking Rai, Yilun Zhou, Shi Feng, Abulhair Saparov, and Ziyu Yao. 2024. A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models. https://doi.org/10.48550/ARXIV.2407.02646 arXiv:2407.02646
[254] Mohaimenul Azam Khan Raiaan, Md. Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. 2024. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE access : practical innovations, open solutions* 12 (2024), 26839–26874. https://doi.org/10.1109/ACCESS.2024.3365742
[255] Mayk Caldas Ramos, Christopher J. Collison, and Andrew D. White. 2024. A Review of Large Language Models and Autonomous Agents in Chemistry. https://doi.org/10.48550/ARXIV.2407.01603 arXiv:2407.01603
[256] Mathieu Ravaut, Bosheng Ding, Fangkai Jiao, Hailin Chen, Xingxuan Li, Ruochen Zhao, Chengwei Qin, Caiming Xiong, and Shafiq Joty. 2024. How Much Are Llms Contaminated? A Comprehensive Survey and the Llmsanitize Library. https://doi.org/10.48550/ARXIV.2404.00699 arXiv:2404.00699
[257] Qingyang Ren, Zilin Jiang, Jinghan Cao, Sijia Li, Chiqu Li, Yiyang Liu, Shuning Huo, and Tiange He. 2024. A Survey on Fairness of Large Language Models in E-Commerce: Progress, Application, and Challenge. https://doi.org/10.48550/ARXIV.2405.13025 arXiv:2405.13025
[258] Xubin Ren, Jiabin Tang, Dawei Yin, Nitesh V. Chawla, and Chao Huang. 2024. A Survey of Large Language Models for Graphs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, Ricardo Baeza-Yates and Francesco Bonchi (Eds.). ACM, Barcelona Spain, 6616–6626. https://doi.org/10.1145/3637528.3671460
[259] David Restrepo, Chenwei Wu, Constanza Vásquez-Venegas, João Matos, Jack Gallifant, and Luis Filipe. 2024. Analyzing Diversity in Healthcare LLM Research: A Scientometric Perspective. https://doi.org/10.48550/ARXIV.2406.13152 arXiv:2406.13152
[260] Zhyar Rzgar K. Rostam, Sándor Szénási, and Gábor Kertész. 2024. Achieving Peak Performance for Large Language Models: A Systematic Review. *IEEE access : practical innovations, open solutions* 12 (2024), 96017–96050. https://doi.org/10.1109/ACCESS.2024.3424945
[261] Paul Röttger, Fabio Pernisi, Bertie Vidgen, and Dirk Hovy. 2024. SafetyPrompts: A Systematic Review of Open Datasets for Evaluating and Improving Large Language Model Safety. https://doi.org/10.48550/ARXIV.2404.05399 arXiv:2404.05399
[262] Tara Safavi and Danai Koutra. 2021. Relational World Knowledge Representation in Contextual Language Models: A Review. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 1053–1067. https://doi.org/10.18653/V1/2021.EMNLP-MAIN.81
[263] Chandan Kumar Sah, Xiaoli Lian, and Muhammad Mirajul Islam. 2024. Unveiling Bias in Fairness Evaluations of Large Language Models: A Critical Literature Review of Music and Movie Recommendation Systems. https://doi.org/10.48550/ARXIV.2401.04057 arXiv:2401.04057
[264] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. https://doi.org/10.48550/ARXIV.2402.07927 arXiv:2402.07927
[265] Rahul Manohar Samant, Mrinal Rahul Bachute, Shilpa Gite, and Ketan Kotecha. 2022. Framework for Deep Learning-Based Language Models Using Multi-Task Learning in Natural Language Understanding: A Systematic Literature Review and Future Directions. *IEEE access : practical innovations, open solutions* 10 (2022), 17078–17097. https://doi.org/10.1109/ACCESS.2022.3149798
[266] Christian A. Schiller. 2024. The Human Factor in Detecting Errors of Large Language Models: A Systematic Literature Review and Future Research Directions. https://doi.org/10.48550/ARXIV.2403.09743 arXiv:2403.09743
[267] Agnia Sergeyuk, Sergey Titov, and Maliheh Izadi. 2024. In-IDE Human-AI Experience in the Era of Large Language Models; A Literature Review. https://doi.org/10.48550/ARXIV.2401.10739 arXiv:2401.10739
[268] Wenbo Shang and Xin Huang. 2024. A Survey of Large Language Models on Generative Graph Analytics: Query, Learning, and Applications. https://doi.org/10.48550/ARXIV.2404.14809 arXiv:2404.14809

[269] Erfan Shayegani, Md Abdullah Al Mamun, Yu Fu, Pedram Zaree, Yue Dong, and Nael B. Abu-Ghazaleh. 2023. Survey of Vulnerabilities in Large Language Models Revealed by Adversarial Attacks. https://doi.org/10.48550/ARXIV.2310.10844 arXiv:2310.10844

[270] Xinyu She, Yue Liu, Yanjie Zhao, Yiling He, Li Li, Chakkrit Tantithamthavorn, Zhan Qin, and Haoyu Wang. 2023. Pitfalls in Language Models for Code Intelligence: A Taxonomy and Survey. https://doi.org/10.48550/ARXIV.2310.17903 arXiv:2310.17903

[271] Tianhao Shen, Renren Jin, Yufei Huang, Chuang Liu, Weilong Dong, Zishan Guo, Xinwei Wu, Yan Liu, and Deyi Xiong. 2023. Large Language Model Alignment: A Survey. https://doi.org/10.48550/ARXIV.2309.15025 arXiv:2309.15025

[272] Haizhou Shi, Zihao Xu, Hengyi Wang, Weiyi Qin, Wenyuan Wang, Yibin Wang, and Hao Wang. 2024. Continual Learning of Large Language Models: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2404.16789 arXiv:2404.16789

[273] Liang Shi, Zhengju Tang, and Zhi Yang. 2024. A Survey on Employing Large Language Models for Text-to-SQL Tasks. https://doi.org/10.48550/ARXIV.2407.15186 arXiv:2407.15186

[274] Luohe Shi, Hongyi Zhang, Yao Yao, Zuchao Li, and Hai Zhao. 2024. Keep the Cost down: A Review on Methods to Optimize LLM' s KV-cache Consumption. https://doi.org/10.48550/ARXIV.2407.18003 arXiv:2407.18003

[275] Tongyue Shi, Jun Ma, Zihan Yu, Haowei Xu, Minqi Xiong, Meirong Xiao, Yilin Li, Huiying Zhao, and Guilan Kong. 2024. Stochastic Parrots or ICU Experts? Large Language Models in Critical Care Medicine: A Scoping Review. https://doi.org/10.48550/ARXIV.2407.19256 arXiv:2407.19256

[276] Zhonghao Shi, Ellen Landrum, Amy O'Connell, Mina J. Kian, Leticia Pinto-Alva, Kaleen Shrestha, Xiaoyuan Zhu, and Maja J. Mataric. 2024. How Can Large Language Models Enable Better Socially Assistive Human-Robot Interaction: A Brief Survey. In *Proceedings of the AAAI Symposium Series*, Ron P. A. Petrick and Christopher W. Geib (Eds.), Vol. 3. AAAI Press, Stanford, California, 401–404. https://doi.org/10.1609/AAAISS.V3I1.31245

