

Concept Paper

Not peer-reviewed version

ShanZhiXingYu: Address Factuality Issues in Large Language Models Through Multi-Dimensions

Juanjuan Yao *

Posted Date: 7 February 2025

doi: 10.20944/preprints202502.0533.v1

Keywords: large language models(LLMs); factual information; ShanZhiXingYu



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Concept Paper

ShanZhiXingYu: Address Factuality Issues in Large Language Models Through Multi-Dimensions

Juanjuan Yao

Shanghai Mingpin Medical Data Technology Co., Ltd., China; lxr@dazhuanjia.com

Abstract: Medical large language models(LLMs) possess huge potential in improving the quality of medical services. However, the problem of factual information has become a severe challenge that urgently needs to be overcome and is hindering the development of medical LLMs. Given this situation, this paper systematically introduces the "ShanZhiXingYu", which is committed to resolving eight key factual information problems, including the reliability of all sources of factual information, the real-time update of factual information, the high quality of the complete set of unique training data, the high efficiency of the complete set of reasoning logic algorithms, the full controllability of model effect verification, the interpretability of the professional knowledge base across the entire domain, the all-dimensional precision of personalized diagnosis and treatment, and the guarantee of data security and compliance throughout the whole chain. A series of innovative solutions have comprehensively strengthened the trustworthiness, controllability and interpretability of LLMs, laid a solid foundation and provided a powerful boost for their practical applications in the medical and health field, and thus served users, medical institutions and pharmaceutical enterprises with higher quality, promoting the steady development of the medical and health industry.

Keywords: large language models(LLMs); factual information; ShanZhiXingYu

1. Introduction

At present, the rapid development of the digital economy is leading the healthcare industry through profound transformations [Yeganeh \(2019\)](#). The extensive application of advanced technologies such as big data, artificial intelligence (AI), and the Internet of Things (IoT) has brought unprecedented development opportunities to the medical field. In recent years, the contradiction between the supply and demand of medical resources has become increasingly prominent, and the demand for efficient and precise medical services has become more and more urgent. Meanwhile, the rapid growth of medical data has provided favorable conditions for the development of medical large language models (LLMs) [Wafae et al. \(2024\)](#), and they play a crucial role in the medical industry and represent the key direction of the current development of AI technology [Mistry et al. \(2024\)](#); [Rajpurkar et al. \(2022\)](#). Medical LLMs utilize complex algorithms and abundant healthcare data to provide doctors and patients with more accurate diagnostic suggestions and treatment plans, enhancing the overall medical care experience. Although medical LLMs have demonstrated their remarkable capabilities and potential, they still face some practical challenges, such as the lack of professional knowledge, insufficient real-time information, and possible AI hallucinations. These issues severely limit the reliability of model applications [Wang et al. \(2024\)](#), which is particularly critical in the rigorous medical research and treatment environment. These non-factual or misleading contents are precisely the greatest obstacles to the wide application of LLMs in the medical field where professionalism, precision, and timeliness are the primary principles. Therefore, effectively addressing these practical problems is of utmost importance for the practical application of medical LLMs. To tackle the challenges regarding the factual issues of LLMs, DAZHUANJIA.COM has carried out in-depth cooperation with multiple enterprises and research institutions. By employing innovative technologies, the general-purpose LLM "Shanzhixingyu" and professional services specifically designed for medical institutions and

pharmaceutical and medical device enterprises have been successfully developed. During the construction of the LLM, the factual problems encountered were classified into eight categories, and specific solution strategies were proposed for each category to ensure the trustworthiness, controllability, and interpretability of the LLMs.

2. Key Factual Issues and Solutions

RQ1. The credible sources of factual information.

During the research and development process of large models, it is of vital importance to ensure the reliability of the sources of factual information. As has been confirmed by relevant literature [Singhal et al. \(2023\)](#), the comprehensive collection of standardized data is the fundamental prerequisite for constructing large language models. If the data collection is not comprehensive, a large amount of valid data will be missing. If the screening is not precise, it will lead to information inconsistency. Subsequently, incorrect, misleading information and biases will be generated. These issues will bring huge risks to the construction of AI models.

