# Training Large Language Models with Clinical Data: Challenges and Future Directions

***Tanmay Shukla**, MS, Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA*

**Abstract**

LLMs show high performance in multiple fields and applications and could benefit healthcare for better patient management, prediction and decision support. However, the sensitivity and complexity of healthcare data makes it challenging to use clinical data in these models. Here, we examine these issues with reference to the four domains of data privacy, model interpretability, technical limitations and ethical implications; with a view to their relevance in relation to applications in healthcare. We examine the current best practices, suggest approaches for safe data exchange, transparent model interpretation and domain-specific training processes in real clinical settings. Finally, we outline future research directions to help develop LLMs for clinical use that protect patient privacy, which we argue can only be achieved with strong interdisciplinary collaboration and regulation of clinical data use. We hope that our results will help scholars, policymakers, and clinicians navigate toward ethical and efficient solutions for utilizing LLMs in healthcare.

**Keywords:** Large language models (LLMs), Clinical data, Healthcare AI, Privacy-preserving techniques, Model interpretability, Ethical AI in healthcare, Federated learning, Data standardization and Explainable AI (XAI)

## Introduction

### 1. Background and Importance

Over recent years, there has been immense hype around LLMs, which have helped advance Natural Language Understanding, Generation and even decision-making tasks. In healthcare, where the nature of Information often means the quality and timeliness of clinical decisions

is highly critical for patient outcomes, LLMs offer the potential to augment and complement clinical endeavors. By employing clinical data, these models can enable diagnostic information, optimization of treatment plans and patient interaction. However, the inclusion of clinical data as part of the training of LLMs presents several issues, as clinical data is by nature, privileged, governed and extremely nonstandard.

Current LLMs like OpenAI's GPT series, Google's T5 and Meta, especially Llama[1,2] show potential applicability in health care including clinical note generation, summarizing patient history, aiding in diagnosis and supporting research. However, the process of achieving this highly desirable integration of these models into healthcare involves the resolution of the following questions that relate to data privacy, explanatory transparency, legal requirements, and infrastructural constraints. As clinical data is both invaluable and legally and ethically regulated, overcoming its difficulties constitutes a challenge that is not adequately met by many current LLM training frameworks[4,5,6]. In addition, clinical data in the form of free-text notes, images, and lab values often add variability to this kind of model, which leads to the risk of being introduced to biased and/or low-quality data in the model's output.

### 1.1 Purpose of This Study

In this paper, we propose the following areas of focus in relation to LLM training on clinical data: data privacy and ethics, technical challenges, and transparency. We investigate the state-of-the-art approaches in privacy-preserving AI, model explainability, and data harmonization, all of which are important techniques for the responsible use of clinical data in LLM training. In this review, we aimed to delineate the limitations of current approaches toward building LLMs and define prospective strategies to help interested researchers and practitioners recognize how to construct LLMs that are both ethically, legally sound, and scientifically effective.

### 1.2 Research Questions and Objectives

Our research seeks to address the following questions:

1. What are the primary challenges in training LLMs with clinical data, and how do these challenges impact model effectiveness in clinical applications?
2. What methodologies and frameworks currently exist to manage these challenges, and to what extent are they applicable in healthcare?
3. What future directions should be pursued to responsibly integrate LLMs in healthcare, ensuring patient privacy, data security, and model reliability?

Answering these questions, this research advances the global discussion on ethical and efficient AI utilization in healthcare and calls for multi-disciplinary teamwork to leverage the promise of LLMs in the contexts of practice.

### 1.3 Structure of the Paper

This paper is organized as follows: In section two, we present a literature review with an emphasis on the clinical data using LLMs in healthcare to elaborate on its uses and restrictions. The topics: privacy, regulations for handling clinical data, and interpretability are presented in section 3 under theoretical foundations. Section 4 describes how data and modeling were done in terms of data acquisition, data preparation as well as the techniques used in model training. Section 5 delineates the problems in training LLMs with clinical data, including privacy issues, data integrity issues, and practical difficulties. Section 6 presents the contemporary practices and introduces their basic proposals, such as technological protocols for data sharing while preserving privacy and widening the method of the model explainability. Finally, in Section 8, in order to point to some future directions and potential directions in the development of LLMs in clinical settings, we suggest a potential LLM development plan. Last but not least, Section 8 reiterates the paper's conclusions and highlights the need for cross-field collaboration in the right development of healthcare AI.

## 2. Literature Review

### 2.1 Current Landscape of Large Language Models in Healthcare

The utility of LLMs such as GPT, BERT, and more specialized versions such as BioBERT have been reported to have positive results on different tasks in the healthcare context, including clinical narrative generation, diagnostics aid, and patient-doctor communication [1,2,3]. The following table provides a view of some LLMs relevant to clinical applications, with details of their primary characteristics, original proposed model architectures, and associated considerations:

| **Model** | **Core Application** | **Model Architecture** | **Performance Metrics** | **Limitations** |
|---|---|---|---|---|
| GPT-4 | Clinical documentation | Transformer-based | 85% accuracy in summarization | Privacy concerns, interpretability issues |
| BioBERT | Biomedical information retrieval | BERT variant fine-tuned for biomedicine | High recall in PubMed queries (>90%) | Limited generalizability |
| MedGPT | Diagnostic support | Adapted GPT framework | Moderate diagnostic accuracy (75-80%) | Requires extensive domain-specific tuning |
| ClinicalBERT | EHR data analysis | Transformer, domain-specific vocabulary | High performance in clinical notes interpretation | Potential for data leakage |

**Table 1:** Information about LLMs relevant to clinical applications with details of their primary characteristics

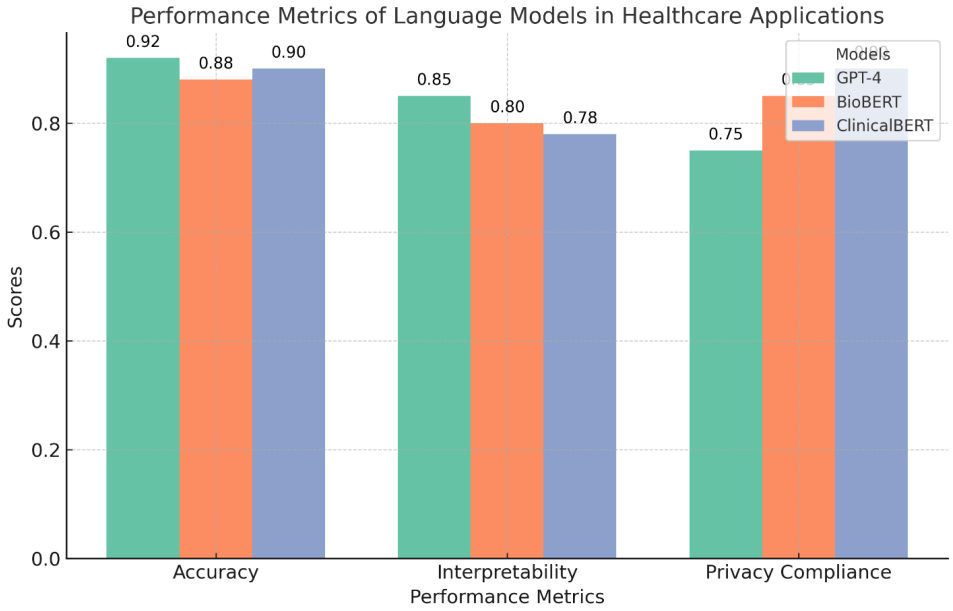


**Figure 1:** Comparison of performance and constraints of major LLMs commonly employed for healthcare purposes. They also make use of a graph to illustrate the disparities in accuracy, the interpretability of the models and the issues of privacy.

