

Review

# Towards a potential paradigm shift in health data collection and analysis: Contemporary challenges of Human-Machine interaction

David J. Herzog<sup>1</sup>, Nitsa J. Herzog<sup>2,\*</sup><sup>1</sup> Computer Science, London campus of Ulster, Solent, and Roehampton Universities, London EC1R 4TF, UK<sup>2</sup> Faculty of Engineering and Environment London campus, Northumbria University, London E1 7HT, UK\* **Corresponding author:** Nitsa J. Herzog, [nitsa.herzog@northumbria.ac.uk](mailto:nitsa.herzog@northumbria.ac.uk)

---

## CITATION

Herzog DJ, Herzog NJ. Towards a potential paradigm shift in health data collection and analysis: Contemporary challenges of Human-Machine interaction. *Metaverse*. 2024; 5(1): 2690.  
<https://doi.org/10.54517/m.v5i1.2690>

---

## ARTICLE INFO

Received: 19 April 2024

Accepted: 15 May 2024

Available online: 7 June 2024

---

## COPYRIGHT



Copyright © 2024 by author(s).  
*Metaverse* is published by Asia Pacific Academy of Science Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.  
<https://creativecommons.org/licenses/by/4.0/>

**Abstract:** Industrial Revolution 4.0 transforms healthcare systems. The first three technological revolutions changed the relationship between human and machine interaction due to the exponential growth of the number of machines. The fourth revolution has placed humans in a scenario where heterogeneous data is generated in unprecedented quantity and quality, not only through traditional methods enhanced by digitization, but also through ubiquitous computing, machine-to-machine interactions, and smart environment. The modern cyber-physical space underlines the role of humans in the expanding context of computerization and big data processing. In healthcare, where data collection and analysis particularly depend on human efforts, the disruptive nature of these developments is evident. Adaptation to this process requires deep scrutiny of the trends and recognition of future medical data technologies` evolution. Significant difficulties arise from discrepancies in requirements by healthcare, administrative and technology stakeholders. Black box and grey box decisions made in medical imaging and diagnostic Decision Support Software are often not transparent enough for the professional, social and medico-legal requirements. While Explainable AI proposes a partial solution for AI applications in medicine, the approach has to be wider and multiplex. LLM potential and limitations are also discussed. This paper lists the most significant issues in these topics and describes possible solutions.

**Keywords:** big data; explainable AI; black and white boxes; large language model; healthcare transformation

---

## 1. Introduction

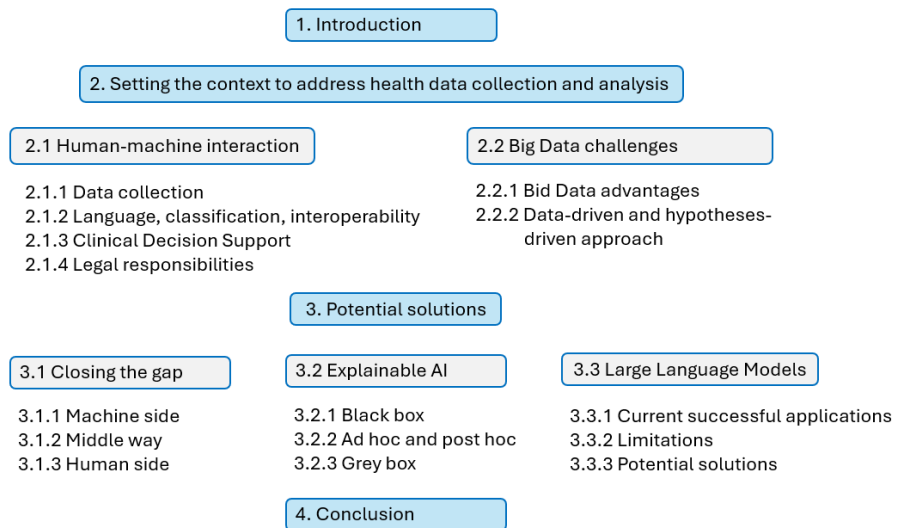
World demographic trends require adjustment of all social institutions. The rapid population growth in expanding economies coexists with lower birth rates and quick ageing in economically developed societies [1]. Healthcare systems have to adapt to the higher demand. Extensive broadening of healthcare services cannot be seen as a primary solution. The fourth technological revolution is projected to be one way to solve the problem of society's growing financial and social healthcare burden [2,3].

New technologies have improved communication, data collection, and management. Artificial intelligence helps analyze voluminous medical data, ubiquitous computing, and Ambient Assisted Living to improve the condition of home-based patients. Telemedicine and robotics connect geographically distant specialists and patients, to name just a few. The SARS-CoV-2 pandemic inadvertently intensified current trends [4].

At the same time, the technological development of healthcare has to overcome not only financial, social and psychological barriers. There are fundamental and methodological problems which cannot be solved by financial and administrative

efforts. Human-Machine interaction specialists are acutely aware of the significant gap in effective synergy [5]. Big data studies are fruitful in medical image processing, laboratory analysis, public health, and some areas of clinical research, but they are somehow less fruitful in the field of general patient-centred clinical healthcare data analysis. The pandemic demonstrated the importance and robustness of prophylaxis on the level of public health by means of quarantine and vaccination. On the personal level, besides prevention, treatment is exhaustively thorough in intensive care, with traditional oxygen, heavy staff workload, and pulmonary and other intensive care protocols of treatment. The highly effective drug for direct treatment has been developed only recently, despite all endeavours with the help of the best technologies [6].

