# Harmonizing foundation models in healthcare: A comprehensive survey of their roles, relationships, and impact in artificial intelligence's advancing terrain

# Harmonizing Foundation Models in Healthcare: A Comprehensive Survey of Their Roles, Relationships, and Impact in Artificial Intelligence's Advancing Terrain

Mohan Timilsina[a,*,1], Samuele Buosi[a], Muhammad Asif Razzaq[a], Rafiqul Haque[a], Conor Judge[b] and Edward Curry[a]

[a]Data Science Institute, Insight Centre for Data Analytics, University of Galway, Ireland
[b]College of Medicine, Nursing, Health Sciences, University of Galway, Ireland

ARTICLE INFO

*Keywords*:
Artificial Intelligence
Healthcare
Data
Prediction
Interoperability
Application

ABSTRACT

The lightning development of artificial intelligence (AI) has revolutionized healthcare, helping significant improvements in various applications. This paper provides a comprehensive review of foundation models in healthcare, highlighting their transformative potential in areas such as diagnostics, personalized treatment, and operational efficiency. We argue the key capabilities of these models, including their ability to process diverse data types such as medical images, clinical notes, and structured health records. Regardless their assurance, difficulties remain, including data privacy concerns, bias in AI algorithms, and the need for extensive computational resources. Our analysis identifies emerging trends and future directions, emphasizing the importance of ethical AI deployment, improved interoperability over healthcare systems, and the development of more robust, domain-specific models. Future research should focus on enhancing model interpretability, ensuring equitable access, and fostering collaboration between AI developers and healthcare professionals to maximize the advantages of these technologies.

## 1. Introduction

AI-driven healthcare applications have evolved with the introduction of foundation models such as *BioBERT* and *GatorTron*, which enhance clinical text analysis, biomedical information retrieval, and predictive diagnostics through large-scale domain-specific training [1, 2]. AI models can derive meaningful insights from healthcare data, which can be used to benefit various healthcare applications. A key driver of this progress is the foundation model—a pre-trained AI framework capable of performing many different tasks. This model helps enhance AI capabilities in healthcare, addressing the mismatch between AI's limitations and the diverse demands of healthcare practices. Developing a specialized foundation model(s) in healthcare increases the potential to improve a wider range of healthcare scenarios, leading to smarter and more sophisticated healthcare services [3]. Due to the diverse nature of real-world data, having robust representation capabilities is essential for adapting Health Informatics Frameworks to healthcare contexts. Multimodal models play a crucial role in addressing this diversity by integrating and analyzing heterogeneous healthcare data sources, such as medical imaging, clinical notes, and structured health records [4]. For instance, a multimodal foundation model can simultaneously process radiology images (e.g., X-rays, CT scans) along with patient history and laboratory test results to enhance diagnostic accuracy and treatment planning. This comprehensive approach enables a more holistic understanding of patient health and supports personalized healthcare interventions.

Foundation models address many of the existing limitations of traditional AI in healthcare, such as their inability to generalize across diverse data types and handle complex, multimodal data. By leveraging large-scale pretraining on heterogeneous datasets, foundation models can bridge the gap between AI's task-specific constraints and healthcare's demand for comprehensive, adaptable solutions. For instance, foundation models can integrate medical imaging, clinical notes, and genomic data to provide holistic insights that would otherwise require multiple specialized systems. This capability enhances diagnostic accuracy, enables personalized treatment, and improves interoperability across

mohan.timilsina@insight-centre.org (M. Timilsina); samuele.buosi@insight-centre.org (S. Buosi); m.asif.razzaq@insight-centre.org (M.A. Razzaq); rafiqul.haque@insight-centre.org (R. Haque); conor.judge@universityofgalway.ie (C. Judge); edward.curry@insight-centre.org (E. Curry)
ORCID(s):

healthcare systems. Due to the diverse nature of real-world data, having robust representation capabilities is essential for adapting Health Informatics Frameworks to healthcare contexts. Developing dynamic neural network structures, such as the increasingly popular attention mechanisms such as the Transformer model, seems to be a promising approach [5]. Unlike traditional models that struggle with sequential dependencies and long-range relationships, Transformers utilize self-attention mechanisms to capture intricate dependencies within and across data modalities such as clinical notes, medical imaging, and genomic data [6]. This capability allows for a more holistic understanding of patient health by linking information from disparate sources, improving diagnostic accuracy and decision-making processes. Furthermore, Transformers facilitate efficient handling of missing or noisy data common in healthcare records by dynamically assigning importance to relevant features while disregarding less informative ones. Their scalability and adaptability make them well-suited for personalized medicine, predictive analytics, and automated medical documentation. For example, in radiology, Transformer models can analyze imaging data in conjunction with textual reports to generate comprehensive diagnostic insights, ultimately enhancing clinical workflows and patient outcomes [7].

The foundation models in healthcare specifically in AI domain are achieving significant success across various sub-fields: Language, Vision, Bioinformatics, and multi-modality. Large Language Models (LLMs) are trained on extensive medical language data, have impressed with their capabilities in medical text processing and dialogues. Although they generate excitement about benefits to patients and clinicians, they also raise concerns about potential risks [8, 9]. The foundation models in the Vision domain are gaining attention for their effectiveness in medical imaging. They are adaptable across different modalities, organs, and tasks, showing significant promise in medical scenarios [10]. Bioinformatics foundation models now enable accurate protein structure prediction, gene sequence analysis, and molecular interaction modeling, facilitating advancements in computational biology and drug discovery. These models offer a pathway to unlock complex biological data, aiding research [11]. Finally, Multimodal foundation models integrate information from various modalities, allowing them to interpret different medical data types and perform tasks that require multi-modal analysis. This approach is essential for generalist healthcare foundation models, enhancing their versatility in clinical contexts [12]. Collectively, the aforementioned foundation models are a powerful resource for addressing complex clinical issues, streamlining healthcare practices, and driving innovation in the healthcare industry [13].

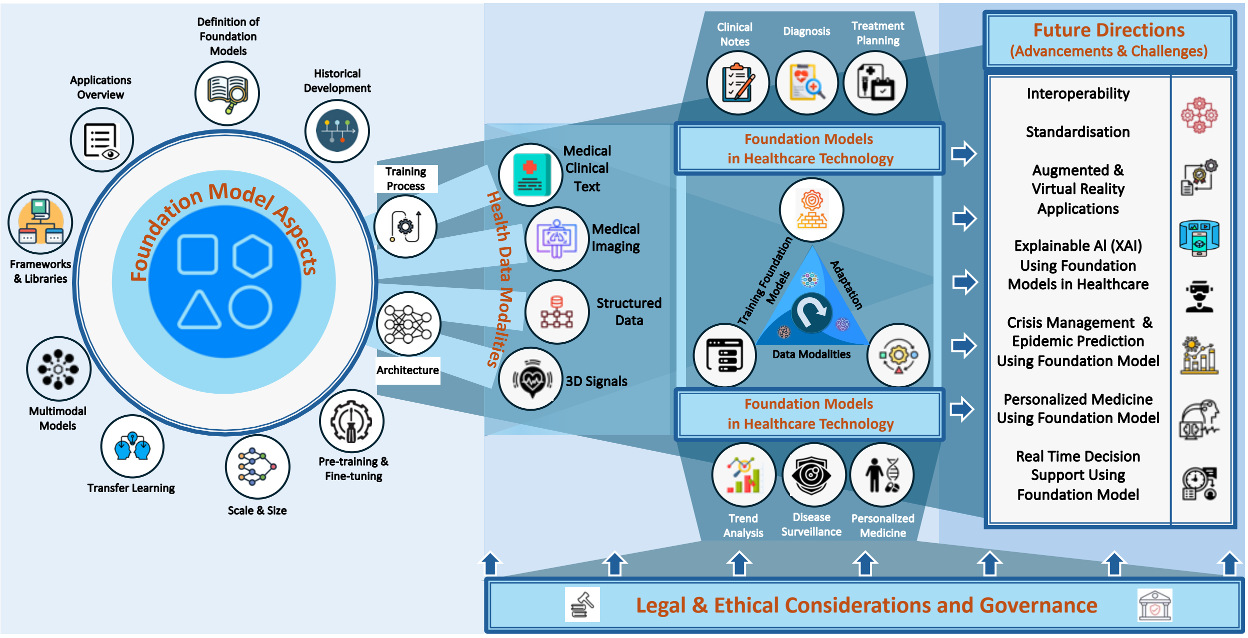


**Figure 1:** An Overview: Foundational models in healthcare, Basics, Modalities, Future Directions (Advancement & Challenges) with Legal & Ethical Considerations along with Governance

Figure 1 provides a comprehensive overview of foundational models in healthcare, illustrating the basic concepts, modalities, advancements, and challenges, all within the context of governance, legal, and ethical considerations. The

| **Aspect** | **Our literature review** | Shentong et al. [14] | Sizhen et al. [15] | Chen et al. [16] | Chen et al. [17] | Liu et al. [18] | Chen et al. [19] | Wasserthal et al. [20] | Butoiet al. [21] | Qiu et al. [22]et | **Positive Aspect of Foundation Model in Healthcare** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Comprehensive Scope | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | + (Broader coverage of various healthcare domains) |
| Applications Coverage | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | + (Wider range of healthcare applications) |
| Methodological Depth | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | + (Incorporates interdisciplinary aspects and governance issues) |
| Interdisciplinary Aspects | ✓ | - | - | - | - | - | - | - | - | - | + (Addresses challenges and s olutions across disciplines) |
| Governance and Ethical Considerations | ✓ | - | - | - | - | - | - | - | - | - | + (Detailed discussion on legal and ethical considerations) |
| Future Directions | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | + (Explores future advancements and potential innovations) |
| Use Cases and Examples | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | + (Diverse range of use cases demonstrating versatility) |

**Table 1**
Comparative Analysis of Literature on Foundation Models in Healthcare

left section of the figure outlines the foundational model basics, including the historical development, architectures, and applications such as transfer learning and multi-modal models. The central section highlights the diverse data modalities used in healthcare, such as medical clinical text, medical imaging, structured data, and 3D (Three-Dimensional) signals, which are integral to training foundational models in healthcare technology. This section also emphasizes the implementation of these models in areas like clinical notes, diagnosis, and treatment planning. The right section focuses on advancements and challenges, detailing the importance of interoperability, standardization, augmented and virtual reality applications, and the integration of explainable AI (XAI) for using foundation models in healthcare. Additionally, it addresses crisis management and epidemic prediction, personalized medicine, and real-time decision support. The entire diagram is underpinned by a foundation of governance and legal and ethical considerations, emphasizing their crucial role in the development and deployment of these technologies. This visual summary effectively encapsulates the multifaceted aspects of foundational models in healthcare, from basic principles to practical applications and the critical governance framework necessary for their ethical and legal implementation.

Table 1 shows the qualitative comparison between our literature survey and these major competitors. Our survey stands out for its comprehensive scope, offering a broader coverage of various healthcare domains and providing a more holistic view compared to the specific focus on gene and protein-related tasks in papers like GeneBERT [14] and CodonBERT [15]. Our review paper addresses a wider range of healthcare applications, enhancing its versatility. It delves deeper into interdisciplinary aspects and governance issues, presenting a well-rounded discussion on the deployment of foundation models in healthcare. Unlike the referenced papers, which primarily focus on their specific domains, this work addresses challenges and solutions across different healthcare disciplines. It includes detailed discussions on legal and ethical considerations, crucial for real-world applications, and less emphasized in the referenced works. Our work also explores future advancements and potential innovations across a broader spectrum of healthcare technologies and showcases a diverse range of use cases, demonstrating the versatility and potential of foundation models in healthcare. These positive aspects highlight the broader, more comprehensive, and interdisciplinary nature of our literature survey.

# 2. Foundation Model Aspects

## 2.1. Definition of Foundation Models

Foundation models are large-scale machine learning models trained on broad and diverse datasets, enabling them to generalize across a wide range of downstream tasks. These models leverage self-supervised learning at scale to capture complex patterns in data and can be adapted to various applications with minimal task-specific training. The term 'foundation model' was first introduced by Stanford's Center for Research on Foundation Models (CRFM) in 2021 [23], emphasizing their role in advancing AI capabilities across multiple domains, including healthcare, finance, and scientific research.

Foundation models are flexible AI systems that can support various applications across domains [23]. These large-scale pre-trained models serve as the fundamental framework for a multitude of AI tasks, eliminating the need for task-specific training [23]. Their advent has transformed the landscape of artificial intelligence, offering an adaptable and effective solution to address diverse challenges [23]. They have been developed across a range of modalities—including

DALL-E and Flamingo for images, MusicGen for music, and RT-2 for robotic control [23]. ChatGPT [23] and GPT-4 [24] are trained on broad, diverse datasets that include, but are not limited to, medical texts. Although recent studies have explored their utility in generating medical summaries and answering clinical queries, they lack the specialized optimization for biomedical text mining, clinical diagnostics, and structured medical data processing that characterizes dedicated healthcare foundation models [25, 26, 27].

## 2.2. Historical Development

The concept of Foundation Models, a class of AI systems that learn general-purpose representations from large amounts of data, has evolved significantly over the past few decades, building upon the foundations of earlier research in machine learning and deep learning. This section provides an overview of the historical development, highlighting key milestones and influential works that have shaped the field.

- **Early Neural Networks (1940s-1980s)** The roots of Foundation Models can be traced back to early work in artificial neural networks (ANNs) by researchers like Warren McCulloch and Walter Pitts in 1943. Although not explicitly focused on learning general representations, these initial efforts laid the groundwork for understanding the potential of interconnected nodes to process information [28]. However, due to computational limitations and lack of large-scale data, progress was slow.

- **Connectionism and Backpropagation (1980s-1990s)** In the 1980s, researchers like David Rumelhart, Geoffrey Hinton, and Ronald Williams introduced backpropagation, a key algorithm for training multi-layer neural networks [29]. This breakthrough allowed for more efficient learning of complex representations, paving the way for the development of deep learning. However, these early systems were often task-specific and did not demonstrate the ability to learn general-purpose knowledge.