[277] Mohamed R. Shoaib, Heba M. Emara, and Jun Zhao. 2024. A Survey on the Applications of Frontier AI, Foundation Models, and Large Language Models to Intelligent Transportation Systems. https://doi.org/10.48550/ARXIV.2401.06831 arXiv:2401.06831

[278] Gabriel Simmons and Christopher Hare. 2023. Large Language Models as Subpopulation Representative Models: A Review. https://doi.org/10.48550/ARXIV.2310.17888 arXiv:2310.17888

[279] Victoria Smith, Ali Shahin Shamsabadi, Carolyn Ashurst, and Adrian Weller. 2023. Identifying and Mitigating Privacy Risks Stemming from Language Models: A Survey. https://doi.org/10.48550/ARXIV.2310.01424 arXiv:2310.01424

[280] Mingyang Song, Yi Feng, and Liping Jing. 2023. A Survey on Recent Advances in Keyphrase Extraction from Pre-Trained Language Models. In *Findings of the Association for Computational Linguistics: EACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 2108–2119. https://doi.org/10.18653/V1/2023.FINDINGS-EACL.161

[281] Shezheng Song, Xiaopeng Li, and Shasha Li. 2023. How to Bridge the Gap between Modalities: A Comprehensive Survey on Multimodal Large Language Model. https://doi.org/10.48550/ARXIV.2311.07594 arXiv:2311.07594

[282] Akchay Srivastava and Atif Memon. 2024. Toward Robust Evaluation: A Comprehensive Taxonomy of Datasets and Metrics for Open Domain Question Answering in the Era of Large Language Models. *IEEE access : practical innovations, open solutions* 12 (2024), 117483–117503. https://doi.org/10.1109/ACCESS.2024.3446854

[283] Kalahasti Ganesh Srivatsa, Sabyasachi Mukhopadhyay, Ganesh Katrapati, and Manish Shrivastava. 2024. A Survey of Using Large Language Models for Generating Infrastructure as Code. https://doi.org/10.48550/ARXIV.2404.00227 arXiv:2404.00227

[284] Jing Su, Chufeng Jiang, Xin Jin, Yuxin Qiao, Tingsong Xiao, Hongda Ma, Rong Wei, Zhi Jing, Jiajun Xu, and Junhong Lin. 2024. Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review. https://doi.org/10.48550/ARXIV.2402.10350 arXiv:2402.10350

[285] Guillem García Subies, Álvaro Barbero Jiménez, and Paloma Martínez Fernández. 2023. A Survey of Spanish Clinical Language Models. https://doi.org/10.48550/ARXIV.2308.02199 arXiv:2308.02199

[286] Kaili Sun, Xudong Luo, and Michael Y. Luo. 2022. A Survey of Pretrained Language Models. In *Knowledge Science, Engineering and Management - 15th International Conference, KSEM 2022, Singapore, August 6-8, 2022, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 13369)*, Gérard Memmi, Baijian Yang, Linghe Kong, Tianwei Zhang, and Meikang Qiu (Eds.). Springer, Cham, 442–456. https://doi.org/10.1007/978-3-031-10986-7_36

[287] Kaili Sun, Xudong Luo, and Michael Y. Luo. 2022. A Survey of Sentiment Analysis Based on Pretrained Language Models. In *34th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2022, Macao, China, October 31 - November 2, 2022*, Marek Z. Reformat, Du Zhang, and Nikolaos G. Bourbakis (Eds.). IEEE, Macao, China, 1239–1244. https://doi.org/10.1109/ICTAI56018.2022.00188

[288] Zhongxiang Sun. 2023. A Short Survey of Viewing Large Language Models in Legal Aspect. https://doi.org/10.48550/ARXIV.2303.09136 arXiv:2303.09136

[289] Penny Sweetser. 2024. Large Language Models and Video Games: A Preliminary Scoping Review. In *ACM Conversational User Interfaces 2024, CUI 2024, Luxembourg, July 8-10, 2024*, Mateusz Dubiel, Luis A. Leiva, Johanne Trippas, Joel E. Fischer, and Ilaria Torre (Eds.). ACM, Luxembourg Luxembourg, 45. https://doi.org/10.1145/3640794.3665582

[290] Ayesha Ayub Syed, Ford Lumban Gaol, Alfred Boediman, Tokuro Matsuo, and Widodo Budiharto. 2022. A Survey of Abstractive Text Summarization Utilising Pretrained Language Models. In *Intelligent Information and Database Systems - 14th Asian Conference, ACIIDS 2022, Ho Chi Minh City, Vietnam, November 28-30, 2022, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 13757)*, Ngoc Thanh Nguyen, Tien Khoa Tran, Ualsher Tukeyev, Tzung-Pei Hong, Bogdan Trawinski, and Edward Szczerbicki (Eds.). Springer, Cham, 532–544. https://doi.org/10.1007/978-3-031-21743-2_42

[291] Zhen Tan, Alimohammad Beigi, Song Wang, Ruocheng Guo, Amrita Bhattacharjee, Bohan Jiang, Mansooreh Karami, Jundong Li, Lu Cheng, and Huan Liu. 2024. Large Language Models for Data Annotation: A Survey. https://doi.org/10.48550/ARXIV.2402.13446 arXiv:2402.13446

[292] Zhaoxuan Tan and Meng Jiang. 2023. User Modeling in the Era of Large Language Models: Current Research and Future Directions. *IEEE Data Eng. Bull.* 46, 4 (2023), 57–96.

[293] Yunlong Tang, Jing Bi, Siting Xu, Luchuan Song, Susan Liang, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, Ali Vosoughi, Chao Huang, Zeliang Zhang, Feng Zheng, Jianguo Zhang, Ping Luo, Jiebo Luo, and Chenliang Xu. 2023. Video Understanding with Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2312.17432 arXiv:2312.17432

[294] Zhengwei Tao, Ting-En Lin, Xiancai Chen, Hangyu Li, Yuchuan Wu, Yongbin Li, Zhi Jin, Fei Huang, Dacheng Tao, and Jingren Zhou. 2024. A Survey on Self-Evolution of Large Language Models. https://doi.org/10.48550/ARXIV.2404.14387 arXiv:2404.14387

[295] S. M. Towhidul Islam Tonmoy, S. M. Mehedi Zaman, Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha, and Amitava Das. 2024. A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models. https://doi.org/10.48550/ARXIV.2401.01313 arXiv:2401.01313

[296] Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. Two Tales of Persona in Llms: A Survey of Role-Playing and Personalization. https://doi.org/10.48550/ARXIV.2406.01171 arXiv:2406.01171

[297] Sean Tucker. 2024. A Systematic Review of Geospatial Location Embedding Approaches in Large Language Models: A Path to Spatial AI Systems. https://doi.org/10.48550/ARXIV.2401.10279 arXiv:2401.10279

[298] Muhammad Umair, Tangina Sultana, and Young-Koo Lee. 2024. Pre-Trained Language Models for Keyphrase Prediction: A Review. *ICT Express* 10, 4 (2024), 871–890. https://doi.org/10.1016/J.ICTE.2024.05.015