To understand the underlying causes, it is important to disentangle fundamental issues of human-centred technologies from temporary “development” hindrances [5]. Appropriate focus on the right type of solutions allows for proper healthcare progress. **Figure 1**, provided below, demonstrates the structure of the represented topics.



**Figure 1.** The article’s roadmap.

## 2. Setting the context to address health data collection and analysis

There are a number of fundamental and applied questions, which are understandably demanding focus before any solution can be found. Human-machine interaction is reshaped by ubiquitous computing and big data. The main areas in these topics hold keys to the balanced vision and rational estimation of future changes in the area of healthcare data handling.

### 2.1. Human-Machine interaction

HMI was formulated in the 1960th and can be distilled into the necessity to interact between user and machine despite their dissimilarities [7]. There are several areas in the HMI: language disparities between human and technical devices, different types of possible errors and bias, and demand for specific interfaces, humans and machines, to list just a few. Ideally, technologies would allow the closing of the gap between humans and machines by creating perfect interactive systems, but this task

has yet to be fulfilled.

### **2.1.1. Data collection**

Historically, information was collected, recorded, and analyzed by humans only, first at the biological and psychological levels and then at the social and civilization levels [7]. Language, writing, mathematical and statistical methods allowed data preprocessing, classification and analysis. The advance of machines and information theory laid the foundation for the computer era. HMI problem was formulated from the start, data collection was prone to human mistakes and bias. Input into the machine was highly technical and required an appropriate interface, which could create additional obstacles for human operators. Then, regardless of the form of data storage, the problem of human readability arises. Health and biometric information are mostly collected by humans, and the records are still mostly human-readable. Machine intervention requires significant effort to optimize data collection as a prerequisite for machine storage and analysis. The dominant role of humans is inevitably being taken more and more by machines with improvements in data collection, archiving, transmission, and analysis. Human control is important in clinical data handling, but technical assistance is clearly beneficial [8].

Today, healthcare data is potentially heterogeneous and comes not only from primary health records. Secondary data is harvested from Electronic Health Records (EHR). The three basic Vs of big data, volume, velocity, and variety, require a different approach. Behavioural data flows from the Internet and surveillance sources, and more information comes from smartphones, smart health devices and applications, smart homes, and smart environments. Clinical data itself, especially biometric data, is often read from sensors before being recorded by medical staff in EHR [9].

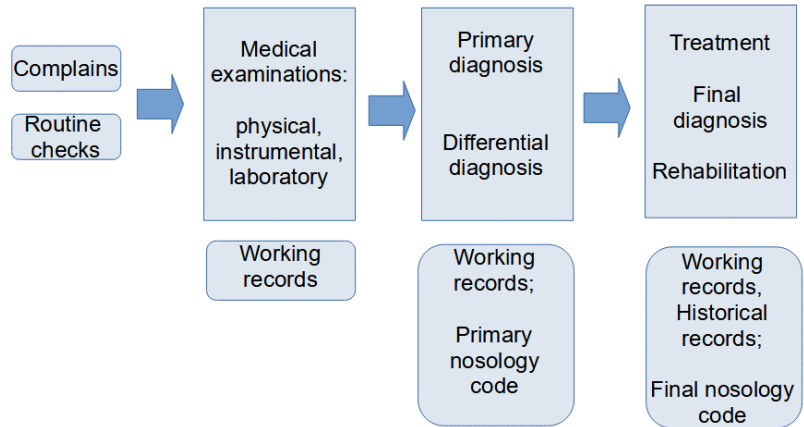
### **2.1.2. Language, classification, interoperability**

There is a well-known fundamental difference between human natural language and machine languages. Automatic data analysis also employs classification techniques. Artificial Intelligence classification of health data significantly differs from the traditional human-based one. Human bias is reflected, for example, in nosology's diseases' classifications. Polysemy in terminology creates confusion even in the different areas of medicine. An unclear definition of the simple term "abdomen" prevents the easy transfer of information [10]. Other healthcare areas, like psychiatry, are known for difficulties in precise classification [11].

Medical nosology classifications historically reflect expert-driven beginnings, objective observations, and further classifications based on symptoms and syndromes. Anatomy, physiology, biometric information, and genetic and biomolecular data today form the basis for more acceptable bottom-up classifications [12]. However, being ontologically more modern, these classifications still lack easy machine readability. **Figure 2** schematically shows the standard medical procedure from obtaining initial information to the final records. From this process, we can see that nosology classifications naturally depend on expert-driven processes.