- **Large Scale Pretraining (2000s-2010s)** The rise of large-scale datasets like ImageNet in 2009 [30] marked a turning point. Researchers began pre-training deep neural networks on these massive datasets, which led to the emergence of models like AlexNet [31], that achieved groundbreaking results in image classification. This shift towards unsupervised learning demonstrated the potential for Foundation Models to learn general representations.

- **Deep Learning and Transfer Learning (2010s-2020s)** The deep learning era saw the development of models like VGG, ResNet, and BERT that achieved state-of-the-art performance across various tasks by leveraging pre-trained representations [32, 33, 34]. These models demonstrated the value of foundation learning in transfer learning, where a model's knowledge can be adapted to new tasks with minimal fine-tuning.

- **Foundation Models: The Modern Era (2020s-present)** Foundation Models, a revolutionary advancement in Artificial Intelligence (AI), are poised to transform our interactions with computers and the world [35]. These models are created through self-supervised learning on vast amounts of unlabeled data, enabling them to grasp patterns and relationships with remarkable accuracy [35]. They excel in tasks such as image classification, natural language processing, and question-answering [35].

  Foundation Models form the basis of generative AI, empowering them to generate text, music, and images by predicting the next item in a sequence based on a given prompt [35]. The future of Foundation Models is bright, driven by factors like the availability of extensive datasets, advancements in computing infrastructure, and the growing demand for AI applications [35].

  A notable milestone in this trajectory is Google's release of LLM Jurassic-1 Jumbo in 2022, boasting an unprecedented 1.75 trillion parameters, making it the largest language model to date [35]. Additionally, OpenAI's unveiling of DALL-E 2 in 2023 marked a breakthrough in text-to-image generation, as it produces realistic visual content from textual descriptions, thereby expanding the horizons of visual content creation [35].

  Looking ahead, Foundation Models hold the potential to significantly disrupt AI development by scaling up deep learning through the augmentation of model size and the breadth and volume of training data [36]. These models consistently achieve state-of-the-art performance across various domains, including natural language processing and computer vision, often requiring minimal further adaptation to excel in diverse tasks [36].

| Key milestones and influential works | Strengths | Weaknesses |
|---|---|---|
| Early Neural Networks (1940s-1980s) | Early efforts in artificial neural networks (ANNs) laid the groundwork for understanding the potential of interconnected nodes to process information. | Progress was slow due to computational limitations and lack of large-scale data. |
| Connectionism and Backpropagation (1980s-1990s) | Introduction of backpropagation algorithm enabled more efficient learning of complex representations, paving the way for deep learning. | Early systems were often task-specific and did not demonstrate the ability to learn general-purpose knowledge. |
| Large Scale Pretraining (2000s-2010s) | Rise of large-scale datasets like ImageNet marked a turning point, leading to the emergence of models like AlexNet that achieved groundbreaking results in image classification. | Computational limitations and lack of large-scale data hindered progress in the early stages. |
| Deep Learning and Transfer Learning (2010s-2020s) | Development of models like VGG, ResNet, and BERT demonstrated the value of transfer learning, where pre-trained representations can be adapted to new tasks with minimal fine-tuning. | Progress in the early stages was slow due to the lack of efficient algorithms for training multi-layer neural networks. |
| Foundation Models: The Modern Era (2020s-present) | Foundation Models, created through self-supervised learning on vast amounts of unlabeled data, consistently achieve state-of-the-art performance across various domains. They hold the potential to significantly disrupt AI development by scaling up deep learning through the augmentation of model size and training data. | Early models lacked the ability to learn general-purpose representations and struggled with scalability due to computational limitations |

**Table 2**
Key Applications of AI-based Healthcare Models across Different Domains

In tracing the historical development of Foundation Models, we observe a progression marked by both remarkable strengths and notable weaknesses. As depicted in Table 2, the evolution from early neural networks to the modern era of Foundation Models showcases a trajectory of advancements characterized by significant milestones and intrinsic challenges. This analysis illuminates the nuanced landscape of AI research, underscoring the iterative nature of innovation and the ongoing quest to address inherent limitations while harnessing the transformative potential of emerging technologies. Building upon this historical evolution, the training process of foundation models is a critical step in realizing their full potential, as it involves leveraging vast amounts of diverse healthcare data and advanced computational techniques. This process ensures that these models can generalize effectively across different medical domains, leading to improved accuracy, efficiency, and adaptability in real-world healthcare applications.

## 2.3. Training Process

In the context of foundation models, their training methodology involves rigorous and sophisticated procedures that harness advanced computational resources and large-scale datasets [33]. As per He et al. [33], these models are meticulously honed on extensive repositories such as ImageNet [30], which serve as the cornerstone for their deep learning architecture.

The training process commences with the acquisition and meticulous preprocessing of diverse data sources, including public image databases, web crawl content, or specialized datasets tailored to the specific task [37]. This data is standardized and cleaned to ensure a consistent learning environment [37].

The hardware infrastructure employed in these endeavors is typically characterized by high-performance computing systems, often incorporating Graphics Processing Units (GPUs) for parallel processing, enabling efficient optimization of model parameters through backpropagation [31]. Techniques like batch normalization and gradient accumulation are also commonly employed to optimize the training process [31].

The core of the training cycle involves iterative updates of the model's weights over multiple epochs, guided by a loss function calculated on a validation set [38]. This process could be computationally intensive, requiring weeks or even months to fully converge due to the need for the models to learn intricate patterns and relationships within the data [38].

## 2.4. Architecture

Foundation models, such as GPT-3 (Generative Pre-trained Transformer 3.) [39], BERT (Bidirectional Encoder Representations from Transformers) [34], and RoBERTa (Robustly Optimized BERT Approach) [40], typically employ the Transformer architecture, which is known for its effectiveness in handling sequential data. The Transformer architecture, introduced by Vaswani et al. in 2017n [5], has revolutionized the field of natural language processing.

The Transformer architecture consists of several key components:

- **Input Embeddings:** The input to the Transformer is a sequence of tokens, which are first converted into vectors, then combined with positional encodings to account for the order of the words [5].
- **Encoder:** The encoder is composed of a stack of identical layers, each with two sub-layers: a multi-head self-attention mechanism, and a position-wise fully connected feed-forward network. The self-attention mechanism allows the model to weigh the relevance of each word in the sequence when encoding a particular word [41].
- **Decoder:** The decoder also consists of a stack of identical layers. In addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack. This allows every position in the decoder to attend over all positions in the input sequence [5].
- **Output Linear Layer and Softmax:** The decoder's output is transformed to the final output sequence by a linear layer followed by a softmax operation [5].

The Transformer's ability to handle long-range dependencies in sequential data, its scalability, and its capacity for parallelization have made it the architecture of choice for foundation models [39]. The Transformer architecture's multi-head attention mechanism is especially beneficial for clinical text because it can capture long-range dependencies within complex medical documents, such as Electronic Health Records and physician notes, where important information may be scattered throughout the text. By catering multiple angles on the same term, multi-head attention helps the model understand distinct medical terminology and context-specific meanings, which is very important for accurate information extraction and clinical decision support. Furthermore, this mechanism helps optimized parallel processing, allowing the model to capture large-scale clinical datasets rapidly, making it applicable for real-time healthcare applications. Its ability to generalize across diverse clinical text formats ranging from structured data to unstructured narratives—further underscores its adaptability and effectiveness in varied healthcare settings.

In assessing the Transformer architecture from the perspective of its architectural components, as depicted in Table 3, we discern a comprehensive evaluation of its strengths and weaknesses. This analysis illuminates the intricate design considerations inherent in the Transformer framework, underscoring its efficacy in handling sequential data while also highlighting potential challenges that warrant further investigation.

## 2.5. Pre-training and Fine-tuning

Foundation models, a significant paradigm in machine learning, undergo a two-step process: pre-training and fine-tuning [39].

Pre-training refers to training a model from scratch and is a very time consuming process and resource intensive [42]. It is commonly used in domains such as natural language processing and computer vision, where models leverage large corpora of text or images to capture essential features. In the **pre-training** phase, models learn from a vast corpus of data [43, 44]. This data is typically unlabelled, encompassing a wide range of topics and styles. The models, often deep neural networks, learn to predict or reconstruct parts of the input data, thereby capturing rich representations of the data distribution [34].

The second phase is **fine-tuning**, where models adapt to specific tasks [45]. Fine-tuning means continuing to train a previously trained model—the model weights are obtained from the previous training process. Because the model already has certain knowledge from pre-training, fine-tuning typically requires fewer resources (e.g., data and compute) than pre-training [42]. Here, the pre-trained model is further trained on a smaller, task-specific dataset. This dataset is usually labelled, providing explicit examples of the desired output. The fine-tuning process adjusts the model parameters to minimize the discrepancy between the model's predictions and the actual labels [46].

To clearly illustrate the notion of pre-training and fine-tuning in the healthcare domain, BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining) is a well-known example of this process [1]. BioBERT is first pre-trained on large-scale biomedical corpora, such as PubMed and PMC articles, to develop a deep understanding of biomedical language and terminology. After pre-training, it is fine-tuned on specific downstream tasks such as named entity recognition (e.g., identifying diseases and drugs), relation extraction, or question answering in clinical settings. This two-step approach significantly improves the model's ability to comprehend domain-specific text and provide accurate predictions in real-world healthcare applications.

| Architecture aspect | Strengths | Weaknesses |
| --- | --- | --- |
| Input Embeddings | Input tokens are efficiently converted into vectors and augmented with positional encodings, enabling the model to capture the sequential nature of the data effectively. | Dependency on tokenization quality may affect the model's ability to accurately represent the input sequence. |
| Encoder | The encoder utilizes multi-head self-attention and fully connected feed-forward networks to encode input sequences, allowing the model to capture complex relationships and dependencies within the data. | Increased computational complexity due to the stacking of identical layers may lead to longer training times and higher resource requirements. |
| Decoder | The decoder employs similar mechanisms as the encoder while also incorporating multi-head attention over the encoder's output, facilitating the generation of output sequences based on the encoded information. | Introducing additional sub-layers in the decoder increases model complexity, potentially leading to overfitting, especially with limited training data. |
| Output Linear Layer and Softmax | The final output sequence is generated by transforming the decoder's output through a linear layer followed by a softmax operation, ensuring that the model's predictions are in the appropriate format for the given task. | The softmax operation may introduce computational overhead, particularly in scenarios with large output vocabularies, impacting inference speed. |

**Table 3**
Challenges and Limitations of Existing Healthcare AI Models

## 2.6. Scale and Size

The scale and size of a foundation model are crucial aspects to consider when evaluating its performance and applicability [47]. A foundation model's scale refers to the number of parameters it contains, which can range from tens of thousands to billions [34]. Larger models typically require more computational resources and memory to train and deploy [47], making them less feasible for deployment on resource-constrained devices. However, larger models have been shown to achieve better performance on certain tasks, such as language translation and image recognition [5]. The size of a foundation model, on the other hand, refers to its capacity to process and store data. Larger models can handle more complex and nuanced patterns in data, but may struggle with sparse or noisy data [48]. Smaller models, while more efficient, may be limited in their ability to capture subtle relationships between features [49]. The size of a foundation model is often measured by its number of parameters, which can range from tens of thousands to millions [50]. The trade-off between scale and size is a critical consideration when designing a foundation model. The choice of scale and size ultimately depends on the specific application and requirements of the foundation model.

## 2.7. Transfer Learning

Transfer learning is a powerful technique in machine learning that allows a model trained on one task to be reused or "transferred" for another related task [3, 51]. This approach is particularly beneficial in healthcare, where labeled data can be scarce and expensive to obtain [3, 51].

In the context of foundation models in healthcare, transfer learning plays a crucial role [3, 51]. Foundation models, such as GPT-4, are pre-trained on large amounts of diverse data, encoding vast knowledge and powerful emergent abilities [3, 51]. These models serve as the foundation for rapidly adapting to new tasks via transfer learning [3, 51].

However, the levels of features to be reused are problem-dependent, and uniformly fine-tuning all layers of pretrained models may be suboptimal [51]. This insight has partly motivated the recent differential transfer learning strategies, which treat the layers in the pretrained models differentially [51].

The application of transfer learning in foundation models can advance our understanding of complex diseases [3]. By reusing and fine-tuning appropriate bottom layers and directly discarding the remaining layers, not only superior performance can be achieved, but also compact models for efficient inference can be obtained [51].

Despite the promising prospects, there are challenges in applying transfer learning to foundation models in healthcare. These include issues related to data privacy, model heterogeneity, and model ownership [3]. Addressing these challenges is crucial for the successful deployment of foundation models in healthcare [3].

In conclusion, transfer learning enables the reuse of pre-trained models, reducing the need for extensive labeled data and training costs. However, careful consideration must be given to the levels of features to be reused and the challenges associated with data privacy and model ownership [3, 51].

## 2.8. Multimodal Models

Multimodal models are a significant advancement in the field of foundation models, particularly in healthcare [3, 52]. These models are capable of processing and integrating information from multiple modalities, such as text, images, and structured data [3, 52]. This ability is crucial in healthcare, where patient data often comes in various forms, including electronic health records, medical images, genomic data, and more [3, 52].

The development of multimodal foundation models has been driven by the need to leverage the rich, diverse data available in healthcare [3, 52]. These models can capture complex relationships across different data types, leading to more accurate predictions and insights [3, 52]. For instance, a multimodal model could use both medical imaging data and patient history to predict disease progression [3, 52].

However, developing effective multimodal foundation models poses several challenges [3, 52]. These include the need for large, diverse, and high-quality datasets, the complexity of integrating different data types, and the requirement for robust and interpretable models [3, 52]. Despite these challenges, the potential benefits of multimodal foundation models in healthcare are immense [3, 52]. Multimodal Foundation Models represent a promising direction in healthcare. They leverage the power of Foundation Models and the richness of multimodal data to provide more accurate and comprehensive insights [53, 54, 55]. However, further research is needed to overcome the challenges associated with their development and deployment [3, 52].