[299] Ashok Urlana, Charaka Vinayak Kumar, Ajeet Kumar Singh, Bala Mallikarjunarao Garlapati, Srinivasa Rao Chalamala, and Rahul Mishra. 2024. LLMs with Industrial Lens: Deciphering the Challenges and Prospects - A Survey. https://doi.org/10.48550/ARXIV.2402.14558 arXiv:2402.14558

[300] Arpita Vats, Vinija Jain, Rahul Raja, and Aman Chadha. 2024. Exploring the Impact of Large Language Models on Recommender Systems: An Extensive Review. https://doi.org/10.48550/ARXIV.2402.18590 arXiv:2402.18590

[301] Mayank Vatsa, Anubhooti Jain, and Richa Singh. 2024. Adventures of Trustworthy Vision-Language Models: A Survey. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada (AAAI'24/IAAI'24/EAAI'24, Vol. 38)*, Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (Eds.). AAAI Press, Vancouver, Canada, 22658–22659. https://doi.org/10.1609/AAAI.V38I20.30275

[302] Shubham Vatsal and Harsh Dubey. 2024. A Survey of Prompt Engineering Methods in Large Language Models for Different NLP Tasks. https://doi.org/10.48550/ARXIV.2407.12994 arXiv:2407.12994

[303] Ivan Vykopal, Matús Pikuliak, Simon Ostermann, and Marián Simko. 2024. Generative Large Language Models in Automated Fact-Checking: A Survey. https://doi.org/10.48550/ARXIV.2407.02351 arXiv:2407.02351

[304] Guangya Wan, Yuqi Wu, Mengxuan Hu, Zhixuan Chu, and Sheng Li. 2024. Bridging Causal Discovery and Large Language Models: A Comprehensive Survey of Integrative Approaches and Future Directions. https://doi.org/10.48550/ARXIV.2402.11068 arXiv:2402.11068

[305] Zhongwei Wan, Xin Wang, Che Liu, Samiul Alam, Yu Zheng, Jiachen Liu, Zhongnan Qu, Shen Yan, Yi Zhu, Quanlu Zhang, Mosharaf Chowdhury, and Mi Zhang. 2024. Efficient Large Language Models: A Survey. *Transactions on Machine Learning Research* 2024 (2024).

[306] Benyou Wang, Qianqian Xie, Jiahuan Pei, Zhihong Chen, Prayag Tiwari, Zhao Li, and Jie Fu. 2024. Pre-Trained Language Models in Biomedical Domain: A Systematic Survey. *Acm Computing Surveys* 56, 3 (2024), 55:1–55:52. https://doi.org/10.1145/3611651

[307] Cunxiang Wang, Xiaoze Liu, Yuanhao Yue, Xiangru Tang, Tianhang Zhang, Cheng Jiayang, Yunzhi Yao, Wenyang Gao, Xuming Hu, Zehan Qi, Yidong Wang, Linyi Yang, Jindong Wang, Xing Xie, Zheng Zhang, and Yue Zhang. 2023. Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity. https:

//doi.org/10.48550/arXiv.2310.07521 arXiv:2310.07521
[308] Can Wang, Bolin Zhang, Dianbo Sui, Zhiying Tu, Xiaoyu Liu, and Jiabao Kang. 2024. A Survey on Effective Invocation Methods of Massive LLM Services. https://doi.org/10.48550/ARXIV.2402.03408 arXiv:2402.03408
[309] Hongru Wang, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. 2023. A Survey of the Evolution of Language Model-Based Dialogue Systems. https://doi.org/10.48550/ARXIV.2311.16789 arXiv:2311.16789
[310] Junjie Wang, Yuchao Huang, Chunyang Chen, Zhe Liu, Song Wang, and Qing Wang. 2024. Software Testing with Large Language Models: Survey, Landscape, and Vision. *IEEE Transactions on Software Engineering* 50, 4 (2024), 911–936. https://doi.org/10.1109/TSE.2024.3368208
[311] Jinqiang Wang, Huansheng Ning, Yi Peng, Qikai Wei, Daniel Tesfai, Wenwei Mao, Tao Zhu, and Runhe Huang. 2024. A Survey on Large Language Models from General Purpose to Medical Applications: Datasets, Methodologies, and Evaluations. https://doi.org/10.48550/ARXIV.2406.10303 arXiv:2406.10303
[312] Jiahao Wang, Bolin Zhang, Qianlong Du, Jiajun Zhang, and Dianhui Chu. 2024. A Survey on Data Selection for LLM Instruction Tuning. https://doi.org/10.48550/ARXIV.2402.05123 arXiv:2402.05123
[313] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Jirong Wen. 2024. A Survey on Large Language Model Based Autonomous Agents. *Frontiers Comput. Sci.* 18, 6 (2024), 186345. https://doi.org/10.1007/S11704-024-40231-1
[314] Lu Wang, Max Song, Rezvaneh Rezapour, Bum Chul Kwon, and Jina Huh-Yoo. 2023. People's Perceptions toward Bias and Related Concepts in Large Language Models: A Systematic Review. https://doi.org/10.48550/ARXIV.2309.14504 arXiv:2309.14504
[315] Mengru Wang, Yunzhi Yao, Ziwen Xu, Shuofei Qiao, Shumin Deng, Peng Wang, Xiang Chen, Jia-Chen Gu, Yong Jiang, Pengjun Xie, Fei Huang, Huajun Chen, and Ningyu Zhang. 2024. Knowledge Mechanisms in Large Language Models: A Survey and Perspective. https://doi.org/10.48550/ARXIV.2407.15017 arXiv:2407.15017
[316] Shen Wang, Tianlong Xu, Hang Li, Chaoli Zhang, Joleen Liang, Jiliang Tang, Philip S. Yu, and Qingsong Wen. 2024. Large Language Models for Education: A Survey and Outlook. https://doi.org/10.48550/ARXIV.2403.18105 arXiv:2403.18105
[317] Shenao Wang, Yanjie Zhao, Xinyi Hou, and Haoyu Wang. 2024. Large Language Model Supply Chain: A Research Agenda. https://doi.org/10.48550/ARXIV.2404.12736 arXiv:2404.12736
[318] Shang Wang, Tianqing Zhu, Bo Liu, Ming Ding, Xu Guo, Dayong Ye, Wanlei Zhou, and Philip S. Yu. 2024. Unique Security and Privacy Threats of Large Language Model: A Comprehensive Survey. https://doi.org/10.48550/ARXIV.2406.07973 arXiv:2406.07973
[319] Song Wang, Yaochen Zhu, Haochen Liu, Zaiyi Zheng, Chen Chen, and Jundong Li. 2024. Knowledge Editing for Large Language Models: A Survey. *Comput. Surveys* 57, 3 (Nov. 2024), 1–37. https://doi.org/10.1145/3698590 arXiv:2310.16218
[320] Wenxiao Wang, Wei Chen, Yicong Luo, Yongliu Long, Zhengkai Lin, Liye Zhang, Binbin Lin, Deng Cai, and Xiaofei He. 2024. Model Compression and Efficient Inference for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.09748 arXiv:2402.09748
[321] Xin Wang, Zirui Chen, Haofen Wang, Leong Hou U, Zhao Li, and Wenbin Guo. 2024. Large Language Model Enhanced Knowledge Representation Learning: A Survey. https://doi.org/10.48550/ARXIV.2407.00936 arXiv:2407.00936
[322] Xuhong Wang, Haoyu Jiang, Yi Yu, Jingru Yu, Yilun Lin, Ping Yi, Yingchun Wang, Yu Qiao, Li Li, and Fei-Yue Wang. 2024. Building Intelligence Identification System via Large Language Model Watermarking: A Survey and Beyond. https://doi.org/10.48550/ARXIV.2407.11100 arXiv:2407.11100
[323] Xindi Wang, Mahsa Salmani, Parsa Omidi, Xiangyu Ren, Mehdi Rezagholizadeh, and Armaghan Eshaghi. 2024. Beyond the Limits: A Survey of Techniques to Extend the Context Length in Large Language Models. https://doi.org/10.48550/ARXIV.2402.02244 arXiv:2402.02244
[324] Yiqi Wang, Wentao Chen, Xiaotian Han, Xudong Lin, Haiteng Zhao, Yongfei Liu, Bohan Zhai, Jianbo Yuan, Quanzeng You, and Hongxia Yang. 2024. Exploring the Reasoning Abilities of Multimodal Large Language Models (Mllms): A Comprehensive Survey on Emerging Trends in Multimodal Reasoning. https://doi.org/10.48550/ARXIV.2401.06805 arXiv:2401.06805
[325] Yufei Wang, Wanjun Zhong, Liangyou Li, Fei Mi, Xingshan Zeng, Wenyong Huang, Lifeng Shang, Xin Jiang, and Qun Liu. 2023. Aligning Large Language Models with Human: A Survey. https://doi.org/10.48550/ARXIV.2307.12966 arXiv:2307.12966
[326] Zhichao Wang, Bin Bi, Shiva Kumar Pentyala, Kiran Ramnath, Sougata Chaudhuri, Shubham Mehrotra, Zixu Zhu, Xiang-Bo Mao, Sitaram Asur, and Na Cheng. 2024. A Comprehensive Survey of LLM Alignment Techniques: RLHF, RLAIF, PPO, DPO and More. https://doi.org/10.48550/ARXIV.2407.16216 arXiv:2407.16216
[327] Zhiruo Wang, Zhoujun Cheng, Hao Zhu, Daniel Fried, and Graham Neubig. 2024. What Are Tools Anyway? A Survey from the Language Model Perspective. https://doi.org/10.48550/ARXIV.2403.15452 arXiv:2403.15452