While the initial trigger can be complaints by patients, relatives or members of the public, the next step is always framed by medical professionals, even when the starting point is a routine examination. After collecting anamnesis, physical, instrumental, and laboratory examinations are performed. The results lead to a primary,

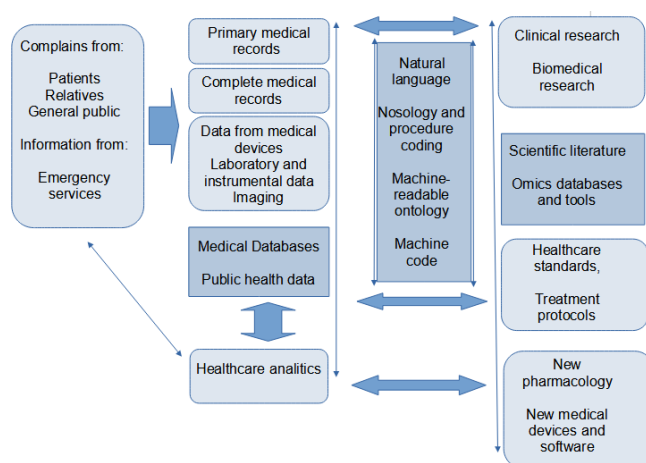
most probable diagnosis. They are registered in medical records. Differential diagnosis may require additional checks. Primary nosology code is added to the records. The treatment can be started at the earliest possible stage or postponed to a later stage. The results of treatment also help to make a more precise diagnosis and final detailed diagnosis. A nosology code is applied for final records and stored in the documentation and database.



**Figure 2.** Standard procedure of diagnosis and treatment, with matching records.

Even top standards differ depending on their classification purpose. The widely used SNOMED-CT terminology is multi-axial, clinically oriented, procedure-oriented, and input-based. The International Classification of Diseases (ICD) is statistically and output-oriented. ICD-10 partially differs in ontology from ICD-11—the last one is closer to SNOMED and multi-axial. However, it does not fully solve the problem of interoperability [13].

**Figure 3** shows simplified data flow in healthcare. The starting point is similar to the scheme in **Figure 2**. Data flow between medical stakeholders' domains is demonstrated by dark blue arrows, while intra-domain data transformation interchange is shown by thin arrows. Light blue blocks reflect generative and transformative nodes. Dark blue blocks show data storage and types of data. We can see complex quasi-horizontal interaction between domains of the general public, healthcare professionals, and biomedical and clinical researchers. The research domain is important for creating treatment standards and protocols, which reflect combined results of analyzed data from the healthcare system and scientific exploratory quasi-horizontal interaction between domains of the general public, healthcare professionals, and the field. Natural language-processed records might be relatively easy to save on the primary level but have to be integrated into machine-readable ontologies on the database level. Analytical results from the databases also cannot be fully machine-adapted because of the inclusion of natural language. The interplay between machine and human readability will exist well in the healthcare system and scientific exploratory quasi-horizontal interaction between domains of the general public, healthcare professionals, and the future.



**Figure 3.** Data flow and the problem of interoperability.

### 2.1.3. Clinical decision support

Clinical Decision Support System is a software application that helps the main stakeholders make optimal decisions. Patients, administrative staff, and healthcare practitioners—each group sees their own benefits in it [14]. For clinicians, the main purpose of CDSS is to optimize the diagnostic procedure. A widely discussed topic is the ability of the automatic system to diagnose and propose treatment without human supervision. However, Clinical Decision Support Systems (CDSS) are widely available. While expert society is aware of its limitations, there is a gap in understanding when we take the viewpoints of different stakeholders. A significant discrepancy exists between technical and healthcare stakeholders.

Technically, there is no obstacle to maximally automatizing the procedure. The patient can contact CDSS before first interacting with clinical staff. If the data stream from the smart environment is informative enough, a system that can point out possible pathology signs and recommend the next step is conceivable.

This approach is more proactive than the contemporary retroactive point-of-care procedure. The process usually starts with subjective complaints by patients or relatives and goes through iterative stages. Knowledge-based CDSS gives significant advantages at every step. However, it is still too early to speak about a more automatic, proactive, highly autonomous system, even with the help of Big Data and Artificial Intelligence.

### 2.1.4. Legal responsibility

Clinical data usually requires approval from several stakeholders for handling. In the case of private entrepreneurship, rules can be stricter or more relaxed, depending on the position of authorities and society. With the entry of Big Data and Artificial Intelligence, the question of responsibility becomes more complex [15].

It is necessary to understand every step in the data handling process clearly. The nature of Big Data itself makes it more complicated. It is important to point out that artificial intelligence tools often work with unlabeled or partially labelled data when analytical processes are at least foggy or treated as a black box. The legal or technical solutions for these problems are not simple.

## **2.2. Big data challenges**

Big Data is changing society's landscape. Some expect it to replace earlier data collection and analysis techniques. It is difficult to deny the advantages of Big Data and new methods of its analysis. On the other hand, there are reservations about the paramount nature of analytical methods based on Artificial Intelligence. It is possible that some areas of life are more suitable for this approach and some less.

### **2.2.1. Big data advantages**

Big Data, with its basic three Vs—volume, velocity, and variety—has many benefits over traditional forms of data. The fourth V, validity, eliminates the negative influence of bias and brings it to a high-quality level [16]. The speed of data influx allows quick response, which is important in emergency situations and can literally save lives in healthcare.