We established the diversity and reliability of medical data resources. Thousands of scientists and experts, based on the thinking patterns formed through their own learning, research and application, carefully screen hundreds of high-value and highly reliable data sources for each discipline. These data sources include medical textbooks, guidelines, papers and expert consensuses published by institutions such as the World Health Organization (WHO), disease prevention and control centers of various countries like China CDC (Center for Disease Control and Prevention), U.S. CDC, European CDC, as well as national drug regulatory authorities. The literature sources include but are not limited to PubMed, Web of Science Core Collection, the Library of Congress of the United States, the Chinese Hospital Digital Library, the Science and Technology Literature Service System of the Chinese Center for Disease Control and Prevention, and the National Science and Technology Library. All medical materials are subject to a second review by the medical department. This is to ensure their accuracy and reliability.

We further enhances the authority, reliability and effectiveness of data sources. It uses professional data quality analysis modules, literature citation analysis modules and so on. In the data quality analysis module, an evaluation system is established. It is based on the national standard "GB/T 36344 - 2018 Information Technology - Data Quality Evaluation Indicators". This system is used to evaluate the quality of the obtained data. The focus is on assessing internal consistency and completeness. This greatly reduces phenomena such as data not conforming to standard specifications and being contradictory. In the literature citation analysis module, a literature quality assessment model is constructed. It considers the duality of the timeliness of literature sources and the dimension of citation volume. This model is used to filter the cited literature. Thus, it ensures that the constantly increasing literature information sources always remain reliable and effective.

Through the strict guarantee of data sources, a solid foundation has been laid for the high-quality construction of ShanZhiXingYu. This enables it to provide accurate, authoritative and cutting-edge knowledge and services.

RQ2. Timeliness of factual information.

The timeliness of knowledge holds great significance in the medical field, especially when it comes to newly emerging diseases, recurrent diseases, and pandemics. In the medical domain, the smaller the time difference of information, the better. When training LLMs, the absence of real-time updates may heighten the risk of significant discrepancies between the model's output and real-world conditions [El-Bouzaidi and Abdoun \(2023\)](#); [Sallam \(2023\)](#); [Schimke \(2009\)](#); [Zhang et al. \(2024\)](#).

To guarantee the full-time update of factual information, ShanZhiXingYu has conducted in-depth optimizations from two aspects: mechanism and technology. In terms of the mechanism, a comprehensive information collection system has been constructed to ensure 24/7 data updates around the clock. This system is designed to maintain the timeliness and accuracy of the data, keeping it in line with the latest information. At the technical level, a team of leading experts has pioneered a

dynamic perception and editing method. This method sustains the accuracy and timeliness of the large language model during the training process. The dynamic perception module precisely locates the parameters related to common sense, and then the knowledge editing module is utilized for updates. Additionally, the independently developed visual automatic data acquisition platform, working in tandem with the Flink stream processing framework for big data, ensures high reliability. It can effectively handle the challenges of massive high-concurrent accesses and the processing of data in multiple formats from diverse sources.

Through the above-mentioned mechanism construction and technological innovation, the full-time update of the factual information of ShanZhiXingYu has been comprehensively and systematically ensured.

RQ3. High Quality of Specialized Training Data.

In the process of training large models, a vast amount of data is required for support, and the quality of this data is directly related to the performance and accuracy of the model. If there are biases in the training data, errors may occur in the model's output in specific fields; if the training data is inaccurate or incomplete, the model may generate hallucinations and produce unfounded or illogical content [Thirunavukarasu et al. \(2023\)](#). For example, in the medical field, if the training data contains incorrect diagnosis results or incomplete medical record information, the large model may provide incorrect diagnostic suggestions when performing medical question answering or writing medical documents [Cai and Zhu \(2015\)](#); [Xie et al. \(2023\)](#).