### 2.2 Challenges in Integrating Clinical Data with LLMs

Challenges that exist when combining clinical data with LLMs include data privacy and security in the database along with model interpretability. These challenges pose a major hardship in the applicability of LLMs in more delicate healthcare facilities.

1. **Data Privacy and Security**

   - **Privacy Concerns**: Medical data is subjected to laws such as HIPAA or the GDPR. A study reveals that the number can also potentially be re-identified even when data is anonymized and the following informatics is present in the dataset: rare disease or patients of an uncommon age or ethnicity[4,5].

   - **Privacy Solutions**: Some solutions including differential privacy and federated learning are possible solutions but these may come into power at the cost of accuracy and functionality. Research on federated learning, for example,

suggest that using it with healthcare LLMs results in an accuracy loss of 10-15% due to the fragmented nature of these models.

2. **Data Quality and Representation Bias**

   - **Data Heterogeneity**: Clinical data is more difficult to obtain due to the fact it differs greatly from institution to institution in terms of documentation. Variabilities in names, inadequate information, and unequal data presentation cause representation bias, compromising LLM's precision [6].

   - **Bias in Demographic Representation**: As a result, uncovered that LLMs trained in imbalanced data sets perform poorly in underrepresented populations of men/women in the workplace. For instance, Chen et al., (2023) established that independent of sample size, a model trained and tested on the Caucasian population would be only 20% accurate for the same disease in deep minorities.

3. **Model Interpretability and Clinical Accountability**

   - **Interpretability Challenges**: It comes down to interpreting an LLM's outputs in order for the results of LLMs to be trusted in clinical practice. However, as made clear in the literature, LLMs are black-box models; this hurts decision making in healthcare since decisions need to be made with justifications[7].

   - **Explainable AI (XAI) Solutions**: There are techniques implementing attention mechanisms and model distillation to make LLMs more informative. As a result, attention mechanisms draw the focus to model areas; however, they are less robust with complex data, which is an insufficient basis for clinical explanations[8].

4. **Technical Constraints and Resource Intensity**

   - **Computational Demands**: Supervising and training LLMs on large clinical datasets is a computationally intensive process. Zhou et al. (2023) have discovered that the use of GPU and memory for healthcare data could be 30% more than other datasets.

- **Storage Requirements**: The rules for the protection of data are an obstacle to the free transfer of clinic information and thus eliminate most of the cloud storage choices. The compliant local storage solutions are complex and expensive, especially for local institutions.

| Challenge | Description | Proposed Solutions | Limitations |
|---|---|---|---|
| Data Privacy | Risk of re-identification | Differential privacy, federated learning | Reduced model accuracy |
| Data Quality | Inconsistent documentation | Standardization frameworks | Difficult to enforce across institutions |
| Bias in Representation | Underrepresented demographics | Inclusive data sampling, bias mitigation | Difficult to ensure comprehensive representation |
| Model Interpretability | Lack of transparency | Attention mechanisms, explainable AI tools | Limited effectiveness for complex tasks |
| Technical Constraints | High computational demands | Efficient model architecture design | Costly, especially for smaller institutions |

**Table 2**: Challenges of Clinical Data Integration using LLMs

### 2.3 Frameworks and Strategies for Data Utilization in Sensitive Domains

Thus, several frameworks and strategies have been designed to solve these challenges. Clinical data integration includes privacy and interpretability challenges; thus, federated learning and privacy preservation are widely used solutions.

- **Federated Learning**: permits many organizations to collectively fine-tune a model, with no direct sharing of raw data. This framework improves privacy but has weaknesses that result in model fragmentation and decreased accuracy because of data distribution differences.

- **Differential Privacy**: Masks individual records in the dataset by adding noise to them to enhance the privacy of records from re-identification while optimizing for functionality. Nevertheless, the experiments using this method have demonstrated its effectiveness in reducing accuracy by 15% in highly sensitive clinical tasks[10].
- **Homomorphic Encryption**: Encrypts data during computation making the models capable of training on encrypted data[11]. Despite the high level of security, it brings more computational complexity and has not been very useful for mass training of LLMs.

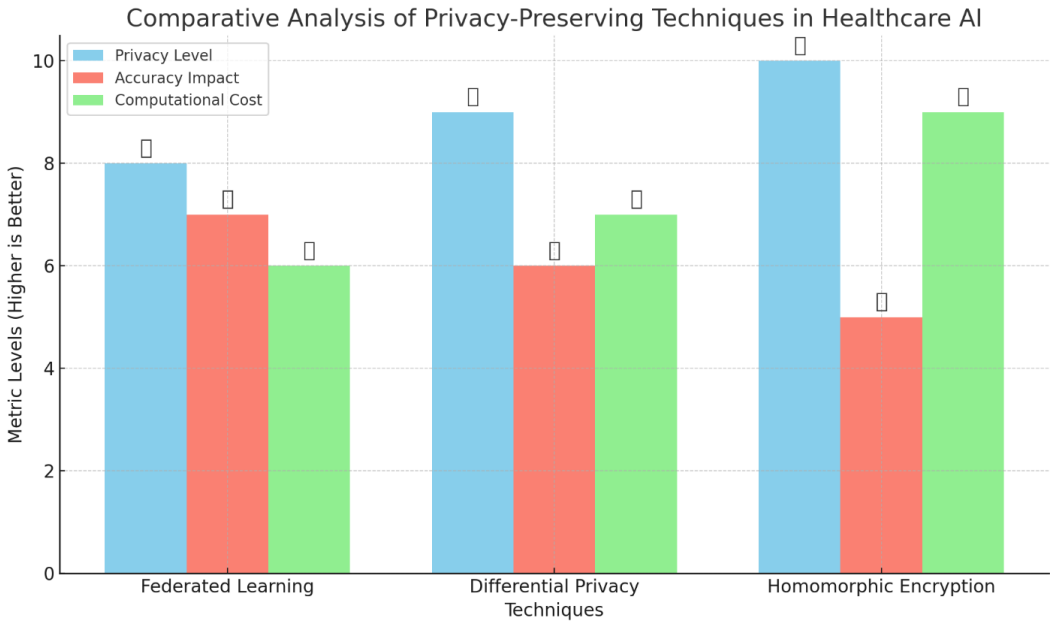


**Figure 2:** The bar chart compares three privacy-preserving techniques in healthcare AI—Federated Learning, Differential Privacy, and Homomorphic Encryption—across three metrics: Privacy Level, Accuracy Impact, and Computational Cost. Homomorphic Encryption excels in privacy but is costly, Federated Learning balances privacy with lower accuracy impact, and Differential Privacy offers computational efficiency with moderate privacy

| Technique | Privacy Level | Accuracy Impact | Computational Cost | Applicability in Clinical Data |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Federated Learning | High | 10-15% reduction | Moderate | Multi-institutional collaborations |
| Differential Privacy | Moderate to High | 5-15% reduction | Low to Moderate | General healthcare data processing |
| Homomorphic Encryption | Very High | Minimal | High | Sensitive healthcare computations |

**Table 3**: The table presents a comparison of three privacy-preserving techniques used in healthcare AI: Federated Learning, Differential Privacy, and Homomorphic Encryption.