Multimodal models enhance clinical analysis by integrating **structured data, 2D images, and 3D signals** into a unified representation. Each data type undergoes specialized feature extraction—structured data (e.g., lab results, demographics) is processed using *transformer-based encoders*, medical images (e.g., echocardiograms, MRIs) are analyzed with *convolutional neural networks (CNNs)*, and 3D signals (e.g., ECGs, echocardiograms) are captured through *time-series or volumetric analysis networks*. The extracted features are then aligned in a **shared latent space** using *attention mechanisms*, ensuring meaningful cross-modal relationships. A multimodal model can combine **structured data** (e.g., blood pressure, cholesterol) to assess baseline risks, **2D echocardiograms** to detect motion anomalies, and **3D MRI scans** to provide volumetric heart assessments. By jointly analyzing these sources, the model **improves diagnostic accuracy**, enabling **early detection** and **personalized treatment planning**. This holistic approach enhances patient care and optimizes healthcare decision-making. By jointly analyzing these sources, the model **improves diagnostic accuracy**, enabling **early detection** and **personalized treatment planning**. This holistic approach enhances patient care and optimizes healthcare decision-making. Figure 2 illustrates the multimodal processing pipeline, where structured data, 2D images, and 3D signals undergo feature extraction before being fused in a shared latent space for clinical predictions, such as cardiac disease diagnosis.

## 2.9. Frameworks and Libraries

In the field of artificial intelligence (AI), particularly in the development and implementation of Foundation Models, several frameworks and libraries have been instrumental. These tools provide a wide range of functionalities, from data preprocessing and model training to evaluation and deployment, thereby streamlining the process of model development and application.

TensorFlow is one of the most popular frameworks used in the development and training of Foundation Models [56]. Developed by the Google Brain team, TensorFlow provides a comprehensive library for designing and training neural networks. It supports a wide range of neural network architectures, including deep learning models, and provides functionalities for automatic differentiation and GPU acceleration, which are crucial for training large-scale models.

PyTorch, developed by Facebook's AI Research lab, is another widely used framework for Foundation models [57]. Like TensorFlow, PyTorch provides comprehensive functionalities for neural network design and training, and supports automatic differentiation and GPU acceleration. However, PyTorch is known for its dynamic computational graph, which allows for more flexibility in model design and training.

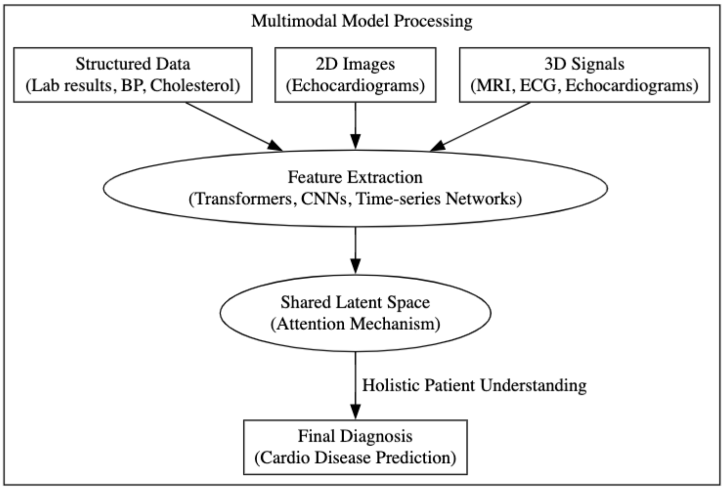


**Figure 2:** Multimodal model processing pipeline for cardiac disease prediction.

Keras is a high-level neural networks API that is widely used for prototyping and building deep learning models [58]. It is user-friendly, modular, and extensible, and can run on top of TensorFlow, allowing users to leverage the power of TensorFlow with a simpler interface.

For transformer-based models like BERT, GPT, and their variants, the Hugging Face's Transformers library is commonly used [59]. It provides thousands of pre-trained models, which can be fine-tuned for various tasks, and is deeply interoperable between TensorFlow 2.0 and PyTorch, allowing users to switch between the two frameworks seamlessly.

Scikit-learn is a machine learning library in Python that is widely used for data mining and data analysis [60]. It features various classification, regression, and clustering algorithms, and provides tools for model selection and evaluation, preprocessing, and more.

These frameworks and libraries have played a significant role in the development and application of Foundation Models, providing researchers and practitioners with powerful tools to design, train, and apply these models.

## 2.10. Applications Overview

Foundation Models have been increasingly applied in the healthcare sector, demonstrating significant potential in various applications ranging from disease diagnosis to drug discovery.

One of the most prominent applications of Foundation Models in healthcare is in medical imaging. These models have been used to analyze medical images such as X-rays, CT scans, and MRI scans to detect and diagnose diseases [61]. For instance, deep learning models have been developed to detect lung cancer from CT scans with a performance that matches or even surpasses that of radiologists [62].

In genomics, Foundation Models have been used to predict gene expression, understand genetic variations, and identify disease-associated genes [63]. These models can learn complex patterns in genomic data, enabling the identification of genetic markers for diseases and the prediction of treatment responses.

Foundation Models have also found applications in electronic health records (EHR). They have been used to predict patient outcomes, identify disease risks, and personalize treatment plans based on patient history [64]. These models can handle the complexity and heterogeneity of EHR data, providing valuable insights for clinical decision-making.

In drug discovery, Foundation Models have been used to predict drug-target interactions, design novel drug molecules, and optimize drug properties [65]. These models can accelerate the drug discovery process, reducing the time and cost associated with traditional methods.

Lastly, in public health, Foundation Models have been used for disease surveillance, outbreak prediction, and health policy planning [66]. These models can analyze large-scale public health data to monitor disease trends, predict future outbreaks, and inform health policy decisions.

Foundation Models have demonstrated significant potential in various healthcare applications, providing valuable tools for disease diagnosis, treatment planning, drug discovery, and public health management.

# 3. Types of Foundation Models Applied in Healthcare

## 3.1. Health Data Modalites

Data is naturally multimodal in health domain [67]. The diagram 3 presents a structured overview of the application of a foundational model within the field of healthcare, illustrating its utility as a central processing unit for a wide array of data types. It begins with the assimilation of data from four primary sources: Clinical Text, Medical Imaging, Structured Data, and use of 3D Signals in Healthcare Analytics [68]. Clinical Text includes Electronic Health Records (EHRs), which provide a detailed patient history; physician's notes, which capture clinical observations and decisions; discharge summaries that encapsulate the results of hospital stays; and medical transcriptions, translating recorded reports into text documents. Medical Imaging encompasses radiology images such as X-rays and CT scans vital for diagnostic processes, MRI scans for their detailed anatomical views, and ultrasound images often used for real-time viewing of bodily organs [69]. Additionally, pathology slide images offer critical insights at the cellular level which are integral to histological analyses.

Structured Data encompasses information that is inherently quantifiable and well-organized, vital for systematized patient care management. This includes laboratory test results that underpin many diagnostic and therapeutic decisions; vital signs, which provide real-time indicators of a patient's physiological status; demographic data, which is fundamental for epidemiological studies; and detailed medication records essential for patient treatment continuity. The category of use of 3D Signals in healthcare analytics comprises complex data from echocardiograms offering dynamic heart function visualization, advanced 4D imaging techniques providing temporal and spatial resolution, electrophysiological recordings like ECGs and EEGs crucial for cardiovascular and neurological assessment, and molecular structure data which is key in biochemical and pharmacological research [70]. This heterogeneous data is then utilized to train the foundational model, a complex artificial intelligence system designed to derive meaningful patterns and interpretations from diverse and high-dimensional data [67].

The adaptation phase, depicted on the right side of the diagram, represents the subsequent stage where the foundational model is tailored to specific healthcare tasks [71, 72]. These include Diagnosis, where the model applies its learned intelligence to identify pathologies; Treatment Planning, which benefits from the model's predictive power to inform therapeutic strategies; Trend Analysis, utilizing the model's analytical strength over large datasets to identify health trends; Disease Surveillance, leveraging the model for public health monitoring; and Personalized Medicine, which uses the model's capacity for integrating genetic, lifestyle, and clinical data to tailor individual care [73]. Symbolized by an intricate neural network, the foundational model stands as a sophisticated framework, converting extensive biomedical data into practical insights and advancements, thus playing a pivotal role in enhancing healthcare outcomes through technology-driven methodologies [1, 74].

## 3.2. Medical Clinical Text

In healthcare, clinical narratives are crucial for documenting patient histories and evaluations, enriching clinical decision-making processes. These narratives are typically recorded by clinicians in free-text clinical notes during patient interactions and are often supplemented with metadata codes from the International Classification of Diseases (ICD). ICD codes provide a standardized framework for logging diagnoses and procedures, serving essential functions such as billing facilitation and patient outcome prediction [75, 76, 77]. Despite their significance, the manual assignment of ICD codes is a time-consuming and error-prone process, driving research into automated coding solutions since the 1990s [78]. The complexity of clinical text, which may include irrelevant content, misspellings, idiosyncratic abbreviations, and a vast range of medical terminology, further complicates automation efforts. This challenge is amplified by the sheer volume of codes, with over 15,000 in ICD-9 and more than 140,000 in the combined ICD-10-CM and ICD-10-PCS systems [79]. Despite these obstacles, natural language processing (NLP) has proven effective in extracting essential information from narratives [80], showcasing its ability to unveil significant evidence within clinical texts for further analysis [81]. Traditionally, evidence identification often utilized rule-based methods, signaling continuous technological advancements in clinical documentation [82]. The study by [83] discusses the evolution of clinical text mining with foundational models, transitioning from rule-based methods to advanced machine learning and deep learning approaches. It underscores the significance of these models in interpreting complex clinical texts, offering insights into their application in healthcare and the challenges of integrating such technologies into

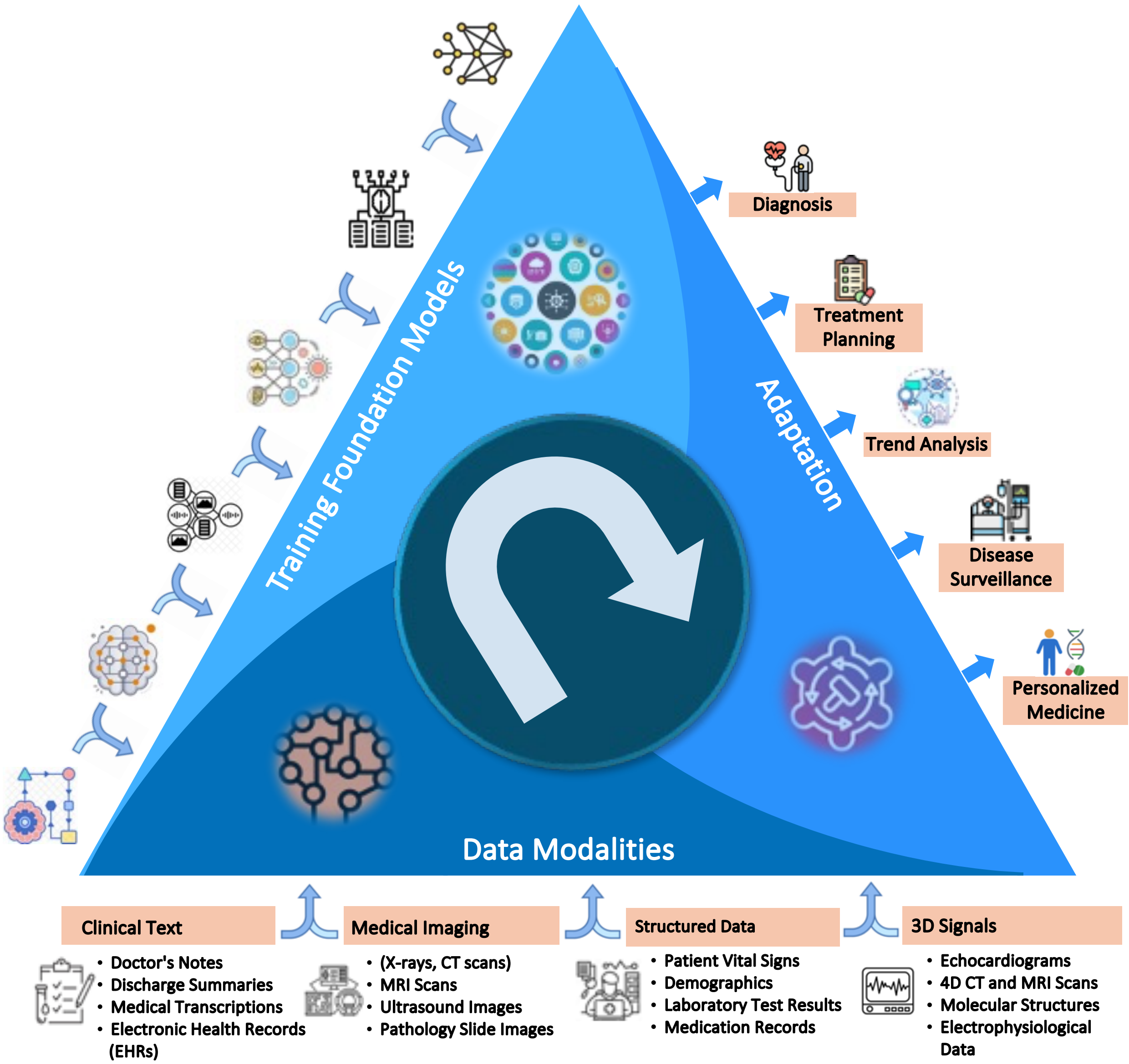


**Figure 3:** The role of a foundation model as an adaptive healthcare engine, integrating diverse medical data modalities including clinical text, medical imaging, structured data, and 3D signals. This model processes multimodal healthcare data to enhance various downstream tasks such as diagnosis, treatment planning, disease surveillance, and personalized medicine by harmonizing disparate data sources, foundation models enable more precise, data-driven decision-making in clinical settings.

health systems for improved care and research outcomes. Leveraging foundational models alongside NLP techniques emerges as a promising strategy for enhancing the efficiency and accuracy of entity extraction from patient health records [67]. These pre-trained models, refined on extensive datasets, excel in general-purpose data representation and can be optimized for specific tasks through task-specific fine-tuning. Among the available pre-trained models, such as BERT [84], GPT-3 [85], and RoBERTa [40], the choice depends on specific domain and task requirements. Similarly, the advent of foundational models like GPT-3, with its vast parameter count and extensive training corpus, has demonstrated remarkable cross-domain performance. In the biomedical field, models like BioBERT [1] and PubMedBERT [86], derived from PubMed literature, and NVIDIA's BioMegatron [87], have pushed the envelope with their specialized parameter sizes. Despite these advances, the clinical domain's unique challenges stem from the sensitive nature of data and computational demands. ClinicalBERT [84], trained on clinical narratives from the MIMIC-III dataset, highlights NLP's potential to enhance healthcare delivery and outcomes by leveraging clinical texts. MIMIC-III [88] is a large, de-identified *electronic health record (EHR)* dataset containing structured patient data, clinical notes, and vital signs from intensive care units. PubMed, on the other hand, serves as an extensive repository of *biomedical literature*, helping models learn medical terminologies, disease relationships, and evidence-based reasoning [89]. These datasets enhance the generalization ability of foundation models by exposing them to real-world clinical text and structured medical records.