[328] Zige Wang, Wanjun Zhong, Yufei Wang, Qi Zhu, Fei Mi, Baojun Wang, Lifeng Shang, Xin Jiang, and Qun Liu. 2023. Data Management for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2312.01700 arXiv:2312.01700

[329] Irene Weber. 2024. Large Language Models as Software Components: A Taxonomy for LLM-integrated Applications. https://doi.org/10.48550/ARXIV.2406.10300 arXiv:2406.10300

[330] Chengwei Wei, Yun-Cheng Wang, Bin Wang, and C.-C. Jay Kuo. 2023. An Overview on Language Models: Recent Developments and Outlook. https://doi.org/10.48550/ARXIV.2303.05759 arXiv:2303.05759

[331] Xiaokai Wei, Shen Wang, Dejiao Zhang, Parminder Bhatia, and Andrew O. Arnold. 2021. Knowledge Enhanced Pretrained Language Models: A Compreshensive Survey. https://doi.org/10.48550/arXiv.2110.08455 arXiv:2110.08455

[332] Bingbing Wen, Jihan Yao, Shangbin Feng, Chenjun Xu, Yulia Tsvetkov, Bill Howe, and Lucy Lu Wang. 2024. The Art of Refusal: A Survey of Abstention in Large Language Models. https://doi.org/10.48550/ARXIV.2407.18418 arXiv:2407.18418

[333] Benjue Weng. 2024. Navigating the Landscape of Large Language Models: A Comprehensive Review and Analysis of Paradigms and Fine-Tuning Strategies. https://doi.org/10.48550/ARXIV.2404.09022 arXiv:2404.09022

[334] Christopher Wolters, Xiaoxuan Yang, Ulf Schlichtmann, and Toyotaro Suzumura. 2024. Memory Is All You Need: An Overview of Compute-in-Memory Architectures for Accelerating Large Language Model Inference. https://doi.org/10.48550/ARXIV.2406.08413 arXiv:2406.08413

[335] Michael Wornow, Yizhe Xu, Rahul Thapa, Birju S. Patel, Ethan Steinberg, Scott L. Fleming, Michael A. Pfeffer, Jason A. Fries, and Nigam H. Shah. 2023. The Shaky Foundations of Clinical Foundation Models: A Survey of Large Language Models and Foundation Models for Emrs. https://doi.org/10.48550/ARXIV.2303.12961 arXiv:2303.12961

[336] Haibin Wu, Xuanjun Chen, Yi-Cheng Lin, Kai-Wei Chang, Ho-Lam Chung, Alexander H. Liu, and Hung-yi Lee. 2024. Towards Audio Language Modeling - an Overview. https://doi.org/10.48550/ARXIV.2402.13236 arXiv:2402.13236

[337] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Philip S. Yu. 2023. Multimodal Large Language Models: A Survey. In *IEEE International Conference on Big Data, BigData 2023, Sorrento, Italy, December 15-18, 2023*, Jingrui He, Themis Palpanas, Xiaohua Hu, Alfredo Cuzzocrea, Dejing Dou, Dominik Slezak, Wei Wang, Aleksandra Gruca, Jerry Chun-Wei Lin, and Rakesh Agrawal (Eds.). IEEE, Sorrento, Italy, 2247–2256. https://doi.org/10.1109/BIGDATA59044.2023.10386743

[338] Junchao Wu, Shu Yang, Runzhe Zhan, Yulin Yuan, Derek F. Wong, and Lidia S. Chao. 2023. A Survey on LLM-generated Text Detection: Necessity, Methods, and Future Directions. https://doi.org/10.48550/ARXIV.2310.14724 arXiv:2310.14724

[339] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, and Enhong Chen. 2024. A Survey on Large Language Models for Recommendation. *World Wide Web (WWW)* 27, 5 (2024), 60. https://doi.org/10.1007/S11280-024-01291-2

[340] Qingyang Wu, Ying Xu, Tingsong Xiao, Yunze Xiao, Yitong Li, Tianyang Wang, Yichi Zhang, Shanghai Zhong, Yuwei Zhang, Wei Lu, and Yifan Yang. 2024. Surveying Attitudinal Alignment between Large Language Models vs. Humans towards 17 Sustainable Development Goals. https://doi.org/10.48550/ARXIV.2404.13885 arXiv:2404.13885

[341] Tongtong Wu, Linhao Luo, Yuan-Fang Li, Shirui Pan, Thuy-Trang Vu, and Gholamreza Haffari. 2024. Continual Learning for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2402.01364 arXiv:2402.01364