In public medicine, Big Data may show tendencies of epidemiological processes and trigger an administrative reaction. Biomedical research gains a lot from hypothesis formulation based on big data analysis. Genomics, proteomics, and metabolomics routinely employ Artificial Intelligence for significant masses of data. Behavioural data obtained from the internet, analysis of search engine requests, smart vehicle data, smart environment data, health applications on a smartphone, and Internet of Things in Ambient Assisted Living assure a promising future for healthcare. Medical image processing in radiology, sonography, MRI, laboratory methods, and automatic processing of skin conditions show robustness and precision.

Clinical decision support, robotic surgery support, and clinical workflow organization are benefiting from Big Data analytics methods. However, there is no consensus about Big Data's solely positive nature and its analysis.

### **2.2.2. Data-driven and hypothesis-driven approach**

Besides the problems with clinical Big Data listed above, there are serious concerns about the Big Data character itself [17]. There is a fundamental difference between traditional clinical data and big clinical data [18]. Traditional data is well prepared, structured, human-readable, and controlled, presented in relatively small sets, and easy to analyze with proven methods. Big data lacks a clear structure. It is often collected and handled without usual human supervision. Sometimes, the data remains unclear for users, even after classification and clusterization. Some researchers strongly believe that big data is more suitable for data-driven hypothesis creation. However, the proof of the hypothesis, especially in clinical research, still requires standard statistical methods and a traditional hypothesis-driven approach [19]. Multidimensionality incorporates large data sets, which are often not the case for clinical data. Sceptical voices reject even the possibility itself of Big Data replacing traditional statistics and metrics [20]. Traditional clinical biometrics and biomarkers leave limited options for completely unsupervised methods of data handling.

## **3. Potential solutions**

This A number of proposals have been made to tackle these issues. One approach is straightforward and oriented toward narrowing the gap between the human and machine sides. The other is called Explainable AI and deals with issues of a black box.

An emerging field of large language models (LLM) can also be considered a prospective solution, at least partially. These options are discussed below.

### **3.1. Closing the gap**

The difference between the human-managed data approach and automated big data collection and analysis can potentially be lessened if not entirely overcome. It can be achieved by making machines able to accept natural language information, by training human side operators in technical input-output specifics or by complex incremental approaches from both sides.

#### **3.1.1. Machine side**

Berners-Lee formulated one of the basic methods as a semantic web concept [21]. The gap between humans and machines can be closed from the machine side when all data becomes machine-readable. Then, only problems of data presentation and human interface will remain for the full solution. The concept still needs to be achieved.

#### **3.1.2. Middle way**

More mundane measures are needed to bring both sides closer. Up-to-date IT education, specifically for healthcare specialists, intuitive UIs, wide data standardization, and gradual, disruption-free technology implementation are practical steps toward workable solutions. CQL, the Clinical Quality Language, is Human-Machine readable and works in both domains [22]. A complex approach, when different sources of data are confluent, leaves a possibility for filtering all clinical data to subdivide it into small data sets for robustness and more straightforward operation.

#### **3.1.3. Human side**

The other data approach is patient oriented. The so-called 4P concept shows the way for future development. Personalized, predictive, preventative, participatory healthcare requires all types of data to be equally machine- and human-readable and analyzable.

The more mundane way to narrow the machine-human gap is to address the issue of practical usability habits. The prior video gaming experience is beneficial for operators in a high-fidelity virtual reality simulated robotic surgery [23]. Similar results are shown in real-day laparoscopic surgery by warming up through the video game prior to it [24]. Explainable AI can be related to the human-side solution. However, it is a matter of balance to put this topic into a separate part to discuss it in more detail.

### **3.2. Explainable AI**

Big Data analysis with AI differentiates from regular data analysis with classical statistical tools. Unsupervised AI/ML neural networks, such as Deep Learning (DL), can be pretty efficient if judged by results but difficult for humans to comprehend. Black box AI poses a problem for high-risk areas of human activity. Explainable AI (XAI) is one way to solve it [25].

Often, ideal DL is envisaged as a tool that accepts raw data and produces high results with ready interpretation. There are methods to raise the accuracy of DL by tuning the neural network architecture and applying the algorithms with non-

transparent machine-driven feature detection and processing. These features are usually hard to categorize for human readability, while the white box is readable and often connected to “real” or physical categories. At the same time, data processing in a black box can be highly effective and fast for large datasets. White box relies on supervised learning, which includes labelled data and supervised feature engineering. Semi-supervised learning is a combination of black-and-white boxes and might utilise partial feature engineering. Complex methods with ensemble meta-algorithms, such as boosting or bagging, belong to the category of white boxes. For example, Random Forest (RF) are often called “self-supervised”. These methods are more explainable than the black boxes. Black box models, white box intrinsic or explainable by design models and grey box combinations have their strengths and weaknesses. For a description of structured characteristics, please see **Table 1**. The table demonstrates a negative relationship between performance and transparency in terms of human readability.

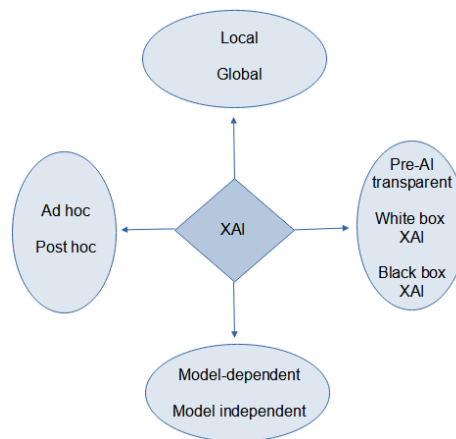
**Table 1.** Comparative characteristics of black, grey and white box models.