Similarly, Xi et al. [90] developed GatorTron [2], a large clinical language model using text from de-identified Health clinical notes, PubMed articles, and Wikipedia. **GatorTron** was developed using a diverse set of biomedical and clinical sources. The inclusion of structured and unstructured medical text allows GatorTron to effectively process clinical narratives, making it more suitable for real-world healthcare applications compared to generic NLP models [2]. **BioBERT** was trained on large-scale biomedical literature, specifically *PubMed abstracts and full-text articles*. This dataset provides a domain-specific corpus tailored for biomedical text processing, enabling BioBERT to outperform general language models in medical terminology comprehension and literature-based information retrieval [1]. These datasets were selected to ensure exposure to diverse biomedical language, improving the models' ability to process, interpret, and generate clinically relevant text. The integration of structured EHR data and unstructured biomedical literature strengthens their performance in downstream applications such as *clinical note summarization, disease prediction, and medical literature retrieval*.

Other notable biomedical text-based models include MedBERT [91], BioMedCLIP [92], BlueBERT [93], and ChemBERT [94], summarized in Table 4. The performance of the models listed in Table 4 was evaluated using well-established metrics suited to their respective tasks. Specifically, **GatorTron, BioMegatron, BioBERT, PubMedBERT, ClinicalBERT, and Med-BERT** were primarily assessed using *precision, recall, and F1-score*, which are widely used in biomedical Named Entity Recognition (NER), relation extraction, and text classification tasks. These metrics ensure a balanced evaluation by considering both the correctness of predicted entities (precision) and the model's ability to retrieve relevant ones (recall), with F1-score providing a harmonic mean of the two. For **BiomedCLIP**, which is designed for multimodal biomedical text-image learning, performance was evaluated using the *recall at top-k* metric. This assesses the model's ability to retrieve the correct medical images or text descriptions within the top-k predictions, making it a suitable evaluation method for image-text alignment tasks.These evaluation criteria were selected to provide a standardized and meaningful comparison of model capabilities across different biomedical NLP and multimodal applications.

The *Strengths/Focus* column in Table 4 highlights key differentiators of biomedical foundation models based on their training data, architecture, and application domains. Strengths were identified by analyzing data diversity, model parameters, and specialization in clinical or biomedical tasks. For instance, *GatorTron* was trained on extensive clinical notes, while *BioBERT* and *PubMedBERT* leveraged large-scale biomedical literature [2, 1]. Architectural specialization also plays a role, as seen in *BioMegatron*, a large-scale biomedical language model optimized for complex NLP tasks [87]. Application domains further distinguish these models: *ClinicalBERT* and *Med-BERT* focus on EHR-based predictions, *BioBERT* and *PubMedBERT* support literature-based knowledge retrieval, and *BiomedCLIP* enables multimodal learning with medical images and text [92, 1]. These criteria, derived from published research and real-world evaluations, highlight each model's strengths in clinical NLP, biomedical text mining, and multimodal healthcare analytics.

## 3.3. Medical Imaging

Among the diverse applications of deep learning in healthcare, medical imaging stands out as a key component of diagnostics and is regarded as the most significant and promising area [61, 95]. In the broader realm of computer vision, the development and testing of essential self-supervised models [96, 97] often rely on benchmarks such as the ImageNet Large Scale Visual Recognition Challenge and other large datasets [98, 99]. These models, known as 'foundation models', are known for their ability to adapt to a wide range of downstream tasks after being pretrained on extensive data collections [67, 100, 101]. For instance, the study by Ke et al.[102] demonstrated that ImageNet pretraining significantly enhances performance in chest x-ray classification. Likewise, foundation models have the potential to serve as central repositories of medical knowledge, trained on diverse sources and modalities of data in medicine[103, 104, 105]. However, in computational pathology, the efficacy of current pretrained models is hindered by the limited size and diversity of pretraining data, primarily consisting of primary cancer histology image slides [106, 96, 107]. The scarcity of quality, publicly available annotations in medical imaging limits the training of large-scale deep learning models

| Model | Data Used | Number of Parameters | Strengths/Focus |
|---|---|---|---|
| GatorTron | Health clinical notes, PubMed articles, Wikipedia, and MIMIC III | 8.9 billion | General-purpose biomedical, handles clinical notes and general text |
| ClinicalBERT | MIMIC-III dataset | 110 million | Clinical text analysis, optimized for EHR data |
| BioBERT | PubMed literature | 110 million | Biomedical research, strong understanding of scientific concepts |
| PubMedBERT | PubMed literature | 110 million | Similar to BioBERT, focused on biomedical research |
| BioMegatron | PubMed-derived text | 345 million to 1.2 billion | Large-scale biomedical language model, good for complex tasks |
| Med-BERT | Large-scale EHR data | 137B parameters | Focused on clinical prediction tasks like disease prediction |
| BlueBERT | Public biomedical text data (PMC) | 137B parameters | Similar to BioMegatron, large-scale for complex tasks |
| BiomedCLIP | Paired biomedical images and text (PMC) | 1.5B parameters (multimodal) | Processes both text and images for tasks like image retrieval or VQA |
| ChemBERT | PubChem text and molecule information | 137B parameters | Understanding chemical structures and properties from text descriptions |

**Table 4**
Summary of Clinical and Biomedical Text Foundation Models and Data Used

for clinical applications. Labeling image data manually is a labor-intensive task for medical professionals, though providing a few distinct samples is practical and aligns with medical training methods. Vision foundation models, trained on diverse large-scale image datasets, could form the foundation for developing medical applications [108].

Recently the Segmentation Anything Model (SAM), introduced by Kirillov et al. [109]), shows promise in natural images but underperforms in complex medical tasks like pancreas and spine segmentation [110]. Different pathologies require specific imaging modalities for diagnosis and treatment, including X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, and Positron Emission Tomography (PET). Modality-specific foundation models are tailored to specific types of imaging, such as radiology images (X-ray, CT, MRI, Ultrasound) [111], 3D image stacks from CT and MRI [19], or specific modalities like X-ray [112] , CT [113], endoscopy [10], and pathology [114, 96]. These models learn features pertinent to their specific modality, such as bone density and tissue contrast for CT models, or soft tissue contrast and motion for MRI models. One of the important aspect of the application of the foundation data model is that it can be customized for specific medical organs [115, 116] or targeted diagnostic tasks like segmentation [21]. This approach is designed to tackle the challenges of organ appearance variability in medical images and the wide array of clinical tasks dependent on image analysis. In medical image analysis, a general AI system is defined as a multitask, multimodal platform that can handle a variety of tasks across multimodal images featuring various organs and diseases. These tasks include classification, detection, segmentation, and registration, all managed by a unified set of model weights [71, 117, 108].

Leveraging large-scale natural image datasets, vision foundation models provide a robust baseline for medical imaging analysis. When adapted to modality-specific foundation models, they can exploit the distinct characteristics of each modality, leading to models optimized for particular imaging tasks. Although this results in higher accuracy and efficiency for specific modalities, these models may not generalize well across different imaging types. Based on the above described models, Table 5 demonstrates the summary of the medical image analysis by the foundation model:

### 3.4. Structured Data

Structured electronic health records are a primary source of input for disease prediction, offering rich and well-structured information that reflects patient disease progression and represent one of the most valuable resources for health data analysis [118]. Foundation models built on this structured data have demonstrated their ability to predict the risk of hospital readmission [119], select appropriate treatments [75], and diagnose rare diseases [120]. Wornow et al. [121] evaluated a foundation model using structured data for clinical prediction tasks and task adaptation. Similarly, MediTab [122], a foundation model trained for medical tabular data prediction, scales up model pre-training with blood testing data. However, there are only a handful of examples of foundation model applications in the healthcare field for structured data, including genomics [123] and tabular data [122]. Multi-task learning and transfer learning have also been applied in the medical domain for electronic health-based predictive modeling [124].

Levin et al. [125] discovered utility in transfer learning in a semi-supervised setting to explore few-shot classification for medical diagnosis. Wiest et al. [129] studied liver cirrhosis using 500 patient medical histories from the MIMIC IV dataset and compared large language models (LLMs) with prompt engineering approaches. Med-BERT [91], which adapts the BERT framework, originally developed for the text domain, to the structured EHR domain, substantially

| Aspect | Strengths | Weaknesses |
|---|---|---|
| **Training Data** | Trained on diverse and extensive data collections from large datasets like ImageNet [98, 99]. | Limited size and diversity of pretraining data in computational pathology, often restricted to specific types of image slides [106, 96, 107]. |
| **Performance** | Enhanced performance in tasks like chest x-ray classification due to pretraining on large datasets [102]. | Underperformance in complex medical tasks such as pancreas and spine segmentation [110]. |
| **Adaptability** | Ability to adapt to a wide range of downstream tasks and modalities in medical imaging [67, 100, 101]. | Poor generalization across different imaging types when using modality-specific foundation models [111, 19]. |
| **Modality-Specific Models** | Tailored to specific imaging modalities (X-ray, CT, MRI, Ultrasound) and capable of learning features pertinent to each modality [111, 19]. | Requires separate models for each modality, which may not efficiently leverage knowledge from other modalities [112, 113]. |
| **Customization** | Can be customized for specific medical organs or diagnostic tasks such as segmentation [115, 116, 21]. | Customized models may face challenges due to the variability in organ appearance across medical images [115]. |
| **Multitasking** | Supports a multitask, multimodal platform managing classification, detection, segmentation, and registration [71, 117, 108]. | The complexity of managing multiple tasks and modalities can lead to challenges in integration and operational efficiency [71, 117]. |

**Table 5**
Strengths and Weaknesses of the Medical Vision Foundation Model

| Aspect | Strengths | Weaknesses |
|---|---|---|
| Prediction | Predicts risk of hospital readmission [119], selects appropriate treatments [75], and identifies rare diseases [120]. | Requires large labeled datasets for effective training. |
| Scalability | Scales up model pre-training with blood testing data [122]. | Demands high computational resources and storage capacity. |
| Transfer Learning | Applies multi-task and transfer learning to improve predictive modeling in healthcare [124]. Levin et al. explored few-shot classification for medical diagnosis [125]. | Needs extensive labeled datasets for effective adaptation. |
| Accuracy | Enhances prediction accuracy in structured EHR data, demonstrated by Med-BERT [91]. | Risks introducing biases if trained on non-representative data. |
| Adaptation | Adapts pre-trained text-based models for structured health data analysis, as seen with Med-BERT [91] and Heath-LLM [126]. | May struggle with preserving clinical context during adaptation. |
| Innovation | Introduces novel approaches like Event Stream GPT for continuous-time event prediction [127] and TEMED-LLM for tabular data extraction [128]. | Requires further validation to ensure clinical reliability. |

**Table 6**
Comparison of Model Architectures in Healthcare Applications

improves prediction accuracy. Heath-LLM combines large-scale feature extraction and medical knowledge into a large model to enhance disease prediction [126]. McDermott et al. [127] introduced Event Stream GPT (ESGPT), an open-source library designed to streamline the end-to-end process for building GPTs for continuous-time event sequences in medical time-series data. TabLLM [130], applied in medical claim tabular data, outperformed ensemble-based models. For timestamped sequences of events in electronic health records and health insurance claims, the MOTOR foundation model is employed, claimed to be the first release of a medical foundation model for time-to-event prediction [131].

| Aspect | Strengths | Weaknesses |
|---|---|---|
| ECG Analysis | Foundation models can detect and predict cardiovascular diseases with high accuracy [132]. SONIC integrates vision models for stress recognition [133]. | Reliance on large annotated datasets for pre-training can be resource-intensive [138]. |
| EEG Analysis | SSL techniques reduce the need for large amounts of labeled data, lowering costs [134]. | Pre-training still requires substantial data, raising issues in settings with limited data availability [135]. |
| Molecular Structures | Capable of predicting properties, optimizing drug candidates, and generating novel molecules [139]. InstructMol aligns molecular structures with natural language for better interaction [140]. | Multimodal data integration is complex and requires advanced algorithms to effectively manage diverse data types [141]. |
| Biosignal Integration | GMAI and METS models show potential in integrating comprehensive patient histories for better diagnostics and treatment plans [71, 142]. | High computational costs and the complexity of integrating multiple modalities and large data volumes [143]. |

**Table 7**
Evaluation Criteria for AI Models in Healthcare

Bisercic et al. [128] proposed a methodology of Tabular Extraction for Enhanced Medical Diagnostics using Large Language Models (TEMED-LLM). This model utilizes LLMs in the extraction of tabular data from medical texts and reports based on predefined features. Based on the above described models, Table 6 demonstrates the summary of the structured medical data by the foundation model.

## 3.5. Use of 3D Signal in Healthcare Analytics

Machine learning models, particularly deep learning models, have shown promise in automatically detecting and predicting cardiovascular diseases from ECG signals. Foundation models trained on large datasets excel at recognizing patterns indicative of conditions such as arrhythmias, myocardial infarction, or heart failure [132]. SONIC [133] focuses on combining different vision foundation models for stress recognition from ECG signals. Self-supervised learning (SSL) approaches for training foundation models have been extensively applied to biosignal modalities such as PPG, ECG, and electroencephalogram (EEG) in both clinical and controlled experimental settings [134, 135]. These models require less labeled data compared to supervised models, significantly reducing the costs of experimental health studies [136]. Abbaspourazad et al. demonstrated the effectiveness of pre-training foundation models using large-scale ECG data collected from wearable consumer devices [137].