Under the framework of bibliometrics methods and the "determinism" concept, We fully utilize AI technology to construct a vast knowledge system. According to clear and detailed annotation criteria, an efficient annotation tool has been developed. It realizes auxiliary labeling through semi-supervised learning based on the expected gradient length strategy, uses text similarity matching for quality inspection, and employs a self-learning mechanism as regular feedback, specifically for annotating medical texts and multimodal data and generating tokens. The usage in recent years shows that this annotation tool ensures the consistency and reliability of medical training data and has the ability to continuously improve data quality and value. A team of thousands of academicians and experts conducts high-quality data annotation based on the above algorithmic tools. In the process of training large basic models, strict data screening techniques are implemented to carefully screen and clean the data, eliminating biased, inaccurate, and incomplete data. At the same time, a data quality monitoring system is established to track changes in data quality in real-time and identify and correct problem data promptly.

Through comprehensive and meticulous operations in multiple aspects such as data collection, screening, quality assessment, and technical annotation, we comprehensively guarantees the high quality of the training data for large models, laying a solid foundation for the efficient and accurate operation of large models.

RQ4. High accuracy for reasoning logic algorithms.

Although existing LLMs are powerful, they often encounter issues regarding algorithmic accuracy. If the training data is skewed or lacks sufficient representativeness, the models may amplify these biases, resulting in unfair or inaccurate outcomes. Overfitting is another challenge. Models with a large number of parameters might learn the noise in the training data instead of the generalized patterns, which limits their performance ability on new data. Meanwhile, large models may face difficulties in deeply understanding the context, which could lead to misinterpretations of sarcasm, subtle meanings or complex contexts, and then cause errors in the model's responses. Moreover, models trained on data from a certain domain may perform poorly when applied to different domains, for example, when transitioning from general web texts to professional fields such as medicine [Ji et al. \(2023\)](#); [Zhou et al. \(2024\)](#).

To improve the accuracy of LLMs, ShanZhiXingYu adopts multiple solutions. Firstly, training with diverse and representative data can reduce biases and enhance the model's generalization ability. Secondly, several regularization techniques are employed to prevent overfitting. Thirdly, external

knowledge sources such as knowledge graphs and semantic databases are integrated, and fine-tuning training is conducted on high-quality datasets for specific tasks. This enables the model to better capture and understand semantic associations, logical structures and implicit information in the text, so that it can make more accurate inferences and responses when facing complex contexts and enhance the model's ability to understand the context. Fourthly, modified RAG (Retrieval-Augmented Generation) framework is introduced, allowing the model to retrieve relevant information from external knowledge bases before generating answers, thereby strengthening the accuracy and timeliness of the output content. Finally, establishing a continuous learning mechanism and a feedback loop enables the model to adapt to new information and improve over time.

On the technical solution level, Dazhuanjia.COM innovatively propose the Adaptive Task-Planning Mixture of Experts (AT-MoE) architecture [Li and Yao \(2024\)](#), aiming to enhance the model's performance in complex tasks, especially in professional fields such as medicine that require high precision. AT-MoE addresses the limitations of traditional Mixture of Experts (MoE) models by introducing a dynamic weight allocation mechanism that optimizes module fusion according to complex task instructions. Firstly, the Parameter-Efficient Fine-Tuning (PEFT) techniques, such as Low-Rank Adaptation (LoRA), are used to train task-specific experts to enhance problem-solving ability and interpretability in specialized domains. Secondly, a layer-by-layer adaptive grouping routing module is trained to optimize the fusion of these expert modules according to the complexity of the task instructions. The routing module conducts overall weight allocation from the dimension of the expert group and then performs local weight normalization adjustments within the group. The AT-MoE architecture improves the traditional MoE by providing a more interpretable and controllable model, especially for complex tasks that require professional knowledge. It allows task-specific learning, which is crucial for fields such as medicine as these fields require domain-specific knowledge. The dynamic weight allocation mechanism enables the model to adapt to different task requirements and ensures optimal task resolution. This architecture is designed to be compatible with any PEFT method, providing flexibility in training and inference. It maintains a balance between specialization and generalization, enabling the model to handle a wide range of tasks while performing well on specific tasks. Through the above technical means, the accuracy and reliability of the ShanZhiXingYu large model algorithm are significantly improved.