[342] Xingyu Wu, Sheng-Hao Wu, Jibin Wu, Liang Feng, and Kay Chen Tan. 2024. Evolutionary Computation in the Era of Large Language Model: Survey and Roadmap. https://doi.org/10.48550/ARXIV.2401.10034 arXiv:2401.10034

[343] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. 2023. The Rise and Potential of Large Language Model Based Agents: A Survey. https://doi.org/10.48550/ARXIV.2309.07864 arXiv:2309.07864

[344] Heming Xia, Zhe Yang, Qingxiu Dong, Peiyi Wang, Yongqi Li, Tao Ge, Tianyu Liu, Wenjie Li, and Zhifang Sui. 2024. Unlocking Efficiency in Large Language Model Inference: A Comprehensive Survey of Speculative Decoding. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and Virtual Meeting, August 11-16, 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 7655–7671. https://doi.org/10.18653/v1/2024.findings-acl.456

[345] Haotian Xia, Zhengbang Yang, Yun Zhao, Yuqing Wang, Jingxi Li, Rhys Tracy, Zhuangdi Zhu, Yuan-Fang Wang, Hanjie Chen, and Weining Shen. 2024. Language and Multimodal Models in Sports: A Survey of Datasets and Applications. https://doi.org/10.48550/ARXIV.2406.12252 arXiv:2406.12252

[346] Yu Xia, Rui Wang, Xu Liu, Mingyan Li, Tong Yu, Xiang Chen, Julian J. McAuley, and Shuai Li. 2024. Beyond Chain-of-Thought: A Survey of Chain-of-x Paradigms for Llms. https://doi.org/10.48550/ARXIV.2404.15676 arXiv:2404.15676

[347] Hanguang Xiao, Feizhong Zhou, Xingyue Liu, Tianqi Liu, Zhipeng Li, Xin Liu, and Xiaoxuan Huang. 2024. A Comprehensive Survey of Large Language Models and Multimodal Large Language Models in Medicine. https:

//doi.org/10.48550/ARXIV.2405.08603 arXiv:2405.08603
[348] Qianqian Xie, Zheheng Luo, Benyou Wang, and Sophia Ananiadou. 2023. A Survey on Biomedical Text Summarization with Pre-Trained Language Model. https://doi.org/10.48550/ARXIV.2304.08763 arXiv:2304.08763
[349] Jialu Xing, Jianping Liu, Jian Wang, Lulu Sun, Xi Chen, Xunxun Gu, and Yingfei Wang. 2024. A Survey of Efficient Fine-Tuning Methods for Vision-Language Models - Prompt and Adapter. *Computers & Graphics* 119 (2024), 103885. https://doi.org/10.1016/J.CAG.2024.01.012
[350] Cheng Xu, Shuhao Guan, Derek Greene, and M. Tahar Kechadi. 2024. Benchmark Data Contamination of Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2406.04244 arXiv:2406.04244
[351] Canwen Xu and Julian J. McAuley. 2023. A Survey on Model Compression and Acceleration for Pretrained Language Models. In *Proceedings of the Thirty-seventh AAAI Conference on Artificial Intelligence and Thirty-fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, Brian Williams, Yiling Chen, and Jennifer Neville (Eds.), Vol. 37. AAAI Press, Washington, DC, USA, 10566–10575. https://doi.org/10.1609/AAAI.V37I9.26255
[352] Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, and Enhong Chen. 2023. Large Language Models for Generative Information Extraction: A Survey. https://doi.org/10.48550/ARXIV.2312.17617 arXiv:2312.17617
[353] Hanyi Xu, Wensheng Gan, Zhenlian Qi, Jiayang Wu, and Philip S. Yu. 2024. Large Language Models for Education: A Survey. https://doi.org/10.48550/ARXIV.2405.13001 arXiv:2405.13001
[354] Hanxiang Xu, Shenao Wang, Ningke Li, Kailong Wang, Yanjie Zhao, Kai Chen, Ting Yu, Yang Liu, and Haoyu Wang. 2024. Large Language Models for Cyber Security: A Systematic Literature Review. https://doi.org/10.48550/ARXIV.2405.04760 arXiv:2405.04760
[355] Lingling Xu, Haoran Xie, Si-Zhao Joe Qin, Xiaohui Tao, and Fu Lee Wang. 2023. Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment. https://doi.org/10.48550/ARXIV.2312.12148 arXiv:2312.12148
[356] Mengwei Xu, Wangsong Yin, Dongqi Cai, Rongjie Yi, Daliang Xu, Qipeng Wang, Bingyang Wu, Yihao Zhao, Chen Yang, Shihe Wang, Qiyang Zhang, Zhenyan Lu, Li Zhang, Shangguang Wang, Yuanchun Li, Yunxin Liu, Xin Jin, and Xuanzhe Liu. 2024. A Survey of Resource-Efficient LLM and Multimodal Foundation Models. https://doi.org/10.48550/ARXIV.2401.08092 arXiv:2401.08092
[357] Rongwu Xu, Zehan Qi, Cunxiang Wang, Hongru Wang, Yue Zhang, and Wei Xu. 2024. Knowledge Conflicts for Llms: A Survey. https://doi.org/10.48550/ARXIV.2403.08319 arXiv:2403.08319
[358] Tiange Xu and Fu Zhang. 2020. A Brief Review of Relation Extraction Based on Pre-Trained Language Models. In *Fuzzy Systems and Data Mining VI - Proceedings of FSDM 2020, Virtual Event, November 13-16, 2020 (Frontiers in Artificial Intelligence and Applications, Vol. 331)*, Antonio J. Tallón-Ballesteros (Ed.). IOS Press, 775–789. https://doi.org/10.3233/FAIA200755
[359] Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou. 2024. A Survey on Knowledge Distillation of Large Language Models. https://doi.org/10.48550/ARXIV.2402.13116 arXiv:2402.13116
[360] Yi Xu. 2024. Machine Unlearning for Traditional Models and Large Language Models: A Short Survey. https://doi.org/10.48550/ARXIV.2404.01206 arXiv:2404.01206
[361] Yuemei Xu, Ling Hu, Jiayi Zhao, Zihan Qiu, Yuqi Ye, and Hanwen Gu. 2024. A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias. https://doi.org/10.48550/ARXIV.2404.00929 arXiv:2404.00929
[362] Yichen Xu and Yanqiao Zhu. 2022. A Survey on Pretrained Language Models for Neural Code Intelligence. https://doi.org/10.48550/ARXIV.2212.10079 arXiv:2212.10079
[363] Biwei Yan, Kun Li, Minghui Xu, Yueyan Dong, Yue Zhang, Zhaochun Ren, and Xiuzhen Cheng. 2024. On Protecting the Data Privacy of Large Language Models (Llms): A Survey. https://doi.org/10.48550/ARXIV.2403.05156 arXiv:2403.05156
[364] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martínez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gasevic. 2024. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology* 55, 1 (2024), 90–112. https://doi.org/10.1111/BJET.13370
[365] Chuanpeng Yang, Wang Lu, Yao Zhu, Yidong Wang, Qian Chen, Chenlong Gao, Bingjie Yan, and Yiqiang Chen. 2024. Survey on Knowledge Distillation for Large Language Models: Methods, Evaluation, and Application. https://doi.org/10.48550/ARXIV.2407.01885 arXiv:2407.01885
[366] Haomiao Yang, Kunlan Xiang, Hongwei Li, and Rongxing Lu. 2023. A Comprehensive Overview of Backdoor Attacks in Large Language Models Within Communication Networks. *IEEE Network* 38 (Aug. 2023), 211–218. https://doi.org/10.1109/MNET.2024.3367788 arXiv:2308.14367
[367] Hao Yang, Yanyan Zhao, Yang Wu, Shilong Wang, Tian Zheng, Hongbo Zhang, Wanxiang Che, and Bing Qin. 2024. Large Language Models Meet Text-Centric Multimodal Sentiment Analysis: A Survey. https://doi.org/10.48550/