"MC-BEC [143] uses a multimodal dataset that combines key health indicators such as heart activity (ECG), blood flow signals (PPG), breathing patterns, and radiology reports. These data sources help assess foundation model performance in analyzing patient health conditions more effectively." The GMAI model [71] integrates a patient's entire history, including reports, waveform signals, laboratory results, genomic profiles, and imaging studies. From the time series perspective, METS [142] employs a trainable ECG encoder alongside a frozen language model to process paired ECG and clinical reports. Recent studies by Yamac et al. [138] and Bhaskarpandit et al. [144] have shown promising results in zero-shot ECG classification tasks, though they still rely on large-scale annotated data for pre-training using supervised learning.

In the field of chemistry and drug discovery, foundation models trained on vast chemical libraries can predict molecular properties, optimize drug candidates, and even generate novel molecules with desired characteristics [139]. InstructMol [140], a multimodal LLM, aligns molecular structures with natural language through an instruction-tuning approach, employing a two-stage training strategy that effectively combines limited domain-specific data with molecular and textual information. Molecule Foundation Models (MolFM) explore various approaches, including training in a single modality or jointly learning from multiple modalities. Research in the single modality domain spans molecule representations from diverse sources, such as 1D sequences [145], 2D molecular graphs [146], and

3D geometric conformations [141]. Integrating multiple types of data, including 3D signal data from ECG, EEG, and molecular structures, foundation models provide a comprehensive understanding of complex biological and chemical systems. These integrative models facilitate cross-disciplinary discoveries and enhance the predictive power of the models. Based on the models described above, Table 7 summarizes the analysis of 3D data by the foundation models.

# 4. Legal and Ethical Considerations and Governance

Fairness is a critical concern in the healthcare sector, and foundation models used in developing downstream applications must comply with legal and ethical regulations ensuring fairness. Historically, AI researchers have relied on the fair use doctrine to avoid copyright concerns about training data, assuming compliance with this legal principle when using copyrighted materials for transformative purposes [67]. However, recent legal challenges, such as lawsuits against Microsoft, GitHub, and OpenAI for using publicly available code without proper attribution, and Getty Images' lawsuit against Stable Diffusion AI for unauthorized use of its photographs, highlight the increasing scrutiny of fair use interpretations [147]. Legal frameworks like the *AI Act* and US law influence the development and use of foundation models, raising issues related to model training, liability for predictions, and data protection [148]. Foundation models often use health data scraped from the Internet, leading to significant legal implications under laws like the U.S. Computer Fraud and Abuse Act (CFAA) [149]. Privacy laws, such as GDPR and HIPAA, mandate strict data handling standards, and non-compliance can result in severe penalties. The California Consumer Privacy Act (CCPA) also raises questions about the removal of training data from models [150].

As foundation models become increasingly integrated into clinical decision-making, determining liability for errors remains a critical legal and ethical concern. Unlike traditional medical tools, these models rely on self-learning algorithms, making it difficult to attribute responsibility for incorrect or biased predictions [151]. Healthcare providers are often held accountable, as final clinical decisions must align with medical expertise and ethical standards, and failure to critically evaluate AI-generated recommendations may result in malpractice claims [152]. Some jurisdictions, such as the European Union's AI Act, are introducing strict liability rules for high-risk AI systems used in healthcare [153]. Additionally, hospitals and healthcare institutions share responsibility for AI validation, staff training, and compliance with medical regulations, ensuring safe AI deployment [154]. To mitigate risks, strategies such as human-in-the-loop AI decision-making, explainability mechanisms, and regulatory compliance frameworks (e.g., GDPR, HIPAA, EU AI Act) are essential to ensure accountability and patient safety [155]. As AI adoption grows, clear liability frameworks must balance innovation with legal and ethical responsibilities, ensuring that developers, healthcare professionals, and institutions share accountability while upholding high standards of medical ethics and patient care [156].

Ensuring data privacy in AI-driven healthcare applications requires strict compliance with the General Data Protection Regulation (GDPR), which classifies healthcare data as sensitive personal information (Article 9) and imposes strict safeguards on its processing [157]. Key principles include lawfulness, fairness, and transparency (Article 5), requiring explicit patient consent, data minimization, and clear disclosure of data usage [158]. Patients are granted rights such as access and data portability (Articles 15, 20)[1], allowing them to retrieve and transfer their health records, and the right to erasure (Article 17)[2], enabling them to request deletion of their personal data. Additionally, automated decision-making restrictions (Article 22)[3] prevent patients from being subjected to AI-driven medical decisions without oversight unless explicitly authorized [151]. To ensure compliance, organizations must integrate privacy-by-design measures, including data encryption, anonymization, federated learning, and synthetic data generation (Article 25)[4], while also conducting Data Protection Impact Assessments (DPIAs) (Article 35)[5] to evaluate privacy risks before deploying AI models [159]. Failure to comply with GDPR can result in severe financial penalties, reaching €20 million or 4% of global annual revenue[6].

Model bias presents a significant ethical and legal challenge in the deployment of foundation models [160] in healthcare. Bias can arise from imbalanced training datasets, where underrepresented populations such as racial and ethnic minorities, low-income groups, and older adults are inadequately represented. This can lead to disparities in healthcare outcomes, where AI models perform less accurately for marginalized groups, potentially reinforcing existing inequities. One major concern is **algorithmic bias** [161], which can manifest in diagnostic models trained primarily

[1] https://gdpr-info.eu/art-20-gdpr/
[2] https://gdpr-info.eu/art-17-gdpr/
[3] https://gdpr-info.eu/art-22-gdpr/
[4] https://gdpr-info.eu/art-25-gdpr/
[5] https://gdpr-info.eu/art-35-gdpr/
[6] https://gdpr.eu/fines/

on data from high-resource healthcare settings, making them less effective for underprivileged communities with different health profiles. [162].For instance, studies have shown that dermatology AI models trained predominantly on lighter skin tones perform poorly in detecting conditions on darker skin tones, magnifying disparities in dermatological care [163]. Similarly, predictive models used for disease risk assessment may overestimate or underestimate risks for certain populations if social determinants of health (e.g., access to healthcare, environmental factors) are not adequately incorporated [164, 165]. To mitigate these biases, regulatory frameworks such as the EU AI Act [166] and the U.S. FDA AI/ML [167] Action Plan emphasize the need for transparency, fairness audits, and continuous monitoring of AI models in healthcare applications. Bias mitigation strategies include algorithmic debiasing [168] techniques, diverse and representative dataset curation, and human-in-the-loop validation by diverse medical professionals. Additionally, explainable AI (XAI) methods can help detect and address biases by making model decisions more interpretable for healthcare providers [169]. Addressing bias is crucial to ensuring that foundation models do not perpetuate healthcare disparities but instead serve as tools for equitable and inclusive medical decision-making.

The creators of foundation models are responsible for their outputs, which can include biased results that perpetuate stereotypes or misinformation, posing legal risks especially in regulated sectors like healthcare [67]. Bias amplification in foundation models presents fresh legal challenges for its identification and mitigation, while issues like **'hallucination'** can lead to the dissemination of false information [170]. Hallucination refers to the generation of incorrect, misleading, or nonsensical information by AI models [171]. These threats have led to calls for stricter regulation [172], emphasizing safety, security, transparency, and accountability. The EU's AI Act, for instance, mandates comprehensive risk assessment and bias mitigation for foundation models [173]. From a governance perspective, sector-specific and general regulations, such as the AI Act, Data Act, and GDPR, should be implemented complementarily across jurisdictions to ensure robust governance throughout the AI value chain. Non-compliance with these varied legal standards can lead to fines, litigation, and restrictions on AI technology use, necessitating thorough legal due diligence by entities utilizing foundation models in healthcare to mitigate potential legal risks. The EU's AI Act also aims to balance innovation with protecting fundamental rights, particularly addressing the needs of marginalized communities to prevent exacerbating health disparities. This includes transparency in AI algorithms and supporting AI developments that aim to reduce health disparities and improve outcomes for underserved populations

# 5. Future Directions

## 5.1. Interoperability and standardisation in foundation model in health

Interoperability and standardization are critical elements in the implementation of foundation models in healthcare. These models, which leverage extensive pre-training on diverse data before being fine-tuned for specific tasks, offer immense potential for enhancing predictive analytics, diagnostics, and patient care management [67]. However, the effectiveness and scalability of these models are heavily contingent upon their ability to operate seamlessly across various healthcare systems and to adhere to universally accepted standards. Interoperability[7] in healthcare refers to the ability of different information technology systems, software applications, and data architectures to connect and communicate effectively without repetitive interventions [174]. This is crucial for foundation models, which must integrate data from disparate sources that often include electronic health records (EHRs), medical imaging databases, genomic databases, and wearable health devices. Effective interoperability ensures that these models can access and synthesize data across systems and platforms, facilitating comprehensive analytics and decision-making processes. For foundation models, standardization involves establishing common protocols for data formats, privacy regulations, and ethical guidelines [175]. Standardization ensures that data collected from multiple sources are compatible with the models, regardless of where or how the data were collected. This is particularly important in healthcare, where data security and patient privacy are paramount. Standards such as Health Level Seven International (HL7), Fast Healthcare Interoperability Resources (FHIR), and Digital Imaging and Communications in Medicine (DICOM) are often employed to facilitate this compatibility in clinical settings [176].

## 5.2. Augmented and Virtual Reality Applications

The integration of augmented and virtual reality (AR/VR) applications using foundation models in healthcare represents a groundbreaking shift towards more immersive and personalized medical treatment and training [177]. These technologies, powered by advanced AI, facilitate realistic simulations and interactive environments that can be used for surgical training, patient education, pain management, and therapy for mental health disorders. Foundation

[7] https://www.himss.org/resources-interoperability

models enhance AR/VR applications by providing a robust, data-driven backbone that can adapt and learn from vast amounts of medical data, improving the accuracy and effectiveness of simulations and interventions [178]. However, significant challenges remain, including the high cost of AR/VR technology development, the need for substantial computational resources, and potential risks related to data privacy [179]. Additionally, the computational requirements for running these advanced AI foundation models in real-time within AR/VR systems are enormous, necessitating significant hardware and processing power, which can escalate costs and limit accessibility. There is also the challenge of bias in foundation models, where systems trained on non-representative data could lead to disparities in healthcare delivery. Overcoming these challenges requires meticulous data management, investment in robust computational infrastructure, and continuous evaluation of the model's decisions to ensure fairness and accuracy.

### 5.3. Explainable AI (XAI) Using Foundation Models in Healthcare

The backbone of the foundation model is the neural networks. The neural networks "blackbox" nature makes understanding their behavior challenging. Therefore, elucidating the underlying rationale behind the outcomes of health foundation model is crucial for fostering trust in their healthcare applications [180]. A key upcoming task involves advancing theoretical research in machine learning, particularly focusing on foundational models. Analyzing the machine learning characteristics of these models will unveil distinctive patterns, empowering researchers to develop more sensible models and enhance research and development efficiency in the healthcare setting. In the healthcare sector, where the stakes of decision-making are particularly high, it is crucial for professionals to comprehend the reasoning behind AI-generated outputs. The development of models that are both interpretable and explainable continues to be a persistent challenge in this field [181]. However, current theories of representation, optimization, or generalization fall short for health foundation model due to their extensive parameters and data volumes, which far surpass the ideal settings for these theories [182]. Another promising avenue lies in discovering more effective means of explanation. Despite existing efforts employing techniques like heatmaps, including attention maps, class activation maps, and uncertainty maps, to elucidate individual behaviors, there's a pressing need to delve into higher levels of abstraction for explanatory evidence [180]. Furthermore, harnessing explainable health foundation models for scientific exploration, such as in drug discovery, shows promise as a future direction [183].

Given the critical nature of healthcare, explainability and transparency are critical to ensuring trust among clinicians, medical professionals, and regulatory bodies. To enhance interpretability, several explainability techniques have been developed, enabling better model understanding and clinical validation. Several techniques have been developed to enhance interpretability, including model-agnostic explainability methods, neural network-based interpretability approaches, and counterfactual explanations. Model-agnostic methods such as SHAP (Shapley Additive Explanations) quantify the contribution of each input feature to the model's prediction using game-theoretic principles. This approach is particularly useful in clinical decision support systems, where understanding the impact of patient biomarkers, symptoms, and lab results on AI-driven outcomes is essential [184]. Similarly, LIME (Local Interpretable Model-Agnostic Explanations) [185] simplifies complex AI models by approximating them with more interpretable models, such as linear regression, making it particularly valuable for electronic health record (EHR) analysis [186]. Beyond model-agnostic approaches, neural network-based interpretability techniques help uncover how deep learning models make predictions. Attention mechanisms, commonly used in transformer-based models like BERT and GPT, highlight key words or image regions that most influence the model's decision. In medical imaging, attention heatmaps assist radiologists in identifying abnormalities, improving diagnostic accuracy [5]. Additionally, counterfactual explanations add another dimension to AI interpretability by generating "what-if" scenarios that identify minimal input changes necessary to alter a model's decision. This is particularly useful in clinical risk assessment and treatment planning, helping healthcare providers explore alternative diagnoses and interventions to ensure that AI-driven recommendations align with best clinical practices [187]. By incorporating these explainability techniques, foundation models in healthcare can become more transparent, interpretable, and trustworthy, ultimately fostering better adoption and confidence among medical professionals.

### 5.4. Crisis Management and Epidemic Prediction Using Foundation Model

The integration of foundation models in crisis management and epidemic prediction is filled with substantial challenges that necessitate careful inspection. One primary issue is the substantial volume of data required for training these models effectively [67]. In the context of a rapidly unfolding epidemic, collecting accurate and comprehensive data can be a formidable task, particularly in regions where healthcare systems are underdeveloped or overwhelmed by the crisis. This scarcity of data can hinder the model's ability to generate reliable predictions and actionable insights.