Type	Performance	Machine readability	Transparency	Human readability
Black box	++++	++++	+	+
Grey box	+++/++++	+++/++++	+/+++	+/+++
Explained White box	+++/++++	+++/++++	+/+++	+/+++
Intrinsic White box	++	++	++++	++++

### 3.2.1. Black box

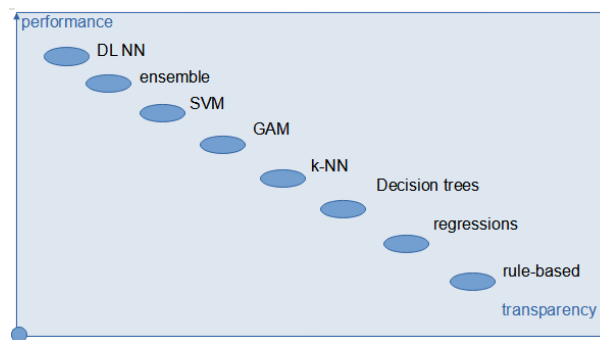
The opacity of AI models starts from the data collection stage. While limited data can be collected and interpreted by humans, big data, especially automatically collected or mixed, is less susceptible to human comprehension or analysis. The neural network was inspired by the formal model of the minimal neural equivalent of human intelligence or neurons. While the artificial “neurons” are pretty simple compared to biological specialised cells, the nature of NN “intelligence” is not much better understood than natural intelligence. Low transparency, in some way, is a result of the high capacity and robustness of the Deep Learning environment, where explainability is not required [26]. The limited control can be only applied in the stage of data collection, model training or output data interpretation. Multidimensional data analysis by unsupervised models is often more precise compared to more transparent ones. Nevertheless, higher precision is traded for obfuscation of analytical clarity for humans. Black box models can be explainable only post hoc, which is different from white box models (see **Figure 4**).





**Figure 4.** Main categories of explainable.

Main models with black box properties are based on neural networks, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Directed Acyclic Graph Neural Networks (DAG-NN), and other types and combinations. More explainable, not black box, models include Generalized Additive Models (GAMs), random forest busting, bagging models, k-Nearest Neighbour (k-NN), and Support Vector Machines (SVM). The explainability of the last ones can be increased by applying the dimensionality reduction techniques where Principal Component Analysis (PCA) is one example of such types of approach. For schematic information, see **Figure 5**. The dependence between performance and transparency is formal and does not need to be strictly linear.



**Figure 5.** AI/ML methods ranged by performance and transparency.

### 3.2.2. Ad hoc and post hoc

In most transparent cases, data is interpretable before AI/ML application. A human-interpretable, transparent by-design statistical AI model is called a white box. It is not a direct equivalent to supervised or semi-supervised AI/ML. White box frequently utilizes monotonic linear regressions (Linear Regression, Logistic Regression, LASSO), rule-based methods, and decision trees. Some of them perform only regression or classification, while others can do both. Probabilistic techniques, such as Naive Bayes, Markov models (can be seen as primary-level Dynamic Bayesian Networks) and other Bayesian networks are readable and often used for explainability.

The XAI approach can be subdivided into several main types [27] (Also see **Figure 4**). It is divided into local and global according to its explainability at the level

of the neuron, layer, or whole system. According to the order of explainable model production, white boxes are ad hoc or intrinsic [26], while black boxes are potentially post hoc interpretable. The first ones are often designed to be interpretable (explainable by design), while the post hoc are produced in the process.

The post hoc [28] techniques of XAI include Local Interpretable Model-agnostic Explanations (LIMEs), which can be applied to any black box, as well as Anchor, another model-independent method [27]; Shapley Additive Explanations (SHAPs), sometimes applied with GAMs [29]; Layer-wise Relevance Propagation (LRP); Deep Taylor Decomposition (DTD); Explainable Graph Neural Networks (XGNN); Prediction Difference Analysis (PDA) and some others.

### 3.2.3. Grey box

The grey box models are usually a combination of black box and white box models. The main goal is to integrate the best characteristics of the black box and the white box, such as high performance and high transparency, respectively [26]. A grey box is suitable in situations where balance is needed between precision, confidentiality and openness. Proprietary algorithm performance can be united with model partial explainability without disclosure of private details. The grey box is a wide description of the model's spectrum with different levels and types of explainability, productivity and ability to focus these characteristics on specific areas of narrow interest.

## 3.3. Large language models

Large Language Models (LLMs) are recent popular objects of fascination for researchers and the general public. LLMs are extensive AI-powered chatbot versions that are able to communicate with users via natural language texts and generate prompted written content on demand. LLMs utilise various NLP techniques and generative AI. The most well-known LLMs, such as ChatGPT, BingGPT, Bart, Chinchilla, LLaMA, and Gopher, are primarily universal. There is also a high demand for specialized domain-focused LLMs.