RQ5. Controlled process of model verification.

In the effectiveness verification of LLM, ShanZhiXingYu has achieved effective and controllable throughout the process through a series of rigorous and scientific methods, ensuring the high quality and reliability. Relying on the team of academicians and experts, we conduct research on the rules of more than ten thousand diseases and the knowledge of more than a thousand pharmaceutical and medical device products according to authoritative medical knowledge and clinical diagnosis and treatment experience, and innovatively establishes a "three-level verification" system. This system includes three core steps: accuracy verification, effectiveness verification, and feasibility verification. Firstly, the disease and product rule models are verified through typical cases uploaded by researchers to ensure the accuracy. Secondly, a second verification of the disease and product rule models is carried out through the massive case library of DAZHUANJIA.COM to ensure the authenticity. A third verification of the disease and product rule models is conducted through real-world scenario applications to ensure the feasibility. Currently, tens of millions of disease rule models and product knowledge models have all undergone strict three-level verification, ensuring the effectiveness of the data and models from different levels and perspectives.

Meanwhile, model verification is a crucial process for confirming whether the model has achieved the established goals, and selecting appropriate model types and parameters is of great significance for constructing an efficient prediction model. ShanZhiXingYu has proposed an effective solution, that is, based on the three-level verification system and a newly designed annotation framework. The unique application feedback verification system combines semi-supervised learning technology and the adaptive learning methodology, and in combination with the large language model, significantly

reduces the verification cost. Moreover, through close cooperation with the expert team composed of thousands of scientists, we are able to efficiently verify the model performance and promote subsequent rapid optimization.

Through the synergistic effect of the above series of measures and mechanisms, ShanZhiXingYu has comprehensively ensured the controllability and reliability of the data and model effect verification, laying a solid foundation for the efficient and accurate application of the model.

RQ6. Interpretability of Training Data.

The interpretability of the knowledge in LLMs is of great significance, especially in the medical field with extremely high professional requirements. Ensuring the global interpretability of its professional knowledge base is even more crucial. There are certain limitations in the practical application of LLMs. Their knowledge reserves may not comprehensively cover the factual knowledge and rules in various fields. This defect is particularly prominent when facing the medical field with strong professionalism and a complex knowledge system. The medical field contains a vast amount of specific terms, complex concepts, and elaborate knowledge architectures. Due to the lack of comprehensive and in-depth medical factual knowledge and rules, LLMs often have difficulty accurately understanding and processing these professional contents [Ding et al. \(2023\)](#).

To effectively ensure the global interpretability, we have constructed a professional medical knowledge base MedBrain, and taken a series of rigorous and comprehensive measures. 1) Strict control is exerted over the knowledge sources to ensure that all input knowledge is sourced from authoritative and reliable medical literature, classic textbooks, clinical guidelines, and rigorously verified case studies. It is carefully reviewed by an experienced expert team to guarantee the accuracy and timeliness of the knowledge, ensuring the quality of knowledge at the source. 2) In the knowledge storage and processing stage, formal representation and standard coding methods (such as International Classification of Diseases, Systematized Nomenclature of Medicine - Clinical Terms) are adopted to enable the computer to process the knowledge efficiently. At the same time, the consistency and interoperability of data in different systems and applications are maintained to enhance the usability and universality of the knowledge. 3) The rule construction is led by domain experts. Rigorous logical rules are written based on profound professional knowledge and repeatedly verified for their effectiveness through actual cases or simulation tests. They are also regularly reviewed and updated according to the latest research results to ensure that the rules are always in line with the forefront of medical development. 4) To enhance the transparency and traceability of knowledge, each knowledge item is annotated with detailed source information, including precise references, author lists, and the review process. At the same time, the knowledge traceability function is realized, facilitating users to clearly track the evidence chain behind the conclusions, making the application of knowledge well-founded. In addition, the knowledge verification team, which brings together medical experts, clinicians, and medical informatics experts, strictly follows the standardized process for the extraction and verification of knowledge and rules, including comprehensive preliminary sorting, strict clinical rationality review, and rigorous academic theory review. Scientific verification methods, such as retrospective studies and prospective verification, are used to effectively ensure the effectiveness and practicality of the diagnosis rules.