Furthermore, when dealing with sensitive health information, data privacy and security concerns are paramount [3]. The handling, storage, and processing of medical data must adhere strictly to regulatory standards such as HIPAA in the United States or GDPR in Europe, to prevent unauthorized access and data breaches [188, 181]. The stakes are incredibly high, as any compromise of patient data can lead to serious legal repercussions and a loss of public trust. Additionally, the inherent complexity of foundation models contributes to the "black box" dilemma. This term refers to the opaque nature of the algorithms, where the processes they use to arrive at conclusions are not clear or understandable to users. This opacity can erode trust among healthcare professionals, who may be reluctant to base their decisions on recommendations whose rationale they cannot comprehend or evaluate. The challenge, therefore, is to enhance the transparency and interpretability of these models, ensuring that they can be audited and understood by those who rely on them for making critical decisions in managing health crises.

Despite these challenges, the use of foundation models in epidemic prediction and crisis management also offers significant opportunities. These models can analyze large datasets quickly, identifying patterns and trends that might be missed by human analysts [3]. This capability allows for real-time monitoring and forecasting of epidemic outbreaks such as COVID [189], providing healthcare systems with critical information that can be used to allocate resources efficiently and strategize intervention plans proactively. Moreover, foundation models can integrate diverse data types—from clinical data to social media content—offering a more holistic view of an epidemic's spread and potential impacts. This integration enables more dynamic and effective responses, potentially saving lives and reducing the economic fallout from health crises [190, 191].

## 5.5. Personalized Medicine Using Foundation Model

Personalized medicine stands on the brink of transformation, with foundation models offering unprecedented opportunities to tailor healthcare to individual needs [192, 44]. These advanced AI models can integrate and analyze vast datasets, including genetic information, electronic health records, and lifestyle data, to predict individual disease risks and response to treatments [193]. This capability enables healthcare providers to design highly specific therapies that are optimized for each patient's unique biological makeup, potentially enhancing treatment efficacy and patient outcomes [194, 189]. Furthermore, foundation models can accelerate drug discovery by simulating and predicting the effects of compounds on various genetic profiles, thus reducing the time and cost associated with clinical trials. As these models learn from diverse and extensive data sources, they continually improve, promising increasingly accurate and effective healthcare solutions.

Despite their potential, the integration of foundation models into personalized medicine also presents significant challenges. One of the primary concerns is data diversity and bias—models trained on datasets that lack variability in ethnicity, gender, or socio-economic status may offer predictions that are less accurate or biased against underrepresented groups [195]. Ensuring that foundation models are equitable and just in their functionality is crucial. There is also the pressing issue of data privacy and security: handling sensitive personal health information necessitates stringent protections to prevent breaches and misuse [196]. Lastly, the regulatory landscape for such advanced AI applications is still evolving, posing further hurdles for implementation in clinical settings where safety and efficacy are paramount [151]. These challenges underscore the need for robust frameworks and continuous oversight to ensure that the benefits of foundation models in personalized medicine are realized without compromising patient trust or safety.

## 5.6. Real Time Decision Support Using Foundation Model

While foundation models hold promise for enhancing real-time decision support in healthcare, their practical implementation is still evolving. These models have the potential to process and analyze large datasets than traditional methods, but they are currently not capable of doing so instantaneously [67, 3]. They can integrate diverse data types, including electronic health records, imaging studies, and real-time biometric monitoring, which could one day help clinicians gain a more holistic view of a patient's condition. This integration is especially useful in critical care scenarios, where rapid decision-making based on comprehensive data analysis can significantly impact outcomes [197]. For instance, foundation models are being explored for their potential to predict severe conditions such as septic shock earlier than conventional approaches, which could allow for timely interventions [198]. However, these applications are in the developmental phase and require further validation. Additionally, while these models may assist in identifying patterns that are not immediately obvious to human clinicians, their current use must be closely monitored to ensure they augment clinical expertise effectively and safely, rather than replacing the nuanced judgment of experienced healthcare professionals. Similarly, deploying foundation models for real-time decision support in healthcare is very

| Feature | Interoperability | AR/VR | XAI | Crisis Management | Personalized Medicine | Real Time Support | Studies |
|---|---|---|---|---|---|---|---|
| Data Diversity | ✓ | - | - | + | ✓ | + | Bommasani et al. [67] |
| Scalability | ✓ | - | + | ✓ | ✓ | + | He et al. [3] |
| Data Integration | ✓ | ✓ | - | ✓ | ✓ | ✓ | Zhou et al. [189] |
| Real-time Processing | - | ✓ | - | + | - | ✓ | Sutton et al. [197] |
| Explainability | - | - | ✓ | - | - | - | Sokol et al. [180] |
| Cost-effectiveness | - | - | + | - | + | - | Qayyum et al. [179] |
| Data Privacy | ✓ | - | - | ✓ | ✓ | ✓ | Sajid et al. [188] |
| Regulatory Compliance | ✓ | - | - | ✓ | ✓ | + | Gerke et al. [151] |
| Clinical Efficacy | + | + | - | + | ✓ | + | Nori et al. [198] |
| Bias Reduction | - | - | + | - | - | - | Selbst and Barocas [195] |
| Robustness | + | - | + | - | + | - | Briganti et al. [181] |
| Training Efficiency | + | - | - | - | ✓ | - | Lee et al. [192] |
| Predictive Accuracy | ✓ | + | - | + | ✓ | + | Verkijk et al. [193] |
| Adaptability | + | + | - | - | + | - | Radford et al. [44] |
| Transparency | - | - | ✓ | - | - | - | Yuan et al. [182] |

**Table 8**
Future Trends and Opportunities in Healthcare AI

challenging [148]. The primary concern is the accuracy and reliability of the model outputs, which must achieve a high standard given the potential consequences of incorrect or delayed medical decisions. The "black box" nature of many AI models also poses a significant barrier, as medical practitioners must understand and trust the AI's recommendations to integrate them into their clinical workflows.
Overall, Table 8 provides the multifaceted landscape of foundation model technologies in healthcare, offering a clear roadmap for addressing current challenges, leveraging opportunities, and focusing on future work to advance the field.

# 6. Conclusion

Foundation model in healthcare present a transformative opportunity across various facets of medical practice. These models have demonstrated significant potential in disease diagnosis, treatment planning, personalized medicine, and crisis management by leveraging extensive pre-training on diverse and heterogeneous datasets. By integrating data from multiple modalities, including electronic health records, medical imaging, genomic data, and real-time sensor inputs, these models can enhance clinical decision-making and improve patient care outcomes. Our review highlights the ability of foundation models to bridge information gaps, streamline workflows, and provide more comprehensive insights that can benefit both clinicians and patients. However, despite these advancements, several challenges must be addressed to ensure the safe, effective, and ethical deployment of these technologies in real-world healthcare settings.

Among the most pressing concerns are issues related to data privacy, regulatory compliance, interoperability, and standardization. The integration of AI-driven systems into clinical workflows necessitates robust frameworks that safeguard patient information while allowing for efficient data utilization. Ensuring transparency and explainability in AI models is particularly crucial, as healthcare professionals must be able to interpret and validate AI-generated recommendations to foster trust and improve adoption. Furthermore, addressing algorithmic bias remains a key priority, as disparities in training data can result in unequal healthcare outcomes across different patient populations. In response to these challenges, future work should focus on developing multimodal data-sharing frameworks that enable secure and federated learning approaches, allowing AI models to improve accuracy while preserving patient confidentiality. Additionally, advances in explainability techniques, such as interpretable neural networks and feature attribution methods, should be explored to enhance model transparency and reliability in clinical applications. Establishing standardized fairness evaluation metrics will be instrumental in ensuring AI systems perform equitably across diverse demographic groups. Moreover, interdisciplinary collaboration between AI researchers, medical practitioners, policymakers, and regulatory agencies is essential to formulate clear governance policies, ethical AI regulations, and standardized deployment guidelines. By integrating technological innovation with ethical considerations and practical implementation strategies, foundation models in healthcare can advance patient-centered care, optimize medical workflows, and ultimately revolutionize the healthcare landscape, making AI-assisted medicine more reliable, accessible, and effective for all.

## Acknowledgment

We would also like to acknowledge Science Foundation Ireland (SFI/12/RC/2289_P2) for supporting this research.

## CRediT authorship contribution statement

**Mohan Timilsina:** M.T wrote and revised the original manuscript.. **Samuele Buosi:** S.M wrote and revised the manuscript.. **Muhammad Asif Razzaq:** M.A.R wrote and revised the manuscript.. **Rafiqul Haque:** R.H revised the manuscript.. **Conor Judge:** C.J revised the manuscript.. **Edward Curry:** E.C revised the manuscript..

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this work.

## References

[1] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[2] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Mona G Flores, Ying Zhang, et al. Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. *arXiv preprint arXiv:2203.03540*, 2022.

[3] Yuting He, Fuxiang Huang, Xinrui Jiang, Yuxiang Nie, Minghao Wang, Jiguang Wang, and Hao Chen. Foundation model for advancing healthcare: Challenges, opportunities, and future directions. *arXiv preprint arXiv:2404.03264*, 2024.

[4] Felix Krones, Umar Marikkar, Guy Parsons, Adam Szmul, and Adam Mahdi. Review of multimodal machine learning approaches in healthcare. *Information Fusion*, 114:102690, 2025.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-December, pages 5999–6009, 2017.

[6] Sumit Madan, Manuel Lentzen, Johannes Brandt, Daniel Rueckert, Martin Hofmann-Apitius, and Holger Fröhlich. Transformer models in biomedicine. *BMC Medical Informatics and Decision Making*, 24(1):214, 2024.

[7] Kerstin Denecke, Richard May, and Octavio Rivera-Romero. Transformer models in healthcare: A survey and thematic analysis of potentials, shortcomings and risks. *Journal of Medical Systems*, 48(1):23, 2024.

[8] Kai He, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *arXiv preprint arXiv:2310.05694*, 2023.

[9] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

[10] Zhao Wang, Chang Liu, Shaoting Zhang, and Qi Dou. Foundation model for endoscopy video analysis via large-scale self-supervised pre-train. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 101–111. Springer, 2023.

[11] Hugo Dalla-Torre, Liam Gonzalez, Javier Mendoza-Revilla, Nicolas Lopez Carranza, Adam Henryk Grzywaczewski, Francesco Oteri, Christian Dallago, Evan Trop, Bernardo P de Almeida, Hassan Sirelkhatim, et al. The nucleotide transformer: Building and evaluating robust foundation models for human genomics. *bioRxiv*, pages 2023–01, 2023.

[12] Julián N Acosta, Guido J Falcone, Pranav Rajpurkar, and Eric J Topol. Multimodal biomedical ai. *Nature Medicine*, 28(9):1773–1784, 2022.

[13] Bobby Azad, Reza Azad, Sania Eskandari, Afshin Bozorgpour, Amirhossein Kazerouni, Islem Rekik, and Dorit Merhof. Foundational models in medical imaging: A comprehensive survey and future vision. *arXiv preprint arXiv:2310.18689*, 2023.

[14] Shentong Mo, Xi Fu, Chenyang Hong, Yizhen Chen, Yuxuan Zheng, Xiangru Tang, Zhiqiang Shen, Eric P Xing, and Yanyan Lan. Multi-modal self-supervised pre-training for regulatory genome across cell types. *arXiv preprint arXiv:2110.05231*, 2021.

[15] Sizhen Li, Saeed Moayedpour, Ruijiang Li, Michael Bailey, Saleh Riahi, Lorenzo Kogler-Anele, Milad Miladi, Jacob Miner, Dinghai Zheng, Jun Wang, et al. Codonbert: Large language models for mrna design and optimization. *bioRxiv*, pages 2023–09, 2023.

[16] Bo Chen, Xingyi Cheng, Pan Li, Yangli-ao Geng, Jing Gong, Shen Li, Zhilei Bei, Xu Tan, Boyan Wang, Xin Zeng, et al. xtrimopglm: unified 100b-scale pre-trained transformer for deciphering the language of protein. *arXiv preprint arXiv:2401.06199*, 2024.

[17] Yiqun Chen and James Zou. Genept: A simple but effective foundation model for genes and cells built from chatgpt. *bioRxiv*, 2023.

[18] Tianyu Liu, Tianqi Chen, Wangjie Zheng, Xiao Luo, and Hongyu Zhao. scelmo: Embeddings from language models are good learners for single-cell data analysis. *bioRxiv*, pages 2023–12, 2023.

[19] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019.

[20] Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5), 2023.

[21] Victor Ion Butoi, Jose Javier Gonzalez Ortiz, Tianyu Ma, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Universeg: Universal medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21438–21451, 2023.

[22] Jianing Qiu, Jian Wu, Hao Wei, Peilun Shi, Minqing Zhang, Yunyun Sun, Lin Li, Hanruo Liu, Hongyi Liu, Simeng Hou, et al. Visionfm: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence. *arXiv preprint arXiv:2310.04992*, 2023.

[23] Wikipedia contributors. Foundation model, 2024.

[24] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[25] Songtao Tan, Xin Xin, and Di Wu. Chatgpt in medicine: prospects and challenges: a review article. *International Journal of Surgery*, 110(6):3701–3706, 2024.

[26] Siru Liu, Aileen P Wright, Barron L Patterson, Jonathan P Wanderer, Robert W Turer, Scott D Nelson, Allison B McCoy, Dean F Sittig, and Adam Wright. Using ai-generated suggestions from chatgpt to optimize clinical decision support. *Journal of the American Medical Informatics Association*, 30(7):1237–1245, 2023.

[27] Jannatul Ferdush, Mahbuba Begum, and Sakib Tanvir Hossain. Chatgpt and clinical decision support: scope, application, and limitations. *Annals of Biomedical Engineering*, 52(5):1119–1124, 2024.

[28] Warren S. McCulloch and Walter M. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

[29] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by backpropagation. In *Neural Networks: Tricks of the Trade*, pages 59–67. Springer-Verlag, 1986.

[30] Jia Deng, Wei Dong, Richard Socher, Lijun Li, Li Fei-Fei, and Daniel Hoiem. Imagenet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009:1–8, 2009.

[31] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:770–778, 2016.

[34] Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:4171–4186, 2019.

[35] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu, and Lichao Sun. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. *arXiv preprint arXiv:2302.09419*, 2023.

[36] Johannes Schneider. Foundation models in brief: A historical, socio-technical focus. *arXiv preprint arXiv:2212.08967*, 2022.