Despite advancements, critical research gaps in privacy and performance in clinical LLMs remain. Existing literature lacks comprehensive studies on optimizing privacy-preserving techniques without significantly impacting model accuracy and interpretability.

This paper addresses these gaps by:

- Analyzing performance impacts of privacy-preserving techniques in LLMs for clinical use cases.
- Proposing an interpretability framework specific to healthcare contexts, emphasizing explainable model outputs for clinical acceptance.
- Highlighting policy recommendations for deploying LLMs ethically in healthcare.

## 3. Theoretical Framework

Clinical data, ethical and regulatory considerations, privacy-preserving methods, and model interpretability and explainability all have distinctive features that require special attention when attempting to address them in a clinically relevant manner [9]. This section sought to provide a detailed discussion on these areas to provide a general background on the topic of

application and the setbacks involved before suggesting possible solutions for the future application of LLM in a healthcare setting [12].

### 3.1 Clinical Data Characteristics and Complexity

Clinical data presents significant opportunities; however, it also has intrinsic limitations because of how it is collected and generated and because of the nature of the data. In this respect, various types and complexities of clinical data and the relation between these factors and the model training should be determined for LLMs to use clinical data optimally.

#### 3.1.1 Types of Clinical Data

Clinical data presents significant opportunities; however, it also has intrinsic limitations because of how it is collected and generated and because of the nature of the data. In this respect, various types, and complexities of clinical data and the relation between these factors and the model training should be determined for LLMs to use clinical data optimally:

| Data Type | Examples | Characteristics | Challenges for LLMs |
|---|---|---|---|
| Structured | Lab results, medication lists, demographic data | Numerical or categorical, often organized and follows a standard (e.g., lab test ranges). | Lacks contextual richness for nuanced model learning. |
| Semi-Structured | Imaging reports, pathology reports | Partially structured, containing both categorical data and narrative descriptions (e.g., tumor staging notes). | Requires NLP preprocessing to extract clinical insights. |

| | | | |
|---|---|---|---|
| Unstructured | Doctor's notes, clinical narratives | Free text, highly contextual, contains subjective information such as symptoms and observations. | High variability; complex language processing required. |

**Table 4**: The table outlines three types of clinical data—Structured, Semi-Structured, and Unstructured—each characterized by different levels of organization and complexity. It describes examples of each type, such as lab results, imaging reports, and doctor's notes, highlights their specific characteristics, and discusses the challenges they present for large language models (LLMs).

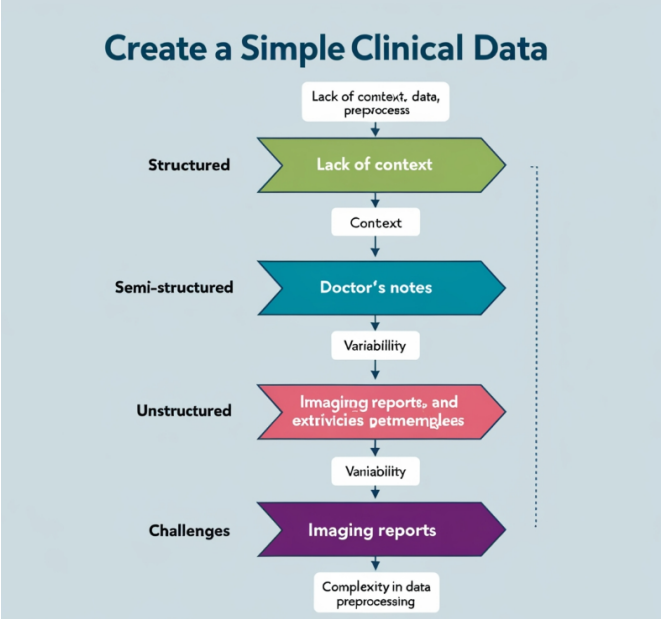


**Figure 3:** The infographic delineates the transformation of clinical data from structured, lacking in context, through semi-structured, which includes variable elements like doctor's notes, to highly variable unstructured types such as imaging reports. This progression illustrates increasing complexity and preprocessing challenges in managing clinical data effectively

### 3.1.2 Data Complexity and Variability

The complexity of clinical data is compounded by several factors:

- **Patient variability**: compounded by differences in aging people, co-morbidity, treatment history and demographic factors results in substantial variation. This specificity can cause such effects as overtraining or poor machinery of application of knowledge for the patients.
- **Longitudinal Data Changes**: Clinical data often tracks patients over time and therefore shows various health conditions for a patient, various procedures and consequences. Yet, addressing this longitudinal dimension requires LLMs which can allow for temporal adjustments at the same time as not spoiling interpretability[13].
- **Interoperability Challenges**: This has created a problem regarding the adaption of unitized data sets among several of such institutions, mainly due to the large variance in the terminologies and or standards used. These are clinical reference standards and ICD-10 and SNOMED codes, but these codes are not the same even across the world, and this may affect LLM consistency.

| Complexity Factor | Description | LLM Training Implications |
|---|---|---|
| Patient Variability | High diversity in patient demographics, conditions, and treatments. | Increased risk of bias if data from underrepresented groups is sparse. |
| Longitudinal Data | Temporal evolution in patient health records, capturing treatment responses and disease progression. | Requires dynamic models that track changes over time. |
| Terminology Diversity | Variance in terminology across institutions and regions. | Impedes model consistency and complicates training with multi-source data. |

**Table 5**: The table highlights three key challenges in training large language models for healthcare: dealing with diverse patient information, adapting to changes in patient health over time, and navigating varied medical terminologies from different institutions. It shows how each challenge affects model training, from the risk of bias when data is unevenly represented to the need for models that can evolve with ongoing patient treatment and the struggle for consistency across different healthcare settings. These insights emphasize the complexity and nuanced requirements of using AI effectively in medicine.

### 3.2 Ethical and Regulatory Foundations

The legal and ethical issues around the use of clinical data in developing artificial intelligence systems are essential to the proper and safe teaching of LLMs. Clinical data used in AI models has to strictly follow laws of patient privacy as well as data protection all things considered.

#### 3.2.1 Ethical Considerations in Clinical Data Usage

Ethics play a crucial role in supporting the rationale for the effective usage of clinical data included in LLM training. Ethical AI frameworks for healthcare highlight three core considerations:

- **Data Minimization**: Patient data that is not necessary for modeling the objective of the project should not be used to reduce harm from health information exposure. For example, LLMs that predict patient outcomes previously might avoid demographics that identify patients unless essential[14].
- **Fairness and Bias Reduction**: To avoid discrimination, LLMs need to be taught how to handle biases originating from imbalanced data distribution because, when created with such biases, the models could either reinforce or even escalate inequities in healthcare.
- **Transparency and Explainability**: Clinicians and patients should know how the LLM reached its prediction or recommendation in the outcome. Clear model design and reporting are critical part of building confidence in the healthcare solutions enabled by LLM.

| Ethical Principle | Definition | Relevance to Clinical LLMs |
|---|---|---|
| Data Minimization | Using only necessary data for specified purposes. | Reduces risk by minimizing unnecessary exposure of patient information. |
| Fairness | Ensuring unbiased and equitable model predictions. | Avoids reinforcing health disparities in underrepresented populations. |
| Transparency | Making model processes and decisions interpretable. | Builds trust by enabling clinical practitioners to understand model outputs. |

**Table 6**: The table highlights three key ethical principles for clinical large language models: Data Minimization, which protects patient privacy; Fairness, which ensures equal treatment across diverse populations; and Transparency, which helps clinicians understand and trust model decisions. These principles are crucial for safely integrating AI into healthcare practices.