Through the above all-round and multi-level construction and guarantee measures, we have successfully built a medical professional knowledge base with authoritative content, full transparency and interpretability, laying a solid foundation for the efficient and accurate application of large models in the medical field.

RQ7. Personalized Diagnosis on Multi-Dimensional Data.

In the medical field, the individual differences of patients play a decisive role in formulating precise diagnosis and treatment plans [Johnson et al. \(2023\)](#). However, when traditional large models face these professional problems, they often fail to fully consider the complex individual factors such as the unique medical history, lifestyle, and health status of each patient. As a result, when

providing diagnosis and treatment plans, it is difficult for them to precisely fit the actual situation of the individual and mostly remain at the level of general suggestions [Li et al. \(2024\)](#); [Stade et al. \(2024\)](#).

In the current process of medical development, achieving all-dimensional precision in personalized diagnosis and treatment has become a key goal. The establishment of the PerHR (Personal Health Record) and ProHR (Product Holistic Record) databases plays a central role in this. The PerHR refers to the standardized and scientific record of the health process of residents. This record takes the individual health of residents as the core, runs through the entire life process, covers various health-related factors, realizes the dynamic collection of information through multiple channels, and meets the needs of residents and health management information resources. Relying on a huge user group and based on comprehensive personal health data, a PerHR database has been established, containing more than 1,500 personal health dimensions, covering basic information, lifestyle, physical examination indicators, and symptom manifestations. The PerHR is dynamically open and jointly constructed by individuals and doctors. This two-way interaction model greatly facilitates doctors' understanding of patients and helps doctors formulate more accurate diagnosis and treatment plans. It also provides strong support for residents to achieve online disease screening and self-diagnosis.

In terms of the construction of the ProHR, DAZHUANJIA.COM has established close cooperative relationships with more than a hundred pharmaceutical companies. Taking pharmaceutical and medical device products as the core, a multidisciplinary expert team composed of Chinese medicine, Western medicine, pharmacy, and other disciplines has been assembled. By fully integrating the professional knowledge and practical experience of different disciplines, an in-depth analysis of various pharmaceutical and medical device products has been carried out to construct a comprehensive and detailed ProHR database, covering more than 150 product dimensions such as drug ingredients, efficacy, side effects, and applicable populations.

By combining the two components, ShanZhiXingYu can gain a deeper understanding of the individual differences of each patient and fully incorporate rich medical product knowledge. Specifically, when generating diagnosis and treatment suggestions, the large model can conduct a comprehensive and precise analysis and judgment based on the individual health status information in the PerHR and the characteristics and applicability of different medical products in the ProHR, thus tailoring truly personalized diagnosis and treatment suggestions for patients, significantly improving the precision and effectiveness of diagnosis and treatment.

RQ8. Security and Compliance of Data.

In the research and development of large models, the flow of data in various links faces numerous risks. Therefore, the security and compliance of the entire data link are of crucial importance [Warnat-Herresthal et al. \(2021\)](#). On the one hand, data leakage may lead to the acquisition of users' sensitive information by lawbreakers, causing serious losses to users. On the other hand, the tampering of training data or data input may affect the accuracy and reliability of large models. Malicious attackers may tamper with training data or input data, causing large models to generate incorrect output results and thus misleading users. In addition, malicious attacks may also lead to the paralysis of the large model system, affecting the normal operation of business. Meanwhile, the use and processing of data must comply with relevant laws, regulations, policy standards, and ethical norms. Different countries and regions have different legal requirements for data security and privacy protection, requiring large model research and development enterprises to strictly abide by these regulations to ensure the legal use of data. Moreover, the processing of data should also conform to ethical norms to avoid discrimination or unfair treatment of specific groups.