### 3.3.1. Current successful applications

**Table 2.** LLMs in medicine and healthcare.

Reference	Domain	Approach	Results and Conclusion
[31]	Consultation in vascular surgery emergencies	Consultation with Five attending surgeons and four LLMs	GPT-4 demonstrated 100% accuracy in emergency; further recommendations were accurate in 68%; 5% of responses were highly likely to cause clinically significant harm
[32]	Detecting Alzheimer's Disease-related signs and symptoms from electronic health records (EHRs)	Created taxonomy progression of AD signs and symptoms was trained using synthetic data generated by LLMs from EHRs	The trained model demonstrated acceptable quality
[33]	Chatbot Systems in Chronic Cardiovascular Disease Self-Management	LLM-based chatbot system that supported chronic patients with hypertension	The system has potential to motivate patients to monitor their blood pressure and adhere to prescriptions
[34]	HealAI: A Healthcare LLM for Effective Medical Documentation	Optimized LLM for specific medical tasks	Optimized LLM outperformed GPT-4, maintaining precision, reducing biases, preventing hallucinations, and enhancing note-writing style

**Table 3.** (Continued).

Reference	Domain	Approach	Results and Conclusion
[35]	LLMs for Healthcare Data Augmentation	Privacy-aware data augmentation approach for LLM-based patient-trial matching (LLM-PTM)	Performance improved by 7.32% on average; the generalizability of new data improved by 12.12%
[36]	Evaluating LLM -Generated Multimodal Diagnosis from Medical Images and symptom analysis	LLM-generated medical diagnosis with multimodal multiple-choice questions (MCQs) consisting of images and text for multiple pathologies	GPT-4-Vision-Preview scored approximately 84% of correct diagnoses
[37]	Implications of LLMs for dental medicine	LLMs for administrative tasks	LLMs could improve administrative work via clinical decision support, text summarization, efficient writing, and multilingual communication, and can be applied to dental telemedicine
[38]	Impact of ChatGPT-3 on Health Services as a Virtual Doctor	Potential benefits from consultation for healthcare services, including the Internet of Orthopaedic Things (IOT)	Potential to become a Virtual Doctor (DocGPT)

The first experiments in this area with ChatGPT [30] were promising when LLM was able to successfully pass the US medical qualification exams (USMLE) and master the Multiple-Choice Question Answering (MCQA) dataset for medical exams. A comparative study on the vascular surgery consultation scenarios was successfully done for GPT 3.5, GPT 4, Bard, and Falcon and compared to the panel of five field specialists [31]. Positive results are shown in the field of biomedical science and public health. **Table 2** provides some examples of tests using LLMs applicable to healthcare.

### 3.3.2. Limitations

Still, it is too early to make assumptions about LLM performance in the healthcare sphere. The well-known problem of LLM hallucination [39] limits not only the reliability and trustworthiness of the generated results but also creates a relatively new problem of responsibility for LLM-generated mistakes, wrong advice and externally coherent but medically meaningless solutions. GPT-4 demonstrated high interpretative abilities but showed reasoning inconsistencies, while Falcon and LLaMA 2 reached significant accuracy yet with insufficient explanatory reasoning [40]. The potential issues connected to the usage of LLMs are listed in **Table 3**.

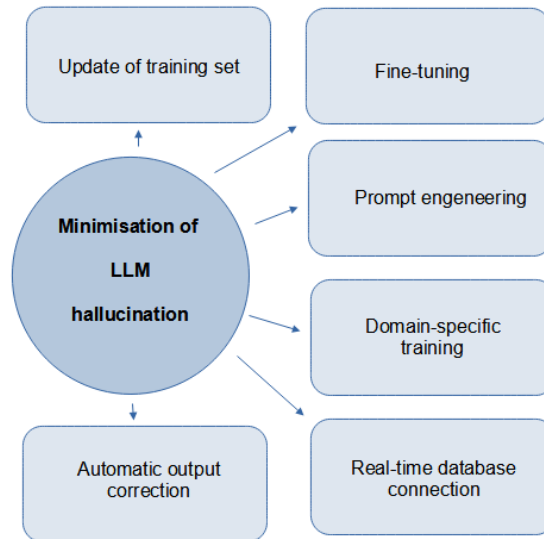
**Table 4.** Limitations of LLMs.

Type	Possible Harm or Issues
Hallucinations	Production of a false output that does not match the user's intent
Security	Leakage of people's private information
Bias	Contribution to the spread of misinformation
Consent	Can expose users to copyright infringement issues
Scaling	Difficulties with time- and resource-consuming for scaling and maintaining the LLMs
Deployment	Requires deep learning, transformer model, distributed software and hardware, and overall technical expertise

### 3.3.3. Potential solutions

There are several ways to address these problems (see **Figure 6**). The reliability of LLM-generated information can be increased by updating training sets, prompt

engineering, domain-specific LLM training, real-time database connection of LLM, automatic output correction, and some other instruments and techniques [41]. For example, improvement in Out-of-Distribution (OoD) detection raises LLM performance [32]. In addition, protecting LLM from adversarial OoD attacks reassures LLM performance stability. Another promising method is the model self-reflection function [42].