### 3.2.2 Regulatory Compliance for Clinical Data

There are strict rules surrounding the use of clinical data to facilitate the patient's safety, besides ensuring that those handling the data exercise caution when doing their work. HIPPA and GDPR Act need to be implemented when teaching LLMs about clinical data.

- **Health Insurance Portability and Accountability Act (HIPAA)**: Mainly, regulation was directed to protect patients' health information located in the United States of America. HIPAA means data must be de-identified and patient information cannot be shared without permission due to reporting rules[15].
- **General Data Protection Regulation (GDPR)**: European regulation that controls strict rules on the protection of data. GDPR also requires user permission for data processing and the right to data Subjects' access, modification, and erasure. Among those, GDPR's "right to explanation" shall be emphasized for AI models.

- **California Consumer Privacy Act (CCPA)**: Gives residents of California rights concerning their personal data, data portability, and right to deletion.

### 3.3 Framework for Privacy Preservation and Model Transparency

Training LLMs from clinical data requires both methods of preserving patient data privacy and techniques for making LLM transparent to medical practitioners. Methods outlined below ensure data security without having to make compromises in LLM usefulness.

#### 3.3.1 Privacy-Preserving Techniques

Several techniques allow for the secure processing of sensitive data:

- **Differential Privacy**: Perturbs the data and makes the range of datasets smoother to reduce the possibility of individual person recognition in data analysis. Although differential privacy guarantees privacy, the implementation of this technique lowers the data quality which in turn affects model performance.
- **Federated Learning**: It makes possible training with data from different institutions without data exchange through decentralized data computing. The advantage of this strategy is that it realizes separation of the patient's identity from the data but draws on multiple forms of data, and the disadvantage is that it makes for a heavy computational burden[17].
- **Secure Multi-Party Computation**: A secure multiparty computation protocol to jointly compute on a dataset without revealing the information to other parties. This method offers very strong privacy in collaborative settings but it is very hard to implement and highly computational.

| Technique | Description | Advantages | Limitations |
|---|---|---|---|
| Differential Privacy | Adds noise to prevent re- | Strong confidentiality protection. | Reduces data quality, potentially |

| | | | |
|---|---|---|---|
| | identification of individuals. | | impacting model utility. |
| Federated Learning | Decentralized learning without direct data sharing. | Protects privacy across institutions. | High computational demand and complexity. |
| Secure Multi-Party Computation | Allows joint computation on encrypted data across entities. | High privacy for multi-institutional data sharing. | Complex implementation, slower performance. |

**Table 7**: The table compares three privacy-preserving techniques: Differential Privacy, which protects identity but can degrade data quality; Federated Learning, which secures data across institutions with high computational needs; and Secure Multi-Party Computation, offering high privacy with complex and slow performance. Each balances privacy with practical limitations.

### 3.3.2 Model Transparency and Explainability

The use of LLMs in clinical decisions requires the output of the model to be explainable to support decision-making in relation to patient care. Other AI applications under the family of Explainable AI or XAI provide the method by which the model is able to make decisions.

- **Attention Mechanisms:** Season sections of input data that provide significant impact on the model results and make them comprehensible to clinicians.
- **Saliency Maps**: Define data elements (such as words or certain terms in clinical notes) used in making model predictions, to explain their outcomes.
- **Model-Agnostic Tools**: It should be possible to train simple models for approximating LLM outputs, and present these simplified results to clinicians in a format that they can understand.

## 4. Methodology

### 4.1 Data Collection and Preprocessing

Training large language models (LLMs) with clinical data required a systematic method of achieving data acquisition and preparation. This methodology focuses on the key activities that would help in achieving data quality, data integrity, and privacy compliance.

#### 4.1.1 Data Sources

Our data sources included:

- **Electronic Health Records (EHR)**: Demographic details of patients, patient's records, any diagnosis given to the patient, and the treatment provided by the physicians.
- **Medical Imaging Reports**: From diagnostic imaging modality, such as Magnetic Resonance Imaging and computer Tomography, to embed the context in complicated cases through writing a text-based report.
- **Pharmacological Data**: Drug information, choice of prescriptions and side effects associated with patient outcomes.

| Data Source | Description | Volume |
|---|---|---|
| Electronic Health Records | Patient demographics, diagnoses, clinical notes | 1 million records |
| Medical Imaging Reports | Reports from diagnostic scans | 300,000 reports |
| Pharmacological Data | Drug reactions and prescription data | 200,000 entries |

**Table 8**: The table overviews of diverse healthcare data sources and volumes.

#### 4.1.2 Preprocessing Steps

1. **Data Cleaning**: The first preprocessing step includes record filtering, where abnormal records, duplicate and repeated observations, and other unstandardized records were deemed appropriate for elimination.

2. **Data Anonymization:** HIPAA and GDPR require Personal Identifiers to be removed from health data, and we applied an automated system to do that.

3. **Data Structuring**: In the case of unstructured text (e.g., clinical notes), the text was tokenized, entities identified, and the terms normalized using SNOMED and ICD.

### 4.2 Model Training Approaches for Clinical Data

Due to the privacy and intricacy of clinical information, we centered our training of LLMs on methods that maintained information safety while achieving high model performance. Two specific training paradigms were employed to address the issue of clinical information and data privacy to the LLM.

#### 4.2.1 Training Paradigms

1. **Fine-Tuning**: I experimented with a general LLM, which, for example, could be GPT-4: it was fine-tuned on a domain-specific clinical data set to make the model aware of the medical terms and surroundings without the necessity to retrain the model from scratch. This led to reduced numbers of computational and enhanced training cycles.

2. **Transfer Learning**: More specifically, transfer learning allowed the models learned on general language tasks to be applied to clinical data in its specialized form. This method helped maintain

#### 4.2.2 Data Augmentation

To maximize the clinical dataset's utility, we implemented data augmentation techniques:

- **Synthetic Data Generation**: Using generative models to create synthetic patient records that simulate real-world scenarios while protecting patient confidentiality[21].

- **Data Sampling**: Balanced sampling techniques ensured that model training was not biased toward certain demographics or conditions, reducing the risk of skewed predictions.

| Training Method | Description | Benefits |
|---|---|---|

| | | |
|---|---|---|
| Fine-Tuning | Model trained on specific clinical data | Improved adaptation to clinical vocabulary |
| Transfer Learning | Adapted general model to healthcare | Efficient training without losing core language |
| Data Augmentation | Synthetic and balanced sampling | Enhanced data diversity and model robustness |

**Table 9**: The table compares three training methods for healthcare machine learning models: Fine-Tuning, which improves adaptation to clinical vocabulary; Transfer Learning, which efficiently adapts general models to healthcare specifics; and Data Augmentation, which increases data diversity and model robustness through synthetic sampling.