To comprehensively ensure the security and compliance of large models, ShanZhiXingYu has implemented strict security safeguard measures throughout the entire link from data acquisition, cleaning, storage to application:

For data acquisition stage, various methods are adopted to ensure the legality and compliance of data sources. Firstly, all data used for training come from legal channels and strictly comply with relevant laws and regulations. Secondly, differential privacy technology is applied to skillfully add

noise when collecting data, which can effectively protect individual information from being leaked while ensuring the accuracy and effectiveness of data analysis results. Thirdly, sensitive information is anonymized and de-identified through encryption or replacement and other means to ensure that personal identities cannot be directly traced, further strengthening data security.

For data cleaning stage, automated data cleaning tools are utilized, and machine learning algorithms are employed to automatically detect and accurately correct error data, such as filling in missing values and removing outliers, significantly improving data quality. Meanwhile, data consistency checks are carried out through hash verification and other methods to firmly prevent malicious tampering and ensure the consistency and integrity of data. In addition, natural language processing technology is used for sensitive information filtering to identify and remove sensitive words or personal information in the text, effectively avoiding potential risks.

In terms of data storage, encrypted storage technology is adopted. The Advanced Encryption Standard is used to encrypt static data. Even if hard disks are stolen, the data content can hardly be easily read. A distributed storage system is constructed to disperse data among multiple nodes, combined with a redundant backup mechanism, which not only improves the fault tolerance ability of the system but also greatly enhances security. Meanwhile, a role-based access control strategy is implemented, allowing only authorized personnel to access specific data resources, effectively reducing internal threats. In addition, the "Three Redundancies and Three Backups" hybrid cloud platform created by DaZhuanJia.COM is quite ingenious. At the hardware level, multiple servers are redundantly deployed. Data is regularly backed up and stored in different geographical locations to prevent data loss. At the software level, advanced encryption technology and access control mechanisms are applied to protect data privacy. Encryption technology converts data into ciphertext, and the access control mechanism limits data access rights to ensure that only authorized users can access sensitive data. Moreover, the hybrid cloud platform is also equipped with a real-time monitoring and early warning system to promptly discover and respond to security incidents and comprehensively safeguard data security.

For data application stage, ShanZhiXingYu adds invisible but verifiable watermarks to large models to track their usage, effectively preventing unauthorized copying and dissemination. Federated learning technology is adopted to allow multiple participants to jointly train shared models without sharing the original data, effectively protecting the data privacy of all parties. Meanwhile, a continuous monitoring and auditing system is deployed to record all operation logs, conduct regular reviews, and promptly detect and respond to any abnormal behaviors or security incidents.

We also deeply participates in standard drafting and system construction in accordance with laws and regulations such as the "Cybersecurity Review Measures", the "Data Security Law", and the "Personal Information Protection Law". Through this measure, our own practical experience and technological advantages are integrated into the standards, which vigorously promotes the healthy development of the industry. Meanwhile, better understand the requirements of laws and regulations to ensure that its own data processing behaviors comply with the standards. In addition, actively cooperates with industry associations, scientific research institutions, etc. to jointly build a data security management system and an evaluation mechanism, standardize the data processing procedures of enterprises, clarify the security responsibilities and operation specifications of each part, and conduct regular evaluations and audits of the security status to promptly discover potential security risks and make rapid rectifications. Through this multi-level approach, the security and compliance level of the entire data link is effectively ensured, providing a solid and reliable guarantee for the research and development and application of LLMs.

3. Conclusion

ShanZhiXingYu is an advanced general-purpose LLM. Relying on trustworthy artificial intelligence technology and the rich medical knowledge base of MedBrain, it is committed to providing accurate AI services for users, medical institutions, and pharmaceutical and medical device enterprises.

Up to now, this platform has successfully served a user group exceeding tens of millions, thousands of medical institutions, and numerous pharmaceutical and medical device enterprises.

Looking ahead, we will continue to build a comprehensive AI medical solution around ShanZhiXingYu, integrating multidisciplinary and multimodal methods. The aim is to serve hundreds of millions of users, thousands of medical institutions, and pharmaceutical enterprises, injecting new vitality and development momentum into the digital medical industry.