[37] Yiding Sun, Feng Wang, Yutao Zhu, Wayne Xin Zhao, and Jiaxin Mao. An Integrated Data Processing Framework for Pretraining Foundation Models. *arXiv preprint arXiv:2402.16358*, 2024.

[38] Yoshua Bengio Ian Goodfellow and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[39] Tom B Brown, Gretchen Krueger, Benjamin Mann, Amanda Askell, Ariel Herbert-voss, Clemens Winter, Daniel M Ziegler, Alec Radford, and Sam Mccandlish. Language Models are Few-Shot Learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[40] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*, 2019.

[41] Zhouhan Lin, Minwei Feng, Cicero Nogueira Dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.

[42] Chip Huyen. Ai engineering: Building applications with foundation models, 2025.

[43] Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. A neural probabilistic language model. *Advances in Neural Information Processing Systems*, 3:1137–1155, 2001.

[44] Halvard Fausk and Daniel C. Isaksen. Improving Language Understanding by Generative Pre-Training. *Homology, Homotopy and Applications*, 9(1):399–438, 2007.

[45] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 1:328–339, 2018.

[46] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, pages 843–852, 2017.

[47] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361*, 2020.

[48] Wenhan Wang, Yanzhou Li, Anran Li, Jian Zhang, Wei Ma, and Yang Liu. An empirical study on noisy label learning for program understanding. In *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE)*, pages 879–879. IEEE Computer Society, 2024.

[49] Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410*, 2016.

[50] Radford Alec, Wu Jeffrey, Child Rewon, Luan David, Amodei Dario, and Sutskever Ilya. Language Models are Unsupervised Multitask Learners | Enhanced Reader. *OpenAI Blog*, 1(8):9, 2019.

[51] L. Peng, H. Liang, G. Luo, T. Li, and J. Sun. Rethinking transfer learning for medical image classification. In *British Machine Vision Conference (BMVC)*, 2023.

[52] Yunkun Zhang, Jin Gao, Zheling Tan, Lingfeng Zhou, Kexin Ding, Mu Zhou, Shaoting Zhang, and Dequan Wang. Data-centric foundation models in computational healthcare: A survey. *arXiv preprint arXiv:2401.02458*, 2024.

[53] Zhenyue Gao, Xiaoli Liu, Yu Kang, Pan Hu, Xiu Zhang, Wei Yan, Muyang Yan, Pengming Yu, Qing Zhang, Wendong Xiao, et al. Improving the prognostic evaluation precision of hospital outcomes for heart failure using admission notes and clinical tabular data: Multimodal deep learning model. *Journal of Medical Internet Research*, 26:e54363, 2024.

[54] Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1):171, 2022.
[55] Weimin Lyu, Xinyu Dong, Rachel Wong, Songzhu Zheng, Kayley Abell-Hart, Fusheng Wang, and Chao Chen. A multimodal transformer: Fusing clinical notes with structured ehr data for interpretable in-hospital mortality prediction. In *AMIA Annual Symposium Proceedings*, volume 2022, page 719. American Medical Informatics Association, 2022.
[56] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous systems. 2015.
[57] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. 2019.
[58] François Chollet et al. Keras. 2015.
[59] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface's transformers: State-of-the-art natural language processing. 2019.
[60] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. 2011.
[61] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
[62] Diego Ardila, Atilla P Kiraly, Sirish Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, William Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961, 2019.
[63] James Zou, David Lähnemann, Artur Poon, Franziska Gruhl, Theodore Alexandrov, Alicia Clum Martin, Wei Chen, Anna Ratner, Anshul Kundaje, Michael Snyder, Howard Y Chang, Paul A Khavari, and William J Greenleaf. A primer on deep learning in genomics. *Nature genetics*, 51(1):12–18, 2019.
[64] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michael Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1):1–10, 2018.
[65] Jessica Vamathevan, Daniel Clark, Paul Czodrowski, Ian Dunham, Eduard Ferran, Grace Lee, Bojian Li, Anant Madabhushi, Prashant Shah, Matthew Spitzer, and Shanrong Zhao. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6):463–477, 2019.
[66] Hsinchun Chen, Daniel Zeng, and Ping Yan. Artificial intelligence in public health surveillance: A systematic review. *Intelligence-Based Medicine*, 1:100001, 2020.
[67] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.
[68] Shweta Agrawal and Sanjiv Kumar Jain. Medical text and image processing: applications, issues and challenges. *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*, pages 237–262, 2020.
[69] Paul Suetens. *Fundamentals of medical imaging*. Cambridge university press, 2017.
[70] Yixian Xu, Xinkai Zheng, Yuanjie Li, Xinmiao Ye, Hongtao Cheng, Hao Wang, and Jun Lyu. Exploring patient medication adherence and data mining methods in clinical big data: A contemporary review. *Journal of Evidence-Based Medicine*, 16(3):342–375, 2023.
[71] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023.
[72] Malwina Anna Wójcik. Foundation models in healthcare: opportunities, biases and regulatory prospects in europe. In *International Conference on Electronic Government and the Information Systems Perspective*, pages 32–46. Springer, 2022.
[73] Jianing Qiu, Lin Li, Jiankai Sun, Jiachuan Peng, Peilun Shi, Ruiyang Zhang, Yinzhao Dong, Kyle Lam, Frank P-W Lo, Bo Xiao, et al. Large ai models in health informatics: Applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics*, 2023.
[74] Sajan B Patel and Kyle Lam. Chatgpt: the future of discharge summaries? *The Lancet Digital Health*, 5(3):e107–e108, 2023.
[75] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference*, pages 301–318. PMLR, 2016.
[76] Rajesh Ranganath, Adler J Perotte, Noémie Elhadad, and David M Blei. The survival filter: Joint survival analysis with a latent time series. In *UAI*, pages 742–751, 2015.
[77] Anand Avati, Kenneth Jung, Stephanie Harman, Lance Downing, Andrew Ng, and Nigam H Shah. Improving palliative care with deep learning. *BMC medical informatics and decision making*, 18:55–64, 2018.
[78] Luciano RS de Lima, Alberto HF Laender, and Berthier A Ribeiro-Neto. A hierarchical approach to the automatic categorization of medical documents. In *Proceedings of the seventh international conference on Information and knowledge management*, pages 132–139, 1998.
[79] Elena Birman-Deych, Amy D Waterman, Yan Yan, David S Nilasena, Martha J Radford, and Brian F Gage. Accuracy of icd-9-cm codes for identifying cardiovascular and stroke risk factors. *Medical care*, 43(5):480–485, 2005.
[80] Irena Spasic, Goran Nenadic, et al. Clinical text data in machine learning: systematic review. *JMIR medical informatics*, 8(3):e17984, 2020.
[81] Irena Spasić, Jacqueline Livsey, John A Keane, and Goran Nenadić. Text mining of cancer-related information: review of current status and future directions. *International journal of medical informatics*, 83(9):605–623, 2014.
[82] Irena Spasic, Özlem Uzuner, and Li Zhou. Emerging clinical applications of text analytics. *International journal of medical informatics*, 134, 2020.
[83] Bethany Percha. Modern clinical text mining: a guide and review. *Annual review of biomedical data science*, 4:165–187, 2021.
[84] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

[85] Bharath Chintagunta, Namit Katariya, Xavier Amatriain, and Anitha Kannan. Medically aware gpt-3 as a data generator for medical dialogue summarization. In *Machine Learning for Healthcare Conference*, pages 354–372. PMLR, 2021.
[86] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.
[87] Hoo-Chang Shin, Yang Zhang, Evelina Bakhturina, Raul Puri, Mostofa Patwary, Mohammad Shoeybi, and Raghav Mani. Biomegatron: larger biomedical domain language model. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4700–4706, 2020.
[88] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.
[89] Kathi Canese and Sarah Weis. Pubmed: the bibliographic database. *The NCBI handbook*, 2(1), 2013.
[90] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, et al. A large language model for electronic health records. *NPJ digital medicine*, 5(1):194, 2022.
[91] Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1):86, 2021.
[92] Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023.
[93] Benyou Wang, Qianqian Xie, Jiahuan Pei, Zhihong Chen, Prayag Tiwari, Zhao Li, and Jie Fu. Pre-trained language models in biomedical domain: A systematic survey. *ACM Computing Surveys*, 56(3):1–52, 2023.
[94] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020.
[95] David A Bluemke, Linda Moy, Miriam A Bredella, Birgit B Ertl-Wagner, Kathryn J Fowler, Vicky J Goh, Elkan F Halpern, Christopher P Hess, Mark L Schiebler, and Clifford R Weiss. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board, 2020.
[96] Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4025, 2021.
[97] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.
[98] Jia Deng. A large-scale hierarchical image database. *Proc. of IEEE Computer Vision and Pattern Recognition, 2009*, 2009.
[99] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.
[100] Lu Yuan, Dongdong Chen, Yi-Ling Chen, Noel Codella, Xiyang Dai, Jianfeng Gao, Houdong Hu, Xuedong Huang, Boxin Li, Chunyuan Li, et al. Florence: A new foundation model for computer vision. *arXiv preprint arXiv:2111.11432*, 2021.
[101] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, pages 1–13, 2024.
[102] Alexander Ke, William Ellsworth, Oishi Banerjee, Andrew Y Ng, and Pranav Rajpurkar. Chextransfer: performance and parameter efficiency of imagenet models for chest x-ray interpretation. In *Proceedings of the conference on health, inference, and learning*, pages 116–124, 2021.
[103] Harlan M Krumholz, Sharon F Terry, and Joanne Waldstreicher. Data acquisition, curation, and use for a continuously learning health system. *Jama*, 316(16):1669–1670, 2016.
[104] Mohammad-Parsa Hosseini, Aaron Lau, Kost Elisevich, and Hamid Soltanian-Zadeh. Multimodal analysis in biomedicine. *Big Data in Multimodal Medical Imaging*, 193, 2019.
[105] Annamalai Suresh, R Udendhran, and S Vimal. *Deep neural networks for multimodal imaging and biomedical applications*. IGI Global, 2020.
[106] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H Loosen, Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, 25(7):1054–1056, 2019.
[107] Jakub R Kaczmarzyk, Rajarsi Gupta, Tahsin M Kurc, Shahira Abousamra, Joel H Saltz, and Peter K Koo. Champkit: A framework for rapid evaluation of deep neural networks for patch-based histopathology classification. *Computer Methods and Programs in Biomedicine*, 239:107631, 2023.
[108] Shaoting Zhang and Dimitris Metaxas. On the challenges and perspectives of foundation models for medical image analysis. *Medical Image Analysis*, page 102996, 2023.
[109] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.
[110] Saikat Roy, Tassilo Wald, Gregor Koehler, Maximilian R Rokuss, Nico Disch, Julius Holzschuh, David Zimmerer, and Klaus H Maier-Hein. Sam. md: Zero-shot medical image segmentation capabilities of the segment anything model. *arXiv preprint arXiv:2304.05396*, 2023.
[111] FC Ghesu, B Georgescu, A Mansoor, Y Yoo, D Neumann, P Patel, RS Vishwanath, JM Balter, Y Cao, S Grbic, et al. Self-supervised learning from 100 million medical images. arxiv 2022. *arXiv preprint arXiv:2201.01283*.
[112] Ekin Tiu, Ellie Talius, Pujan Patel, Curtis P Langlotz, Andrew Y Ng, and Pranav Rajpurkar. Expert-level detection of pathologies from unannotated chest x-ray images via self-supervised learning. *Nature Biomedical Engineering*, 6(12):1399–1406, 2022.

[113] Ziyan Huang, Haoyu Wang, Zhongying Deng, Jin Ye, Yanzhou Su, Hui Sun, Junjun He, Yun Gu, Lixu Gu, Shaoting Zhang, et al. Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training. *arXiv preprint arXiv:2304.06716*, 2023.

[114] Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Siqi Liu, Philippe Mathieu, Alexander van Eck, Donghun Lee, Julian Viret, et al. Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778*, 2023.

[115] Yukun Zhou, Mark A Chia, Siegfried K Wagner, Murat S Ayhan, Dominic J Williamson, Robbert R Struyven, Timing Liu, Moucheng Xu, Mateo G Lozano, Peter Woodward-Court, et al. A foundation model for generalizable disease detection from retinal images. *Nature*, 622(7981):156–163, 2023.

[116] Xiangde Luo, Wenjun Liao, Jianghong Xiao, Jieneng Chen, Tao Song, Xiaofan Zhang, Kang Li, Dimitris N Metaxas, Guotai Wang, and Shaoting Zhang. Word: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from ct image. *arXiv preprint arXiv:2111.02403*, 2021.

[117] Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Towards generalist foundation model for radiology. *arXiv preprint arXiv:2308.02463*, 2023.

[118] Ashish K Jha, Catherine M DesRoches, Eric G Campbell, Karen Donelan, Sowmya R Rao, Timothy G Ferris, Alexandra Shields, Sara Rosenbaum, and David Blumenthal. Use of electronic health records in us hospitals. *New England Journal of Medicine*, 360(16):1628–1638, 2009.

[119] Chao Pang, Xinzhuo Jiang, Krishna S Kalluri, Matthew Spotnitz, RuiJun Chen, Adler Perotte, and Karthik Natarajan. Cehr-bert: Incorporating temporal information from structured ehr data to improve prediction tasks. In *Machine Learning for Health*, pages 239–260. PMLR, 2021.

[120] PKS Prakash, Srinivas Chilukuri, Nikhil Ranade, and Shankar Viswanathan. Rarebert: transformer architecture for rare disease patient identification using administrative claims. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 453–460, 2021.

[121] Michael Wornow, Rahul Thapa, Ethan Steinberg, Jason Fries, and Nigam Shah. Ehrshot: An ehr benchmark for few-shot evaluation of foundation models. *Advances in Neural Information Processing Systems*, 36, 2024.

[122] Zifeng Wang, Chufan Gao, Cao Xiao, and Jimeng Sun. Anypredict: Foundation model for tabular prediction. *arXiv preprint arXiv:2305.12081*, 2023.

[123] Kexin Ding, Mu Zhou, Dimitris N Metaxas, and Shaoting Zhang. Pathology-and-genomics multimodal transformer for survival outcome prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 622–631. Springer, 2023.