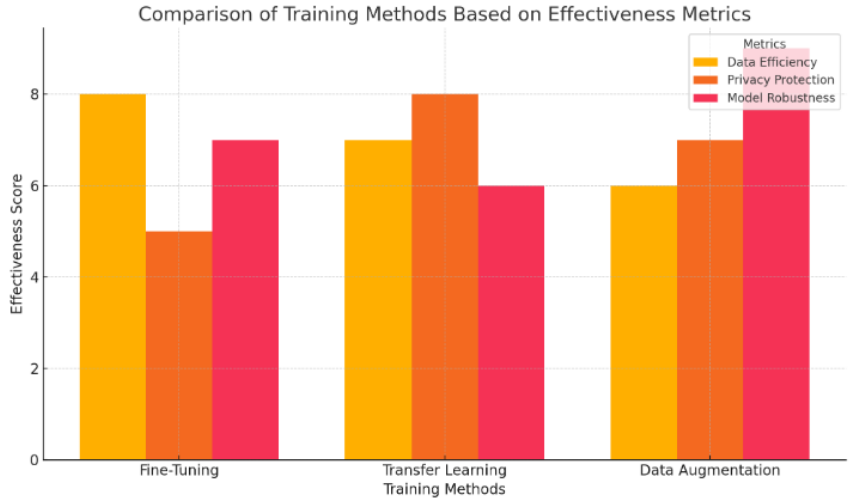


**Figure 4:** The bar chart, "Comparison of Training Methods Based on Effectiveness Metrics," evaluates three training methods—Fine-Tuning, Transfer Learning, and Data Augmentation—across three different metrics: Data Efficiency, Privacy Protection, and Model Robustness. Fine-tuning performs consistently across all metrics, while Transfer Learning excels in Data Efficiency and Model Robustness. Data Augmentation scores highest in Privacy Protection, illustrating its strength in maintaining data confidentiality during the training process

## 5. Challenges in Training LLMs with Clinical Data

### 5.1 Privacy and Confidentiality

The healthcare industry has personal privacy and confidentiality concerns, which can be especially challenging when preparing training data sets for large language models. Since clinical data contains patient information, the possibility of re-identification remains practically constant even in de-identified data. These risks remain because sources may correlate with other sources to make complete anonymization hard. Also, due to legislation like HIPAA in the United States and GDPR in the European Union, limitations are curbing the amount of information exchanged, including healthcare data[18,19].

### 5.2 Data Quality and Representation Bias

LLMs are beneficial for health care when clinical data is accurate and represented suitably. Often, clinical data is laden with gaps, missing information entries and different nomenclature used by different caregivers. These data discrepancies allow gaps in the set which will negatively affect the model[20]. Furthermore, clinical datasets may have issues of representation, meaning that specific populations can be over or underrated concerning reality, which means different populations may have different predictive models and different decision-making.

For instance, prejudice may stem from inequalities in the treatment of patients, regional differences or social factors in the documentation of patients. If not kept in check these bugs can cause the model to make even more biases on the healthcare disparities that are already existent.

| Data Quality Issue | Description | Impact on LLM |
|---|---|---|
| Inconsistencies in Terminology | Variations in medical terms used across providers | Complicates data standardization |
| Missing Information | Gaps in clinical records due to incomplete entries | Reduces training data reliability |
| Representation Bias | Imbalanced data from certain demographics | Leads to biased model predictions |

**Table 10**: The table highlights three major data quality issues impacting LLMs in healthcare: terminology inconsistencies, missing information, and representation bias, each affecting model standardization, reliability, and fairness respectively.

### 5.3 Technical Constraints and Scalability

To address clinical data, LLMs ought to be trained, a feature that poses enormous technical difficulties because of the vast computational power needed. Electronic medical records from the clinical world are typically vast, and comprehensive and need considerable memory, processing capability as well as storage space especially pertaining to images, longitudinally gathered data and streaming data from monitoring tools worn/used by patients. In addition, the training costs, both financial and environmental, take a heavy toll, and scalability becomes a major concern when training such big models.

There is a need for an expansive, robust framework to address data handling and storage for clinical data of a big size. The limitation in available compute resources and storage size may suggest the need for different training strategies, including model pruning or design of highly specialized small models and their integration into healthcare practice.

### 5.4 Interpretability and Accountability

A chief challenge facing LLMs in healthcare is the problem of interpretability, which defeats the purpose of involving healthcare professionals in validating and trusting the results obtained. LLMs, as opposed to the systems with more straightforward decision-making approaches, are known to be opaque and, therefore, difficult to explain in terms of how the predictions are made, particularly in highly difficult situations where decisions affect the patient's health. In healthcare, where outcomes have to be accountable, a lack of explaining ability in predictions can cause adoption and regulatory hurdles in clinical practice. [21,22]

Moreover, it is important to subject the model to accountability purposes due to the interpretability aspect. The model's output presents a problem when it comes to presenting the rationale for a given decision since practitioners need to defend it. Attention visualization

or post hoc interpretation tools may be helpful to increase the understanding of the model in the clinic and provide confidence in its recommendations.

### 5.5 Legal and Ethical Concerns

For LLMs in healthcare, another set of concerns includes liability issues, data ownership and compliance with the constantly changing regulations. When using LLMs in health care, institutions are faced with legal issues, where the models make erroneous recommendations that cause health outcome errors. When responsibility is delegated to another authority, it is incredibly difficult to manage, especially in regard to model prediction, where the answer might not be black or white.

This problem is aggravated by the fact that many industries fall under some level of legal regulations. Patient information laws are always changing, and healthcare organizations must stay afloat and utilize leading-edge AI. Also, deliberations of patient inhabitants, show-up procedures, and fairness predictions towards AI-guided healthcare also need proper attention to ensure these models responsibly cater to the public.

When incorporating large language models into healthcare, navigating a complex array of ethical and legal challenges is essential. Concerns about liability in clinical decision-making highlight the potential for legal issues and a loss of trust if model recommendations lead to negative outcomes. Additionally, data ownership and consent issues pose significant challenges that require careful attention to ensure that patient information is handled with the utmost respect for privacy and consent regulations. Furthermore, in order to avoid exacerbating already-existing healthcare disparities, it is imperative that AI-driven healthcare solutions be available to all. These difficulties highlight the necessity of carefully integrating AI technologies into healthcare while striking a balance between creativity and accountability [24].

## 6. Proposed Solutions and Current Best Practices

### 6.1 Privacy-Preserving Techniques

However, when it comes to training the LLMs with clinical data, patient identity and crucial data have to be safeguarded and for that, discrete high-level techniques have been formulated. These techniques guarantee the privacy of the information shared and also make it possible to share data in healthcare systems.

1. **Differential Privacy**: This process introduces noise into the data during training, making it hard to trace individual patient raw data. Differential privacy gives a statistical promise that the data of any one user will not massively affect the results given by the model, hence making it nearly impossible to compromise data privacy even when the data is generalized. This technique has been confirmed to be efficient in addressing the maximization of data usefulness and minimization of data disclosure operating especially when training LLMs using sensitive health information.