ARXIV.2406.08068 arXiv:2406.08068
[368] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Ben Hu. 2024. Harnessing the Power of Llms in Practice: A Survey on ChatGPT and Beyond. *ACM Trans. Knowl. Discov. Data* 18, 6 (2024), 160:1–160:32. https://doi.org/10.1145/3649506
[369] Kichang Yang. 2021. Transformer-Based Korean Pretrained Language Models: A Survey on Three Years of Progress. arXiv:2112.03014
[370] Ke Yang, Jiateng Liu, John Wu, Chaoqi Yang, Yi R. Fung, Sha Li, Zixuan Huang, Xu Cao, Xingyao Wang, Yiquan Wang, Heng Ji, and Chengxiang Zhai. 2024. If LLM Is the Wizard, Then Code Is the Wand: A Survey on How Code Empowers Large Language Models to Serve as Intelligent Agents. https://doi.org/10.48550/ARXIV.2401.00812 arXiv:2401.00812
[371] Linying Yang, Vik Shirvaikar, Oscar Clivio, and Fabian Falck. 2024. A Critical Review of Causal Reasoning Benchmarks for Large Language Models. https://doi.org/10.48550/ARXIV.2407.08029 arXiv:2407.08029
[372] Xianjun Yang, Liangming Pan, Xuandong Zhao, Haifeng Chen, Linda R. Petzold, William Yang Wang, and Wei Cheng. 2023. A Survey on Detection of Llms-Generated Content. https://doi.org/10.48550/ARXIV.2310.15654 arXiv:2310.15654
[373] Xiaoxian Yang, Zhifeng Wang, Qi Wang, Ke Wei, Kaiqi Zhang, and Jiangang Shi. 2024. Large Language Models for Automated Q&A Involving Legal Documents: A Survey on Algorithms, Frameworks and Applications. *Int. J. Web Inf. Syst.* 20, 4 (2024), 413–435. https://doi.org/10.1108/IJWIS-12-2023-0256
[374] Yutao Yang, Jie Zhou, Xuanwen Ding, Tianyu Huai, Shunyu Liu, Qin Chen, Liang He, and Yuan Xie. 2024. Recent Advances of Foundation Language Models-Based Continual Learning: A Survey. https://doi.org/10.48550/ARXIV.2405.18653 arXiv:2405.18653
[375] Zhenjie Yang, Xiaosong Jia, Hongyang Li, and Junchi Yan. 2023. LLM4Drive: A Survey of Large Language Models for Autonomous Driving. https://doi.org/10.48550/ARXIV.2311.01043 arXiv:2311.01043
[376] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Eric Sun, and Yue Zhang. 2023. A Survey on Large Language Model (LLM) Security and Privacy: The Good, the Bad, and the Ugly. https://doi.org/10.48550/ARXIV.2312.02003 arXiv:2312.02003
[377] Hongbin Ye, Tong Liu, Aijia Zhang, Wei Hua, and Weiqiang Jia. 2023. Cognitive Mirage: A Review of Hallucinations in Large Language Models. https://doi.org/10.48550/ARXIV.2309.06794 arXiv:2309.06794
[378] Jiexia Ye, Weiqi Zhang, Ke Yi, Yongzi Yu, Ziyue Li, Jia Li, and Fugee Tsung. 2024. A Survey of Time Series Foundation Models: Generalizing Time Series Representation with Large Language Model. https://doi.org/10.48550/ARXIV.2405.02358 arXiv:2405.02358
[379] Sibo Yi, Yule Liu, Zhen Sun, Tianshuo Cong, Xinlei He, Jiaxing Song, Ke Xu, and Qi Li. 2024. Jailbreak Attacks and Defenses against Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2407.04295 arXiv:2407.04295
[380] Zihao Yi, Jiarui Ouyang, Yuwen Liu, Tianhao Liao, Zhe Xu, and Ying Shen. 2024. A Survey on Recent Advances in LLM-based Multi-Turn Dialogue Systems. https://doi.org/10.48550/ARXIV.2402.18013 arXiv:2402.18013
[381] Da Yin, Li Dong, Hao Cheng, Xiaodong Liu, Kai-Wei Chang, Furu Wei, and Jianfeng Gao. 2022. A Survey of Knowledge-Intensive NLP with Pre-Trained Language Models. arXiv:2202.08772
[382] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. 2023. A Survey on Multimodal Large Language Models. https://doi.org/10.48550/ARXIV.2306.13549 arXiv:2306.13549
[383] Paul Youssef, Osman Alperen Koras, Meijie Li, Jörg Schlötterer, and Christin Seifert. 2023. Give Me the Facts! A Survey on Factual Knowledge Probing in Pre-Trained Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 15588–15605. https://doi.org/10.18653/V1/2023.FINDINGS-EMNLP.1043
[384] Huizi Yu, Lizhou Fan, Lingyao Li, Jiayan Zhou, Zihui Ma, Lu Xian, Wenyue Hua, Sijia He, Mingyu Jin, Yongfeng Zhang, Ashvin Gandhi, and Xin Ma. 2024. Large Language Models in Biomedical and Health Informatics: A Bibliometric Review. https://doi.org/10.48550/ARXIV.2403.16303 arXiv:2403.16303
[385] Mingze Yuan, Peng Bao, Jiajia Yuan, Yunhao Shen, Zifan Chen, Yi Xie, Jie Zhao, Yang Chen, Li Zhang, Lin Shen, and Bin Dong. 2023. Large Language Models Illuminate a Progressive Pathway to Artificial Healthcare Assistant: A Review. https://doi.org/10.48550/ARXIV.2311.01918 arXiv:2311.01918
[386] Zhihang Yuan, Yuzhang Shang, Yang Zhou, Zhen Dong, Zhe Zhou, Chenhao Xue, Bingzhe Wu, Zhikai Li, Qingyi Gu, Yong Jae Lee, Yan Yan, Beidi Chen, Guangyu Sun, and Kurt Keutzer. 2024. LLM Inference Unveiled: Survey and Roofline Model Insights. https://doi.org/10.48550/ARXIV.2402.16363 arXiv:2402.16363
[387] Munazza Zaib, Quan Z. Sheng, and Wei Emma Zhang. 2020. A Short Survey of Pre-Trained Language Models for Conversational AI-A New Age in NLP. In *Proceedings of the Australasian Computer Science Week, ACSW 2020, Melbourne, VIC, Australia, February 3-7, 2020*, Prem Prakash Jayaraman, Dimitrios Georgakopoulos, Timos K. Sellis, and Abdur Forkan (Eds.). ACM, Melbourne VIC Australia, 11:1–11:4. https://doi.org/10.1145/3373017.3373028
[388] Daoguang Zan, Bei Chen, Fengji Zhang, Dianjie Lu, Bingchao Wu, Bei Guan, Yongji Wang, and Jian-Guang Lou. 2023. Large Language Models Meet Nl2code: A Survey. In *Proceedings of the 61st Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 7443–7464. https://doi.org/10.18653/V1/2023.ACL-LONG.411