**Figure 6.** Main methods of lowering LLM hallucination.

#### 4. Discussion and conclusion

There are significant gaps between automatic and human data collection and analysis. These differences manifest themselves in the issues of Big Data and traditional statistical methods. Clinical data and its analysis are one of the testing grounds where questions about human control and responsibility reflect the complicated nature of these matters. Solutions are in the area of using more composite, heterogeneous human-collected and automatically collected data, mixed modes of analysis and dual readability for humans and machines. The tendency is complex: from one side, there is significant demand, based on technical capability, for automatically collected, analyzed and generated data. On the other side, there is no way to exclude human methods, which implies the necessity to enhance education and practical experience in the technical sphere not only for healthcare workers but for all members of modern society. Moreover, the implementation of modern LLMs might improve the diagnostic and treatment processes if the external reasoning agent is added to the models' architecture. It will increase the models' explainability and overall performance.

Thus, the benefits of advanced AI technologies are numerous, such as enhancing diagnostic accuracy, personalizing treatment plans, and smooth-running administrative processes. Modern LLMs can quickly analyze the vast amount of clinical data, identify specific patterns in it, and provide recommendations that improve outcomes. However, AI technologies also present challenges, such as data privacy concerns, the possibilities of algorithmic biases, and the need for significant computational resources.

The future of AI in healthcare lies in the continued development of sophisticated machine learning models, integrated evaluation and prediction systems and complex hierarchical architectures. Improved natural language processing will lead to better patient-system and specialist-system interactions. Enhanced computational frameworks can ensure ethical and unbiased AI use. Healthcare systems will achieve more comprehensive and accurate data analysis by combining human expertise with AI capabilities. Education and practical experience, enriched with AI-related technologies for healthcare professionals and the broader society, will maximise these benefits and ensure the responsible use of AI applications.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Vollset SE, Goren E, Yuan CW, et al. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *The Lancet*. 2020; 396(10258): 1285-1306. doi: 10.1016/S0140-6736(20)30677-2
2. Lee JY, Lim JY. The Prospect of the Fourth Industrial Revolution and Home Healthcare in Super-Aged Society. *Annals of Geriatric Medicine and Research*. 2017; 21(3): 95-100. doi: 10.4235/agmr.2017.21.3.95
3. Kim G, Lee HK, Lee B, et al. Fourth Industrial Revolution in Japan: Technology to Address Social Challenges. *SSRN Electronic Journal*. Published online 2021. doi: 10.2139/ssrn.3815376
4. Quer G, Radin JM, Gadaleta M, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nature Medicine*. 2020; 27(1): 73-77. doi: 10.1038/s41591-020-1123-x
5. Dix A. *Human-computer interaction*. Pearson Education Limited; 2003.
6. Shorten C, Khoshgoftaar TM, Furht B. Deep Learning applications for COVID-19. *Journal of Big Data*. 2021; 8(1). doi: 10.1186/s40537-020-00392-9
7. Norman KL, Kirakowski J. *The Wiley Handbook of Human Computer Interaction*. Published online December 28, 2017. doi: 10.1002/9781118976005
8. Ottens J, Baker RA, Newland RF, et al. The Future of the Perfusion Record: Automated Data Collection vs. Manual Recording. *The Journal of ExtraCorporeal Technology*. 2005; 37(4): 355-359. doi: 10.1051/ject/200537355
9. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Research and Clinical Practice*. 2017; 36(1): 3-11. doi: 10.23876/j.krcp.2017.36.1.3
10. Brown, Philip JB, Yongsheng G, et al. What is the abdomen? Rationalising clinical and anatomical perspectives using formal semantics. *Journal of Anatomy*. 2021; 238(6): 1472-1491. doi: 10.1111/joa.13384
11. Kendler KS. An historical framework for psychiatric nosology. *Psychological Medicine*. 2009; 39(12): 1935-1941. doi: 10.1017/s0033291709005753
12. Kendler KS. Reflections on the Relationship Between Psychiatric Genetics and Psychiatric Nosology. *American Journal of Psychiatry*. 2006; 163(7): 1138-1146. doi: 10.1176/ajp.2006.163.7.1138
13. Rodrigues JM, Schulz S, Mizen B, et al. Is the application of SNOMED CT concept model sufficiently quality assured? *AMIA Annual Symposium Proceedings*. American Medical Informatics Association. 2018; 16: 1488-1497.
14. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine*. 2020; 3(1). doi: 10.1038/s41746-020-0221-y
15. Hoffman S, Podgurski A. The Use and Misuse of Biomedical Data: Is Bigger Really Better? *American Journal of Law & Medicine*. 2013; 39(4): 497-538. doi: 10.1177/009885881303900401
16. Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*. 2019; 6(1). doi: 10.1186/s40537-019-0217-0
17. Kitchin R, McArdle G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*. 2016; 3(1): 2053951716631130. doi: 10.1177/2053951716631130
18. Batko K, Slezak A. The use of Big Data Analytics in healthcare. *Journal of Big Data*. 2022; 9(1): 3. doi: 10.1186/s40537-021-00553-4