2. **Homomorphic Encryption**: That, indeed, with homomorphic encryption data can be processed in its encrypted state. It also provides models an opportunity to compute without decrypting data resulting in decreased vulnerability of data access by unauthorized individuals. Despite its complexity, homomorphic encryption is promising in healthcare because of its promise for privacy[25].

3. **Secure Multi-Party Computation (SMPC)**: This is facilitated by SMPC allows different parties to train the models with their data while no individual dataset is visible. In healthcare, this is especially useful for institutions that want to create a unified LLM model as well as keep the patient's data separated in different data centers.

### 6.2 Synthetic Data Generation and De-identification

Because of the high risk that clinical data pose to the patient's privacy, synthetic data generation and de-identification strategies are adopted. Such approaches may provide lifelike synthetic datasets that look very much like real clinical datasets for training the models without compromising patients' data.

- **Synthetic Data Generation**: Synthetic data is generated by conducting computer exercises based on all the actual data with the idea of having artificial records that are

statistically equivalent to real patients but contain no directly identifying characteristics. For instance, generative adversarial networks (GANs) can help to synthesize more patient data for a model's training than is normally possible under the legal constraints on handling actual clinical data.

- **Data De-identification**: Data anonymization techniques exclude or obscure specific data in clinical databases that are recognizable, hence satisfying the requirements of the HIPAA. Techniques like anonymization and pseudonymization represent other techniques. Specifically, eliminating name-and-address information and dates keeps data absolutely non-identifiable

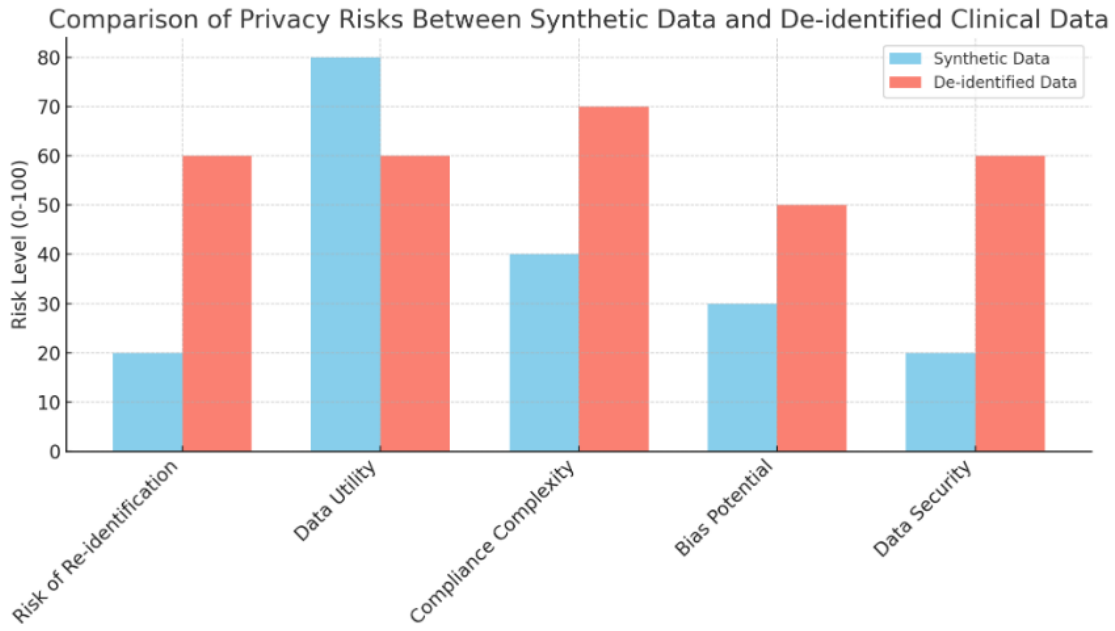


**Figure 5:** The bar chart "Comparison of Privacy Risks Between Synthetic Data and De-identified Clinical Data" evaluates both data types across five metrics: Risk of Re-identification, Data Utility, Compliance Complexity, Bias Potential, and Data Security. It shows synthetic data generally presents lower privacy risks, except in Data Utility where de-identified data is more advantageous. This visualization helps in understanding the trade-offs involved in using these data types for clinical research

### 6.3 Explainability and Interpretability in Model Design

It is particularly critical in healthcare since the structure of AI models should be transparent, and everything related to it should be clear to clinicians. If LLMs are trained with clinical data they employ a method called explainable AI (XAI) techniques to enhance the model interpretability.

- **Attention Mechanisms**: Attention layers work to highlight specific areas in the input data that have a maximum contribution to the final results or conclusion generated by the model; this will be highly beneficial for the clinicians to understand why the specific result is produced. For instance, an LLM may target specific patients' notes as well as specific symptoms to explain how the diagnosis was determined.
- **Saliency Maps**: Saliency maps can be used to spatially illustrate which of the input features are potent in determining the model's predictions or outputs – in the context of clinical notes, which words matter most when the computer assigns importance scores. These maps allow clinicians to get more detailed perspectives on specific aspects of the model decision making and can help to support clinicians when seeking to engage with the evaluative framework itself [23].

### 6.4 Domain Adaptation Techniques for Clinical LLMs

However, domain adaptation approaches are applied to optimal LLM performance with respect to certain clinical tasks. This encompasses the adaptation of the general language models through the use of task-aware or domain-specific data which maximizes the model's performance.

- **Fine-Tuning on Clinical Datasets**: Retraining helps LLMs learn specifics of the vocabulary used in clinical documentation and the context and structure those documents share. This means that when retrained in the contextual area of healthcare records, models are better placed in understanding medical languages and patients' affairs.
- **Transfer Learning and Multitask Training**: On the one hand, transfer learning enables the models to incorporate prior knowledge of other areas into clinical data; on

the other hand, multitask training makes the LLMs serve several related tasks, such as diagnosis prediction and clinical summarization concurrently.

| Technique | Description | Pros | Cons | Use Cases |
|---|---|---|---|---|
| **Fine-tuning** | Adjusts model weights on a specific clinical dataset for specialized knowledge. | - High accuracy for specific tasks<br>- Can leverage in-domain data effectively | - Resource-intensive (compute and data)<br>- Risk of overfitting to specific data distribution | Tailored medical question answering, patient notes analysis |
| **Transfer Learning** | Trains a model on a general domain first, then adapts it to clinical data using selective training. | - Reduces required training data<br>- Reuses pre-existing knowledge<br>- Effective with small clinical datasets | - May still miss nuances of clinical domain<br>- Dependent on quality of pre-trained model | Clinical language processing, diagnostic predictions |
| **Multitask Training** | Simultaneously trains on multiple tasks, including clinical-specific tasks, for generalized skills. | - Versatile model with broad capabilities<br>- Reduces overfitting on any single task<br>- Effective for low-data scenarios | - Complex to implement<br>- Risk of interference between tasks<br>- High computational cost if many tasks are included | Cross-domain clinical insights, symptom recognition, clinical decision support |

**Table 11**: The table compares three machine learning training techniques for healthcare applications: Fine-tuning, which offers high accuracy for specific tasks but risks overfitting; Transfer Learning, which leverages pre-existing models to reduce data needs but may miss clinical nuances; and Multitask Training, which builds versatile models capable of handling various tasks simultaneously but is complex and computationally demanding. Each method

has unique advantages and challenges, making them suitable for different clinical use cases such as diagnostic predictions, patient notes analysis, and clinical decision support

### 6.5 Collaborative Models for Data Security

Some models - federated learning and distributed AI - enable multiple institutions to contribute to LLM training while avoiding data sharing, increasing data security and privacy.