[389] Fanlong Zeng, Wensheng Gan, Yongheng Wang, Ning Liu, and Philip S. Yu. 2023. Large Language Models for Robotics: A Survey. https://doi.org/10.48550/ARXIV.2311.07226 arXiv:2311.07226

[390] Pai Zeng, Zhenyu Ning, Jieru Zhao, Weihao Cui, Mengwei Xu, Liwei Guo, Xusheng Chen, and Yizhou Shan. 2024. The CAP Principle for LLM Serving: A Survey of Long-Context Large Language Model Serving. https://doi.org/10.48550/ARXIV.2405.11299 arXiv:2405.11299

[391] Chen Zhang, Zhuorui Liu, and Dawei Song. 2024. Beyond the Speculative Game: A Survey of Speculative Execution in Large Language Models. https://doi.org/10.48550/ARXIV.2404.14897 arXiv:2404.14897

[392] Dingkai Zhang, Huanran Zheng, Wenjing Yue, and Xiaoling Wang. 2024. Advancing ITS Applications with Llms: A Survey on Traffic Management, Transportation Safety, and Autonomous Driving. In *Rough Sets - International Joint Conference, IJCRS 2024, Halifax, NS, Canada, May 17-20, 2024, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 14840)*, Mengjun Hu, Chris Cornelis, Yan Zhang, Pawan Lingras, Dominik Slezak, and JingTao Yao (Eds.). Springer, Cham, 295–309. https://doi.org/10.1007/978-3-031-65668-2_20

[393] Hanqing Zhang, Haolin Song, Shaoyu Li, Ming Zhou, and Dawei Song. 2024. A Survey of Controllable Text Generation Using Transformer-Based Pre-Trained Language Models. *Acm Computing Surveys* 56, 3 (2024), 64:1–64:37. https://doi.org/10.1145/3617680

[394] Jie Zhang, Haoyu Bu, Hui Wen, Yu Chen, Lun Li, and Hongsong Zhu. 2024. When Llms Meet Cybersecurity: A Systematic Literature Review. https://doi.org/10.48550/ARXIV.2405.03644 arXiv:2405.03644

[395] Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024. Vision-Language Models for Vision Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (2024), 5625–5644. https://doi.org/10.1109/TPAMI.2024.3369699

[396] Liang Zhang and Zhelun Chen. 2023. Opportunities and Challenges of Applying Large Language Models in Building Energy Efficiency and Decarbonization Studies: An Exploratory Overview. https://doi.org/10.48550/ARXIV.2312.11701 arXiv:2312.11701

[397] Lingzhe Zhang, Tong Jia, Mengxi Jia, Yifan Wu, Aiwei Liu, Yong Yang, Zhonghai Wu, Xuming Hu, Philip S. Yu, and Ying Li. 2024. A Survey of Aiops for Failure Management in the Era of Large Language Models. https://doi.org/10.48550/ARXIV.2406.11213 arXiv:2406.11213

[398] Qiang Zhang, Keyang Ding, Tianwen Lyv, Xinda Wang, Qingyu Yin, Yiwen Zhang, Jing Yu, Yuhao Wang, Xiaotong Li, Zhuoyi Xiang, Xiang Zhuang, Zeyuan Wang, Ming Qin, Mengyao Zhang, Jinlu Zhang, Jiyu Cui, Renjun Xu, Hongyang Chen, Xiaohui Fan, Huabin Xing, and Huajun Chen. 2024. Scientific Large Language Models: A Survey on Biological & Chemical Domains. https://doi.org/10.48550/ARXIV.2401.14656 arXiv:2401.14656

[399] Quanjun Zhang, Chunrong Fang, Yang Xie, Yuxiang Ma, Weisong Sun, Yun Yang, and Zhenyu Chen. 2024. A Systematic Literature Review on Large Language Models for Automated Program Repair. https://doi.org/10.48550/ARXIV.2405.01466 arXiv:2405.01466

[400] Quanjun Zhang, Chunrong Fang, Yang Xie, Yaxin Zhang, Yun Yang, Weisong Sun, Shengcheng Yu, and Zhenyu Chen. 2023. A Survey on Large Language Models for Software Engineering. https://doi.org/10.48550/ARXIV.2312.15223 arXiv:2312.15223

[401] Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, and Guoyin Wang. 2023. Instruction Tuning for Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2308.10792 arXiv:2308.10792

[402] Xiyuan Zhang, Ranak Roy Chowdhury, Rajesh K. Gupta, and Jingbo Shang. 2024. Large Language Models for Time Series: A Survey. https://doi.org/10.48550/ARXIV.2402.01801 arXiv:2402.01801

[403] Xuanliang Zhang, Dingzirui Wang, Longxu Dou, Qingfu Zhu, and Wanxiang Che. 2024. A Survey of Table Reasoning with Large Language Models. https://doi.org/10.48550/ARXIV.2402.08259 arXiv:2402.08259

[404] Yu Zhang, Xiusi Chen, Bowen Jin, Sheng Wang, Shuiwang Ji, Wei Wang, and Jiawei Han. 2024. A Comprehensive Survey of Scientific Large Language Models and Their Applications in Scientific Discovery. https://doi.org/10.48550/ARXIV.2406.10833 arXiv:2406.10833

[405] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, Longyue Wang, Anh Tuan Luu, Wei Bi, Freda Shi, and Shuming Shi. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. https://doi.org/10.48550/ARXIV.2309.01219 arXiv:2309.01219

[406] Yue Zhang, Ziqiao Ma, Jialu Li, Yanyuan Qiao, Zun Wang, Joyce Chai, Qi Wu, Mohit Bansal, and Parisa Kordjamshidi. 2024. Vision-and-Language Navigation Today and Tomorrow: A Survey in the Era of Foundation Models. https://doi.org/10.48550/ARXIV.2407.07035 arXiv:2407.07035

[407] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Adrian de Wynter, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. 2024. LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models.

https://doi.org/10.48550/ARXIV.2404.01230 arXiv:2404.01230

[408] Yang Zhang, Yufei Wang, Kai Wang, Quan Z. Sheng, Lina Yao, Adnan Mahmood, Wei Emma Zhang, and Rongying Zhao. 2023. When Large Language Models Meet Citation: A Survey. https://doi.org/10.48550/ARXIV.2309.09727 arXiv:2309.09727

[409] Zeyu Zhang, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Quanyu Dai, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. 2024. A Survey on the Memory Mechanism of Large Language Model Based Agents. https://doi.org/10.48550/ARXIV.2404.13501 arXiv:2404.13501