[124] Kyunghoon Hur, Jungwoo Oh, Junu Kim, Jiyoun Kim, Min Jae Lee, Eunbyeol Cho, Seong-Eun Moon, Young-Hak Kim, Louis Atallah, and Edward Choi. Genhpf: General healthcare predictive framework for multi-task multi-source learning. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[125] Roman Levin, Valeriia Cherepanova, Avi Schwarzschild, Arpit Bansal, C Bayan Bruss, Tom Goldstein, Andrew Gordon Wilson, and Micah Goldblum. Transfer learning with deep tabular models. *arXiv preprint arXiv:2206.15306*, 2022.

[126] Mingyu Jin, Qinkai Yu, Dong Shu, Chong Zhang, Suiyuan Zhu, Mengnan Du, Yanda Meng, and Yongfeng Zhang. Health-llm: Personalized retrieval-augmented disease prediction system.

[127] Matthew McDermott, Bret Nestor, Peniel Argaw, and Isaac S Kohane. Event stream gpt: a data pre-processing and modeling library for generative, pre-trained transformers over continuous-time sequences of complex events. *Advances in Neural Information Processing Systems*, 36, 2024.

[128] Aleksa Bisercic, Mladen Nikolic, Mihaela van der Schaar, Boris Delibasic, Pietro Lio, and Andrija Petrovic. Interpretable medical diagnostics with structured data extraction by large language models. *arXiv preprint arXiv:2306.05052*, 2023.

[129] Isabella Catharina Wiest, Dyke Ferber, Jiefu Zhu, Marko Van Treeck, Sonja Katharina Meyer, Radhika Juglan, Zunamys I Carrero, Daniel Paech, Jens Kleesiek, Matthias P Ebert, et al. From text to tables: a local privacy preserving large language model for structured information retrieval from medical documents. *medRxiv*, pages 2023–12, 2023.

[130] Stefan Hegselmann, Alejandro Buendia, Hunter Lang, Monica Agrawal, Xiaoyi Jiang, and David Sontag. Tabllm: Few-shot classification of tabular data with large language models. In *International Conference on Artificial Intelligence and Statistics*, pages 5549–5581. PMLR, 2023.

[131] Ethan Steinberg, Jason Fries, Yizhe Xu, and Nigam Shah. Motor: A time-to-event foundation model for structured medical records. *arXiv preprint arXiv:2301.03150*, 2023.

[132] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65–69, 2019.

[133] Orchid Chetia Phukan, Ankita Das, Arun Balaji Buduru, and Rajesh Sharma. Sonic: Synergizing vision foundation models for stress recognition from ecg signals. *arXiv preprint arXiv:2404.00827*, 2024.

[134] Jiewei Lai, Huixin Tan, Jinliang Wang, Lei Ji, Jun Guo, Baoshi Han, Yajun Shi, Qianjin Feng, and Wei Yang. Practical intelligent diagnostic algorithm for wearable 12-lead ecg via self-supervised learning on large-scale dataset. *Nature Communications*, 14(1):3741, 2023.

[135] Nathaniel Diamant, Erik Reinertsen, Steven Song, Aaron D Aguirre, Collin M Stultz, and Puneet Batra. Patient contrastive learning: A performant, expressive, and practical approach to electrocardiogram modeling. *PLoS computational biology*, 18(2):e1009862, 2022.

[136] Joseph Y Cheng, Hanlin Goh, Kaan Dogrusoz, Oncel Tuzel, and Erdrin Azemi. Subject-aware contrastive learning for biosignals. *arXiv preprint arXiv:2007.04871*, 2020.

[137] Salar Abbaspourazad, Oussama Elachqar, Andrew C Miller, Saba Emrani, Udhyakumar Nallasamy, and Ian Shapiro. Large-scale training of foundation models for wearable biosignals. *arXiv preprint arXiv:2312.05409*, 2023.

[138] Mehmet Yamaç, Mert Duman, İlke Adalıoğlu, Serkan Kiranyaz, and Moncef Gabbouj. A personalized zero-shot ecg arrhythmia monitoring system: From sparse representation based domain adaption to energy efficient abnormal beat detection for practical ecg surveillance. *arXiv preprint arXiv:2207.07089*, 2022.

[139] Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

[140] He Cao, Zijing Liu, Xingyu Lu, Yuan Yao, and Yu Li. Instructmol: Multi-modal integration for building a versatile and reliable molecular assistant in drug discovery. *arXiv preprint arXiv:2311.16208*, 2023.

[141] Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Liò. 3d infomax improves gnns for molecular property prediction. In *International Conference on Machine Learning*, pages 20479–20502. PMLR, 2022.

[142] Jun Li, Che Liu, Sibo Cheng, Rossella Arcucci, and Shenda Hong. Frozen language model helps ecg zero-shot learning. In *Medical Imaging with Deep Learning*, pages 402–415. PMLR, 2024.

[143] Emma Chen, Aman Kansal, Julie Chen, Boyang Tom Jin, Julia Reisler, David E Kim, and Pranav Rajpurkar. Multimodal clinical benchmark for emergency care (mc-bec): A comprehensive benchmark for evaluating foundation models in emergency medicine. *Advances in Neural Information Processing Systems*, 36, 2024.

[144] Sathvik Bhaskarpandit, Priyanka Gupta, and Manik Gupta. Lets-gzsl: A latent embedding model for time series generalized zero shot learning. *arXiv preprint arXiv:2207.12007*, 2022.

[145] Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3(1):015022, 2022.

[146] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.

[147] Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A Lemley, and Percy Liang. Foundation models and fair use. 2023.

[148] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248):1–43, 2020.

[149] Does scraping violate the Computer Fraud and Abuse Act? Federal appeals court says no. — rcfp.org. `https://www.rcfp.org/scraping-not-violation-cfaa/`. [Accessed 29-05-2024].

[150] Eduard Fosch Villaronga, Peter Kieseberg, and Tiffany Li. Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law & Security Review*, 34(2):304–313, 2018.

[151] Sara Gerke, Timo Minssen, and Glenn Cohen. Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare*, pages 295–336. Elsevier, 2020.

[152] Danton S Char, Nigam H Shah, and David Magnus. Implementing machine learning in health care—addressing ethical challenges. *New England Journal of Medicine*, 378(11):981–983, 2018.

[153] Michael Veale and Frederik Zuiderveen Borgesius. Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4):97–112, 2021.

[154] Helen Smith. Clinical ai: opacity, accountability, responsibility and liability. *Ai & Society*, 36(2):535–545, 2021.

[155] Mohammad Amir Khusru Akhtar, Mohit Kumar, and Anand Nayyar. Transparency and accountability in explainable ai: Best practices. In *Towards Ethical and Socially Responsible Explainable AI: Challenges and Opportunities*, pages 127–164. Springer, 2024.

[156] Anna Jobin, Marcello Ienca, and Effy Vayena. The global landscape of ai ethics guidelines. *Nature machine intelligence*, 1(9):389–399, 2019.

[157] Xuejuan Zhang. Healthcare regulation and governance: Big data analytics and healthcare data protection. 2020.

[158] Cecile De Terwangne. Principles relating to processing of personal data. In *The EU general data protection (GDPR): a commentary*, pages 309–320. Oxford University Press, 2020.

[159] Biwei Yan, Kun Li, Minghui Xu, Yueyan Dong, Yue Zhang, Zhaochun Ren, and Xiuzhen Cheng. On protecting the data privacy of large language models (llms): A survey. *arXiv preprint arXiv:2403.05156*, 2024.

[160] Shuzhou Sun, Li Liu, Yongxiang Liu, Zhen Liu, Shuanghui Zhang, Janne Heikkilä, and Xiang Li. Uncovering bias in foundation models: Impact, testing, harm, and mitigation. *arXiv preprint arXiv:2501.10453*, 2025.

[161] Richard J Chen, Tiffany Y Chen, Jana Lipkova, Judy J Wang, Drew FK Williamson, Ming Y Lu, Sharifa Sahai, and Faisal Mahmood. Algorithm fairness in ai for medicine and healthcare. *arXiv preprint arXiv:2110.00603*, 2021.

[162] Khiara M Bridges. Race in the machine: Racial disparities in health and medical ai. *Virginia Law Review*, 110(2), 2024.

[163] Roxana Daneshjou, Kailas Vodrahalli, Roberto A Novoa, Melissa Jenkins, Weixin Liang, Veronica Rotemberg, Justin Ko, Susan M Swetter, Elizabeth E Bailey, Olivier Gevaert, et al. Disparities in dermatology ai performance on a diverse, curated clinical image set. *Science advances*, 8(31):eabq6147, 2022.

[164] Benjamin P Chapman, Feng Lin, Shumita Roy, Ralph HB Benedict, and Jeffrey M Lyness. Health risk prediction models incorporating personality data: Motivation, challenges, and illustration. *Personality Disorders: Theory, Research, and Treatment*, 10(1):46, 2019.

[165] Yongkang Zhang, Yiye Zhang, Evan Sholle, Sajjad Abedian, Marianne Sharko, Meghan Reading Turchioe, Yiyuan Wu, and Jessica S Ancker. Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death. *PloS one*, 15(6):e0235064, 2020.

[166] Hannah van Kolfschooten and Janneke van Oirschot. The eu artificial intelligence act (2024): implications for healthcare. *Health Policy*, 149:105152, 2024.

[167] Stan Benjamens, Pranavsingh Dhunnoo, and Bertalan Meskó. The state of artificial intelligence-based fda-approved medical devices and algorithms: an online database. *NPJ digital medicine*, 3(1):118, 2020.

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

[169] Ahmed Shihab Albahri, Ali M Duhaim, Mohammed A Fadhel, Alhamzah Alnoor, Noor S Baqer, Laith Alzubaidi, Osamah Shihab Albahri, Abdullah Hussein Alamoodi, Jinshuai Bai, Asma Salhi, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96:156–191, 2023.

[170] Carolyn Ashurst. How to Regulate Foundation Models.

[171] Vipula Rawte, Amit Sheth, and Amitava Das. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922*, 2023.

[172] Future Society. Towards Effective Governance of Foundation Models and Generative AI, 4 2024.
[173] Paul Foley. Foundation Models, Generative AI and the EU AI Act, 8 2023.
[174] Shalini Bhartiya and Deepti Mehrotra. Challenges and recommendations to healthcare data exchange in an interoperable environment. *Electronic Journal of Health Informatics*, 8(2):16, 2014.
[175] Frank M Aarestrup, Abdullah Albeyatti, W John Armitage, Charles Auffray, Luca Augello, Rudi Balling, Nora Benhabiles, Guido Bertolini, Jan G Bjaalie, Michaela Black, et al. Towards a european health research and innovation cloud (hric). *Genome medicine*, 12:1–14, 2020.
[176] Linda T Li, Lauren C Haley, Alexandra K Boyd, and Elmer V Bernstam. Technical/algorithm, stakeholder, and society (tass) barriers to the application of artificial intelligence in medicine: A systematic review. *Journal of Biomedical Informatics*, page 104531, 2023.
[177] Zhenliang Zhang, Zeyu Zhang, Ziyuan Jiao, Yao Su, Hangxin Liu, Wei Wang, and Song-Chun Zhu. On the emergence of symmetrical reality. *arXiv preprint arXiv:2401.15132*, 2024.
[178] Zhixin Lai, Jing Wu, Suiyao Chen, Yucheng Zhou, Anna Hovakimyan, and Naira Hovakimyan. Language models are free boosters for biomedical imaging tasks. *arXiv preprint arXiv:2403.17343*, 2024.
[179] Adnan Qayyum, Junaid Qadir, Muhammad Bilal, and Ala Al-Fuqaha. Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, 14:156–180, 2020.
[180] Kacper Sokol and Peter Flach. One explanation does not fit all: The promise of interactive explanations for machine learning transparency. *KI-Künstliche Intelligenz*, 34(2):235–250, 2020.
[181] Giovanni Briganti. A doctor's guide to foundation models. 2023.
[182] Yang Yuan. On the power of foundation models. In *International Conference on Machine Learning*, pages 40519–40530. PMLR, 2023.
[183] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10):573–584, 2020.
[184] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.
[185] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.
[186] RA Ribeiro, EN Baungratz, G Vaccaro, PS Schmitz, AK Fernandes, V Colpani, and M Falavigna. Factors influencing door-to-doctor time in an emergency department: discrete event simulation model. *Value in Health*, 19(7):A624, 2016.
[187] Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841, 2017.
[188] Anam Sajid and Haider Abbas. Data privacy in cloud-assisted healthcare systems: state of the art and future challenges. *Journal of medical systems*, 40(6):155, 2016.
[189] Zhihan Zhou, Yanrong Ji, Weijian Li, Pratik Dutta, Ramana Davuluri, and Han Liu. Dnabert-2: Efficient foundation model and benchmark for multi-species genome. *arXiv preprint arXiv:2306.15006*, 2023.
[190] Thomas Grote and Oliver Buchholz. Machine learning in public health and the prediction-intervention gap. 2024.
[191] BY Anom. Ethics of big data and artificial intelligence in medicine. *Ethics, Medicine and Public Health*, 15:100568, 2020.
[192] Seunghyun Lee, Da Young Lee, Sujeong Im, Nan Hee Kim, and Sung-Min Park. Clinical decision transformer: intended treatment recommendation through goal prompting. *arXiv preprint arXiv:2302.00612*, 2023.
[193] Stella Verkijk and Piek Vossen. Medroberta. nl: a language model for dutch electronic health records. In *Computational Linguistics in the Netherlands*, volume 11, pages 141–159, 2021.
[194] Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Le Song, and Xuegong Zhang. Large scale foundation model on single-cell transcriptomics. *bioRxiv*, pages 2023–05, 2023.
[195] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 59–68, 2019.
[196] Reihaneh Torkzadehmahani, Reza Nasirigerdeh, David B Blumenthal, Tim Kacprowski, Markus List, Julian Matschinske, Julian Spaeth, Nina Kerstin Wenke, and Jan Baumbach. Privacy-preserving artificial intelligence techniques in biomedicine. *Methods of information in medicine*, 61(S 01):e12–e27, 2022.
[197] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1):17, 2020.
[198] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*, 2023.