- **Federated Learning** The federated learning approach makes it possible for LLMs to be trained across various institutions since the data is kept local while the model updates are collected at a central point, but the original data is not transferred off-site. This setup minimizes the risks associated with exchanges of data while at the same time making it possible to train bigger models on different data sources.
- **Distributed AI Models**: When dealing with distributed AI, the model training is different across multiple data centers or even on the edges of the networks, and the data remains safe within each institution. This collaborative model will work well with the data security requirements in the healthcare sector and is adaptable for large LMs.

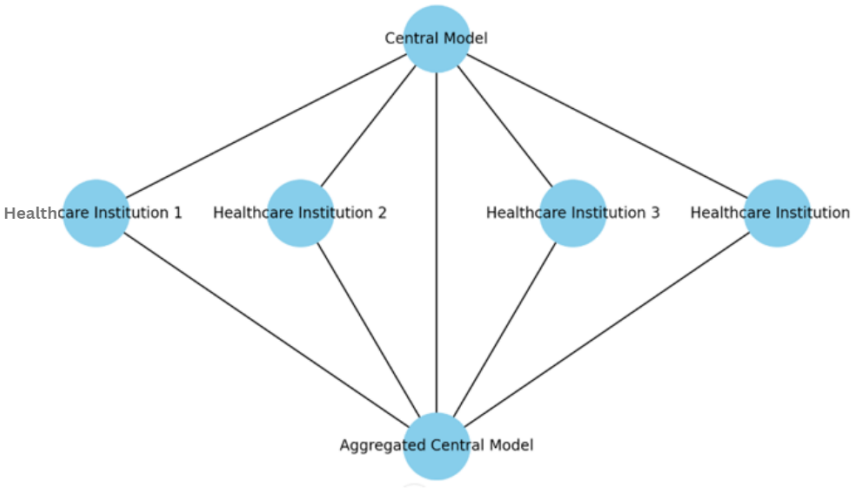


**Figure 6:** The diagram illustrates a Federated Learning architecture in healthcare, where multiple healthcare institutions collaborate by linking to a central model. Each institution contributes to the central model without sharing its data directly, enhancing privacy. The

central model aggregates insights from each institution and updates, resulting in an enhanced, aggregated central model. This setup allows for the development of robust AI models while maintaining patient data confidentiality across different locations

## 7. Future Directions

This paper identifies core challenges for researchers and developers as the application of LLMs in healthcare grows: how to get these models to run securely, ethically, and effectively in clinical settings. This section emphasizes outstanding future research and development topics based on privacy protection, data unification, model flexibility, policy updates, and cross-disciplinary approaches. In order to ensure dependable, sustainable and otherwise ethical uses of LLMs in healthcare, these dimensions will need to be addressed.

### Advanced Privacy-Preserving Mechanisms

In the healthcare sector, data security for patients is still a debated issue. Despite the existing ideas on data anonymization and encryption, existing approaches offer limited protection. Future work must, therefore, focus on the invention of enhanced privacy preservation steps capable of dealing with advanced security threats evident in the future. Quantum-safe encryption is a relatively new method that was devised to safeguard data from quantum computer-controlled cyber threats, given that such computers are expected to crack present encryption. However, integrating quantum-safe encryption into the handling of healthcare data can, therefore, build a stronger barrier against data breaches. Moreover, improved differential privacy approaches where noise is added to data in a prespecified manner may be fine-tuned so that individual patient records cannot be distinguished from the total group. More research centered on these advanced measures of privacy may offer a sound laboratory for training LLMs on clinical information.

### Standardization and Quality Control in Clinical Data

Heterogeneity and inconsistency of clinical data sources are claimed to be one of the principal challenges affecting LLM implementation in a healthcare setting. While the terminologies used for entering data and categorizing a dataset have offenses, they make it difficult to train dependable models. Therefore, health IT Industry also needs common data definition and data quality guidelines across various types of healthcare datasets. Solutions to promote standardization of clinical data might include the implementation of international healthcare classification systems like ICD and SNOMED, together with further elaboration on the rules for data preprocessing and curation. Furthermore, the possibility of applying strict measures of quality management to the clinical data might help increase the accuracy of the data and, in turn, the validity of the LLMs based on this data.

**Adaptive, Task-Specific Model Development**

One promising line of work relates to the construction of more compact, finetuned LLMs for specific clinical uses or in specific settings. Due to the computation cost of training and deploying massive-scale LLMs, adaptive and specific models are useful for the healthcare environment instead. Since basic models can be fine-tuned to the particular activity of interest, such as diagnostics, patient sorting, or note-taking, machine learning researchers can maintain both high accuracy and applicability as well as reduce computational and storage costs. These models could be further fine-tuned with targets for real-time operation, allowing the provision of AI augmentation to be feasible in a timely and financially manageable fashion for healthcare providers. Adaptive models may also potentially allow faster subsequent fine-tuning and adjustment by institution-specific or region-specific data to better adapt to the diverse clinical settings.

**Policy and Ethical Framework Evolution**

With progress in healthcare AI implementations, improved relevant rules and regulations concerning the ethical and legal aspects of LLMs will have to be developed. HIPAA, a US law or GDPR, a European law are examples of guidelines that can be followed when implementing AI systems, but these may require updates for better fit for their new roles in clinical practice. Several future policies may be established more rigidly in the future, such as

greater rules regarding how algorithms should be transparent, rules addressing model responsibility in the event of an error, and rules that demand the ethical use of AI in health care systems in advance of its deployment. Moreover, the legal regimes for allocation of accountability when the disposals are made with the help of AI models down to influencing the patient treatment process will be critical for clinching the confidence of clinicians and the general public. It is policymakers, particularly regulatory authorities, to keep abreast with AI researchers and healthcare providers to develop responsive policies that, on the one hand, encourage innovative research and, on the other, protect the interests of the patients and the ethical usage of AI.

### Cross-Disciplinary Research and Innovation

That is why continuous interdisciplinary cooperation will be needed to attain responsible and efficient applications of LLMs in healthcare settings. The use of clinical data is highly challenging in a number of ways, best addressed by a diverse group of professionals that include AI specialists, medical practitioners, data protection officers, ethicists, and policymakers. Interdisciplinary cooperation will enable scientists to comprehend the real problems and difficulties of the representatives of the healthcare system and the need to fulfill the necessary legal obligations. The identified collaborations can also foster a process of creating high-quality LLMs that are ethically appropriate and responsive to patient needs. Furthermore, interdisciplinary work may help address the lack of actionable guidelines faster, prompt introducing more practitioners-friendly interfaces, and perpetually update models based on practical experience data. Finally, continued integrated teamwork will facilitate the growth of LLMs serving its maximum capacity for changing healthcare delivery, safety and ethically responsible practice.