[410] Ziyin Zhang, Chaoyu Chen, Bingchang Liu, Cong Liao, Zi Gong, Hang Yu, Jianguo Li, and Rui Wang. 2023. A Survey on Language Models for Code. https://doi.org/10.48550/ARXIV.2311.07989 arXiv:2311.07989

[411] Zihan Zhang, Meng Fang, Ling Chen, Mohammad-Reza Namazi-Rad, and Jun Wang. 2023. How Do Large Language Models Capture the Ever-Changing World Knowledge? A Review of Recent Advances. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 8289–8311. https://doi.org/10.18653/V1/2023.EMNLP-MAIN.516

[412] Zijian Zhang, Yujie Sun, Zepu Wang, Yuqi Nie, Xiaobo Ma, Peng Sun, and Ruolin Li. 2024. Large Language Models for Mobility in Transportation Systems: A Survey on Forecasting Tasks. https://doi.org/10.48550/ARXIV.2405.02357 arXiv:2405.02357

[413] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. 2024. Explainability for Large Language Models: A Survey. *ACM Transactions on Intelligent Systems and Technology* 15, 2 (2024), 20:1–20:38. https://doi.org/10.1145/3639372

[414] Pengyu Zhao, Zijian Jin, and Ning Cheng. 2023. An In-Depth Survey of Large Language Model-Based Artificial Intelligence Agents. https://doi.org/10.48550/ARXIV.2309.14365 arXiv:2309.14365

[415] Shuai Zhao, Meihuizi Jia, Zhongliang Guo, Leilei Gan, Jie Fu, Yichao Feng, Fengjun Pan, and Luu Anh Tuan. 2024. A Survey of Backdoor Attacks and Defenses on Large Language Models: Implications for Security Measures. https://doi.org/10.48550/ARXIV.2406.06852 arXiv:2406.06852

[416] Wayne Xin Zhao, Jing Liu, Ruiyang Ren, and Ji-Rong Wen. 2024. Dense Text Retrieval Based on Pretrained Language Models: A Survey. *ACM Trans. Inf. Syst.* 42, 4 (2024), 89:1–89:60. https://doi.org/10.1145/3637870

[417] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. https://doi.org/10.48550/ARXIV.2303.18223 arXiv:2303.18223

[418] Chaoqi Zhen, Yanlei Shang, Xiangyu Liu, Yifei Li, Yong Chen, and Dell Zhang. 2022. A Survey on Knowledge-Enhanced Pre-Trained Language Models. https://doi.org/10.48550/ARXIV.2212.13428 arXiv:2212.13428

[419] Junhao Zheng, Shengjie Qiu, Chengming Shi, and Qianli Ma. 2024. Towards Lifelong Learning of Large Language Models: A Survey. https://doi.org/10.48550/ARXIV.2406.06391 arXiv:2406.06391

[420] Yanxin Zheng, Wensheng Gan, Zefeng Chen, Zhenlian Qi, Qian Liang, and Philip S. Yu. 2024. Large Language Models for Medicine: A Survey. https://doi.org/10.48550/ARXIV.2405.13055 arXiv:2405.13055

[421] Zibin Zheng, Kaiwen Ning, Yanlin Wang, Jingwen Zhang, Dewu Zheng, Mingxi Ye, and Jiachi Chen. 2023. A Survey of Large Language Models for Code: Evolution, Benchmarking, and Future Trends. https://doi.org/10.48550/ARXIV.2311.10372 arXiv:2311.10372

[422] Hongjian Zhou, Boyang Gu, Xinyu Zou, Yiru Li, Sam S. Chen, Peilin Zhou, Junling Liu, Yining Hua, Chengfeng Mao, Xian Wu, Zheng Li, and Fenglin Liu. 2023. A Survey of Large Language Models in Medicine: Progress, Application, and Challenge. https://doi.org/10.48550/ARXIV.2311.05112 arXiv:2311.05112

[423] Hao Zhou, Chengming Hu, Ye Yuan, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, Xue Liu, Charlie Jianzhong Zhang, Xianbin Wang, and Jiangchuan Liu. 2024. Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities. https://doi.org/10.48550/ARXIV.2405.10825 arXiv:2405.10825

[424] Pengyuan Zhou, Lin Wang, Zhi Liu, Yanbin Hao, Pan Hui, Sasu Tarkoma, and Jussi Kangasharju. 2024. A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming. https://doi.org/10.48550/ARXIV.2404.16038 arXiv:2404.16038

[425] Xin Zhou, Sicong Cao, Xiaobing Sun, and David Lo. 2024. Large Language Model for Vulnerability Detection and Repair: Literature Review and the Road Ahead. https://doi.org/10.48550/ARXIV.2404.02525 arXiv:2404.02525

[426] Yue Zhou, Litong Feng, Yiping Ke, Xue Jiang, Junchi Yan, Xue Yang, and Wayne Zhang. 2024. Towards Vision-Language Geo-Foundation Model: A Survey. https://doi.org/10.48550/ARXIV.2406.09385 arXiv:2406.09385

[427] Yuxiang Zhou, Jiazheng Li, Yanzheng Xiang, Hanqi Yan, Lin Gui, and Yulan He. 2023. The Mystery and Fascination of Llms: A Comprehensive Survey on the Interpretation and Analysis of Emergent Abilities. https://doi.org/10.48550/ARXIV.2311.00237 arXiv:2311.00237

[428] Zixuan Zhou, Xuefei Ning, Ke Hong, Tianyu Fu, Jiaming Xu, Shiyao Li, Yuming Lou, Luning Wang, Zhihang Yuan, Xiuhong Li, Shengen Yan, Guohao Dai, Xiao-Ping Zhang, Yuhan Dong, and Yu Wang. 2024. A Survey on Efficient Inference for Large Language Models. https://doi.org/10.48550/ARXIV.2404.14294 arXiv:2404.14294

[429] Hongyan Zhu, Shuai Qin, Min Su, Chengzhi Lin, Anjie Li, and Junfeng Gao. 2024. Harnessing Large Vision and Language Models in Agriculture: A Review. https://doi.org/10.48550/ARXIV.2407.19679 arXiv:2407.19679

[430] Xunyu Zhu, Jian Li, Yong Liu, Can Ma, and Weiping Wang. 2023. A Survey on Model Compression for Large Language Models. https://doi.org/10.48550/ARXIV.2308.07633 arXiv:2308.07633

[431] Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. 2023. Large Language Models for Information Retrieval: A Survey. https://doi.org/10.48550/ARXIV.2308.07107 arXiv:2308.07107

[432] Ziyu Zhuang, Qiguang Chen, Longxuan Ma, Mingda Li, Yi Han, Yushan Qian, Haopeng Bai, Zixian Feng, Weinan Zhang, and Ting Liu. 2023. Through the Lens of Core Competency: Survey on Evaluation of Large Language Models. https://doi.org/10.48550/ARXIV.2308.07902 arXiv:2308.07902

[433] Shi Zong and Jimmy Lin. 2024. Categorical Syllogisms Revisited: A Review of the Logical Reasoning Abilities of Llms for Analyzing Categorical Syllogism. https://doi.org/10.48550/ARXIV.2406.18762 arXiv:2406.18762