References

- Wafae Abbaoui, Sara Retal, Brahim El Bhiri, Nassim Kharmoum, and Soumia Ziti. 2024. Towards revolutionizing precision healthcare: A systematic literature review of artificial intelligence methods in precision medicine. *Informatics in Medicine Unlocked*, page 101475.
- Li Cai and Yangyong Zhu. 2015. The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14:2–2.
- Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, Zonghan Yang, Yusheng Su, Shengding Hu, Yulin Chen, Chi-Min Chan, Weize Chen, et al. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3):220–235.
- Youssra El Idrissi El-Bouzaidi and Otman Abdoun. 2023. Advances in artificial intelligence for accurate and timely diagnosis of covid-19: A comprehensive review of medical imaging analysis. *Scientific African*, page e01961.
- Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.
- Douglas Johnson, Rachel Goodman, J Patrinely, Cosby Stone, Eli Zimmerman, Rebecca Donald, Sam Chang, Sean Berkowitz, Avni Finn, Eiman Jahangir, et al. 2023. Assessing the accuracy and reliability of ai-generated medical responses: an evaluation of the chat-gpt model. *Research square*.
- Jiajia Li, Zhouyu Guan, Jing Wang, Carol Y Cheung, Yingfeng Zheng, Lee-Ling Lim, Cynthia Ciwei Lim, Paisan Ruamviboonsuk, Rajiv Raman, Leonor Corsino, et al. 2024. Integrated image-based deep learning and language models for primary diabetes care. *Nature medicine*, pages 1–11.
- Xurui Li and Juanjuan Yao. 2024. At-moe: Adaptive task-planning mixture of experts via lora approach. *arXiv preprint arXiv:2410.10896*.
- Hirenkumar Kamleshbhai Mistry, Chirag Mavani, Amit Goswami, and Ripalkumar Patel. 2024. The impact of cloud computing and ai on industry dynamics and competition. *Educational Administration: Theory and Practice*, 30(7):797–804.
- Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. 2022. Ai in health and medicine. *Nature medicine*, 28(1):31–38.
- Malik Sallam. 2023. Chatgpt utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*, volume 11, page 887. MDPI.
- Ingolf Schimke. 2009. Quality and timeliness in medical laboratory testing. *Analytical and bioanalytical chemistry*, 393:1499–1504.
- Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. 2023. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180.
- Elizabeth C Stade, Shannon Wiltsey Stirman, Lyle H Ungar, Cody L Boland, H Andrew Schwartz, David B Yaden, João Sedoc, Robert J DeRubeis, Robb Willer, and Johannes C Eichstaedt. 2024. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, 3(1):12.
- Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine*, 29(8):1930–1940.
- Yuxia Wang, Minghan Wang, Muhammad Arslan Manzoor, Fei Liu, Georgi Georgiev, Rocktim Jyoti Das, and Preslav Nakov. 2024. Factuality of large language models in the year 2024. *arXiv preprint arXiv:2402.02420*.
- Stefanie Warnat-Herresthal, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara, Kristian Händler, Peter Pickkers, N Ahmad Aziz, et al. 2021. Swarm learning for decentralized and confidential clinical machine learning. *Nature*, 594(7862):265–270.

- Sang Michael Xie, Shibani Santurkar, Tengyu Ma, and Percy S Liang. 2023. Data selection for language models via importance resampling. *Advances in Neural Information Processing Systems*, 36:34201–34227.
- Hamid Yeganeh. 2019. An analysis of emerging trends and transformations in global healthcare. *International Journal of Health Governance*, 24(2):169–180.
- Gongbo Zhang, Qiao Jin, Yiliang Zhou, Song Wang, Betina Idnay, Yiming Luo, Elizabeth Park, Jordan G Nestor, Matthew E Spotnitz, Ali Soroush, et al. 2024. Closing the gap between open source and commercial large language models for medical evidence summarization. *npj Digital Medicine*, 7(1):239.
- Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri, and José Hernández-Orallo. 2024. Larger and more instructable language models become less reliable. *Nature*, pages 1–8.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.