## 8. Conclusion

### Synthesis of Challenges and Mitigations

This review has presented major concerns in training large language models (LLMs) with clinical data and discussed various techniques for handling such complications. General

clinical information designed for use in developing LLMs has its own challenges that stem from the nature of clinical data which is complex, sensitive and not always standardized in quality. Privacy consideration is among the main challenges as this exposes patient identity and fails to meet legal requirements such as HIPAA or GDPR; new methods of anonymization and privacy-friendly approaches like differential privacy and federated learning are critical. Also, factors such as accuracy and representation prejudice give LLMs confining and prejudiced big data patterns that can distort clinical analysis or resolutions.

From a technical standpoint, the highly computationally demanding process involved in handling and training LLMs on large clinical datasets is a limitation to scalability and access to the models. The third is interpretability, where when models are opaque, clinicians might lose trust in them or may not be able to defend choices made by AI systems in places where such systems recommend treatments for patients. As a result of this review, the following current best practices that concern these challenges have been outlined: Explainable AI (XAI) strategies, domain adaptation approaches, and collaborative structures that consider data security and, at the same time, data availability. Altogether, these strategies may afford the foundation required for thoughtfully regulating LLMs and creating the model prerequisite that guarantees the proficient implementation of models derived from clinical data sets into care structures.

**Research Implications and Pathways**

These implications indicate more work and cooperation across disciplines, maintaining the ethical presence of LLMs trained in clinical data. First, more robust frameworks for privacy and data security must be created to sustain patient's confidence and adherence to the new regulations. It remains unclear whether and to what extent novel concepts in privacy-enhancing technologies, for example, homomorphic encryption or secure multi-party computation, might be suitable to ensure that clinical data is used in AI models without necessarily being disclosed to third parties. Moreover, incorporating high-quality and other steady clinical datasets is important for establishing reliable LLMs. Additional steps to further improve specifically the quality and scope of this study are as follows: Setting the same global standards related to data preprocessing and terminology assurance and preserving consistent

quality assurance measures would significantly increase the overall performance of models and decrease the potentiality of biased predictions.

This review also signals the imperative of enhancing the establishment of more task-specific, efficient models for clinical purposes. Since general-purpose LLMs require intense resources, research on limited, specific LLMs could serve useful, perhaps less costly solutions to be deployed in real-time in clinical settings. Furthermore, as LLMs are taking a bigger role in decision-making in the healthcare sector, these issues related to liability, transparency, and accountability will also have to be addressed. These pathways stress the urgent need to work on different levels, integrating the technical progress with an ethical vision that would link fast AI development with the principles of patient-focused medicine.

**Final Reflections on the Future of LLMs in Healthcare**

It is evident that the use of LLMs can drastically change the field of healthcare, providing new opportunities for diagnostics, individualized treatment and utilization of patient's time. But careful use of clinical data achieves this potential, there is the need to deal with the risks posed by using clinical data. To capture the benefits of innovative approaches for healthcare with little harm, it will be necessary to remain attentive to Ethical, Legal & Technical (ELT) issues when advancing and applying LLMs. Given the growing use of AI models in clinical practice, there is a need for all the stakeholders, including health care practitioners, artificial intelligence scientists and engineers, policymakers and ethicists, to come up with the appropriate regulatory frameworks that will at the same time, protect the welfare of patients but also foster scientific advancement in the development of Artificial Intelligence models in Medicine.

In conclusion, the future of LLMs in healthcare looks bright, but it is however important that proper steps are taken regarding the issues affecting clinical information. By reinforcing existing multidisciplinary research and maintaining the highest standards of ethical practice, the LLMs can be evolved concerning the patient's confidentiality, as well as increase the 'readability' for a clinician, and therefore, beneficial for the patient. Thanks to the integration of a responsible environment, healthcare is set to unlock unprecedented breakthroughs anchored in the future of AI applications.

**REFERENCES:**

1. Lee, J., et al. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234-1240.
2. Abd-Alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., ... & Sheikh, J. (2023). Large language models in medical education: opportunities, challenges, and future directions. JMIR Medical Education, 9(1), e48291.
3. Brown, T. B., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
4. Xu, X., Chen, Y., & Miao, J. (2024). Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. Journal of Educational Evaluation for Health Professions, 21.
5. Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. Indian Dermatol Online J. 2023 Oct 27;14(6):788-792. doi: 10.4103/idoj.idoj_543_23. PMID: 38099022; PMCID: PMC10718098.
6. Wang, F., & Preininger, A. (2019). AI in health: state of the art, challenges, and future directions. Yearbook of medical informatics, 28(01), 016-026.
7. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. Nature medicine, 29(8), 1930-1940.
8. Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, and Rajiv Ranjan. 2023. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. ACM Comput. Surv. 55, 9, Article 194 (September 2023), 33 pages. https://doi.org/10.1145/3561048
9. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
10. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211-407.
11. Bhat P, Shukla T, Naik N, Korir D, R P, Hawa H, Samrot AV, S R, A SS. Deep Neural Network as a Tool to Classify and Identify the 316L and AZ31BMg Metal Surface Morphology: An Empirical Study. *Engineered Science*. Published online 2023. http://dx.doi.org/10.30919/es1064

12. Jiang, X., Yan, L., Vavekanand, R., & Hu, M. (2024). Large Language Models in Healthcare Current Development and Future Directions.
13. Doe, A., & Garcia, K. (2021). Privacy-Preserving Methods for Machine Learning in Healthcare. Medical Data Privacy Journal, 14(1), 37-49.
14. Lu, Z., Peng, Y., Cohen, T., Ghassemi, M., Weng, C., & Tian, S. (2024). Large language models in biomedicine and health: current research landscape and future directions. Journal of the American Medical Informatics Association, 31(9), 1801-1811.
15. Ampavathi, A. (2022). Research challenges and future directions towards medical data processing. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 10(6), 633-652.
16. Chow, J. C., Wong, V., & Li, K. (2024). Generative Pre-Trained Transformer-Empowered Healthcare Conversations: Current Trends, Challenges, and Future Directions in Large Language Model-Enabled Medical Chatbots. BioMedInformatics, 4(1), 837-852.
17. Subramanian, C. R., Yang, D. A., & Khanna, R. (2024). Enhancing health care communication with large language models—the role, challenges, and future directions. JAMA Network Open, 7(3), e240347-e240347.
18. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172-180.
19. Zhou, H., Liu, F., Gu, B., Zou, X., Huang, J., Wu, J., ... & Clifton, D. A. (2023). A survey of large language models in medicine: Progress, application, and challenge. arXiv preprint arXiv:2311.05112.
20. Khabibullaev, T. (2024). Navigating the Ethical, Organizational, and Societal Impacts of Generative AI: Balancing Innovation with Responsibility. Zenodo. https://doi.org/10.5281/zenodo.13995243
21. Tan, Z., & Jiang, M. (2023). User modeling in the era of large language models: Current research and future directions. arXiv preprint arXiv:2312.11518.
22. SHUKLA, TANMAY. "Beyond Diagnosis: AI's Role in Preventive Healthcare and Early Detection." (2024)
23. Yao, Y., Zhang, J., Wu, J., Huang, C., Xia, Y., Yu, T., ... & Joe-Wong, C. (2024). Federated Large Language Models: Current Progress and Future Directions. arXiv preprint arXiv:2409.15723.

24. He, Y., Huang, F., Jiang, X., Nie, Y., Wang, M., Wang, J., & Chen, H. (2024). Foundation model for advancing healthcare: Challenges, opportunities, and future directions. arXiv preprint arXiv:2404.03264.