19. Caliebe A, Leverkus F, Antes G, et al. Does big data require a methodological change in medical research? *BMC Medical Research Methodology*. 2019; 19(1). doi: 10.1186/s12874-019-0774-0
20. Apgar D. The False Promise of Big Data: Can Data Mining Replace Hypothesis-Driven Learning in the Identification of Predictive Performance Metrics? *Systems Research and Behavioral Science*. 2013; 32(1): 28-49. doi: 10.1002/sres.2219
21. Machado LMO, Souza RR, da Graça Simões M. Semantic web or web of data? a diachronic study (1999 to 2017) of the publications of tim berners-lee and the world wide web consortium. *Journal of the Association for Information Science and Technology*. 2019; 70(7): 701-714. doi: 10.1002/asi.24111
22. Odigie E, Lacson R, Raja A, et al. Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine—Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study. *JMIR Medical Informatics*. 2019; 7(2): e13590. doi: 10.2196/13590
23. Hvolbek AP, Nilsson PM, Sanguedolce F, et al. A prospective study of the effect of video games on robotic surgery skills using the high-fidelity virtual reality RobotiX simulator. *Advances in Medical Education and Practice*. 2019; 10: 627-634. doi: 10.2147/amep.s199323
24. Rosser JC, Gentile DA, Hanigan K, et al. The Effect of Video Game “Warm-up” on Performance of Laparoscopic Surgery Tasks. *JSL: Journal of the Society of Laparoendoscopic Surgeons*. 2012; 16(1): 3-9. doi: 10.4293/108680812x13291597715664
25. Pasquale F. The Black Box Society. Published online January 5, 2015. doi: 10.4159/harvard.9780674736061
26. Pintelas E, Livieris IE, Pintelas P. A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability. *Algorithms*. 2020; 13(1): 17. doi: 10.3390/a13010017
27. Holzinger A, Saranti A, Molnar C, et al. Explainable AI methods-a brief overview. Available online: [https://link.springer.com/chapter/10.1007/978-3-031-04083-2\\_2](https://link.springer.com/chapter/10.1007/978-3-031-04083-2_2) (accessed on 7 February 2024).
28. Pedreschi D, Giannotti F, Guidotti R, et al. Meaningful Explanations of Black Box AI Decision Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019; 33(01): 9780-9784. doi: 10.1609/aaai.v33i01.33019780
29. Sebastian B, von Luxburg U. From Shapley values to generalized additive models and back. Available online: <https://arxiv.org/abs/2209.04012> (accessed on 6 February 2024).
30. Cascella M, Montomoli J, Bellini V, et al. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*. 2023; 47(1). doi: 10.1007/s10916-023-01925-4
31. Le Q, Lavingia KS, Amendola M. The performance of large language models on fictional consult queries indicates favorable potential for AI-assisted vascular surgery consult handling. *JVS-Vascular Insights*. 2024; 2: 100052. doi: 10.1016/j.jvsvi.2023.100052
32. Li R, Wang X, Yu H. Two Directions for Clinical Data Generation with Large Language Models: Data-to-Label and Label-to-Data. *Findings of the Association for Computational Linguistics: EMNLP 2023*. Published online 2023. doi: 10.18653/v1/2023.findings-emnlp.474
33. Montagna S, Ferretti S, Klopfenstein LC, et al. Data Decentralisation of LLM-Based Chatbot Systems in Chronic Disease Self-Management. In: *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*; 6 September 2023. doi: 10.1145/3582515.3609536
34. Goyal S, Rastogi E, Rajagopal SP, et al. HealAI: A Healthcare LLM for Effective Medical Documentation. In: *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*; 4 March 2024. doi: 10.1145/3616855.3635739
35. Yuan J, Tang R, Jiang X, et al. Large language models for healthcare data augmentation: An example on patient-trial matching. Available online: <https://arxiv.org/abs/2303.16756> (accessed on 21 January 2024).
36. Panagoulas DP, Virvou M, Tsihrintzis GA. Evaluating LLM--Generated Multimodal Diagnosis from Medical Images and Symptom Analysis. Available online: <https://arxiv.org/abs/2402.01730> (accessed on 16 January 2024).
37. Eggmann F, Weiger R, Zitzmann NU, et al. Implications of large language models such as ChatGPT for dental medicine. *Journal of Esthetic and Restorative Dentistry*. 2023; 35(7): 1098-1102. doi: 10.1111/jerd.13046
38. Iftikhar L, Iftikhar MF, Hanif MI. Docgpt: Impact of chatgpt-3 on health services as a virtual doctor. *EC Paediatrics*. 2023; 12(1): 45-55.
39. Berenguer A, Morejón A, Tomás D, et al. Using Large Language Models to Enhance the Reusability of Sensor Data. *Sensors*. 2024; 24(2): 347. doi: 10.3390/s24020347

40. Zhang X, Vemulapalli S, Talukdar N, et al. Large Language Models in Medical Term Classification and Unexpected Misalignment Between Response and Reasoning. Available online: <https://arxiv.org/abs/2312.14184> (accessed on 16 January 2024).
41. Yao JY, Ning KP, Liu ZH, et al. Llm lies: Hallucinations are not bugs, but features as adversarial examples. Available online: <https://arxiv.org/abs/2310.01469> (accessed on 6 January 2024).
42. Ji Z, Yu T, Xu Y, et al. Towards Mitigating LLM Hallucination via Self Reflection. Findings of the Association for Computational Linguistics: EMNLP 2023. Published online 2023. doi: 10.18653/v1/2023.findings-emnlp.123